

Open data business models for media industry - Finnish case study

Information Systems Science

Master's thesis

Tomi Kinnari

2013

ABSTRACT

Governments and private companies have begun to make vast amounts of data resources available to the public without usage restrictions, in the form of open data. For example, Finnish governmental bureaus have made legal documents, statistics, geographical data, traffic data, and environmental data freely available for public use. These new data sources have enabled innovative services in several areas, and create a lucrative opportunity for media companies. Open data can enrich media content, for example, with live data streams, advanced visualizations, and context and location dependent information.

This thesis identifies opportunities open data provides for media companies by conducting an extensive field study of the Finnish open data landscape. First, 15 companies pioneering in open data use are analysed to determine their offering, revenue model and resources, and the general value network in which they operate. These findings are then considered from the media company perspective in order to identify opportunities that open data provides for them.

The open data industry in Finland is still in its early stages, but some commercial success can already be identified. This study grouped the examined companies into five profiles in an open data value network: (1) data analysers, (2) data extractors and transformers, (3) user experience providers, (4) commercial data publishers, and (5) support services and consultancy. These five profiles are grounded on both; the empirical findings of this study as well as the theoretical frameworks established by preceding academic papers.

For media companies this research found three opportunity avenues; (1) use open data as a source in data journalism, (2) gather article ideas and content from the visual and numerical data analyses conducted by third-party analysers, or (3) achieve costs savings by publishing private data and using crowds to analyse it or creating user interfaces on top of it.

Keywords: open data, value network analysis, business models, media

ABSTRAKTI

Valtion yksiköt ja yksityiset yritykset ovat alkaneet avata omia tietovarantojaan yleisölle ilman käyttörajoituksia, eli avoimena datana. Esimerkiksi Suomen hallinnolliset virastot ovat avanneet lakitekstejä, tilastotietoa, maantieteellistä dataa, liikennedatata ja ympäristödataa vapaasti käytettäväksi. Nämä uudet tietolähteet ovat mahdollistaneet innovatiivisia, uusia palveluja ja sovelluksia useilla alueilla, ja ne ovat myös houkutteleva mahdollisuus mediayhtiöille. Avointa dataa voidaan käyttää esimerkiksi mediasisällön rikastamiseen reaaliaikaisilla datavirroilla, kehittyneillä visualisoinneilla tai paikka- ja tilanneriippuvalla tiedolla.

Tämä tutkielma kartoittaa laajan suomalaisen kenttätutkimuksen avulla mahdollisuuksia, joita avoin data tarjoaa mediayhtiöille. Aluksi analysoidaan 15 avoimen datan edelläkävijäyrityksen tarjoama, ansaintamalli ja resurssit sekä yleinen arverkko, jossa yritykset toimivat. Sen jälkeen nämä edelläkävijöiltä kerätyt löydökset tulkitaan mediayhtiöiden näkökulmasta, jotta voidaan ymmärtää avoimen datan tarjoamat mahdollisuudet medialle.

Avoimen datan toimiala on Suomessa edelleen varhaisessa vaiheessa, mutta ensimmäisiä esimerkkejä kaupallisesta menestyksestä on kuitenkin nähtävissä. Tutkielma jaotteli nämä yritykset viiteen profiiliin, jotka yhdessä muodostavat avoimen tiedon arverkon: (1) datan analysoijat, (2) raakadatan käsittelijät, (3) loppukäyttäjän käyttökokemuksen tarjoajat, (4) kaupalliset datan avaajat ja (5) tukitoiminnot ja konsultointi. Nämä viisi profiilia pohjautuvat sekä tässä tutkielmassa tehtyyn kenttätutkimukseen että aikaisempiin akateemisiin tuloksiin.

Medialle tämä tutkielma löysi kolme mahdollisuutta hyödyntää avointa dataa: (1) käyttää avointa dataa datajournalismin lähteenä, (2) hyödyntää kolmansien osapuolten tekemiä visuaalisia tai numeerisia analyysejä uusien artikkeleiden innoittajana ja sisältönä tai (3) tavoitella kustannussäästöjä joukkoistamalla oman tiedon analysointia ja uusien käyttöliittymien luontia.

Avainsanat: avoin data, arverkkoanalyysi, liiketoimintamalli, media

ACKNOWLEDGEMENTS

I would like to express my very great appreciation to Matti Rossi for giving me the opportunity to write this Thesis and for the support and help throughout the research and writing process. I would also like to offer my sincerest thankfulness for Tekes Next Media project for providing the financial support to conduct this research. I would also like to thank Sanoma Oy and Helsinki Region Infoshare for their assistance and help. I am truly indebted and thankful for the advice and feedback given by Juho Lindman throughout the process. I owe sincere and earnest thankfulness for all the interviewees, without your helpful comments and insight this Thesis could not have been done. I wish to thank Hannu Kivijärvi and the entire Master's Thesis seminar group for your valuable feedback. I would like to offer my special thanks to for Virpi Tuunainen, Aalto University School of Business, and Aalto Service Factory for providing the facilities to write the Thesis. In addition, I am obliged to all of my colleagues who supported me during the writing process. Finally, I would like to thank my family and friends for their support and inspiration.

TABLE OF CONTENTS

Abstract	1
Abstrakti	2
Aknowledgements	3
Table of Tables	7
Table of Figures	7
List of Acronyms and Abbreviations	8
1. Introduction	9
1.1 Motivation for the research	9
1.2 Open data as a research topic.....	10
1.3 Structure of the study	13
2. Open data and business models in literature	13
2.1 Open data.....	13
2.2 Business model elements.....	14
2.3 Open data business models in research	15
2.4 Open data definition in this paper	20
2.5 Media definition in this paper	21
2.6 Economics of the media business	21
2.7 Open data and media business	22
3. Research design	24
3.1 Philosophical worldview.....	24
3.2 Research question.....	25
3.3 Proposition and purpose of the research.....	26
3.4 Research strategy	26
3.5 Organizing the research fieldwork	28
3.5.1 Selecting the companies.....	29
3.5.2 Interview process and questions	31
3.6 Estimating the quality of the research design	32
3.6.1 Construct validity	33
3.6.2 External validity	36
3.6.3 Reliability	37
4. Case studies	38
4.1 HSL Reittiopas API.....	39
4.1.1 Description.....	39
4.1.2 Analysis: Crowd-sourced client development	40
4.2 Case Duunitori.fi	42
4.2.1 Description.....	42

4.2.2	Analysis: Create valuable user experience and monetize with advertising (two-sided markets).....	43
4.3	Case Mitamukaanlennolle.fi.....	43
4.3.1	Description.....	43
4.3.2	Analysis: Create valuable user experience and monetize with advertising and licensing	44
4.4	Case pikkuparlamentti.fi	45
4.4.1	Case description.....	45
4.4.2	Analysis: Create valuable user experience and monetize with advertising	45
4.5	Case ReittiGPS and Reitit	46
4.5.1	Description.....	46
4.5.2	Analysis: Create valuable user experience and monetize with one-time fee.....	47
4.6	Case Hilmappi	47
4.6.1	Description.....	47
4.6.2	Analysis: Create a valuable user experience and monetize with annual subscription.....	48
4.7	Case Kansanmuisti.fi.....	48
4.7.1	Description.....	48
4.7.2	Analysis: Create a valuable user experience, and monetize with crowd-funding	49
4.8	Case Hahmota Oy Tax-tree.....	49
4.8.1	Description.....	49
4.8.2	Analysis: Create visualizations and monetize by selling project work	50
4.9	Case Asiakastieto.....	50
4.9.1	Description.....	50
4.9.2	Analysis: Algorithm-based analysing	51
4.10	Case Cloud'N'Sci.....	52
4.10.1	Description.....	52
4.10.2	Analysis: Algorithm-based analysing	53
4.11	Case HS Open	54
4.11.1	Description.....	54
4.11.2	Analysis: Crowd-sourced data analysing	56
4.12	Case Louhos	57
4.12.1	Description.....	57
4.12.2	Analysis: Extract and transform	58
4.13	Case Flo Apps	59
4.13.1	Description.....	59
4.13.2	Analysis: Consultation and software projects	60
4.14	Case Logica.....	61
4.14.1	Description.....	61
4.14.2	Analysis: Better services with machine-to-machine communication..	62
4.15	General observations from the interviews.....	63

5. Value network analysis	64
5.1 Data analysers.....	65
5.1.1 Data visualizers	65
5.1.2 Algorithm based analysis.....	65
5.2 Extract and transform	66
5.2.1 Extract and transform as integrated part of data analysing.....	67
5.2.2 Extract & transform as separate business	67
5.3 User experience provider.....	68
5.4 Commercial open data publishers	69
5.4.1 Co-creation under open license.....	70
5.4.2 Co-creation under restricted license	70
5.5 Support services and consultation.....	71
5.6 Summary of the value network analysis	72
6. Conclusions for media companies	74
6.1 Opportunity 1: Raw data as a source in data journalism and transparency ..	76
6.2 Opportunity 2: Third-party created analysis as a source for new content and article ideas	77
6.3 Opportunity 3: Publish commercial open data	77
6.3.1 Publish data with no limitations for re-use	79
6.3.2 Publish data with limited re-use	80
6.4 Summary	82
6.5 Feedback from the interviewees (feedback round #2).....	83
7. Discussion of the results	84
7.1 Theoretical contributions.....	84
7.2 Limitations of the results.....	88
7.3 Future research	89
Appendixes	90
References.....	91

TABLE OF TABLES

Table 1 List of analysed companies.....	30
Table 2 Case study tactics for the four design tests (COSMOS Corporation, as cited in Yin, 2003)	33
Table 3 Different sources of evidence according to Yin (2003, pp. 85-96), and summary of the evidence utilized in this Thesis	34
Table 4 Recapitulation of the value network profiles.....	73
Table 5 Value network comparison to the previous research.....	85

TABLE OF FIGURES

Figure 1 Data catalogues around the world (Open Government Data Dashboard).....	9
Figure 2 Relative search volumes for “open data” in Google search engine. The number 100 represents the peak search volume (Google Trends)	11
Figure 3 Number of academic documents published per year (Web of Knowledge).....	12
Figure 4 Business model elements as defined by Rajala (2009)	15
Figure 5 Linked data value chain, reproduced from “The Linked Data Value Chain: A Lightweight Model for Business Engineers “ by Latif et al., 2009, p. 3	16
Figure 6 Linked data value chain, reproduced from “Open Data Business Models” by Tammisto and Lindman, 2011, p. 9	17
Figure 7 Assemblage of open data complementarities, reproduced from “The Roles of Agency and Artifacts in Assembling Open Data Complementarities” by Kuk and Davies, 2011, p. 11.....	19
Figure 8 Data hub model adopted from Aitamurto et al. (2011), originally presented by Kayser-Bril (2011).....	24
Figure 9 Research strategy	28
Figure 10 Apps 4 Finland submission screening process	29
Figure 11 Research process	35
Figure 12 An example newspaper article based on crowd-sourced analyzing. Source: HS Next-blog post (Mäkinen, 10.2.2012).....	55
Figure 13 Data analysers (profile 1)	65
Figure 14 Extract & Transform (profile 2)	66
Figure 15 User experience provider (profile 3)	69
Figure 16 Commercial open data publisher (profile 4).....	69
Figure 17 Support services and consultation (profile 5).....	71
Figure 18 Open data value network from the media context.....	76
Figure 19 Monetizing value of data – limitations applied matrix.....	79
Figure 20 Re-representation of the value network found out in this Thesis.....	87

LIST OF ACRONYMS AND ABBREVIATIONS

A4F	Apps 4 Finland (pronounced “apps for Finland”)
API	Application programming interface
HTTP	Hypertext Transfer Protocol
HS	Helsingin Sanomat
M2M	Machine-to-machine
RDF	Resource Description Framework
URI	Uniform Resource Identifier

1. INTRODUCTION

1.1 Motivation for the research

Open data is an ideology where governments and companies place their datasets on the internet for anyone to use freely. Open data's promises of increased transparency and innovative services are attracting developers, managers and business people (Poikola et. al, 2010; Kuk & Davies, 2011). Governments and cities all over the world are considering what datasets to open, and what effect the opening would have for the society. According to Open Government Data Dashboard, as of October 5, 2012, there were 268 data catalogues listing open data resources on country, state and city levels around the world (see Figure 1). These catalogues help developers to find open data, but how the huge amount of data should be utilized and what effect does it have to business? Especially, what kind of sustainable businesses can be built on top of open data?

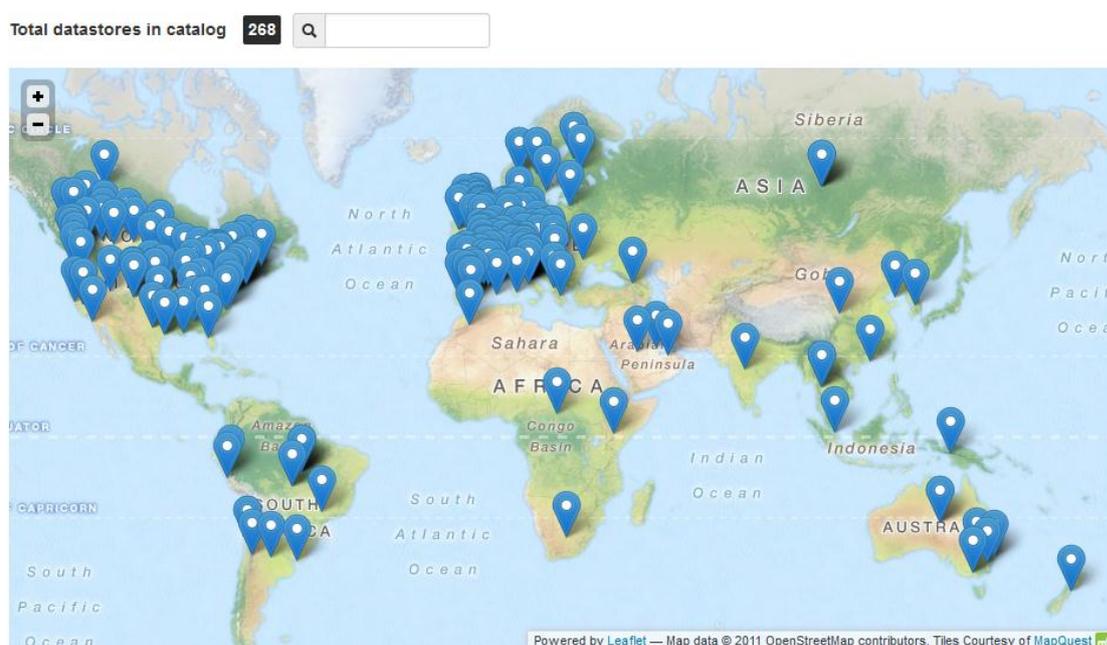


Figure 1 Data catalogues around the world (Open Government Data Dashboard)

Open data services and applications are largely provided by hobbyists and enthusiasts, who are working with pro-bono mind-set (Kuk & Davies, 2011, pp. 8-10). It is important that these services would be operating with viable business model and revenue logic. Without sustainable business models, these services risk having a shorter life span, which, in the long run, might endanger the entire open data ecosystem. Therefore, the goal of this paper is to

conduct a systematic business model analysis in order to understand open data phenomena from the business model perspective.

The particular research question is how media companies could utilize the benefits of open data in their business. Media companies are a fascinating research subject because they are facing several extraordinary transformations in the coming years. For example, decreasing circulation of newspapers, paywalls in internet news portals, on-demand content streaming and shifting of advertiser behaviour are just a few of the transformational forces that will be shaping the industry.

The research question will be answered by first studying open data companies in general, regardless of their industry, and figuring out the underlying value network and business model on which they operate. Data for the business model analysis is gathered in the empirical part of this paper, where 15 pioneer Finnish companies dealing with open data are examined. These general findings are then reflected from media companies' perspective in order to understand their opportunities in the open data sphere. This top-down approach ensures that open data business opportunities are understood objectively and holistically, because the phenomenon is first studied in its own right without an interference from industry specific limitations. Although the conclusions focus on media companies, the general findings of the empirical part should benefit other industries as well.

1.2 Open data as a research topic

Open data is a popular topic at the moment, and a lot of hype and expectations have been built around it. The increasing attention towards open data can be witnessed, for example, by examining search activity in Google Trends. The amount of weekly "open data" searches in Google search engine is plotted over the past 8 years in Figure 2 below. Eight years is the maximum time scale Google Trends provides, and no data before this is available. Number 100 represents the peak search volume and number 0 no searches, respectively.

The graph can be interpreted by overlooking the short-time search volume fluctuation, and focusing on changes in long-term search activity in the yearly average curve (see Figure 2). First, between 2004 and 2007 the search activity has somewhat declined. Then between 2007 and 2010 it has remained quite constant. After 2010 the search activity has increased almost in a linear fashion, and coincidentally, the last week included in the sample (the week

beginning at 23.9.2012) had more search activity than ever before. Overall, the graph illustrates well how the general awareness of open data has evolved over time.

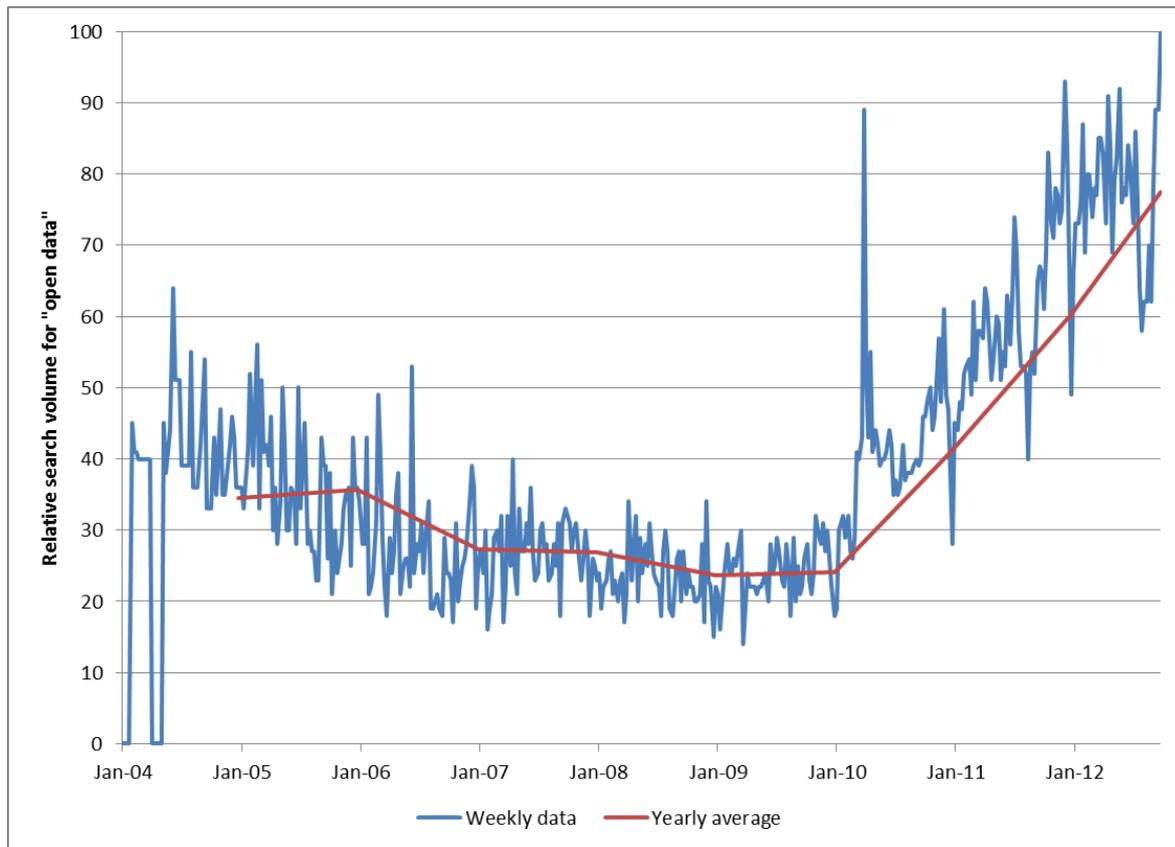


Figure 2 Relative search volumes for “open data” in Google search engine. The number 100 represents the peak search volume (Google Trends)

Academic interest towards open data can be analyzed by examining the annual number of published academic records relating to open data (see Figure 3 below). The data was acquired from Web of Knowledge by counting the amount of annually released documents with “open data” either in their title or abstract. Since no qualitative evaluation was done, this approach is prone to ambiguity because the documents might have utilized the phrase “open data” in another context as well. Regardless of the ambiguity, the graph is still a good general indicator of the awareness and popularity of the topic in academic research.

According to the figure, the term open data has emerged in academic research in 1996 (Figure 3: “open data” in document abstract). Before 1996 there have been only couple mentions, but after the year there have been publications released each year. In 2001 the number of annual documents has slightly increased, but in the year 2006 and after it has started to grow annually. Out of the documents released in 2011, almost 60 covered open data in their abstract, and the trend seems to continue growing.

First documents mentioning open data in their title have been released in 1978, but these old records have used the term in a different context (Figure 3: “open data” in document title). Along the years there have been occasional documents published, but the term has not started to appear constantly until the year 2007. Since then the topic has grown in popularity, and in 2011 20 articles with open data in their title was found from Web of Knowledge.

The rising attention towards open data after the mid-00s correlates also with academic interest towards linked data (Figure 3: “linked data” in document title). Linked data is a specific technical form of open data, which was described and made well-known by Tim Berners-Lee in his *Linked data - Design issues* (2006) article. After the article was published, the linked data-related research has apparently skyrocketed. The rising interest towards linked data might explain some of the attention towards open data and vice versa.

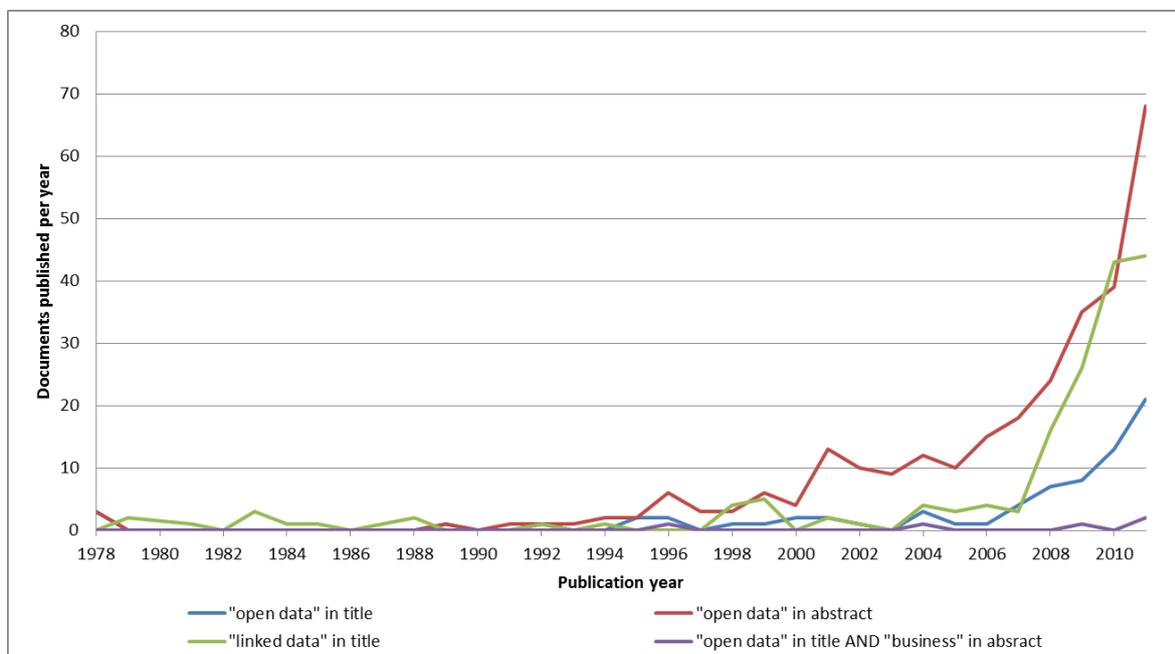


Figure 3 Number of academic documents published per year (Web of Knowledge)

However, when searching records with “open data” in their title and “business” within its abstract (Figure 3: “open data” in document title and “business” in abstract), only 5 documents were found from Web of Knowledge altogether. For some reason, the business research has not caught up with the speed of the open data movement yet. This Thesis aims to start filling the gap of lacking business research of the open data ideology.

1.3 Structure of the study

Chapter 2 conducts a literature study from open data and business model analysis in general. Chapter 3 introduces the research design including research question, research method, and organization of the fieldwork. Chapter 3 also introduces methods on how to analyse quality of the research design. Chapter 4 describes the case studies of the companies under examination, and iteratively analyses their business models. This chapter establishes the evidence on top of which the following analysis is founded on. Chapter 5 sketches the underlying value network in which the companies operate. Chapter 6 answers to the research question by reflecting the findings from the media companies' perspective. Chapter 7 finalizes the Thesis with a discussion of the results and introducing topics for future research.

2. OPEN DATA AND BUSINESS MODELS IN LITERATURE

2.1 Open data

Open data is defined by the Open Knowledge Foundation as being accessible as a whole, free-of-charge or at most with a reasonable reproduction costs, redistributable, reusable, in a data format which does not cause technological obstacles, and without discrimination against persons or groups nor against any particular fields of endeavor (Open Definition). This definition is very popular, and it is utilized e.g. in a landmark Finnish book training to open up data (Poikola, Kola, & Hintikka, 2010).

In addition to availability and licensing issues, there is also a technical dimension in the data openness. Tim Berners-Lee in his (2006) World Wide Web Consortium paper outlines the concept of linked open data. Linked data is constructed to include relations to other linked data, thus forming a mesh of interrelated data. According to Berners-Lee, in order to create linked data; the data should use Hypertext Transfer Protocol (HTTP) Uniform Resource Identifiers (URI) as names for things; it should provide the information in a standardized technical format, such as Resource Description Framework (RDF); and it should contain links to other URIs. These universal references to other linked data make it easier to combine larger sets of data from several different sources together. In 2010 he updated his paper to include a 5-point starring scheme in order to encourage government data officers to open their datasets, and perhaps to even compete with their level of data openness.

2.2 Business model elements

There is a wealth of academic papers written about business models over the decades, but still – or perhaps because of that – there remains some ambiguity around the definition of the business model concept. For example, Afuah (2004, p. 2) defines business model as a framework for making money. In more detail “it is the set of activities which a firm performs, how it performs them, and when it performs them so as to offer its customers benefits they want and to earn profit” (ibid.). Another definition by Chesbrough and Rosenbloom (2002, p. 532) states that “business model provides a coherent framework that takes technological characteristics and potentials as inputs, and converts them through customers and markets into economic outputs”. These are just two of the numerous business model definitions. Despite the several definitions, a consensus that “business model is a conceptual and theoretical layer between strategy and business processes” can be generalized (Rajala and Westerlund, 2007, p. 118).

The business model definition alone, however, does not answer to the question what exactly should be examined when comparing companies in order to understand what differentiates them from competition. Osterwalder in his (2004, p. 43) dissertation outlines nine building blocks under four pillars, which together constitute the business model of a company. Other authors have had similar views, but with different wording (Rajala and Westerlund, 2007). In his (2009) dissertation *Determinants of Business Model Performance in Software Firms*, Rajala proposes five business model elements based on an extensive literature study of prior research. These elements are (see also Figure 4):

- **Offering** is a value proposition that a software firm offers its customers and other stakeholders, and with which it positions itself in the market.
- **Resources** are the assets and capabilities that are needed to develop and implement a given business model. They can be tangible (personnel, equipment, etc.) or intangible (brand name, relationships, etc.). In essence, they are the internal source of advantage, or the core competency of a company.
- **Relationships** are the means to access external resources and capabilities.
- **Revenue model** includes the revenue sources, pricing policy, cost structure, and revenue velocity. It is the firm’s means to capture value out of its offerings.

- **Management mind-set** distinguishes business model as something that stems from the values, emotions, and attitudes of management; instead of cognitive, rational thinking and planning.

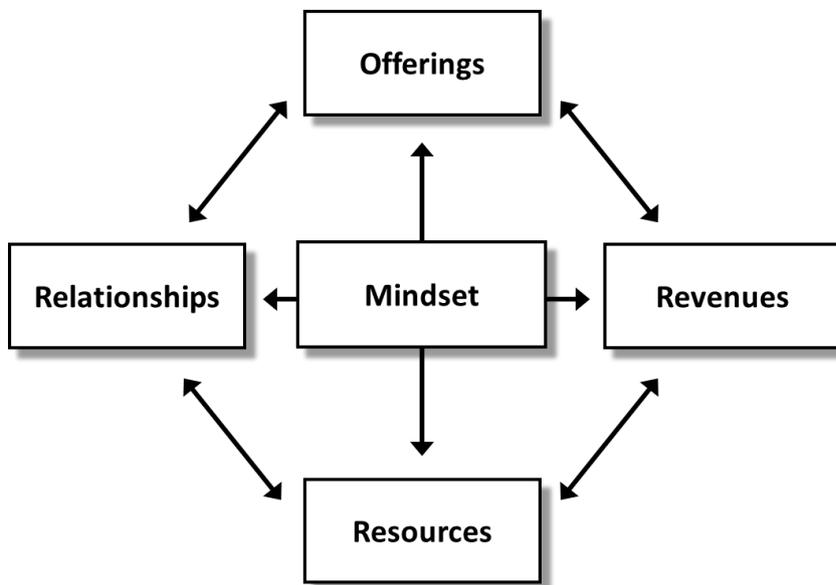


Figure 4 Business model elements as defined by Rajala (2009)

Based on these elements, Rajala (ibid.) defines business model as

“concise representation of how an interrelated set of elements – the offering, relationships, resources, revenue model and management mind-set – are addressed to create and capture value in defined markets“.

Because Rajala’s (ibid.) work is specific to software companies, and because it composes together a broad perspective from several authors, this Thesis will utilize Rajala’s five business model elements to analyse and compare the open data companies.

2.3 Open data business models in research

There is some academic work done also from the business models and value chains relating to open data. Latif et al. in their (2009) conference paper depict a linked data value chain which has four entities: raw data provider, linked data provider, linked data application provider and end-user (Figure 5). Raw data provider publishes raw data, linked data producer utilizes the raw data to produce linked data, and finally the application provider utilizes the linked data to produce a human readable output for human end-users, respectively. These

roles are closely connected by three types of data artifacts: raw data, linked data, and human-readable data. (Latif, Saeed, Hoefler, & Stocker, 2009)

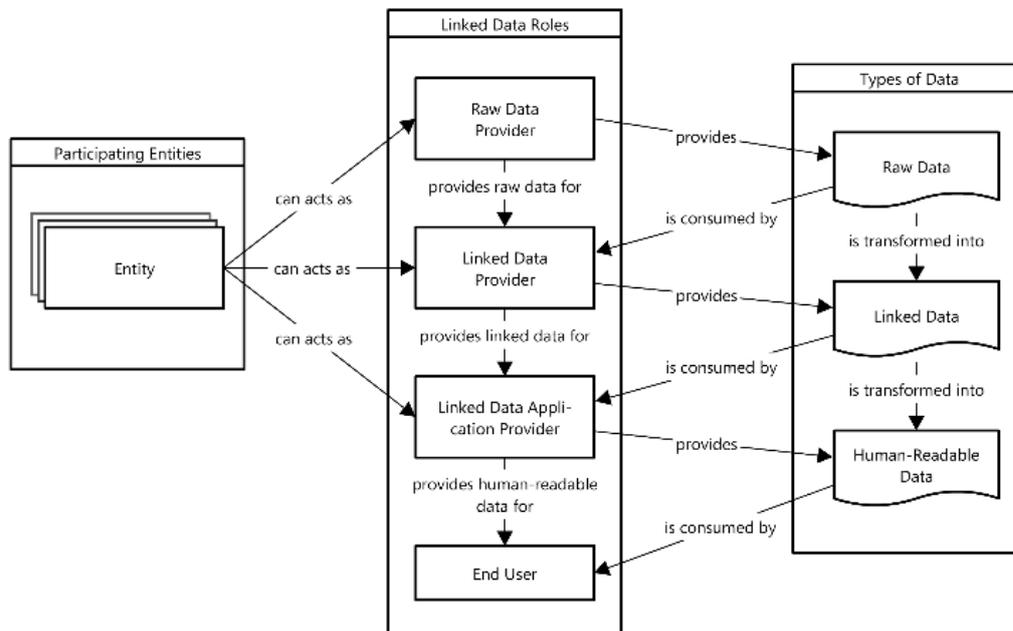


Figure 5 Linked data value chain, reproduced from “*The Linked Data Value Chain: A Lightweight Model for Business Engineers* “ by Latif et al., 2009, p. 3

Tammisto and Lindman (2011) researched how data service providers capture value by conducting an explorative case study interviewing four respondents from three companies. They utilized the roles proposed by Latif et al. (2009) as a foundation, but found that consulting was an additional source of revenue for the open data companies (see Figure 6). Therefore, according to Tammisto and Lindman (2011), the main revenue sources of open data related activities were open data consulting, transforming the data into linked open data, and developing applications on top of the data. The data can be published at any stage of the data development process (see Figure 6). The data development process can also include an additional stage – “data filtering” – that refers to removing the pieces of data that contain private or other sensitive information from the datasets before publishing. (Tammisto and Lindman, 2011)

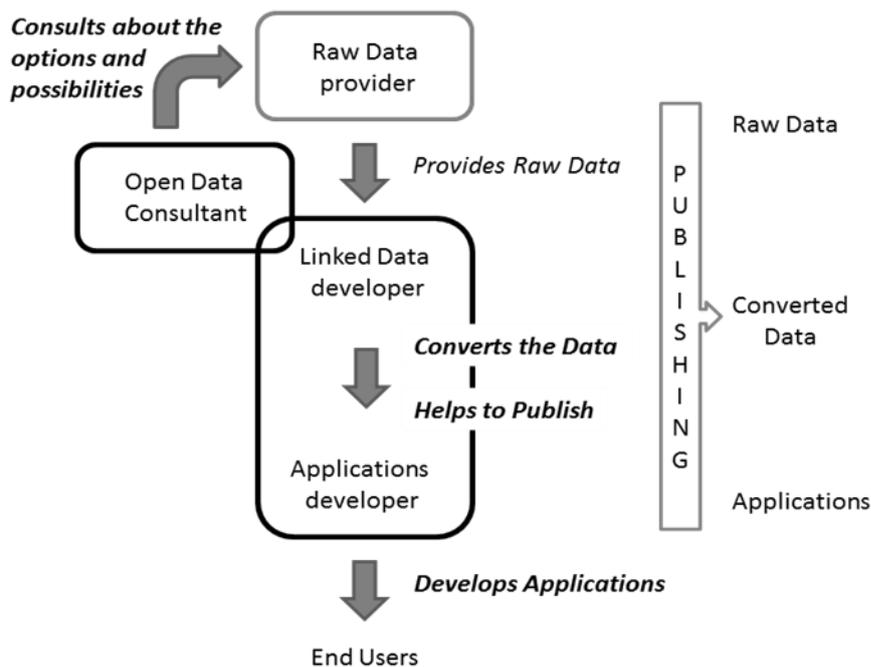


Figure 6 Linked data value chain, reproduced from “Open Data Business Models” by Tammisto and Lindman, 2011, p. 9

Poikola et al. in their (2010) book have a more extensive approach, listing 10 roles in the open data value chain. Seven of these roles are considered from the data publishing perspective, and freely translated from Finnish they are: Data recorder, data refiner, data aggregator, data harmonizer, data updater, data publisher, and registry maintainer. In addition, they see three end-users for the data: Application developer utilizing the data as part of his service; data interpreter utilizing data in his research, commercial, or democratic activities; and finally a human, a company, or an organization as an end-user utilizing these applications or interpretations.

Compared to Tammisto and Lindman (2011), Poikola et al. have used finer grain in their value network representation. In addition to the roles mentioned by Tammisto and Lindman, Poikola mentions also data updater, registry maintainer, data aggregator, data harmonizer, and data interpreter as an end-user. Some of these roles could be seen to be included in the Tammisto and Lindman (ibid.) value network as well, depending on the exact definition. Tammisto and Lindman, on the other hand, pay emphasis on consultancy companies’ role in the value network as an adviser, especially in the phases relating to data publishing.

Lehtonen in (2011) report *Open data in Finland – Public sector perspectives on open data* depicts a process for open data utilization. Lehtonen does not explicitly call this process a value chain, but in essence the meaning is the same. Lehtonen lists data filtering / data

mining, data organizing, data visualization, and data interpretation and production as the four steps in data utilization. This model is in line with the data aggregator, data harmonizer, and data interpreter roles from the model presented by Poikola et al. (2010). Lehtonen only uses different names for these activities.

Kuk and Davies in their (2011) research paper studied the role of agency and artifacts in assembling open data complementarities, where the theory of complementarities suggests that certain activities, when brought together, are more than the sum of their parts. The research was conducted by examining hack-day events and their participants' motivations. One of their results is the assemblage of open data complementarities (see Figure 7), where the resulting artifacts create a recursively independent artifact stack. The phases in this artifact stack are cleaning of data, making the data linkable, writing software to analyse or visualise the data, sharing the source code of the software in a revision control system such as github, and finally letting other developers to innovate new services on top of the source code. Kuk and Davies argue that the open data utilization process has similarities to the one witnessed in open source projects, but they see some differences as well, especially in the licensing of the outputs. Whereas open source projects focus on openly licensed output (the source code), open data hacking focuses on openly licensed input (the data). The output of open data processing, however, does not necessarily need to be openly licensed. (Kuk & Davies, 2011)

The difference with the approach taken by Kuk and Davies to the previously mentioned scholars is the emphasis on the uncoordinated co-operation between the agents involved in the process. This co-operation is organized around the intermediary artifacts (cleaned data, linkable data, source code, shared source code, and service technologies), each of which reinforces the value of the previous artifact. (Kuk & Davies, 2011)

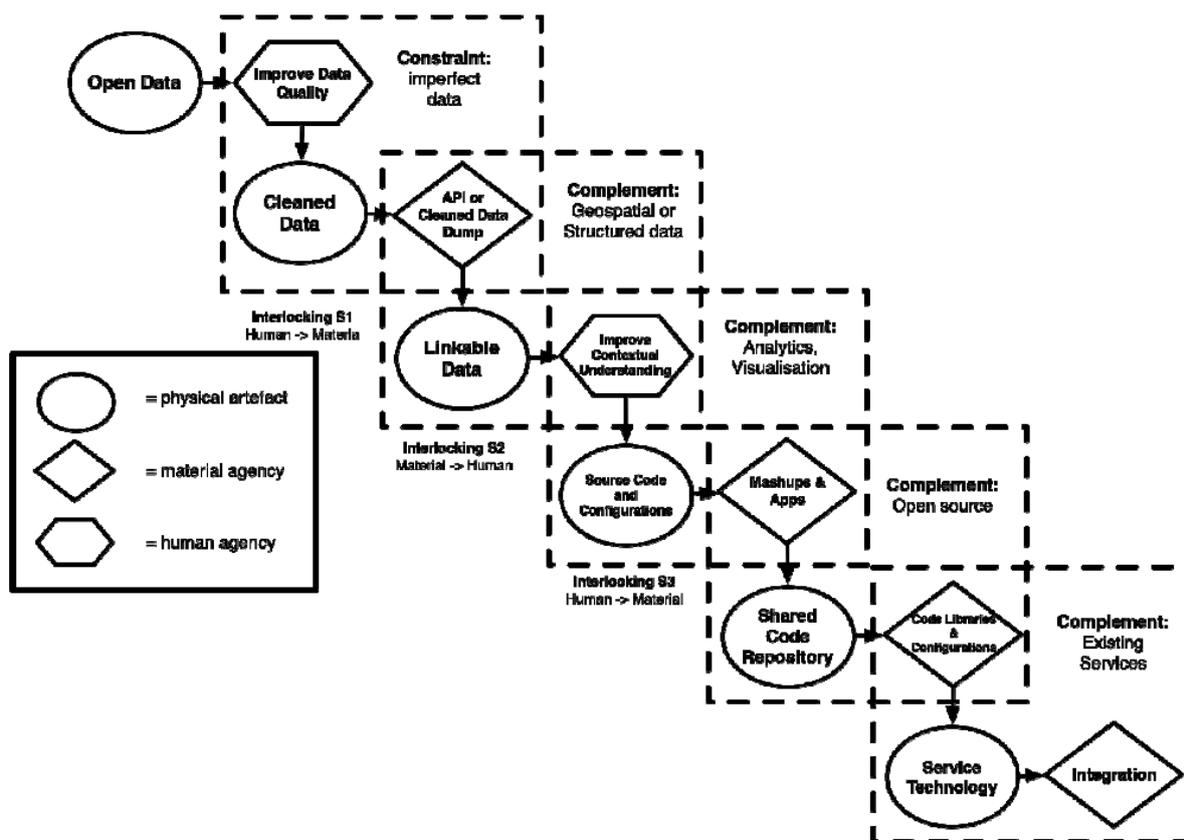


Figure 7 Assemblage of open data complementarities, reproduced from “The Roles of Agency and Artifacts in Assembling Open Data Complementarities” by Kuk and Davies, 2011, p. 11

Some minor definitional differences and naming conventions aside, the value networks proposed by Latif et al. (2009), Poikola et al. (2010), Tammisto and Lindman (2011), Lehtonen (2011), and Kuk and Davies (2011) include similar roles. Poikola et al. (2010) have described their value network using a greater technical detail, and thus have more elements in their value chain. Tammisto and Lindman (2011) mentioned the important role of consultancy within the value network, especially for the open data publishers. Kuk and Davies (2011) proposed that the value chain elements (complementarities in their language) are recursively independent and coordinated with help of the intermediary artifacts (see Figure 7). Compared to the value network established by the previous scholars, Kuk and Davies (ibid.) add the activity of sharing the source code in a public repository, and letting other developers innovate on top of it.

This research will utilize the value network profiles established by these authors as a foundation when interpreting the results from the interviews.

2.4 Open data definition in this paper

The open data definition introduced in Chapter 2.1 is problematic from the perspective of this paper, because it would outline some case companies out of scope. For example, some of the case companies scraped data from a website without an explicit legal permission from the data owner. Many companies seamlessly mashed up commercial and open data together from several sources in order to create a better user experience. Strictly speaking, these companies would not be dealing with open data.

However, in some instances this “hacky” usage of the data has helped the data owner to see the potential of its data re-usage, and ultimately steered the data owners to alter their position on data re-usage permissions and even to create application programming interfaces (APIs) to let developers access the data easier. This do-it-yourself or even hacker-type of activism is very common in the open data community, and since it also has business consequences, it should be included in this study (Kuk & Davies, 2011). Therefore, despite the contradiction with (Open Definition), this study considers a broader definition for the open data.

Thus, in this Thesis open data is defined as:

Data, which is accessible through Internet in a machine readable format. It does not necessary have to be completely free of charge or free or licenses, but it should allow experimenting with the data, and even running a small-scale-business without restrictions. Technically the data can be in a linked- or in any other machine readable format.

Machine readable, in this context, means any format, which is readable by a computer. This includes, for example, comma-separated values (.csv), Excel spread sheet (.xls), or even PC-axis (.px) formats. In addition, all websites and text documents are considered machine readable as well. However, a scanned paper document (.pdf) or any image are not machine readable, because a computer can only show these files, but cannot easily make sense of its contents.

This definition is a bit different from the one adapted by Poikola et al. (2010), but it is very useful in the context of this paper since business consequences of open data are being studied.

2.5 Media definition in this paper

Oxford English Dictionary defines media as “the main means of mass communication, esp. newspapers, radio, and television, regarded collectively; the reporters, journalists, etc., working for organizations engaged in such communication” (Oxford English Dictionary).

According to Denis McQuail (2010, p. 4), mass media and mass communication were coined in the early twentieth century to describe what was then a new social phenomena of communicating to many in a short space of time from a distance. The early forms of mass media (newspapers, magazines, phonogram, cinema, and radio) are still largely recognizable today, only the scale has increased and more diversification has emerged. (McQuail, 2010)

However, mass media is no longer the only means in society-wide mass communication; they have been supplemented by a new type of media. McQuail recognizes that Internet and mobile technologies have establishes an alternative network of mass communication. McQuail describes this new media being more extensive, less structured, often interactive as well as private and individualized. (McQuail, 2010)

This document recognizes media as all channels for mass communication, that is, traditional mass media and new media. Media company is therefore any organization involved in mass communication through these channels, be it television, radio, newspaper, magazine, social media, outdoor media, etc.

2.6 Economics of the media business

According to Picard (1989, as cited in Albarran, 2010) a unique aspect of the media industry is the two separate but interrelated markets that need to be catered: Audience and advertisers. Albarran (2010) refers this as dual product market, but a more commonly used term would be two-sided market (Rochet & Tirole, 2003; Parker & Van Alstyne, 2005). The two-sided market is affecting the revenue models of most media industries. For example, local broadcast TV, radio, newspapers and magazines all have advertising as their primary income (Albarran, 2010, p. 42).

Another peculiar aspect of the media industry is strong cross-elasticity of demand. Cross-elasticity simply means consumers’ tendency to settle for a comparable substitute product in case their primary choice is not available. For example, if we drive to local movie theatre to find out that the film we want to see is full, we can either drive back to home or to find another activity. Many people will buy tickets to another movie, since they already are in the

theatre. Since there are also many choices for different media content, the cross-elasticity of demand leads in to fragmentation of the audience in to smaller and smaller segments. The problem is further magnified with digital platforms, making it even easier for the audience to choose when and how they consume the content. (Albarran, 2010, p. 41)

Third defining aspect of the media business is the newly emerged multi-platform media enterprises. The modern horizontally and vertically integrated media enterprises are no longer restricted to one distribution channel, but instead they deliver the same content through an array of new platforms including Internet, Video on demand, mobile platforms, and social media sites. This transition has been driven by a change in audiences' behaviours including adoption of new technology and demand for cross-platform services. (Albarran, 2010, pp. 69-83)

These multi-platform media enterprises utilize similar revenue models as the traditional media: advertiser-supported, subscriptions, or pay-per-use. In the advertiser-supported model the content is usually free, but advertisements are placed within the content (Albarran, 2010). This is a two-sided markets business model, where free content is subsidized with advertisements (Andersson, 2009). Subscription based model has been recently applied to Wall Street Journal Online and The New York Times online, and it will be employed to Helsingin Sanomat online content as well (Albarran, 2010; The New York Times Company, 2011; Helsingin Sanomat, 2012). In the pay-per-use model the user pays only for the obtained content, such archival content of newspapers or magazines (Albarran, 2010).

In order to ensure cross-media content for the consumers, some media companies have made strategic alliances with internet ventures including web portals, niche websites and Internet service providers. In addition, alliances with technology partners such as On2 Technologies or Akamai Technologies Inc. have become more popular to provide services in broadband video management and online media content syndication and distribution. (Albarran, 2010, pp. 69-83)

2.7 Open data and media business

Media companies usually relate open data in to data journalism. Data journalism utilizes public information sources in enhancing articles and even creating new article ideas. Data journalism is said to be a “new camera” for the journalists (McCandless, 2012 p. 4) or “equipping yourself with the tools to analyse it [data] and pick up what’s interesting”

(Berners-Lee, 2012 p. 6). Data journalism is another way to scrutinize the world, and it is becoming more and more important as the amount of available data has surged (O'Murchu, 2012, p. 10).

Aitamurto, Sirkkunen and Lehtonen in their *Trends in data journalism* (2011) report state, that reporters at US daily newspapers routinely turn to local, state and federal government websites to hunt for data that they can use in their stories. The journalists see data journalism as a way to find hidden stories and to increase transparency in the journalistic process. Aitamurto et al. say that news organizations are searching for sustainable business models to support data journalism. Many have visions of becoming a number-one data store.

Lehtonen in her (2011) report *Open data in Finland – Public sector perspectives on open data* sees that the role of media in the open data ecosystem is to work as a mediator. Media was seen to gather and filter diverse information, and then winnow out the parts serving the needs of public. The benefit of open data was seen on the one hand to provide better and more reliable stories; and on the other hand to improve transparency in journalism, administration and decision making. In addition, Lehtonen also proposes that media could open its own data for wider re-use.

The idea of media as a data publisher was taken onwards in the report by Aitamurto et al. (2011). They describe a data hub model (Figure 8), originally presented by Kayser-Bril (2011), where the media house collects data from different sources and makes them accessible to outside end-users, developers and organizations interested in data. The data should be open for re-use through application programming interfaces. According to an article by Lorenz, Kayser-Bril, and McGhee (2011), by becoming this hub of data, media companies would turn themselves into a center of trusted data, able to do complex analysis. Lorenz et al. propose that instead of “attention market”, media should think themselves to be at “trust market”.

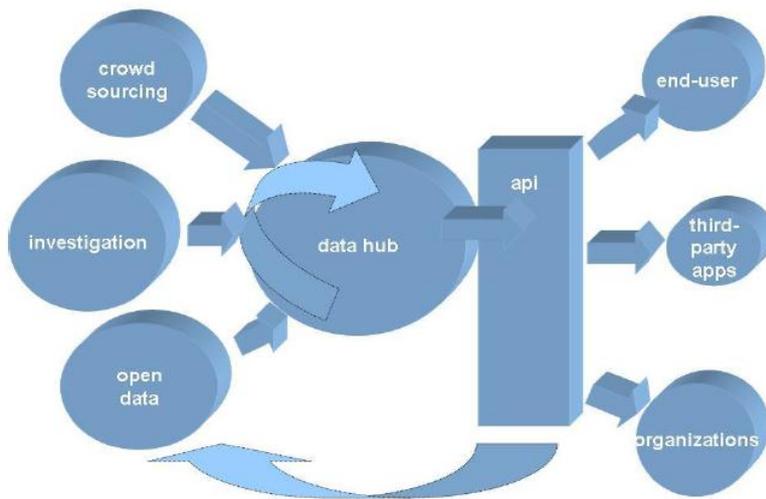


Figure 8 Data hub model adopted from Aitamurto et al. (2011), originally presented by Kayser-Bril (2011)

3. RESEARCH DESIGN

Research design is in a way a blueprint of the research to be done. It deals with four problems: what questions to study, what data is relevant, what data to collect, and how to analyse the results (Philliber, Schwab, & Samsloss, 1980, as cited in Yin 2003).

The chapter starts by introducing a general philosophical worldview on which the entire research is leaning on. Then the chapter continues to define the research question, purpose of the research, research method, and organization of the empirical fieldwork. Finally, the chapter establishes a criterion for evaluating the quality of the research design.

3.1 Philosophical worldview

Philosophical worldview, as explained by John W. Creswell in his (2009) book *Research design*, is the set of philosophical ideas which are often hidden in research, but still influence its results. Some call them epistemologies and ontologies, others call them research paradigms or research methodologies (Creswell, 2009, p.6). Philosophical worldview represents a distillation of what we think about the world, but cannot prove (Lincoln, & Guba, 1985).

This Thesis is founded on social constructivism philosophical worldview. The constructivist philosophy assumes that individuals seek understanding of the world in which they live and

work. Creswell explains (2009 p.8), that the more open-ended the questioning, the better, because then the researcher listens to what people say or do in their life setting. Creswell says that the researcher's intent is to make sense of (or interpret) the meanings other have about the world. Rather than starting with a theory, inquirers generate or inductively develop a theory. The benefit of a constructivist school of thought is the ability to describe and learn from a real world social phenomenon, without the need to exhaustively understand and model it. (Creswell, 2009).

In general, the social constructivism belongs to postpositivistic worldview and admits that, especially in social sciences, there are (1) no ultimate truths to be found and that the (2) beliefs of the inquirer always affect to the end-results. Sometimes these two aspects are intertwined as well; two inquirers with different social and cultural backgrounds observing the same social phenomena might end up to two different conclusions.

This bias might be caused by prejudice of the inquirer towards his research topic, and it should be taken into account when designing the research. Since this Thesis is written by one individual, the risk of biased observations and conclusions is high. Chapter 3.6 will introduce measures that were taken in order to reduce bias in the research results.

3.2 Research question

The research question of the Thesis is “how media companies can utilize open data in their business”. The research question is tackled by first studying Finnish open data companies in general, and then relating these findings to the context of media companies. Thus, first two sub-questions need to answered:

- 1) *What business models are the open data companies in Finland utilizing?*
- 2) *How have the companies utilized the benefits of open data in their business?*

In other words, in order to understand the effects of open data phenomenon for the media industry, the phenomena will be first explored from an overall perspective and the media viewpoint will be considered after the overall analysis. This wide-to-narrow approach ensures that wide variety of business models are considered, and then either applied or discarded depending on their applicability to media business.

An alternative research strategy would have been to start from the media's current business practices, and then applying open data thinking within those premises. This research strategy,

however, is rejected because it might constrain thinking and ignore some more radical business opportunities. Therefore, the wide-to-narrow explorative strategy is utilized instead.

3.3 Proposition and purpose of the research

Since the research is exploratory in nature, it does not carry any particular hypothesis or proposition with it. Instead, the Thesis constructively examines what effects the availability of open data has had to Finnish companies, and tries to report the results as objectively as possible. The purpose of the Thesis is therefore to first explore the business effects of the open data phenomenon in general, and then make conclusions from the more specific media viewpoint.

The only hypothesis, which can be thought of, is that open data has had at least *some* effect in the businesses of the companies utilizing it. What exactly this effect has been will be examined during the fieldwork, and to support objectivity, no hypothesis or propositions of it will be presented beforehand.

Some questions, which still need to be answered before a case study can be conducted, are a) how to define the case being studied, b) how to determine the relevant data to be collected, and c) what should be done with the data once collected. (Yin, 2003)

3.4 Research strategy

Since the research questions are “what” and “how” -type, they are qualitative in nature. They could be studied with a number of research methods, including experiment, survey, archival analysis, history, or case study. Because the open data phenomenon is contemporary, only experiment, survey and case study can be considered. Experiment would require control of the events, which is not feasible for this study, leaving either survey or case study as proper research methods. (Yin, 2003)

The research strategy selected for this Thesis is exploratory multiple case study. In his (2003) book, *Case study research*, Robert K. Yin explains that case studies are particularly useful when studying contemporary phenomena, when the investigator has little control over the events and when “how” or “why” questions are being posed. According to Yin, case studies are useful in that they offer direct observations of the events being studied, and also interviews of the persons involved in the events.

The unit of analysis in this research needs to be considered very carefully. The research question is to find the implications of open data for the media companies' businesses. Therefore, one might hastily presume that the unit of analysis would be a media company or a group of media companies, but this is not the case. Since the paper wishes to learn from other companies and their affiliation with open data, the unit of analysis in this study should instead be Finnish companies that are somehow utilizing open data in their businesses.

Because the unit of analysis includes many companies, the next question is whether to do an embedded single-case design or a holistic multiple-case design (Yin, 2003, p. 40). If the single-case design is utilized, "open data in Finland" should be put as the unit of analysis and different companies as sub-units. This design, however, would put the emphasis on the open data phenomena in general, instead of the individual companies and their benefits from open data. Therefore, a holistic multiple case study analysing the business models of a group of Finnish open data companies fits better to the objectives of this research.

The multiple-case design, according to Yin (*ibid.*), is an iterative process that develops an initial theory, conducts the first case study and writes an individual report out of it; then draws conclusions, modifies the theory, and conducts a second case study et cetera. This is called literal replication. It should not be confused with survey's sampling logic, in which a researcher tries to establish a certain level of confidence by choosing a representative sample and asking exactly the same questions from all the informants. Literal replication, on the contrary, does not aim for a certain level of confidence, but instead iteratively enhances the principal theory until the theory seems to explain the phenomena under study. The difference to the sampling logic is that a much smaller amount of evidence can be used, and yet theories that are more complicated can be confirmed.

The research strategy is visualized below in Figure 9. On the left-hand side are the case companies which utilize open data in their current business, and whose resources, relationships, management mind-set, offering and revenue model will be studied empirically. On the right hand side in Figure 9 (dotted line with a question mark), the research question of how media companies could utilize the open data resources in their line of business can be answered based on the empirical results.

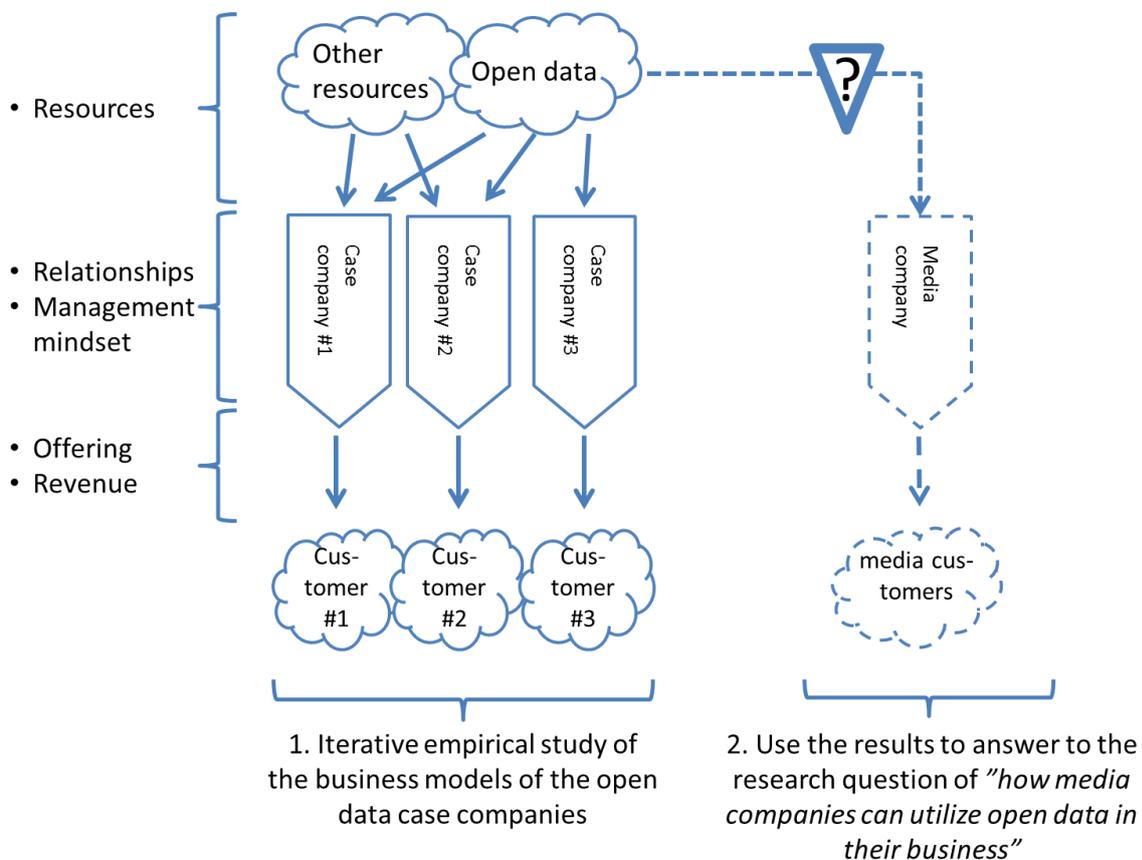


Figure 9 Research strategy

3.5 Organizing the research fieldwork

The research strategy states that case companies' business models should be examined in order to build an open data business model framework, which can then be extended in to media companies' businesses. But, on which basis the case companies should be selected? The optimal solution would have been to find media companies already utilizing open data as case companies, in order to make externalization of the results more compatible. However, there is very little usage of open data in the media industry at the moment. In addition, if the examination would be constrained only to existing media companies, the body of findings would be deliberately limited, possibly leaving some important open data business aspects out of scope. Therefore, no limitations for the industries of the case companies were placed.

3.5.1 Selecting the companies

To find and select the case companies, Apps 4 Finland competition was taken as a starting point. Apps 4 Finland (A4F) is an application contest run by Forum Virium, and it aims to encourage developers of any background to create new applications for open data. The contest has been orchestrated three times, and the amount of submissions has increased annually. The first competition in 2009 received 23 submissions, whereas in 2011 already 140 contestants were registered. The 2011 competition had four categories for the submissions: Visualization, Data opening, Application, and Concept.

There were too many submissions to analyse everything, so finding developers who have continued the development of their idea and founded a company around it was the priority. When going through the submissions the concept-category was omitted, because it contained only idea-level submissions and not actual working applications or visualizations. To distil the submissions even further, only the ones which had founded a business around their idea were picked.

In the A4F contest years 2009, 2010 and 2011 there were altogether 193 submissions posted. Out of these 193 submissions 29 had continued the development after the contest. From these, 17 had any business activity. Six of these 17 works were developed by Flo Apps Oy and two by Hahmota Oy, thus giving 11 separate companies. These 11 companies are the main focus of this study, and they were contacted for an interview. Three companies did not answer and one refused from an interview inquiry, thus seven companies were interviewed out of the 193 original submissions. These seven companies met the initial requirement of continued development and business intentions. See Figure 10 for an overview of the screening process.



Figure 10 Apps 4 Finland submission screening process

To make sure the research is on the right path and interviewing the right people, a snowball sampling technique was employed. In the snowball sampling, each interviewee was asked who else should be interviewed as well. In addition, Ville Meloni, an open data expert and

one of the organizers of the A4F contest from ForumVirium, was interviewed and asked for guidance in selecting the right interview subjects. The snowball sampling technique and Ville Meloni gave altogether eight additional interview subjects, thus increasing the total amount of companies to 15. The full list of the analysed companies, contact persons and interview dates is summarized below in Table 1. The table indicates also, where the interview lead was acquired: Apps 4 Finland contest and year, Ville Meloni, or snowball sampling.

Please note that with Helsingin Sanomat no interview was conducted, because enough information was available through their website.

Table 1 List of analysed companies

#	COMPANY	CONTACT PERSON	INTERVIEW DATE	A4F SUBMISSION	FROM
1	-	Leo Lahti, Co-founder; Juuso Parkkinen, Co-founder	9.3.2012	soRvi- avoimen datan työkalupakki R-kielille	Snowball sampling
2	Hahmota Oy	Peter Tattersall, CEO	29.3.2012	Suomen kansantaloudellinen Elämänpuu	A4F 2011
3	Fresh Bits	Pasi Kolkkala, Software developer	30.3.2012	Reitit for iPhone	A4F 2011
4	Gemilo Oy	Arto Liukkonen, Social network developer	30.3.2012	Hilmappi.fi - Implementation read more	A4F 2009
5	Skyhood Oy	Thomas Grönholm, CEO	3.4.2012	Duunitori.fi	A4F 2010
6	Forum Innovations Oy	Jaakko Hilke	4.4.2012	Pikkuparlamentti.fi	A4F 2010
7	Suomen Turvaprojektit Oy	Panu Häikiö, CEO	23.4.2012	Mitä mukaan lennolle - hakupalvelu	A4F 2011
8	Cloud’N’Sci Ltd	Pauli Misikangas, CEO	24.4.2012	-	Ville Meloni
9	Essentia Solutions Oy	Markus Halttunen, CEO	24.4.2012	-	Snowball sampling
10	KAMU Ry	Juha Yrjölä, Chairman of the association	25.5.2012	Kansanmuisti.fi	Snowball sampling
11	Logica	Jukka Ahtikari, Development Director	4.5.2012	-	Ville Meloni
12	Flo Apps Ltd	Tapio Nurminen, CEO	22.5.2012	Several	A4F 2011
13	Suomen Asiakastieto Oy	Heikki Koivula, Director	22.5.2012	-	Ville Meloni
14	Helsingin Seudun Liikenne	Jari Honkonen, Project manager	28.5.2012	-	Snowball sampling
15	Helsingin Sanomat	-	-	-	Snowball sampling

3.5.2 Interview process and questions

The purpose of the interviews was to examine on with what business model each company is operating. As discussed in Section 2.2, according to Rajala (2009) the business model can be broken down to five interlinked elements. These elements are the offering, the resources, the relationships, the revenue model, and the management mind-set. Thus, the interview questions were written to reflect these five aspects of the business model. A full list of interview questions can be found in the Appendix A.

The interview technique was an open-ended interview. The purpose of an open-ended interview is to avoid unintentionally guiding the interviewee into a predefined conclusion. The questions should be made general enough to leave room for unexpected answers. The interview questions were used more as a backbone to guide the conversation, as opposed to a question-quoted answer type of dialogue. The questions also required some adaptation to each particular case company, because the companies varied quite a lot in size. The smallest companies were one-man endeavours and the largest a 4.5 billion euro global corporation.

Most of the interviews were recorded with a portable recorder for a further reference. During the interviews, written notes were also taken to cover the most important answers. In addition, an interview diary was kept during the entire interview process to shelter more general thoughts from each interview. After the interview process, the recorded interviews were lettered. The interviews were conducted between 9th of March and 28th of May 2012. Most interviews were done face-to-face, but two had to be done over a Skype-call and one over e-mail due to logistical problems. Each interview took between one to one and half hour of time, except the e-mail interview in which only key questions were sent and answered, respectively.

Several peer evaluations were employed in order to ensure that the phrasing of the questions did not highlight or lead to any particular preconceptions, and also that all the relevant questions regarding the business model framework are being asked. In practise, two professionals: Auli Harju, researcher in Tampere research Centre for Journalism, Media and Communication COMET; and Juho Lindman, assistant professor from HANKEN School of Economics, Helsinki; went through the questions. In addition, the questions were exposed to a peer review at Sanoma Oy and Aalto University School of Business.

3.6 Estimating the quality of the research design

This chapter has so far presented the general philosophical worldview, the research question, the research method, and also the organization of the research fieldwork. However, a convincing research design should also evaluate itself against an established design criterion. According to (Kidder & Judd, 1986, as cited in Yin 2003), four tests have been commonly used to ensure the quality of an empirical social research:

- 1) **Construct validity**, which establishes correct operational measures for the concepts being studied. It ensures that the case study reflects important real phenomena, and not only investigator's impressions.
- 2) **Internal validity**, which establishes causal relationships. However, it relates only to explanatory studies, and since this Thesis is exploratory, evaluation of internal validity is omitted in this paper.
- 3) **External validity**, which establishes the domain in to which a study's findings can be generalized.
- 4) **Reliability**, which ensures that the operations of a study can be repeated with the same results.

Yin gives in his (2003) book several tactics on how these four criteria can be fulfilled while doing case studies. These tactics; along with the test they relate to, and the phase of research in which they occur; are summarized in Table 2 below. The table is a re-presentation of a table in Yin (2003) p. 34, but the original source is COSMOS Corporation. The last column of the table represents the status of the test tactic in this Thesis. Since this is an exploratory research, the internal validity test is omitted, but the remaining three tests and their fulfilment is described in the following chapters.

Table 2 Case study tactics for the four design tests (COSMOS Corporation, as cited in Yin, 2003)

Test	Case Study Tactic	Phase of research in which tactic occurs	Status in this thesis
Construct validity	• Use multiple sources of evidence	Data collection	OK
	• Establish chain of evidence	Data collection	OK
	• Have key informants review draft case study report	Composition	OK
Internal validity	• Do pattern matching	Data analysis	Omitted
	• Do explanation-building	Data analysis	Omitted
	• Address rival explanations	Data analysis	Omitted
	• Use logic models	Data analysis	Omitted
External validity	• Use theory in single-case studies	Research design	N/A
	• Use replication logic in multiple-case studies	Research design	OK
Reliability	• Use case study protocol	Data collection	OK
	• Develop case study database	Data collection	OK

3.6.1 Construct validity

Construct validity is ensured by using multiple sources of evidence, by establishing a traceable chain of evidence, and by reviewing the draft report with the informants (Yin, 2003). These three steps will be explained in more detail below.

(1) Multiple sources of evidence

Yin in his (2003) book lists six possible sources of evidence which can be used in a case study. These sources, along with their existence in this Thesis, are presented in Table 3 below. The sources of evidence are listed in the same order as they were presented in (Yin, 2003), but this order does not correspond to the importance of the evidence. From Table 3 can be seen that in this Thesis the construct validity has been enhanced by utilizing multiple sources of evidence.

Table 3 Different sources of evidence according to Yin (2003, pp. 85-96), and summary of the evidence utilized in this Thesis

Source of evidence	Utilization in this thesis
1) Documentation , such as letters, memorandums, agendas, announcements, etc.	Company websites, such as news, event announcements and press releases were utilized.
2) Archival records , such as computer files and records, maps, charts, lists of names, survey data etc.	A list of all Apps 4 Finland submissions between years 2009 and 2011 was scraped from the contest's website. This list included a detailed description of the application, as well as, the contact information.
3) Interviews , according to Yin (2003), are one of the most important sources of case study evidence. The interviews should be guided conversations, instead of structured queries.	Interviews are the main source of information in this Thesis. They were conducted as open-ended interviews between March and May, 2012. Most of the interviews were recorded and lettered.
4) Direct observations , such as field visit to the case study site. The observations can range from formal to casual data collection activity.	Direct observations were made during the interview visits in the company premises. These observations were written down to an interview diary after each interview.
5) Participant observations , where the observant is not passive, but assumes certain roles within a case study situation and participates to the events being studied. For example, having casual social interaction with residents of a neighbourhood under study.	I participated to several open data community events, such as summits and unofficial sauna evenings, to get a feeling of the general atmosphere of the scene, and to make new connections.
6) Physical artefacts , such as technological devices, tools or instruments, works of art, or some other physical evidence.	Most of the open data software applications, which were submitted to Apps 4 Finland competition, were tested to get acquainted to the open data phenomena and its possibilities.

(2) Establish chain of evidence

Yin (2003) also recommends that case researchers should establish a chain of evidence, so that the reader can follow and understand the reasoning process from the conclusions back to the initial evidence, or from the initial evidence to up to the conclusions. In other words, the

reader must be able to validate whether the argumentation of the findings is based on real evidence, and that no evidence has been lost in the process.

The principal research process is explained in Figure 11 below. In short, it consists of evidence gathering, analysis of the evidence, conclusions for the media industry, final discussion, and two feedback rounds to increase reliability of the evidence and its analysis. In order to make the boundaries of these research steps more transparent, cross-references have been included whenever an important link between the chapters is being made. This will hopefully make the reasoning and thought process easier to follow.

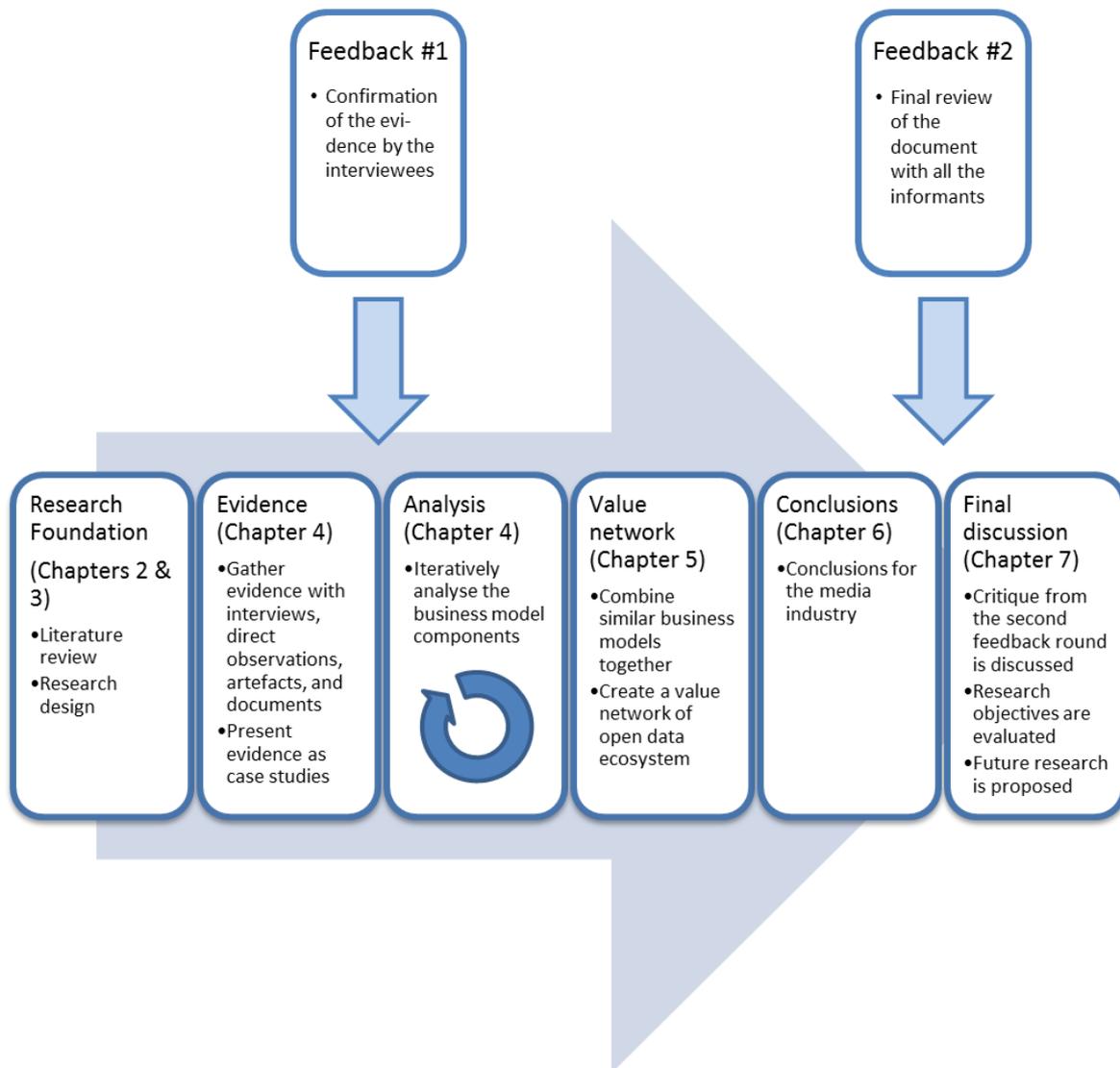


Figure 11 Research process

Since the interviews form bedrock for the entire research, a lot of emphasis was put in selecting which companies should be interviewed in the first place. The aim was to minimize biasedness in the selection process by systematically picking a representative group of

Finnish open data companies. As is described in more detail in Chapter 3.5.1, this was achieved with the participation list of A4F competition, snowball sampling, and expert insight.

The biggest challenge with the interviews, however, is how to present the interview evidence for the reader. This is especially important, because most of the subsequent analysis is founded on top of the interviews. Most of the interviews were recorded with a portable recorder and lettered. In addition, an interview diary including the date, place, summary of discussion, and some initial thoughts was kept (see Chapter 3.5.2 for details). This information, however, is lengthy and in Finnish and therefore it would be difficult to include it as such in this Thesis. In addition, the informants were promised that direct quotes would not be published without their permission.

To solve this problem; and to convince the reader that the case studies presented in Chapter 4 are authentic, and in line with the reality of the companies; a feedback round #1 was organized (see Figure 11). In the feedback round the interviewees reviewed the evidence gathered from their company. In practise, they were sent the part of Chapter 4 relating to their business by e-mail. Most of the interviewees replied in May or June, and some amendments and changes to the case description were made based on these comments. This feedback process validates the evidence presented in Chapter 4, and so it can be utilized in analysis and conclusions made in the subsequent chapters.

(3) Key informants review draft case study report

When the Thesis was in its final stages in November 2012, it was sent for all the interviewees via e-mail for another feedback round. Most of the interviewees (13 out of 14) responded and gave additional feedback in this stage. Comments from the second feedback round are collected and discussed in Chapter 6.5. The motivation for the second feedback round was to give the informants an opportunity to review and comment the results of the research holistically.

3.6.2 External validity

External validity establishes the domain in to which a study's findings can be generalized (Yin, 2003, p. 34). This study leans on multiple case replication logic. By individually studying the 15 companies listed in Table 1, the research tries to objectively map the current business practises in Finland relating to open data. Even though case methodology in general

does not utilize sampling logic, the case companies included in this study most likely represent a very good cross-section of the open data business practises in Finland at the moment of research.

There are, however, some limitations in generalizing these results. First, the research focused solely on Finnish companies, thus completely overlooking the global open data phenomena. The decision to focus on Finland was done mostly because of lack of resources (only one researcher). But then again, focusing on Finland gave an opportunity to conduct personal interviews, which can deliver more deep-rooted results than an overall outlook would have done.

Another limitation is the moment of time the field research took place. At the time of research, open data was a very young phenomenon, but it was evolving at breath-taking speed. Since the companies were examined during spring 2012, this research represents a snapshot of the open data scene at that time. Without doubt, more companies and more business ideas will rise and others will fade as the industry matures. Therefore, when reading this Thesis the reader should keep in mind the era in which the research was conducted.

Third limitation in generalizing the results relates to silent usage of open data. Although the method of selecting the case companies was well structured, there exists a possibility that not all usages of open data were found and examined. This silent usage of open data could occur, for example, internally between companies' siloes or between governmental facilities. This usage would not be visible in the Apps 4 Finland application contests, and even many of the current open data experts might be unaware of it. Yet this silent usage might still hold remarkable business impacts. Thus, in spite of the rigorous attempts to cover the Finnish open data industry in its entirety, there still might be a lot of silent usage which just did not get caught on the windshield of this research.

3.6.3 Reliability

Reliability is ensured by establishing a case study protocol for the overall research, and by maintaining a case study database for the evidence gathered during the research (Yin, 2003).

(1) Case study protocol

Case study protocol is a major way of increasing the research reliability. A case study protocol should have the following sections (Yin, 2003, p. 69):

- Overview of the study project (Chapter 1.3 & Chapter 3)
- Field procedures (Chapter 3.5)
- Case study questions (Chapter 3.5.2 and Appendix A)
- A guide for case study report (Outline presented in Figure 11)

In parenthesis after each item is the part of Thesis where the corresponding section of the case study protocol has been described. Thus, judging on the criteria established by Yin, the case study protocol seems to be in good order in the Thesis.

(2) Case study database

The case study material, which accumulated over the research process, was stored digitally on a personal computer. This includes the interview audio records from most of the interviews, 60 pages of interview lettering, and an interview diary with personal observations and impressions written after each interview. These are organized by each case company, and include also the date and place of interview.

In addition, there is also a length of tabular material stored in the case study database. This includes textual descriptions of the Apps 4 Finland submissions including their contact information, contact dates, and contact persons; and company websites including news, press releases and other related material. All this material together establishes the case study database, which was utilized as source of evidence in the research.

4. CASE STUDIES

In this chapter, 15 case study companies will be introduced and analysed. Each case begins with a company description followed by an analysis of its business model. In each analysis, the new findings are compared to the information established by the previous cases, thus iteratively building business model information. The number of data points would not be sufficient for a quantitative analysis, but according to Yin (2003), case analysis is efficient even with a small number of cases because of its iterative nature. The business models charted in this chapter are analysed further in Chapter 5, where they are linked together and placed within an open data value network.

The material for the 15 cases was collected by interviewing 14 informants from 14 companies and combining the interviews with secondary sources of information. Helsingin Sanomat (HS) Open is the only company case where no interviews were made, and thus it is based

solely on secondary sources of information. HSL Reittiopas case, on the other hand, involves several companies and thus has several informants as well. All the rest of the cases are created based on one informant from the case company in question.

4.1 HSL Reittiopas API

4.1.1 Description

Reittiopas is a popular Finnish service, **offering** point-to-point public transport instructions within the Helsinki-region for over 150 000 daily users (HSL website news). Reittiopas is a free service offered by HSL (Helsinki Regional Transport Authority), which runs the commuter traffic service in the greater Helsinki region. The service is officially available through a web-browser interface with both desktop and mobile instances, but without native mobile applications. According to the respondent from HSL, an application programming interface (API) to the Reittiopas service was built at the same time the service was launched in 2001, but it was not opened for the public until 2009. Before 2009, the API was used internally and in some occasional partnership projects. In addition, it was given for third party developers on request, but according to the respondent from HSL, it did not raise much interest. HSL decided to publish the API in 2009, because at that point the amount of third party requests had risen and the general awareness of open data possibilities had increased as well.

After opening the API, the developers have been very interested about it. According to HSL respondent, in May 2012 over 650 developers have already registered to get access to the APIs. HSL provides other APIs as well, but Reittiopas is the most popular amongst them. HSL is listing over 30 third-party applications utilizing the API in their webpage (HSL Palvelut muissa kanavissa).

The respondent admits that developing and updating similar service offering for this amount of platforms would have been, in practise, an impossible task for HSL to do in-house. The respondent said that in the beginning they did not think so much of cost savings, but were more interested in seeing what new could be achieved. Transparency of governance was also one of the arguments, that is, since the information was generated with taxpayers' money, taxpayers should have also a free access to it. (Project manager, HSL)

One of the most known applications is ReittiGPS by Essentia Solutions Oy, providing a native iPhone application for the journey planner service. The CEO of Essentia Solutions said

the project was started in 2008 to satisfy the founder's personal need of checking the public transportation schedule easier on the road. At that time there was not yet a public API released from HSL, so the information had to be scraped from the HSL website. ReittiGPS and BusWatch were among the first applications to show journey planner information in a native mobile client combined with GPS coordinates from the mobile phone's GPS receiver. (CEO, Essentia Solutions Oy; Project manager, HSL)

The popularity and success of ReittiGPS was a strong indicator for HSL, that it might be worthwhile to release the API for the third-party developers. Quickly after HSL released the API, other similar applications started to emerge. Thus, although ReittiGPS sprang up without support from HSL, the official API release lowered the bar and encouraged several developers to create their own version of the mobile journey planner.

These new applications have increased competition and brought innovation to the marketplace. The newcomers forced the incumbent ReittiGPS to implement new features as well. A good example is when Reitit (previously Reitit for iPhone) by Fresh Bits integrated the Helsinki service guide interface in to its application, ReittiGPS had to implement it as well. (CEO, Essentia Solutions; Software developer, Fresh Bits)

The increased competition has even started a price war in the Apple's App Store. In the interview with Essentia Solutions, the CEO said that they had to answer the increased price pressure by dropping ReittiGPS price from 4 € to 3 €. The respondent from Fresh Bits said that they purposely challenged the incumbent ReittiGPS by carefully pricing the client at approximately 2.5 €.

The respondent from HSL said that after the initial release of the API in 2009, they have continuously improved it in order to better answer the needs of the third-party developers. In 2011, HSL organized their own developer challenge, HSL Mobiilikisa, which invited people to innovate new uses for the API. HSL received 63 submissions in their contest, out of which eight were rewarded (HSL Mobiilikisa).

4.1.2 Analysis: Crowd-sourced client development

A company can achieve remarkable savings by, in effect, outsourcing the client development to third-party developers. The core idea is that native mobile application development and updating is very expensive for a company whose core competence is somewhere else. Yet, a well working mobile application generates a lot of value for the company's customers. Thus,

the company can either hire IT-professionals to do the development of the application in-house, source the client from a subcontractor, or, as is presented above, publish the necessary data and interfaces in the internet and let the third-party enthusiasts develop the application.

The benefits of crowd-sourced client development are obviously related to cost savings. In fact, since the only cost is publishing and updating the necessary data and supporting the developers, crowd-sourced clients can appear in unthinkable environments and platforms, where traditional outsourcing would not be feasible. However, initially HSL motivation was not cost savings or user interface crowdsourcing, but only to see what new could be achieved and to open the API since it was produced with taxpayer money.

What were the success factors behind crowd-sourcing Reittiopas user interface? Three observations can be made from the case; (1) importance of releasing an API, (2) personal need and motivation of the developers, and (3) facilitation of the developer community with competitions and support services.

The need for a good mobile user interface for HSL Reittiopas has existed a long time before HSL opened an official API for the developers. First application, ReittiGPS, was initially created without an API, only by scraping the necessary content from HSL website. The fact that someone made a client without an official API indicates that the demand for such clients was incredibly high. However, after opening the API, the amount of Reittiopas mobile clients increased dramatically. Thus, although ReittiGPS was available without the API, the API opening was necessary to ignite development in wider scale.

Another important observation is the personal need and motivation of the developers. Both of the interviewed developers had a strong personal need of the route planner service they had created. Therefore, motivating and getting the developers interested in the API is a key element in successful user interface crowd-sourcing.

Final observation is that competitions seem to increase the awareness and interest towards APIs substantially. Some good examples are HSL Mobiilikisa and Apps 4 Finland competitions. HSL Mobiilikisa was designed to encourage developers create more applications and for them to compete with each other. Since the competition received over 60 submissions, it has been a great success.

4.2 Case Duunitori.fi

4.2.1 Description

Duunitori.fi, by Skyhood Oy, scrapes job openings from the Finnish Government's Employment and Economic Development Office (mol.fi) website and plots them on a map. These visualized job openings are then enriched with data from several resources, including Tilastokeskus, Reittiopas, Yritystele, Great Place to Work, Facebook, etc. The result is an interesting mash-up of data fetched from different sources, **offering** jobseekers a hub to find all the relevant information from the employer with the job opening. Duunitori.fi is a great example of the data hub model Aitamurto et al. re-presented in their (2011) report, albeit it is lacking an API to let developers to re-use the information.

Resources to run and maintain the service are kept minimal; the service runs on virtual servers at external partners' facilities. According to the CEO, most work went in to development and building stage of the service, which took tens of person-months altogether.

Relationships are built with labour unions and municipalities, and Duunitori.fi has some pilot cases on the webpages of these partners. The respondent from the company stated that they are better partner for these organizations, because Duunitori.fi has all the job offerings also from the blue-collar segment.

When asked about the **revenue model**, the CEO remarks that since open data is free, and developing services on top of it is reasonably cheap, it leads to situation that users are not ready to pay for them. How these services could then generate revenue, ponders the CEO of Skyhood. He answers to his own question that it could be by selling the users similarly as Facebook, which exploits its user data as raw material and sells it onwards for advertisers. In a way, Duunitori.fi has a similar subtext; it is bringing job creators and job seekers together. Even if the end users would be reluctant to pay, the businesses are willing to open up their wallet because they want to find new employees. (CEO, Skyhood)

The revenue model of Duunitori.fi is based on advertisements; the more visitors the page attracts, the more advertisement revenues are possible. In addition, the company is doing custom advertisement campaigns with key partners by offering them increased visibility.

4.2.2 Analysis: Create valuable user experience and monetize with advertising (two-sided markets)

The core idea behind the business of Duunitori.fi is to combine different sources of data, both public and commercial, in order to create an eye-catching user experience where the raw data has been enriched and made valuable for the end-user. The revenue model is based on attracting as many users as possible, and then selling the user masses for advertisers and businesses who are hiring employees. This two-sided market business model is typical for internet portals and media companies (Hagiu and Wright, 2011, Rochet and Tirole, 2003, p. 992).

The idea of a two-sided market is to design separable products and under-price one component in order to implement price discrimination in markets with positive network externalities (Parker and Van Alstyne, 2005). In other words, on the right-hand side a company can give away products with prices at or below zero in order to increase the customer base, and then charge the left-hand side for the expenses. The reasoning is that if the markets are coupled, the network externalities on the right-hand market affect also the left-hand market's demand curve. Therefore, by inducing high demand with under-priced products on right-hand market, the demand curve of the coupled left-hand market moves outward. Outward moving demand curve grows the revenue by making it possible to raise the price of a product while increasing the amount of sold products at the same time. (Parker and Van Alstyne, 2005).

4.3 Case Mitamukaanlennolle.fi

4.3.1 Description

Mitamukaanlennolle.fi **offers** a web service where airline passengers can check what items are allowed to take on to the plane either in carry-on luggage or in cargo hold luggage. The passengers are more satisfied when they know beforehand whether certain items, such as medicine, are allowed on board.

The main **resource** of the service is a database consisting of 1600 items and their security information. The database is a combination of open and closed data, which has been gathered from International Air Transport Association Dangerous Goods Regulations (IATA DGR) manuals, International Civil Aviation Organization (ICAO) data, European Union

regulations, and their own information based on several years of security training of airport officials. (CEO, Suomen Turvaprojektit Oy)

Revenue model of the mitamukaanlennolle.fi service consists of advertisements and licensing. In addition to advertisements, other airports, European Union, and IATA have been interested in the service, and Suomen Turvaprojektit has already licensed the service to Norwegian and German airports. They are expecting the licensing revenue to increase when they expand to even more countries in the future. (CEO, Suomen Turvaprojektit Oy)

In addition, the CEO reveals that there are several researches, which show that if the security check at the airport goes smoothly the passenger is more likely to spend money in the tax-free shops before departure. Certainly, if the luggage is packed correctly the security check will be easier and less stressful.

In addition, by informing the passengers on how to pack, mitamukaanlennolle.fi is creating a lot of value for the airport security as well. Less security personnel will be needed, because the amount of unnecessary luggage openings and item confiscation will be reduced. In fact, the CEO has calculated that in Helsinki Airport alone, there are 9000 cigarette lighters confiscated each month from the cargo luggage. Since it takes between 5 and 10 minutes to open a luggage, it will save between 9000 and 18000 hours of work from the security officials annually.

Relationships are used very effectively. The company has outsourced the advertisement selling to a partner, leaving them more time to focus on the core service. In addition, the service itself was developed by a subcontractor, and it is maintained by another subcontractor. Efficient outsourcing has let Suomen Turvaprojektit Oy to focus on updating the item database.

4.3.2 Analysis: Create valuable user experience and monetize with advertising and licensing

The business model of mitamukaanlennolle.fi is very similar as with Duunitori.fi. Suomen Turvaprojektit Oy is essentially combining raw data from several sources and creating a valuable user experience on top of it. At the moment the revenue is gathered from advertisements (two-sided markets), and also by licensing the software and the database for foreign airports. The relationships are used very efficiently, and all non-core elements of the service have been outsourced to external partners.

4.4 Case pikkuparlamentti.fi

4.4.1 Case description

Pikkuparlamentti.fi **offers** an objective and independent web page, which brings citizens together to discuss about topics of their interest. According to the respondent from Forum Innovations Oy, their website offers more quality, independency, and objectivity than other discussion forums. The idea is that when people search information relating to a certain topic they probably have also some insight from the topic, which they can share with others through the forum.

Resource of the site, in addition to the discussion forum itself, is data gathered from Parliament of Finland. At the time of interview, the data was gathered manually by posting a link to a particular decision proposal and letting users to comment and discuss it. The site has been created by the founders themselves, and thus no subcontracting has been used. (Founder, Forum Innovations)

Revenue model of the site was at the time of interview still at a start-up phase, but they had plans on advertisement-based revenues with banners. They had also plans on selling software for municipalities and government bureaus to help them clarify their decision making process. For example, they were participating in a tender from Ministry of Justice, Finland for an online debating module.

Management mind-set was entrepreneurial oriented, as the respondent stated that they had intentions to build a start-up company already at the idea generation phase. Third place in Apps 4 Finland 2010 competition gave the encouragement and starting capital to find the private limited company.

4.4.2 Analysis: Create valuable user experience and monetize with advertising

Pikkuparlamentti.fi has similar business model as the two previous cases; combine open data with other data sources in order to create a valuable user experience and monetize it with advertisements. However, the representative from Forum Innovations said that at the moment of interview their user base was not large enough to attract advertisers, but yet he saw advertising as one of the monetizing options. In addition, they have also plans on licensing and selling their platform for other companies, similarly as with mitamukaanlennolle.fi has done.

4.5 Case ReittiGPS and Reitit

4.5.1 Description

ReittiGPS and Reitit mobile clients were introduced in HSL Reittiopas case (Chapter 4.1) from HSL perspective, but they deserve another inspection focusing on their business model. Why exactly are the third-party developers building clients on top of the HSL Reittiopas API? The representatives from both ReittiGPS and Reitit (formerly known as Reitit for iPhone) responded that the initial reason to start developing was their own frustration in the usability of Reittiopas website with mobile phone's browser.

Offering of both services is a better user experience of Reittiopas service for mobile terminals. Their applications utilize GPS information from the handset to determine current location of end-user, and deliver fast route check-up in to a desired destination. Both of the services were developed for Apple iPhone, but Reitit had in addition an iPad version.

Resources of the services are mainly the public transportation APIs offered by Helsinki and Tampere transportation authorities. In addition, both have integrated Helsinki service map in order to provide a directory of services and points of interest within the city. The client does not require a back-end server, because it connects directly to the public transportation APIs, and thus it is easy to maintain.

Relationships are limited to HSL and Apple's App Store, as both perform all activities in-house. Since the offerings of the two companies are very similar, the competition has forced the companies in to a price war. Both respondents said that price is an important decisive factor when a customer is selecting an application from the App Store.

Revenue model for both companies is simply a one-time fee for buying the client from Apple's App Store. Initially, however, neither of the founders had plans on making money with the clients, just to create a better Reittiopas service. However, despite the initial expectations of the founders, the applications have turned out to be quite popular in App Store. Apple does not give out sales figures of the applications, but representatives from both companies said that they have at least momentarily reached top-10 in Finland. In spite of the success, both developers stated that the revenue is not enough to quit their day jobs.

Both companies offer also a free version of the product with some features removed and with an additional banner for advertisement. The companies have used the free version's banner space only to advertise their premium version, but neither ruled out the possibility that at

some point of time in the future the banner space might be sold to outside advertisers as well thus generating additional revenue source.

4.5.2 Analysis: Create valuable user experience and monetize with one-time fee

Both companies utilize open data provided public transportation authorities to create a valuable user experience for mobile phones. Whereas the previous case companies have monetized their service through advertisements or licensing, ReittiGPS and Reitti create revenue with sales of the premium version. Both companies have considered advertisement revenues in addition to the one-time fee, but so far have not pursued them. However, they have launched a free of charge version with stripped functionalities and an advertisement banner pointing towards the premium software. This could be described also as a freemium business model (Andersson, 2009).

4.6 Case Hilmappi

4.6.1 Description

Gemilo's Hilmappi is a website that **offers** a better user interface to Finnish Government's procurement announcement service, named HILMA, by plotting them on a map and offering tools to manage and tag the announcements. According to a representative from Gemilo Oy, their user interface can save end-users' time remarkably. He said that before the service, their own employers spent 30 minutes daily just to browse the new announcements with the government's user interface. With help of Hilmappi, they can perform the same task in about 5 minutes. These 25 minutes saved every day add up in a substantial figure on annual level.

Hilmappi has been built with Gemilo's own **resources**, and according to the representative from the company, they spent altogether two weeks in developing the service. Since then it has required only some maintenance and administration work. Therefore, the service has not required large investment from Gemilo. The main data source of the service is an API to the HILMA database, which is operated by Ministry of Employment and the Economy in Finland. The database is comprehensive, because all public procurements over 30000 € must be listed.

Revenue model of Gemilo Oy is to sell Hilmappi with a 50 € annual subscription fee. The service has also a one-month free trial period to attract users. However, the respondent stated

that since Hilmappi is not their core service and they do not have much time to develop and market it, in future they might remove the subscription fee altogether and use the service only to generate public relations to the company's other services.

4.6.2 Analysis: Create a valuable user experience and monetize with annual subscription

Hilmappi creates a valuable user experience on top of open data, just as the previous cases have done, but instead of one-time fee or advertisements, they monetize it with annual subscription fee. They offer a one-month's free trial period after which an annual subscription is required. Although Gemilo is considering removing the subscription fee and making the service free altogether, it is still a case example of the subscription revenue model.

4.7 Case Kansanmuisti.fi

4.7.1 Description

Kansanmuisti.fi is a journalistic website **offering** citizens an easier way to follow parliament activity with the help of public information sources in a non-partisan way. In their website they state that their mission is to:

“provide citizens with the opportunity to track parliamentary performance in an easily understandable and politically transparent fashion. Kamu collects information about the voting behavior of the MPs, members' statements made at plenary sessions (full-sitting sessions of parliament), as well as members' proposal of initiatives, and election funding.” (KAMU Ry background)

In their rules the association states that they collect donations, heritages, and grants to fund their activities (KAMU Ry rules). However, in a discussion on 25.4.2012 with the Chairman of KAMU Ry, he stated that at the moment most of their income comes from speaker fees, not donations. Thus, they have still a long way ahead of them to reach a truly crowd-funded status, and only time will show if Kansanmuisti.fi will grow to be the first journalistic website funded by the crowds in Finland.

4.7.2 Analysis: Create a valuable user experience, and monetize with crowd-funding

In Finland crowd-funding is still in its very early stages. This is largely due to a strict Finnish law about collecting funding from the crowds. The Money Collection Act 31.3.2006/255 dictates, that in order to arrange a money collection activity in which the money is collected by appealing to the public, a money collection permit needs to be acquired. The permit is granted only for non-profit purposes (Finlex 31.3.2006/255), thus completely outlining e.g. Kickstarter-type commercial crowd-funding activities in Finland.

Despite the strict Finnish laws, Kansanmuisti.fi is aiming to monetize valuable user experience with direct donations from the crowds. This revenue model requires transparent governance and a strong cause in which the end-users can relate to, so that they feel it is worth supporting.

4.8 Case Hahmota Oy Tax-tree

4.8.1 Description

Offering of Hahmota Oy is a visualization of financial data in a tree-like shape. The CEO of Hahmota Oy explains that their visualization principally offers a new metaphor for the basis of conversation. The tree metaphor has allowed their clients to invent a new terminology in their discussion; they talk from leaves, branches, roots etc. In a way, the visualization is like Google Maps for financial data. Their customers could be private companies, governments, governmental bureaus and public utilities alike. The CEO envisions that if they could create a new visualization method to be used with financial reports, it would open a completely new market for the product. (CEO, Hahmota Oy).

Resources utilized are the financial data to be analysed, a proprietary analysing engine generating the visualization, and personal work to pick the best figures to be analysed. The CEO states that typically their clients offer the necessary data to be analysed, thus Hahmota Oy does not need to extract or transform any data. However, sometimes they have made example visualizations for prospect customers as a starting point of the sales process, and this has required some data extraction as well.

Relationships are formed with various companies who wish to use the Finance Tree visualisation engine in conjunction with their own product or service, for example a finance

analytic company. The respondent from Hahmota ponders also, that an international partner would increase their visibility.

Revenue of Hahmota Oy comes from one-time project fees charged directly from their customers. Pricing of the project is composed of two parts: Consultation hours and creation of the actual visualization. Their CEO states that at the moment 70 % of their work time goes to consultation regarding how to present the data and 30 % to the final visualization, but in the future the consultation time should become smaller as Hahmota gains a larger portfolio of previous works which help new customers to decide how to present their own figures.

Management mind-set was entrepreneurial; the newly found company was searching for growth and looking forward to make real business with their concept.

4.8.2 Analysis: Create visualizations and monetize by selling project work

Hahmota collects data from a client, and uses its unique visualization engine to create a tree-like visualization out of the data. The data does not have to be open, but Hahmota has become famous for the concept of visualizing open data from municipal authorities named tax-tree. The tax-tree visualization, however, was based on imaginary municipal financial data, and worked as an advertisement for the company. The actual revenue comes from tailored projects for other companies and government bureaus. Their project-based business model differs from the business models of the user experience providers analysed earlier. The project work is done for one customer, requires lengthy sales work, private contracts, committing to timetables etc. Thus, in case of Hahmota Oy, the business model is closer to a software subcontractor that has a special asset of visualizing vast amounts of data.

4.9 Case Asiakastieto

4.9.1 Description

Offering of Asiakastieto is information from Finnish companies, private citizens, and properties. The respondent from Asiakastieto states that they collect precise data on individual level, with the accuracy of personal identity number, business ID or real estate number. By cross-analysing this data with advanced algorithms, Asiakastieto can give a risk rating for each individual and company in Finland. The respondent said that 95 % of their business is based on this individual level knowledge. Since this knowledge is used in important credit decisions, there is no room for mistakes in the data.

Resources of the company are raw data sources, and both employees and algorithms which make the data analysis. Asiakastieto has been extracting data from various public data sources for almost 100 years. Nowadays, the extraction is usually done over digital interfaces with automatized processes, but some data sources still require manually scanning paper documents into a digital format. Typical data sources include, but are not limited to, Finnish Business Information System (YTJ), Trade Register (Kaupparekisteri), National Board of Patents and Registration of Finland (PRH), and Statistics Finland (Tilastokeskus). In addition to public data, they also collect unique data directly from companies with questionnaires and financial statements. Thus, not all of their data sources are open, as open data should be accessible over the internet, but majority of their data sources are publicly available nevertheless.

Relationships include partnerships with EU-wide and global information providers for information exchange, partnerships with Finnish companies for balance sheet exchange, and collaboration with Finnish authorities in various work groups relating to legislative preparation.

Revenue comes from selling information products to customers with transaction-based pricing. The company generates 40 million euros annual revenue from 25000 customers. The majority of the revenue comes from small amount of big customers, but the Finnish entrepreneurial scene as whole is still well represented as a customer base.

Asiakastieto's products are priced in relation to the benefit that the client can achieve with the information. For example, the efficacy of a credit risk analysis can be tested with historical data, and thus the potential benefit can be proven for the customer.

4.9.2 Analysis: Algorithm-based analysing

Asiakastieto is perhaps the oldest player to monetize public data in Finland. The company's roots go all the way back to Suomen Luotonantajyhdistys, which was founded in 1905. Today Asiakastieto generates 40 M€ annual revenue, mostly by utilizing publicly available data sources. Currently they employ 150 employees, and during the almost 100 years of operation, they have earned reputation of a trustworthy information provider. The respondent from Asiakastieto said that if a credit is applied in Finland, it is likely that at some point the credit request goes through Asiakastieto's information systems. How does their business model compare to the other companies examined so far?

The first case of HSL Reittiopas related to API release to gain service level improvements and cost savings in UI development. The seven next cases involved presenting open data in an attractive user interface and then generating revenue with either advertisements, one-time fees, subscriptions or donations. The eighth case of Hahmota Oy provided altogether different business model by visualizing data for made-to-order projects sold for corporate customers. Now Asiakastieto has presented again a different way of executing business.

Asiakastieto analyses the data with mathematical algorithms, similarly as Hahmota Oy, but does not necessary provide eye-catching visualizations or user interfaces. Instead, they combine several data sources and refine the data in order to give new knowledge and new valuable insight for its customers. Thus, they provide analysing on algorithm-level. Of course they also provide an easy to use user interface to let their customers access these analyses, but although important, the user interface is not the key part of their offering. The user interface has changed over time, and embodied the leading technologies of each era. Initially postal service and telephone was used as an interface, then text based terminal connections, and nowadays internet and information system integration are utilized (Parpola & Kiljala, 2005).

The biggest difference, however, is Asiakastieto's pricing model, which in essence is product-based transaction pricing. Each bit of information they have in their databases is productized and charged at fixed fee based on how many times it has been requested by customers. Thus, the entire business model is based on "create once, sell many times"-type of information product. This pricing model has proven to be very efficient and profitable for Asiakastieto.

4.10 Case Cloud'N'Sci

4.10.1 Description

Offering of Cloud'N'Sci Ltd is an algorithm-as-a-service platform where third-party algorithm developers and business world problems are connected. That is, a third-party developer with ingenious algorithm can sign in to the platform and offer his solution to the market. These solutions are then packaged and sold as a service for businesses with various algorithm needs. For the businesses Cloud'N'Sci offers a selection of risk-free algorithm solutions whose worth can be calculated before the investment. This differs remarkably from traditional ground-up algorithm development, which utility is usually unknown before the algorithm is ready.

Resources of the service are the actual platform on top of which the algorithm modules can run, the third-party algorithm developers creating new modules, sources of data, and an algorithm architect who has a responsibility from the entire service towards the business customer. The algorithm architect knows what modules are on offer in the platform and takes responsibility that the system delivers promised results for the business customer. The data source can be any public or private, as long as there is a module that extracts the data from its source in to the platform.

Relationships are used to market and increase knowledge of the platform. The respondent from Cloud’N’Sci says that they have considered collaboration with Helsinki Region Infoshare-project, because it would result in obvious synergies. However, their focus is first to prove the service concept, and then continue finding new partners.

Revenue comes from a revenue share model between Cloud’N’Sci, algorithm developers, data sources, user interface providers, and the algorithm architect, who is responsible of the whole value chain. The actual split is decided per algorithm solution basis, and it varies depending on the importance of the different players in the solution. Some algorithms might be so central to the solution, that their share of the profit will be proportionally larger.

Management mind-set of the company is very business oriented. The CEO says that the service is born global; there is no reason to limit the service only to Finland. In addition, the fixed costs are kept minimal by making the service as self-served as possible. Thus adding a new algorithm module, creating a new algorithm solution by combining the available modules and data sources, and splitting the revenue of a certain service can all be made by third parties, without interference from the maintenance.

4.10.2 Analysis: Algorithm-based analysing

This is the second case relating to algorithm-based analysing. However, Cloud’N’Sci differs quite a lot from Asiakastieto in their business model. Whereas Asiakastieto sells its own information products with the transaction based pricing, Cloud’N’Sci is providing a marketplace where third-party algorithm providers can sell their services onwards. Cloud’N’Sci has prudently productized their platform, including the revenue share models, but their company is still too young to make a judgement about their business model.

From the open data perspective, the Cloud’N’Sci platform is agnostic to the type of data the algorithms are calculating; it could be open or private. The CEO states that for them open

data is just one data source among others, and that if an open data source proves to be vital they are willing to compensate for data provider. In fact, the CEO saw the whole freeness of data as an issue, because an ecosystem where one can freely reap the benefits of the data, which someone else has published, does not necessary encourage to publish more data.

4.11 Case HS Open

4.11.1 Description

HS Open **offers** an event that brings journalists, graphical designers and programmers under a same roof and encourages them to brainstorm and create new purposes for open data. The event is organized by Helsingin Sanomat, and first HS Open was held on 14.3.2011 (Mäkinen, 16.3.2011). It has been a very successful activity, producing tens of prototypes utilizing open data in visualizations and innovative user interfaces. It has been organized regularly, and the fifth HS Open was held on 21.5.2012.

HS Open events have encouraged a crowd of people to make data analysing on their spare time for free. Some of these hobbyist analysers have used very advanced statistical methods, such as factor analysis, in their analyses. Often they have written a custom program that analyses the data and creates visualization or other interpretation out of it. Most of the analyses created during HS Open events were published in the HS Next blog or in their creators' private website or blog. However, some of the best visualizations inspired articles that were published also in the paper version of the newspaper.

One example where a user-generated visualization has been published comes from an article released on 24.6.2011 about electoral funding relations (see Figure 12 for demonstration). The data analyser is a bioinformation technology student, who noticed that an algorithm made for the network analysis of genes could be utilized in electoral funding as well. The visualization inspired an article examining the power structure and connections behind electoral funding. (Mäkinen, 10.2.2012)

Vaaliraha paljastaa vallan verkoston

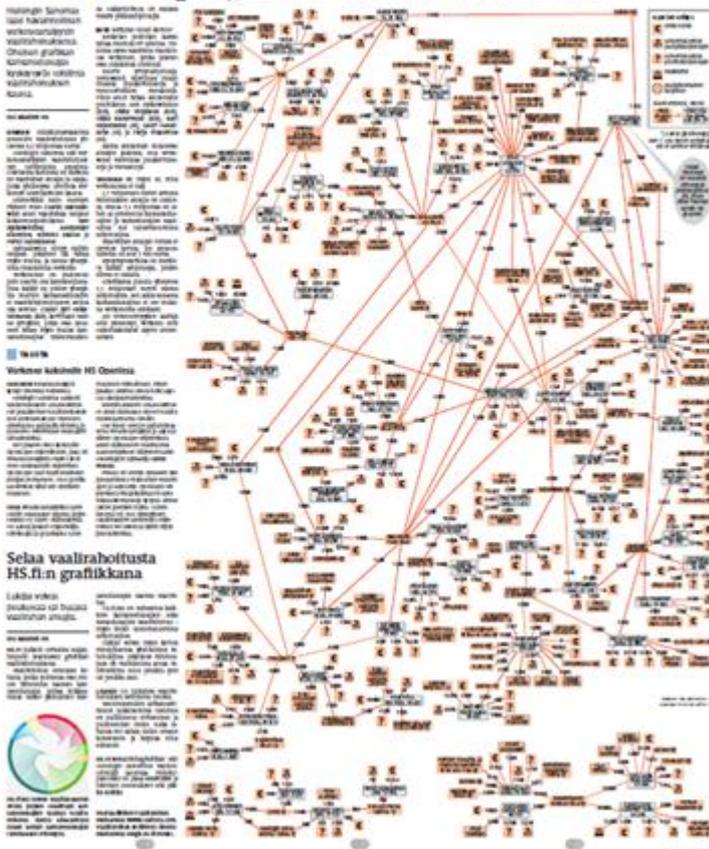


Figure 12 An example newspaper article based on crowd-sourced analyzing. Source: HS Next-blog post (Mäkinen, 10.2.2012)

In addition to organizing HS Open events and blog, Helsingin Sanomat has also published its own data for others to analyse. A landmark data opening was parliamentary election Vaalikone data release on 6.4.2011 (Mäkinen, 28.3.2011). Vaalikone is a web-based service aiding voters to select a favourable candidate in the elections by first asking a set of questions from all the candidates, then letting voters answer the same questions, and finally making suggestions by comparing the answers. By releasing the data, HS hoped to find new news and visualizations, and to better fulfil journalistic values by conveying more information from the government to the public (ibid.). A week after the data release, HS had received already 15 applications and visualizations utilizing the data (Mäkinen, 12.4.2011). In the following year, HS decided to open also the presidential election's Vaalikone data for the public (Mäkinen, 3.1.2012). In both Vaalikone data openings the data was released under Creative Commons license, which prohibits commercial re-use (Mäkinen, 12.4.2011; Mäkinen, 3.1.2012).

4.11.2 Analysis: Crowd-sourced data analysing

This case has two interesting aspects, 1) HS Open events, and 2) the release of private corporate data in the HS Vaalikone example.

Starting with the HS Open events, in practise HS encourages crowds to analyse data and then use the results in newspaper articles and blog posts. The crowd has been facilitated by organizing an event with a certain theme, by bringing up interesting data to be analysed, and by inviting capable individuals from various backgrounds to come and do the work. Motivation for the crowd is the general interest towards transparency and data analysis, the opportunity to get their work published in a national newspaper, and possibly the complimentary beverages available. This concept has proven to be very successful, resulting in tens of visualizations and data analyses. For HS the cost of the events in terms of money or effort has been minimal.

HS Vaalikone data opening, on the other hand, is a good demonstration of the benefits of opening up corporate data. Endorsed together with a dedicated HS Open event, the data opening got enough publicity to catch the attention of the masses and as a result found surprisingly innovative usages. This data opening represents a large shift in the management mind-set; a data, which during previous elections has been considered a private asset, was now released to the public free of charge.

What was the reason behind the shift in management thinking? For the public sector, the reason to open data is typically to advance citizens' participation in democratic decision-making process, increasing government transparency, and a general pressure from the crowds to open up data sources. For corporations, however, similar reasoning does not apply. Quite often, the data is a core competency of the company, which makes the companies understandably cautious when opening the datasets.

According to Mäkinen (28.3.2011), the motivation was similar as with HS Open events; use crowds to analyse the data to achieve new insight. However, this cannot explain the entire action, because HS has its own professional data analysers and reporters working with the data as well.

This Thesis believes that by releasing the raw data, HS increases readers trust in the newspaper's data collection and analysis process. In addition, by inviting crowds to browse through the data, HS gets several "second opinions" in case their journalists missed something. Therefore, unlike with the HSL case where the aim was to involve crowds in the

user experience development, HS did not expect to outsource the analysis. Instead, they planned to increase readers' trust into the data integrity and also to get a second opinion in to their internal analyses.

4.12 Case Louhos

4.12.1 Description

Offering of Louhos is a comprehensive software library for R-language, named soRvi-toolkit, assisting analysts to extract and analyse open data from several sources. The toolkit offers automatic data fetching-routines supporting several open data sources ranging from municipalities to World Bank and from Finnish postal numbers to OpenStreetMaps (Louhos website). In addition to data fetching, the toolkit offers analysis routines to process the data onwards in R. For example, plotting county-level information on top of a Finnish map is made very simple with the toolkit. However, the toolkit does not have any central storage for the data; it is a script which extracts the data from the original source each time the toolkit is run.

In addition to the toolkit, Louhos also creates plenty of analysis and visualizations and publishes those in their Louhos-blog along with the example R-source code to replicate the visualization.

Resources required to build and maintain the service are light, principally just the programming knowledge of the founders. The source code has been released in GitHub, and thus anyone can continue the development of the toolkit (soRvi GitHub). The founders said they are hoping that other programmers would get interested in the project and start adding new features, datasets, and countries in to the toolkit.

The founders are very active in the open data scene in Finland. These **relations** are used to increase knowledge of the toolkit and to invite other active developers into the community. They are co-operating, among others, with Helsinki Region Infoshare, HS Open, Apps 4 Finland, Open Knowledge Foundation, and Kansan Muisti.

Revenue at the time of interview was zero, and the project ran on the founders will to advance open data. Since there was no revenue, the resources have been kept minimal as well, and the founders are hoping the open source community would help with the

development. Although they did not generate revenue at the time of interview, the founders had ideas about future income from consulting and other supporting tasks.

Since there was no revenue generation, the **mind-set** of the founders is pro-bono. However, they pondered that in the future a combination of commercial and volunteer activity might be the most feasible path onwards.

4.12.2 Analysis: Extract and transform

Louhos saves a lot of other companies' time, because now everyone does not have to go through the laborious work of figuring out the source data formats and parsing out the relevant data into tables in order to execute the actual data analysis. With help of Louhos, the analysers can focus on real analysis, instead of spending time in mundane data extraction and transformation tasks. In addition, since integrating a new data source has been made effortless, the analysers are more likely to employ multiple data sources resulting in a more extensive combinatorial analysis. Practically all the interviewed companies expressed a need to extract and transform the data before processing it onwards. Many of the companies did this work in-house before the actual analysis or visualization, but it seems evident that there is a need for a separate entity extracting and transforming as well.

Considering how much added value extract & transform delivers, there is surprisingly little commercial activity in this field. Louhos and their soRvi-toolkit is perhaps the only corresponding entity in Finland. However, at the moment they do not have commercial activity either. It can be even questioned why Louhos was included in the analysis, since the intent of the empirical part was to interview companies with revenue intentions (see Chapter 3.5.1 for details on criteria). An exception was made in the case of Louhos, because their name came up in so many discussions and interviews that their role within the open data scene seems to be indisputable, and thus they could not be left without evaluation. Also, they won the Apps4Finland 2011 data opening competition.

Louhos is also the only interviewed entity publishing their source code freely in an Internet repository. This activity is described by Kuk and Davies (2011) as an important part of the overall open data complementarities assemblage, and according to them, enables further innovation (see Chapter 2.3). The respondents from Louhos said they are hoping for more open source developers to join the project, in order to keep adding new data sources nationally and globally.

4.13 Case Flo Apps

4.13.1 Description

Flo Apps is a Finnish software company **offering** technical implementations of open data visualisations, thus helping their customers to present their data in a more visual and appealing form. The CEO of Flo Apps stated that the customers usually require tailored solutions, including information design, user interface design, and software implementation.

Resources are used efficiently, as Flo Apps produces only the technical implementation in-house, and outsources most of the graphical and information design to partners. This approach has kept the company lean and cost-effective, and able to engage in a variety of different projects. Open data resources have been used especially in their Apps 4 Finland submissions, but also in several customer projects.

Relationships. Flo Apps has been part of the open data scene from its early stages. They have participated to A4F competitions in order to build reputation for their company. Flo Apps has sent altogether six submissions to the competition, all with very good results. However, the competition rarely brings new customers directly. Instead, it positions the company in the spotlight of open data scene, and thus makes the subsequent sales work easier.

Revenue comes from projects sold directly for the customers, and in year 2010 approximately 30 % of the revenue came from open data related projects. The projects are priced based on estimated work, which holds the risk of the project being more complex than presumed in the calculations. The CEO says that the projects require substantial amount of tailored work, which rises the price tag easily so high that the customers are hesitating to commit. In order to close the deal, both parties often need to make compromises, which reduce profits.

According to the CEO, one solution to make the projects more profitable would be to document and productize the visualization process, thus making it more transparent and efficient. When the process is standardized, the cost can be estimated more accurately, the project can be completed faster, and the customer gets a more professional impression, all of which increase the profit.

Flo Apps has also investigated the possibility of creating its own information product, but the Finnish market is so small that creating and supporting it would be inefficient. The product-

based business would require at least European wide distribution, which again would require standardized open data interfaces across European Union countries. At the moment such standardization does not exist, therefore cross-border software would require extensive localization for each country.

Management mind-set is very entrepreneurial, and they were one of the first Finnish companies to commercially exploit the possibilities of open data. However, their business is not limited to open data. Instead, they are ready to engage in any software project as long as it is realizable with their expertise. This creates an opportunity cost for their open data related activities as well; they need to be at least as profitable as the rest of the business. This background has ensured that all their open data activities have strong business intention, and is probably one of the key factors why at the moment Flo Apps holds one of the biggest open data related turnovers in Finland.

4.13.2 Analysis: Consultation and software projects

The business of Flo Apps is based on tailored project work sold directly to business and government customers, similarly as with Hahmota Oy earlier. Open data is primarily used to gain visibility in Apps 4 Finland and similar application contests. The reputation, which Flo Apps has gained through these competitions, has eventually brought new contacts and new customers. In essence, it is the same marketing strategy, which also Hahmota Oy uses: Utilize visualizations created out of open data as a marketing tool for the company.

What differentiates Flo Apps from many other open data companies is their profit oriented managerial mind-set. Since they have started from general IT projects with no relation to open data, the activities concerning open data need to be profitable as well. This background in general software business has helped Flo Apps to focus only on the profitable open data projects, and discard the ones without potential to generate revenue.

Another observation is the importance of products and productization. Asiakastieto case demonstrated that easy reproducibility and transaction based pricing are key elements in information-based business. Flo Apps has experience with product-based business, does not yet have its own open data products. They have been considering the option to pursue to the product-based business, but so far have not made the switch because they are concerned that the Finnish market is too small for open data product business to thrive.

If this statement is true, does the Finnish market size affect other companies and types of products as well, or is it present only in some segments? That is, can the Flo Apps's experiences of insufficient market size be generalized in to entire open data business? To assess this question, counter-examples can be taken from the product-based case companies examined earlier in this Thesis.

ReittiGPS and Reitti mobile clients, according to their creators (see Chapter 4.5), sell quite well, even though they are restricted to Finnish market only. However, this revenue has not been enough for the founders to quit their day jobs. Duunitori.fi (see Chapter 4.2) is generating revenue with advertisements and Mitamukaanlennolle.fi (see Chapter 4.3) with advertisements and licensing, but they refuse to reveal their turnover figures. However, Asiakastieto (see Chapter 4.9) is probably one of the most profitable open data related companies in Finland, and it operates with product-based strategy. This one example alone is enough to revoke the too small markets statement, and thus it can be said that open data related product based business is conceivable in Finnish markets. However, since only one successful example was found, one should be careful not to generalize the result too much.

4.14 Case Logica

4.14.1 Description

Logica is an IT-services company, which **offers** its customers help following the levels of European Interoperability Framework (EIF). This means especially organization, semantics, and technology layers. Relating to open data, they have been involved in HSL Reittiopas API project, Paikkatietoikkuna, and several other governmental open data projects. All these projects have had emphasis on machine-to-machine communication with the organization's service and process development.

Resources of Logica as a big global ICT-player are vast, and they are easily the largest company interviewed in this Thesis. Logica employs 3000 workers in Finland alone, and globally they are part of CGI Group with about 70000 employees.

Relationships are formed with both big and small players. They have searched small companies from A4F competition to do modules or components in their offering. In addition, they collaborate, for example, with Microsoft, IBM and Oracle, and utilize the Azure platform. They are also involved in the Data to Intelligence (D2I) program hosted by TIVIT.

D2I program the aims to put organizations first, and to see how services and processes could be built in a new way.

Revenue model is moving towards service-based pricing. For example, transaction pricing using customers' environment as pricing point is getting more popular. So instead of having artificial pricing of CPU hours or license agreement, the pricing can be based on customers' business environment, which is easier for the customers to understand and estimate the practical benefits.

4.14.2 Analysis: Better services with machine-to-machine communication

Machine-to-machine (M2M) communication is related to open data through system integration and information systems' back-end architectures in general. Although M2M communication is not something end-users can directly see or relate to, it has an important role in service design through information logistics. The respondent from Logica emphasized this aspect of open data.

The interviewee stated that M2M communication is especially important when designing user centric services and reaching better productivity. For example, government bureaus' data is often stored in vertically integrated and non-interoperable information systems. The consequence is that an end-user may need to re-enter trivial data, such as name and address, repeatedly in different electronic forms. This can be very frustrating. In addition, if the users' address changes, in some cases, the user needs to re-enter the data to each system again. Systems are not synchronizing data as efficiently as modern services and processes require.

The respondent sees M2M communication as one answer to this problem. By opening the data internally within a bureau and between governmental bureaus, such problems could be avoided. Information sharing could bring other benefits as well. For example, by collecting data from several bureaus and utilizing advanced algorithms to analyse it, predictions from citizens' future behaviour and service needs could be made more accurate. It would be important to identify possible critical privacy policies and legislations, so that in the future, for example, predictions could be based on a combination of different data reserves. According to the respondent, a Social Services Department (Sosiaalivirasto) worker could, for example, proactively approach a long-term unemployed citizen if they would coordinate their datasets with Employment and Economic Development Office (Työ- ja elinkeinotoimisto). This could prevent social exclusion and bring inclusive service experience for the end-user. Similar user-centric and organizationally valuable services could be built

also between public and private business organizations, based on interoperable and secured automation information logistics. (Development director, Logica)

Youngin Yoo, a professor at Temple University Philadelphia, in his (2012) speech at Aalto School of Economics, took an example of Philadelphia's fire department not having access to the infrastructure information relating to water and gas pipelines, electrical cables etc., although this information is electrically available in another city's bureau. This is a good example where M2M communication could solve a problem relating to everyday life.

However, to be consistent with the previous analysis, M2M is not a business model with revenue model, distribution channels etc., but more like an area of application for open data in general. Within the M2M area, there can be found several individual business models. Probably the most obvious is the one of Logica, which works as a systems integrator helping its customers to succeed in this sphere and enable, for example, new service and revenue models together with organization development and user-centric benefits.

4.15 General observations from the interviews

Commercial activity within the area of open data is still in early phase. Out of the 193 proposals submitted to Apps 4 Finland contests in three years, only 11 were backed up by a company with commercial ambitions. This works well as an indicator of the commercial activity in general: There is a lot of interest towards open data, but large-scale commercial utilization is missing. Remote commercial activity could be either a consequence of novelty of the phenomena or a failure of open data to supply real business opportunities. Since the cases examined within this Chapter illustrate that there can be business opportunities, it might be that open data requires more time before businesses start to capitalize on it.

One specific comment, which rose from several interviewees, was that the value of the raw data is zero euros without an application utilizing it in a meaningful way. The philosophy behind open data, on the other hand, suggests that opening data to the public is valuable in itself regardless of its predicted re-use potential. This contradiction can backfire if no discretion is imposed on where the data opening resources are directed. Opening data is not free, and if the data does not generate valuable applications, the whole phenomena could be denounced vain. The situation is worsened by the fact that it is usually difficult for the data owners to foresee which data will be perceived interesting and which will remain unused by the community. Therefore, to avoid wasting scarce resources on opening inessential data, a

dialogue with the community is strongly encouraged. With help of a dialogue, the data owner can recognize exciting data reserves and focus their resources on opening those.

Another observation is that many companies find it difficult to create product-based business on small Finnish markets. For example, ReittiGPS, although a popular mobile application, is employing its founder only part-time. Standardization of open data interfaces was seen as one answer to this problem. If, for instance, EU would have a common standard for public traffic information, a same application could be easily scaled in to tens of cities around Europe. However, not all product-based businesses were struggling. For example, Asiakastieto was running a very successful product-based business in Finland.

5. VALUE NETWORK ANALYSIS

This chapter will group similar companies together based on the analysis done in the previous chapter; and the earlier work done by Poikola et al. (2010), Tammisto and Lindman (2011), Lehtonen (2011), and Kuk and Davies (2011).

The grouping was done by placing companies with similar offering together. Thus, for example, all the seven user experience providers are grouped under one profile, even though they have different revenue models. Although offering was decided to be the pivotal factor for categorization of the companies, it was not straightforward as many companies had so versatile business practises that they occupied several positions. These borderline cases were decided based on their primary value adding functionality. Extract & transform profile ended up having only one company representing it, but all the other profiles have several ones.

Altogether five distinct profiles were identified and jointly these profiles establish the Finnish open data value network. The name of the profile and the corresponding companies placed under it are listed below, and they will be described further in the subsequent chapters.

1. data analysers (Hahmota Oy, Cloud’N’Sci Ltd, Suomen Asiakastieto Oy)
2. extract & transform (Louhos)
3. user experience providers (Skyhood Oy, Suomen turvaprojektit Oy, Essentia solutions Oy, Reitit, Forum innovations Oy, Gemilo Oy, KAMU Ry)
4. commercial open data publishers (Helsingin Sanomat, HSL)
5. support service and consultation (Logica, Flo Apps Ltd)

5.1 Data analysers

Data analysis is an obvious part of the open data value network. The interviews revealed multiple types of data analysers. Some were analysing the data to create new visualisations, others were cross-analysing different data sources with advanced algorithms in order to provide valuable knowledge. Some analysers did their job only to serve the common good, while others had strong business model. Figure 13 below summarizes this profile; it stands between the open data providers and end users. The case companies operating within this profile are: Hahmota Oy, Asiakastiето, and Cloud’N’Sci.

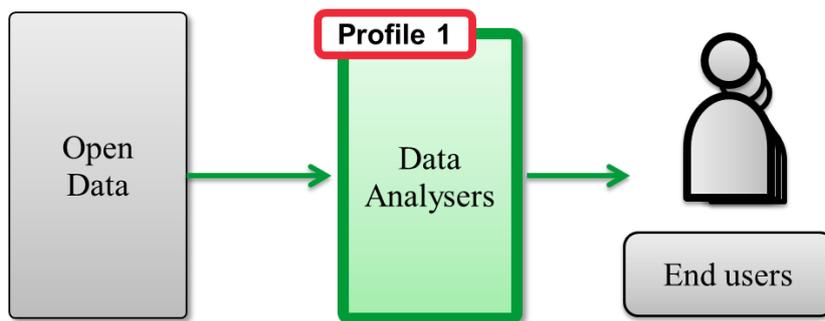


Figure 13 Data analysers (profile 1)

5.1.1 Data visualizers

Visualization is a powerful way to communicate key points of the data for general crowd. Likewise, for a typical end-user the raw data is basically worthless without an appealing interpretation of it. Therefore, it comes as no surprise that visualizations are very popular within the open data community; there are lots of different visualizations all over the web done by hobbyist visualizers. However, this Thesis found only one company that created open data visualizations as their core offering (see Chapter 4.8).

5.1.2 Algorithm based analysis

Another way to analyse the data is to utilize advanced algorithms and scrutinize the raw data on numerical level. Our study found two cases companies operating on algorithm level; Asiakastiето and Cloud’N’Sci. Both had strong business intentions, but very different business models. Whereas Asiakastiето leaned on transaction based pricing of information products, Cloud’N’Sci offered a platform where third-party algorithm providers could connect to business customers. Yet, in both cases their *raison d’être* was to utilize algorithms to refine raw data into something which has value to their customers.

5.2 Extract and transform

In order for the raw data to be analysed, it must be available in a format allowing further processing and handling. Thus, in order to analyse any data, the data needs to be extracted from its original source and transformed into a meaningful format. This activity is what the “extract and transform” entity does. To be clear, no analysis of the data is done at this stage, only extraction and transformation, respectively.

Open data is typically published in a number of different forms, ranging from Excel files to proprietary formats, which are not necessarily compatible with each other. For example, governmental bureaus often publish their data in various formats, and make no effort to standardize it between the bureaus. In addition, a data publisher might alter the data structure over time, thus making inconsistency with the previously released historical data. Also, the data sources have no guarantee on how long the historical data will be available in the first place, thus storing it in a third-party database would solve many problems.

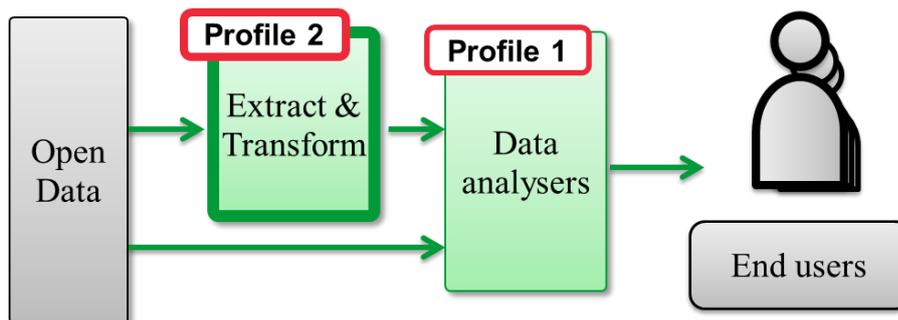


Figure 14 Extract & Transform (profile 2)

The transformation process is even more important if the analysis includes several data sources; the data must be arrayed the same format and in the same scale. This is very cumbersome as there are no guarantees in which scale the data publishers have decided to release their data, and typically, some conversions are necessary.

Usually the data must be also administered in order to ensure its integrity. The data might include double records, missing information or otherwise incorrect information, which needs to be corrected. Part of this work can be done with clever algorithms, but often it requires hours of labour.

5.2.1 Extract and transform as integrated part of data analysing

The boundary between data analysers and data extractors and transformers is not always clear. In practise, most of the data analysers, which were introduced in Chapter 5.1, operate also as data extractors and transformers. This is simply because the raw data is rarely available in a meaningful format, and the analysers need to convert the data by themselves. Analysers often use external providers, such as Sorvi, as one data source but the data available through these providers is limited. For example, the respondent from Hahmota Oy said that sometimes the customers bring proprietary data that needs to be analysed. Since this is private data, it needs to be transformed by the analyser himself.

Asiakastieto's business model could be also described as extract, transform and analyse, as extraction and transformation are a big part of their overall process. The respondent from Asiakastieto said they still have to digitalize paper documents in order to gather enough data to perform analysis. In addition, they also store historical data thus making time-series analysis possible.

5.2.2 Extract & transform as separate business

Since any data analysis or visualization requires data extraction, it is obvious that there is a need for a separate player as well. Sorvi toolkit by Louhos is probably the best-known example of data extraction and transformation in Finland. However, their company case in Chapter 4.12 revealed that, at the moment, they do not have business intentions. In addition, they have no plans to store historical data.

This leaves open the question whether there would be room for a commercial operator within this value network profile. On the one hand, many interviewees emphasized that raw data as such has no value before it is made valuable by novel analysis and visualization. Based on this, it would be unlikely that someone would be willing to pay for extracted and transformed raw data. On the other hand, several analysers and user experience providers expressed the concern that considerable amount of their time is consumed in data transformation and administration tasks. Most of the public data sources are of surprisingly poor quality, and until the time-consuming transformation and administration process, they are principally useless. Therefore, it would be easy to imagine that refined and trustworthy raw data service would create value for the analysers.

With commercial resources, the service could be improved further. For example, in addition to extraction and transformation, this they could also store the data for further reference. Over time, the stored data would make a time-series analysis possible. In addition, there is no guarantee how long the original data publisher keeps the records available, but on a third-party database the information would be accessible from the entire time span, even if the data publisher would remove some of the older data. Storing the data can be useful also in cases where the data publisher changes its data structure frequently; the interface to the third-party data storage can be kept unaltered.

5.3 User experience provider

User experience providers are the only entity directly in contact with consumer end-users. The core idea is to utilize open data sources to create a valuable application for the end-users. The interaction can be done either through a mobile or through a web user interface. The case analyses found three revenue models for this player: advertisements, subscription and donations. As with data analysers, sometimes the user experience providers also need to extract the data from its original source. In addition, they often process and analyse the data as well, so in some cases they perform three types of activities in the value network.

User experience providers are by far the most popular part of the value network. This study found seven companies operating in this role: Duunitori, Reitit, ReittiGPS, Hilmappi, Kansanmuisti, Pikkuparlamentti, Mitamukaanlennolle.fi. This might be because of diverse revenue possibilities, the offering is easier to conceive and assess because the creator is also a consumer himself, and also the entry barrier is lower because company relationships are not required to sell anything. In addition, the application markets provided by all the major mobile phone operating systems make the sales process easier.

The value network including the user experience provider as well as the two previous profiles is sketched in Figure 15 below. Most of the user experience providers utilized the raw data directly from its source, thus performing extract & transform and analysis tasks as well. However, some companies, such as Duunitori.fi and KAMU utilized also data generated by other analysers or extractors & transformers.

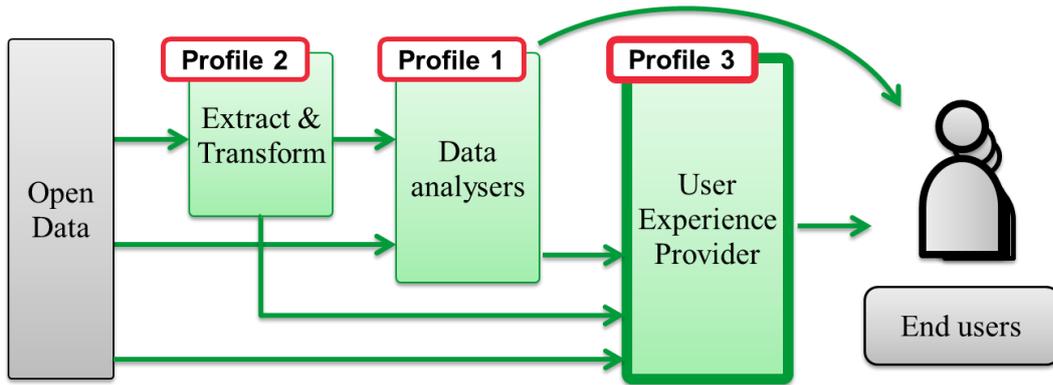


Figure 15 User experience provider (profile 3)

5.4 Commercial open data publishers

Commercial open data publishers are especially interesting, because they bring a new horizon in the open data value network – instead of utilizing data from other open data publishers, a company can publish its own resources and achieve concrete business benefits in doing so. The commercial data publishers are portrayed within the open data sphere in Figure 16 below. By releasing data, they join among other open data publishers, and enlarge the open data offering in the net. It is then up to the community to decide how this data will be utilized.

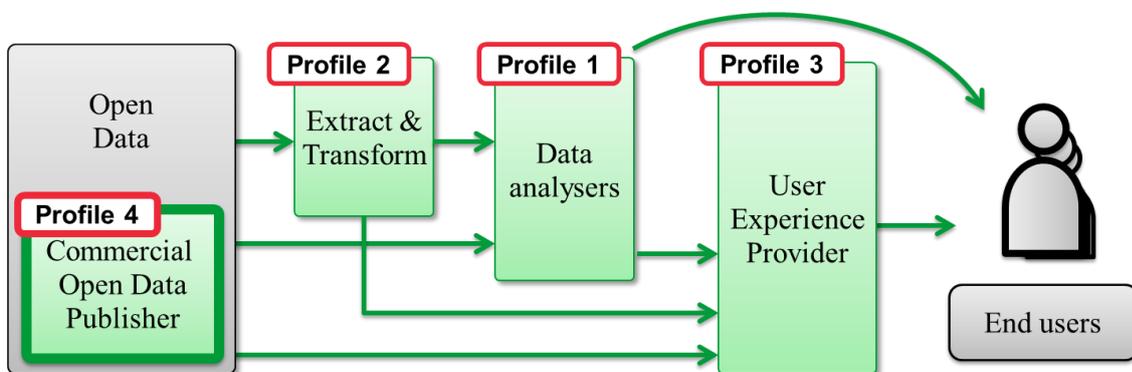


Figure 16 Commercial open data publisher (profile 4)

Opening up private corporate data should be, however, carefully thought and planned. For many companies the data assets constitute core competency of the company, and releasing it might jeopardize the entire business. Nevertheless, in some cases opening the data can be proven to show remarkable business benefits. The argumentation is similar as with the public sector; the owner of the data might not be its best exploiter. Thus, by opening up the data, a

third-party extractor & transformer, analyser or user experience generator could create a valuable service utilizing it.

Aitamurto and Lewis have studied open APIs from four big news organizations in their (2012) article. They find that open APIs accelerate R&D process and generate new means of commercializing content, especially in niche segments otherwise difficult to serve. They call this process “extended product portfolio”, in which the products are built on news organizations’ content, but with external developer’s user interface. According to Aitamurto and Lewis, this co-creation model generates the most value in open API ecosystem.

5.4.1 Co-creation under open license

A perfect example of the benefits of co-creation is Reittiopas API (see Chapter 4.1 for details). They have achieved remarkable savings in client development costs, while still providing a better user experience, by releasing the API to the public transportation schedule and routing data. As found out in Chapter 4.1, developers have contributed tens of mobile applications for several platforms based on the API.

The co-creation model has proven to be beneficial with non-profit companies, such as HSL, but could it be applied in to commercial companies as well?

5.4.2 Co-creation under restricted license

Commercial companies have a risk of losing profits or customers if they release wrong data. They need to be very cautious on what exactly should be opened and with what licenses. At the same time, they need to inspire the developer community in order to induce activity around the data release. Thus, companies need to select the right data to open, and to put effort in encouraging crowds to utilize the data as well. In a way companies need to balance between what data can be opened in the first place, what would bring the most benefits for the company, and what would inspire the developer community.

HS Vaalikone data opening and HS Open event is a good example of well-done commercial co-creation (see Chapter 4.11 for details). Helsingin Sanomat chose Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported (CC BY-NC-SA 3.0) license for its Vaalikone data, which restricts commercial re-usage, requires attributing the author, and sharing altered or transformed works with alike license (Mäkinen, 28.3.2011; Mäkinen

3.1.2012; Creative commons website, n.d.). Regardless of the license restrictions, the data has been very popular amongst analysts resulting in tens of visualizations and analyses.

Especially the non-commercial requirement in the license utilized by HS is in contradiction with the open data definition, which states that the data should be free of restrictions and permit unlimited commercial re-use (Open Definition; Poikola et al., 2010). To resolve the contradiction, a less strict open data definition was introduced back in Chapter 2.4. Thus, this research reckons HS Vaalikone data as open data, because it is available free of charge, albeit with some restrictions.

5.5 Support services and consultation

Support services and consultation are the fifth value network profile. Fitzgerald in his (2006) paper describes that open source software has long established its business model on a combination of volunteer work, free-to-download software, support services, and consultation (Fitzgerald, 2006). Since open data is a close relative to open source software, similar business models can be applied. Support services and consultation is portrayed in the value network (Figure 17) as a separate entity, assisting all other profiles (1-4) in their business. Thus, this entity is not directly involved in the value chain from raw data to end user, but instead assists the players in it. Please note that the illustration in Figure 17 is not entirely accurate, as for example, extract & transform (soRvi toolkit by Louhos) did not receive any assistance from neither of the companies portrayed in this profile.

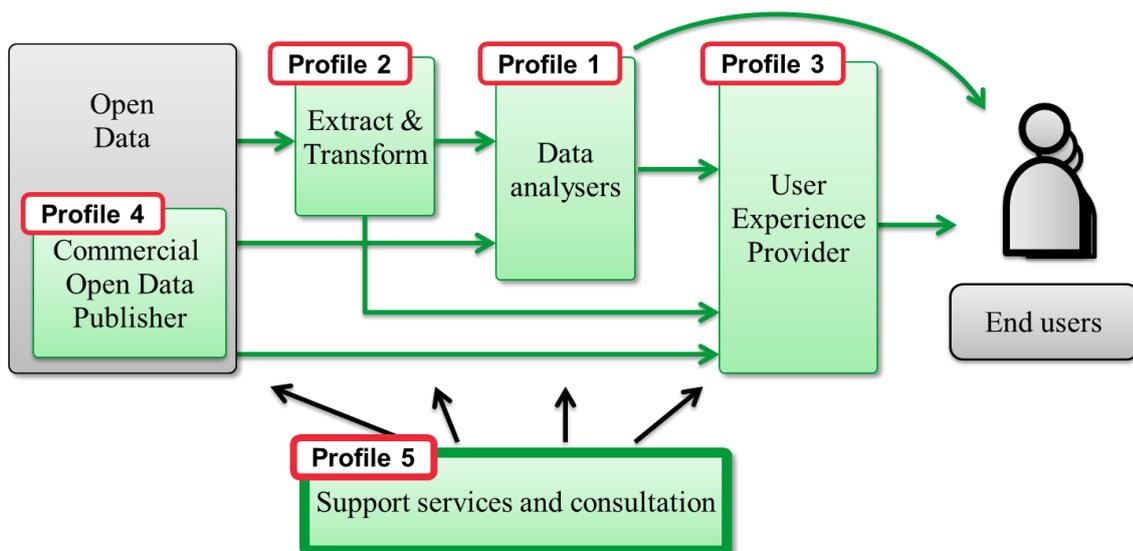


Figure 17 Support services and consultation (profile 5)

This study found two companies belonging in to this category, Flo Apps (see Chapter 4.13) and Logica (see Chapter 4.14). They both have project-based business, where revenue is generated from tailored projects sold directly to corporate customers. They both perform consultation and subcontracting. Logica was more positioned towards M2M communication and system integrations, whereas Flo Apps was focusing more in visualizations. Flo Apps could be also portrayed in the data analyser box (value network profile 1), but since they do not have their own products, but instead perform project work for other clients they are categorized here.

5.6 Summary of the value network analysis

The process of defining the value network profiles started by first conducting the interviews and describing the company cases (Chapter 4), then analysing the business model each company is utilizing (Chapter 4), and finally in Chapter 5 the companies were categorized under five distinctive profiles in the value network. The companies were grouped together using their offerings as a determining factor. Thus, the companies under one value network profile share similar offerings, but might differ in other aspects of their business models. These five profiles with their corresponding companies, offerings, resources, relationships, revenue models, and management mind-sets are recapitulated in the table below.

Table 4 Recapitulation of the value network profiles

	Commercial open data publisher	Extract and transform	Data analyser	User experience provider	Support services and consultancy
Companies	<ul style="list-style-type: none"> HSL Reittiopas HS Open 	<ul style="list-style-type: none"> Louhos 	<ul style="list-style-type: none"> Hahmota Cloud'n'Sci Asiakastieto 	<ul style="list-style-type: none"> Duunitori.fi Suomen turvaprojektit Gemilo ReittiGPS Reitit for iPhone Kansanmuisti Pikkuparlamentti 	<ul style="list-style-type: none"> Flo Apps Logica
Offering	<ul style="list-style-type: none"> Data for others to analyse Facilitate events to encourage activity 	<ul style="list-style-type: none"> Convert data in to an easier format for further analysis 	<ul style="list-style-type: none"> Data visualizations Algorithm based data analysis 	<ul style="list-style-type: none"> User experience created with help of open data sources 	<ul style="list-style-type: none"> Consultation and subcontracting to help clients benefit from open data
Resources	<ul style="list-style-type: none"> Maintenance of the data 	<ul style="list-style-type: none"> Data sources Open source community developing the source code 	<ul style="list-style-type: none"> Data sources Algorithms for visualization or numerical analysis 	<ul style="list-style-type: none"> Data sources Development and maintenance of the user experience 	<ul style="list-style-type: none"> In-house work Subcontracting
Relationships	<ul style="list-style-type: none"> Open data scene Support community of active developers to increase usage of the data 	<ul style="list-style-type: none"> Open source community Open data scene Governmental and private data publishers 	<ul style="list-style-type: none"> Open data scene Finnish authorities Global information providers 	<ul style="list-style-type: none"> Open data scene Labour unions Municipalities Subcontractors 	<ul style="list-style-type: none"> Open data scene Small open data companies as partners Big technology companies
Revenue model	<ul style="list-style-type: none"> Cost savings with crowd sourcing 	<ul style="list-style-type: none"> Pro-bono (open source) 	<ul style="list-style-type: none"> Project work Product based transaction pricing Modular-ecosystem Pro-bono Business customers 	<ul style="list-style-type: none"> Advertisement (two-sided market) One-time fee Subscription Crowd funded Licensing Freemium Consumer customers 	<ul style="list-style-type: none"> Project work Service based pricing Business customers
Management mindset	<ul style="list-style-type: none"> Entrepreneurial 	<ul style="list-style-type: none"> Pro-bono 	<ul style="list-style-type: none"> Entrepreneurial 	<ul style="list-style-type: none"> Entrepreneurial Pro-bono 	<ul style="list-style-type: none"> Entrepreneurial

The amount of companies within the five profiles is distributed very unevenly. The extract and transform activities were conducted by all of the interviewed companies as part of their analysis, but only one entity focusing solely on extraction and transformation was found. Since the need for the activity is so high, there might be room for more players collecting and storing data, and offering it for third-party analysers. Data analysers and user experience providers, on the other hand, were very popular. Half of the studied companies were categorized under user experience provision, and three under data analysis. This might be because these roles have a smaller entry barrier, but also because they have more versatile revenue model possibilities.

Resources usage differs quite a lot between the profiles; the commercial data publisher does not need data sources because it publishes its own, while extract & transform, data analyser, user experience provider all rely on external data to be analysed. Support services and consultancy focuses on its employee time management and make or buy decisions, much like typical consultancy or manufacturer companies do.

Relationships were seen as; (1) a sales channel, (2) a way to increase public relations and knowledge, (3) a way to drive and encourage the developer and analyser community in co-creation, (4) a way to influence government to open up preferred data-sets, and (5) a method in managing big projects by utilizing help and resources from partners. The relationships ranged from independent hacker-like developers to small companies, governments, municipalities, labour unions and publicly funded open data projects.

Revenue models between the five profiles are very different, varying from pro-bono open source model to project work model. The revenue model selection is influenced by several aspects of the business operation, but mostly it correlates with the underlying customer base. For example, the user experience providers work directly with end-users, and hence apply consumer oriented revenue models such as advertising, subscription, or one-time fee. The data analysers, on the other hand, work typically with business customers, thus applying business-to-business revenue models such as project work or transaction based pricing. The choice of the revenue model is affected also by the external market forces such as competitors' actions and the uniqueness of the offering.

Management mind-set affects strongly to the chosen revenue model and ambitions of the company. Some managers were quite happy with their product being a secondary income whilst keeping their day job somewhere else. Others were pursuing start-up strategy with external investments and hiring new employees. Louhos was purely pro-bono, having released their toolkit as open source free for anyone to use. Many companies used free visualizations or applications as a marketing tool for their other products.

6. CONCLUSIONS FOR MEDIA COMPANIES

This chapter answers to the original research question of “*how media companies can utilize open data in their business*”. Chapters 4 and 5 examined what business models the open data companies are utilizing in Finland, how have they capitalized the benefits of open data in

their operation, and what kind of value network is driving the ecosystem. These findings are essential in order to answer to the original research question.

The general open data value network presented in the end of Chapter 5 is redrawn from media perspective in Figure 18 below. Media companies are placed in analogous role as user experience providers in the earlier value networks; they have a direct connection with end users. The figure has made also some simplifications in order to better visualize media companies' position in the value network. First, the support services and consultation has been omitted because they are not directly linked in the value chain. Secondly, the data flow between the value network profiles has been simplified by leaving out some connecting arrows. The general idea and relation between the value network entities is nevertheless kept the same.

The value network offers three opportunities for the media companies. Firstly, raw open data resources can be used to create data journalistic content for the end users and to increase transparency in the articles. This activity is emphasized with “opportunity 1” box in the figure below, which encompasses the arrows coming from the open data source and from the extract & transform entity. Secondly, media companies can utilize data analysis done by third-party analysers and create new articles or content based on those. This activity is emphasized with “opportunity 2” box below, and it encompasses the arrow coming from the data analysers-profile. Thirdly, media companies could publish their own data to be analysed and refined by the open data community. This activity is emphasized with “opportunity 3” box in the figure. These three opportunities will be described in more detail in the following chapters.

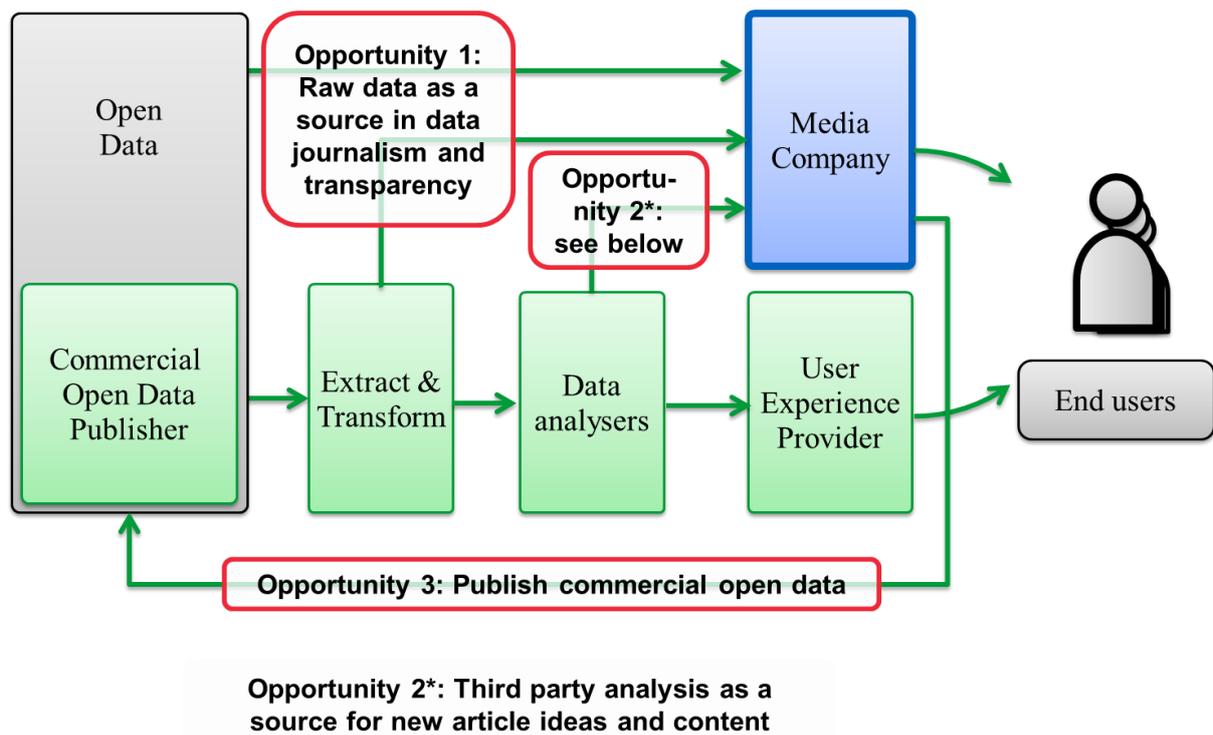


Figure 18 Open data value network from the media context

6.1 Opportunity 1: Raw data as a source in data journalism and transparency

One of the most palpable ways for media companies to utilize open data is to use it as a resource in data journalism. Data journalism is a form of reporting, which derives its article ideas from novel data interpretations, such as visualizations or numerical analysis. Typically the analysis is done by professional data journalists working on a media companies' payroll. They retrieve the data from a data source, manipulate and analyse the data with various tools and finally make an interpretation and write an article based on it. Data journalism can give more depth in to the articles. If open data has been used as a source, the readers can also confirm the results and argumentation themselves, thus leveraging the transparency and credibility of the article. Alternatively, if the analysis is done from restricted data, the reader has no way of knowing whether the data journalist interpreted it correctly or not.

The business rational for data journalistic stories is increased traffic and advertisements. For example, The Guardian has discovered that readers spend more time with data-journalistic articles than with regular articles. However, this is not always the case, and often time and resources investment in data-journalistic stories does not pay off. (Aitamurto, 2011, p. 14).

6.2 Opportunity 2: Third-party created analysis as a source for new content and article ideas

The analysis of the data can also be performed by professional or non-professional third-party freelance analysers instead of the media company's own staff. HS Open case demonstrated how the non-professional analysers can be nudged and encouraged to perform analysis on preferred topics. For example, as we saw in Chapter 4.11, a bioinformation technology student had utilized an algorithm made for network analysis of genes in generating a visualization of electoral funding. These types of multidisciplinary analyses would be very difficult to conduct with in-house personnel.

However, the Finnish freelance data analysers lack a marketplace where they could sell their work onwards to media clients. It is very cumbersome to contact and deal with several newspapers every time a new visualization needs to be sold. The lack of marketplace has led to a situation where most analysers work on their spare time for pro-bono causes, typically publishing their visualizations on their website or blog for free. The situation is very different with e.g. journalistic photographs, where photo agency STT-Lehtikuva connects freelance photographers to media companies. Similar intermediaries or exchanges could be beneficial for the data visualization ecosystem as well.

The circumstances might change in the future as Helsingin Sanomat considers launching a data visualization ecosystem, where they would compensate X euros per publication for the author (Mäkinen, 28.2.2012). Accompanied with inspiring events, such as HS Open, the reward-convention might increase the interest towards data visualizations in general. However, this would solve the issue only from the perspective of one newspaper; the market would still lack a marketplace connecting analysers to several different media companies.

6.3 Opportunity 3: Publish commercial open data

Third opportunity for media companies is to open up their own commercial data for others to analyse and utilize. Most media companies operate on a two-sided markets business model; on the one side are the readers who consume the articles and possibly pay for the content with subscriptions or pay-per-use, and on the other side are the advertisers who typically contribute majority of the media's revenue (see Chapter 2.6). Opening up commercial data should not conflict with either of these revenue sources, unless the company changes their entire business model on which they operate. Taking this into account, could there still be situations where publishing data would be beneficial?

This research found out that the companies have two basic options to publish the data: with or without restrictions for re-use. Only the option without restrictions represents open data as defined by Open Knowledge Foundation (Open Definition). However, as stated in Chapter 2.4, this paper uses more relaxed definition for open data, thus allowing some restrictions in the data re-use as well.

In addition to licensing issues, the technical form of the released data needs to be decided as well. In its simplest form the data can be released, e.g., as a comma separated text file and placed on a web server. A more advanced option would be to offer a custom API, letting developers access the data faster and thus allowing more complex application areas. API is more suitable in large databases or in cases where the data is constantly changing, such as the HSL Live position tracking of busses and trams. However, many interviewees expressed that the technical form of the data is not important as long as the data is simple to access and utilize, and it has lucid license allowing its re-usage.

The most difficult choice is to select whether the re-usability of data is somehow limited. The research done within this Thesis is not sufficient to offer an exhaustive answer on selecting the best option. However, some remarks can be made based on the case companies studied in Chapter 4. These findings are summarized in monetary value of data for the releasing company – limitations applied matrix in Figure 19 below. The next two chapters will discuss both limited and un-limited approaches in order to explain Figure 19 in more detail.

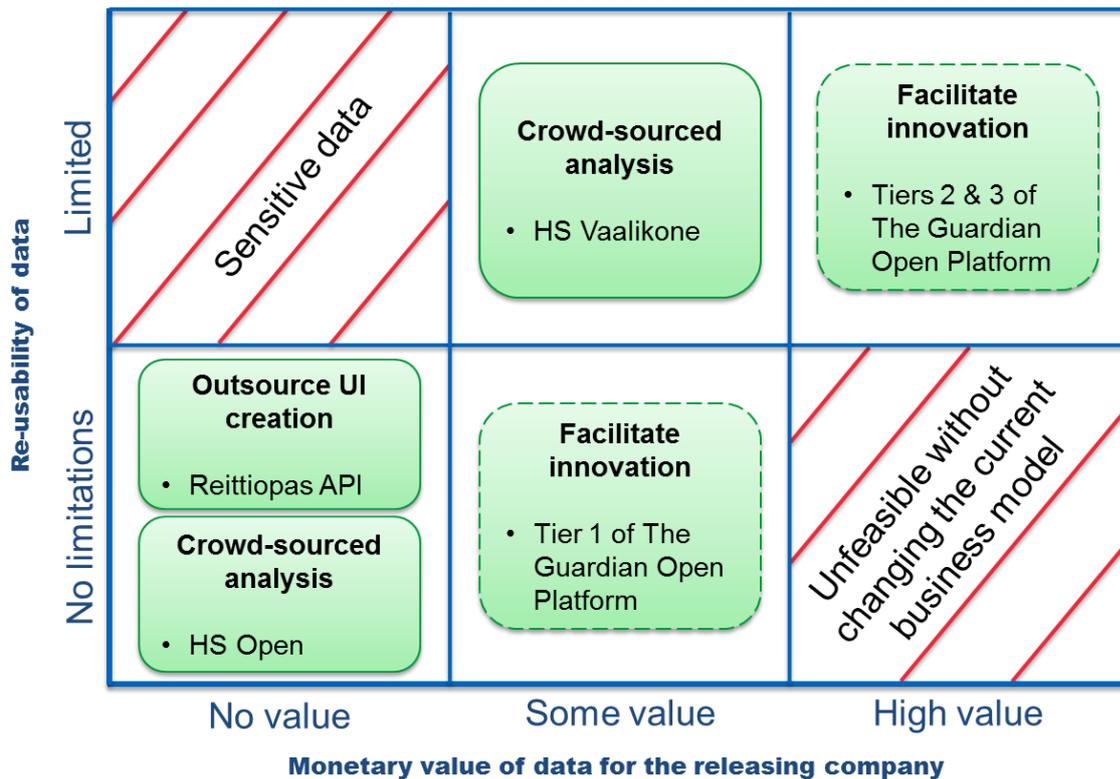


Figure 19 Monetizing value of data – limitations applied matrix

6.3.1 Publish data with no limitations for re-use

No limitations are practical in situations where the data does not offer direct monetizing opportunities for the company. Several open data cases have demonstrated that often data which seems worthless for its owner might find valuable application areas when released and someone else gets a look at it. In this situation it would be reasonable to release the data with no limitations, since it would be worthless for its owner to begin with. This option is represented by HS Open case in lower-left corner in Figure 19. By releasing the data, the publisher would gain general goodwill and possibly some third-party analyses to be used as content in new articles. Furthermore, in case the data has already been used as a source in a data journalistic article, releasing the raw data increases transparency and trust towards the journalistic process and the entire newspaper. While these benefits are vague, it is good to keep in mind that the data was worthless for the company to begin with, and thus, even if a little value is created this approach is vindicated.

Another situation when open license would be reasonable is when there is a need to create a new user interface for some information. This is because creating a new user interface, for example a new mobile client, is very expensive. Scaling the client to multiple platforms and

handset models, and updating the client as new operating system versions come along is a cumbersome task. It requires investments, business case calculations, and large enough predicted user base for the project to be reasonable. Clearly not all potential projects fulfill these requirements, and thus many user experiences are left undone. For example, when HSL launched its application contest, it received 63 submissions utilizing Reittiopas API (lower-left corner in Figure 19) in new mobile client or web user interface (HSL Mobiilikisa). Generalizing this in to media companies' circumstances, releasing data does not necessarily create direct revenues, but in the long run it might gravitate masses towards media house's other services and thus be beneficial.

Even if the data has some value, it might be justifiable to release it without restrictions, if it would be too expensive or too risky for the company to monetize it with in-house products. This option is represented by The Guardian Open Platform Tier 1 example in the lower-middle compartment in Figure 19. Developers trying, failing, and trying again different user interfaces and mashing up the data with other data sources creates an innovation environment which would be very difficult to replicate with in-house resources. The Guardian example will be described in more detail in the next chapter.

Finally, the lower-right corner represents a situation where the data is very valuable for the company. In this case, publishing data without any restrictions would be unfeasible without changing also the underlying business model. It is, however, possible to operate in this corner of the matrix. For example, typical open source companies, such as Canonical or Arduino, give out their source code for free, because their revenue model is founded on, e.g., support services.

In general, any data which does not directly pose a threat to the existing revenue of the media company and which cannot directly be monetized with in-house products could and should be released as open data. The benefit for releasing the data generates goodwill for the media house, increased transparency, and engenders crowd-sourced analyses and user interface innovation.

6.3.2 Publish data with limited re-use

Another option is to release the data with some re-use limitations. For example, HS Vaalikone (see Chapters 4.11 and 5.4.2 for details) made good results by publishing the data with commercial re-use prohibition, while still allowing developers to use the data for other purposes. In addition to several crowd-sourced analyses, releasing the Vaalikone data also

increased transparency and trust towards the entire Helsingin Sanomat Vaalikone system. This strategy represents the upper-middle compartment in Figure 19. The approach is quite safe for the data publisher, because in order to go commercial and to make profit, the developers need to negotiate with the company. Therefore, the publisher maintains control over the data commercialization, while still leaving enough room for developers to experiment and innovate with the data.

In case the data is highly valuable for the company, the non-commercial limitation alone might not be sufficient to protect the core business. The research corpus examined within the empirical part of this Thesis did not cover this situation, and thus cannot provide guidelines or best practises. However, a quick glance to the international market revealed The Guardian's Open Platform, which is a good example of very valuable "*open data*" released with license restrictions (upper-right corner in Figure 19). Open Platform lets developers access The Guardian's articles using a three-tier admission system. The first tier lets the developers access the headlines, but not the article body. The second tier, which requires registration and an API key provided by The Guardian, grants the developers a full access to the article body as well. In both cases, the usage of the data costs nothing for the developers, and they can even keep all the profits from their commercial activities. The only catch is that the developer is required to show The Guardian's advertisements in the article body. The third tier offers ad-free access to all The Guardian's content, but requires a contract with the newspaper. (The Guardian Open Platform). Since The Guardian case was not in the main corpus of this research, it is drawn with dashed lines in the upper-right corner and the lower-middle compartment of Figure 19.

Open data is placed within quotation marks in the above example because the restrictions imposed on the data are in direct contradiction with open data's definition. Since the definition does not capitulate well in situations where companies are releasing essential data on which their entire business is relying on, a more relaxed definition of open data was introduced back in Chapter 2.4. The point is not to practise terminology acrobatics, but to encourage enterprises to become part of the open data sphere while still guarding their intellectual properties. A multi-tier licensing model, such as the one employed by The Guardian, would let third-party developers to experiment and even run small-scale businesses, while still keeping the data in the hands of businesses. Without this exemption, it is difficult to see how corporations could release data on which their core business is dependent.

Whether the limitations are prohibiting commercial re-use, requiring authentication from the developers, or something else, it is important to keep the data open enough to maintain developers' interest towards it. The entire point of opening up data is to make it easily accessible for the developers, so they can make quick mash-ups, visualizations or other analysis on it. On the other hand, companies need to protect their intellectual properties in order to sustain their business. Therefore, opening up valuable data is balancing between restrictions and openness to benefit both needs.

Finally, the upper-left corner in Figure 19 represents data which does not necessarily have direct monetizable value for the company, but which still might jeopardize its operations if released. This is typically sensitive data which should to be kept within the company premises.

6.4 Summary

Media companies operate in a position comparable to user experience providers in the open value network. This observation leads to three opportunity avenues for media companies within the value network (visualized in Figure 18):

- (1) Use raw data as a source in data journalism and transparency. This requires more effort and data analysing skills from the newsroom journalists, but also makes the stories more interesting and prolongs the time readers spend with the article. However, not always the data-journalistic articles become successful, and therefore, there is a risk in putting the extra effort to write a data-journalistic article.
- (2) Use third-party analysis as source for new article ideas and content. This is closely related to the previous opportunity, with the exception that the analysis is done by third-party analysers. They can be motivated and guided with hackathon events, such as HS Open, but they can work on their own as well. The third-party analysers can be hobbyist, creating visualizations on their spare-time, or they can be professional freelance data journalists working for several newspapers. Utilizing third-party analyses reduces the risk associated in creating the in-house data-journalistic articles.
- (3) Pursue cost savings, transparency, and goodwill by publishing commercial data. Cost savings can be achieved by either crowd-sourcing data analysis or by letting developers innovate new user interfaces based on the data. When publishing

commercial data, media companies should pay attention to re-usability restrictions of the data. Restrictions are necessary in situations where the data is essential for the company's core business, but if releasing the data does not inflict a direct threat to the revenue, applying limitations would be futile. Opening media's own resources, for example as was done in HS Vaalikone case, increases transparency and trust towards the newsroom because the users can replicate the analysis done in the paper or make their own ones. These third-party analyses can give a second opinion to the newsroom's internal analyses, or might even discover a new perspective to the entire story.

6.5 Feedback from the interviewees (feedback round #2)

When the Thesis was in its final stages, it was sent for the interviewees for a second round of feedback on 14.11.2012. The respondents were given several weeks to reply, and all except one informant responded. The response rate is very good, and even the company which did not respond has approved the first draft concerning their company back in June 2012.

In general, the feedback was positive, and the interviewees considered the subject very topical and interesting. Many interviewees defined their comments, especially relating to the co-operation partners. Some respondents stated that they had gotten new revenue sources or otherwise grown their operations, but these comments were omitted in order to keep the report as a consistent snapshot of the research timeframe between March to May, 2012. Therefore, any new contracts or changes to the business model, which had occurred after the given timeframe, were not included in the analysis.

Some respondents wished to emphasize certain aspect of the interview. For example, the representative from Hahmota emphasized that the Apps 4 Finland 2009 submission was based on fictional municipal financial data. The representative from Cloud'N'Sci emphasized that their algorithm solution works as-a-service, and thus it cannot be packaged, installed or sold like regular software license. Reitit for iPhone had changed its name to Reitit, and wished that the company would be addressed with the new brand. These emphasis and changes were incorporated in to the corresponding company cases.

The representative from HSL gave plenty of valuable feedback, corrections, and suggestions in to the Chapter describing HSL Reittiopas API. Most importantly the informant stated that there has been a mobile version of the Reittiopas available at m.reittiopas.fi for 10 years, and

they are currently working with new version of it. Also, the informant stated that HSL offers other APIs as well, but Reittiopas API is the most popular. HSL has had the API from the beginning of the Reittiopas service in 2001, but they did not open the API for public until 2009. Before this the API was used internally, and also in occasional partnership campaigns. The informant also clarified their motivations; initially HSL did not search for cost savings by releasing the API, but instead wished to see what new could be achieved. In addition, the information produced with citizens money should be free to use. The informant also clarified that they do not plan to crowd-source the entire UI development; their own resources are still directed to service development of web-based services.

Practically all of the feedback related to the respondents' corresponding case companies, and very little general comments or comments towards other case companies or their analysis were given. The only exception came from the CEO of Flo Apps, who commented "*I would be careful in using Asiakastiето as an example, because a true "open data" requires machine readable interfaces and Asiakastiето utilizes several non-open data sources; albeit this is a line drawn in to the water.*"

This is important and accurate comment. In addition to open data, Asiakastiето uses also public data and privately collected data. To respond in to the comment, the reason why Asiakastiето was considered to be an open data company is simply because open data is part of its offering, despite the fact that it uses other data sources as well. Similar argument could be applied to Duunitori.fi, since they also combine open and private data together in order to create a seamless service experience. Therefore, as long as at least one part of a company's offering was based on open data, it was included in this Thesis.

7. DISCUSSION OF THE RESULTS

7.1 Theoretical contributions

The purpose of this paper was to study the Finnish open data landscape in order to reveal the business models the companies are operating with, the value network in which they function, and to analyse how the findings can be reflected into a media company context. As a result, the Thesis identified five value network profiles, which led to three opportunity avenues for media companies.

Comparing to the previous academic research (Table 5), this Thesis inspected the value network profiles on an individual company level. That is, for each value network profile, at least one corresponding case company can be pointed out as evidence that the profile exists. This approach differs from the one taken by Kuk and Davies (2011) and Poikola et al. (2010), who focused on open data artifact processing. Tammisto and Lindman (2011) had similar research approach than this Thesis, building the value network from case companies and previous academic research, but this research utilized a larger and more balanced research corpus.

Table 5 Value network comparison to the previous research

	Research question	Approach	Value network/chain employed
This research	Media companies' opportunities in open data	Explorative case study of 15 Finnish open data companies in order to build a value network and to make interpretations for media companies	<ul style="list-style-type: none"> A. Open data publishers B. Extract and transform C. Data analysers D. User experience providers E. Support services and consultation F. End users
Latif et al. (2009)	Commercial uptake of the Semantic Web vision	They presented a linked data value chain, and applied it to BBC case example	<ul style="list-style-type: none"> A. Raw data provider B. Linked data provider C. Linked data application provider D. End user

	Research question	Approach	Value network/chain employed
Poikola et al. (2010)	(1) How data resources can be opened in a controlled fashion (2) Reveal the building blocks around the open data ecosystem (3) Guidelines for open data principles	Extensive amount of expert interviews and material	A. Data recorder B. Data refiner C. Data aggregator D. Data harmonizer E. Data updater F. Data publisher G. Registry maintainer H. Application developer I. Data interpreter J. End user
Tammisto and Lindman (2011)	How data service providers capture value	Explorative case study with four interviews in three open data companies	A. Raw data provider B. Open Data consultant C. Linked Data developer D. Applications developer E. End users
Lehtonen (2011)	Overview of the development in the field of open data in Finland	Expert interviews and a selection of open data initiatives	A. Data filtering / data mining B. Data organizing C. Data visualization D. Data interpretation and production
Kuk and Davies (2011)	How hackers create and reshape services out of public datasets	A multimethod study of the open data hackers in UK	A. Cleaned data (producer) B. Linked data (producer) C. Software source code (developer) D. Software source code (sharer) E. Service technologies (innovator)

Despite the different research methodologies, the basic functionalities of the proposed value network profiles are very similar between all the scholars. Data publisher, data cleaner and refiner, data analyser or visualizer, and human end-user were present in all value networks

(see Table 5 and Chapter 2.3 for details). The naming conventions and exact definitions varied, but the basic principle of these activities remained the same.

Therefore, the research contribution of this Thesis is confirming the value networks presented by the preceding scholars with a representative sample of Finnish companies. Because of the explorative case methodology, the high number of analysed companies, and the unbiased selection process of the interview companies, the resulting value network is founded on a representative cross-section of the Finnish open data companies and it can be said to be inclusive. That is, with a very high likelihood the Finnish open data value network has been wholly depicted in this paper. This value network is re-represented in Figure 20 below.

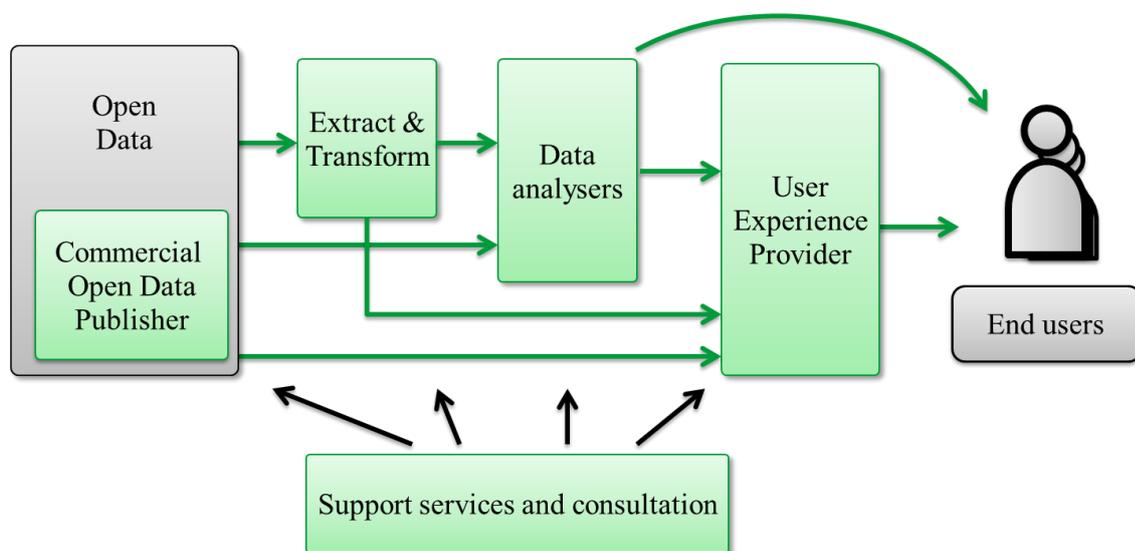


Figure 20 Re-representation of the value network found out in this Thesis

Another contribution of this Thesis is the descriptions of the 15 Finnish case companies' business models. These descriptions can be used by other scholars as a comparison point in their research. That is, subsequent open data research from Finland can compare their results back to the findings of this Thesis, in order to see how the industry has evolved over time. Foreign researchers, on the other hand, can utilize the results as a comparison point across countries.

This research also contributed three open data opportunity avenues (Figure 18, Chapter 6) for media companies: (1) utilize raw data as a source in data journalism and transparency, (2) use third-party analysis as a source for new article ideas and content, and (3) publish commercial open data to let third parties innovate new user experiences and analyses. These three avenues are grounded on the value network profiles found in the empirical part of the Thesis.

Finally, the Thesis explored decision criteria for companies to select which data to open. This criterion was explained in Chapter 6.3 and visualized in Figure 19. It divided the data in to no value, some value and high value data and proposed either releasing it with open license or restricted license based on the examples found during the research process. The decision criteria came as a by-product of the value network research, and therefore it did not receive more emphasis in this Thesis. However, the data publishing decision criteria is an interesting topic for further studies.

7.2 Limitations of the results

The results were acquired with a process described in Chapter 3. In short, the companies to be examined in the empirical part were acquired with a non-biased process utilizing Apps 4 Finland competition submissions and snowball sampling method. The process ensures that the set of companies interviewed and examined is well balanced, and represents the current Finnish open data scene. Despite the rigorous selection process, however, there is a risk that some key companies might have been left without inspection, resulting in a missing object in the value network. This risk has been minimized by examining the results of the preceding open data value network research (see Chapter 2.3 for details), and reckoning in the findings of these scholars as a foundation for this research. The likelihood that all the previous researchers would have by accident omitted some part of the value network is fairly low, and thus it can be said with high confidence that the value network presented here is wholesome and additional interview companies would not have affected the results. However, interviewing more companies probably would have brought alternative business models within the existing value network profiles.

A real blind spot for this research is the possible “*hidden usage*” of open data within companies’ internal applications. For example, if open data is utilized between company siloes or governmental units, it would not have shown up in Apps 4 Finland competition, and the interviewed experts might be unaware of it as well. Yet, the hidden usage would generate value for its users, and possibly affect the value network presented in this Thesis. The size and type of this hidden usage, and whether it exists in the first place, remains unknown. Therefore, when generalizing results from this research, one should keep in mind the possibility of such activity.

In addition, the results of this research represent only Finnish market. The market situation is somewhat different in, e.g., US or UK, where the open data movement has longer roots. This

deficiency has most likely caused regionally biased results in this research. The Guardian's Open Platform example alone proves this point, since the Finnish market was completely lacking such activity. On the other hand, neglecting abroad companies has made it possible to explore the Finnish open data scene in more depth, thus giving better results. The regionally biased results of this Thesis could be improved by conducting another case study abroad, which would replicate the research process of this study and compare the results back to the findings from Finland. Together these studies would form more comprehensive outlook on the open data value network.

7.3 Future research

The international comparison study is an obvious continuation for this research. Replicating the research in another country would give an excellent opportunity to compare the development phases of the open data markets between the countries. It would also test the value network found in this Thesis in another country in order to find out whether all the value network profiles are included. Even if no new value network profiles would be found, the international comparison would be beneficial because the additional company cases would accompany the business models found within this research.

Another future research topic would be the conditions, licenses and practises relating to commercial data openings. This research gave strong indication that corporations have a lot of potential to open up their data reserves, but could not give an exhaustive answer on how exactly this should be done. Further research could examine, for example, the licensing practises of successful companies, and create a licensing framework and guidelines to make it easier for others to follow. The licensing framework would help a company, which is new to open data, to consider which part of its data assets to publish and under what conditions.

APPENDIXES

APPENDIX A: QUESTIONS FOR OPEN DATA BUSINESS MODEL INTERVIEWS

I. Offering

1. For what do your customers use your service?
2. Are there any additional services offered to complement the product?
3. Who are the customers using your service? *Find out at least:*
 - a. *If consumers, then what age, income class, tech awareness*
 - b. *If businesses, then what industry, what size of companies, who is the buying business unit*
 - c. *Where do they live, Finland, Nordic, Europic, Global?*
4. Do you customize your service for different geographical areas or customer segments?
5. How do you distribute the service to the customers?
6. How would you characterize the scalability of your service offering?

II. Resources

1. What are the key open data sources you are utilizing?
2. What other resources do you utilize to provide your service?
3. Have you encountered any obstacles in employing open data sources? E.g. with technicalities, licenses, etc.
4. Are there some type of data sources you would like to use, but which are unavailable or under too strict commercial license?
5. Can you think of any threats involved utilizing open data sources in business?

III. Relationships

1. Who are the key commercial actors in your business network, and what activities do they perform?
2. Who are your key partners in the open data community?
3. Who in your business network owns the end user information?
4. Do you have any competitors?
5. Have you stimulated any open data community involvement?
 - a. *If, then how have you leveraged from these activities?*

IV. Revenue Model

1. What are the main sources of revenue? *Find out at least:*
 - a. *Who pays to you (from whom do you get the revenues)?*
 - b. *At which point of the transaction do you get paid?*
 - c. *How frequent and recurring the payments are?*
2. How (on what basis) is the service priced?
3. Have you considered other potential revenue flows you could utilize in the future?

V. Other questions or comments emerged during the interview

1. Can you think of other open data companies, which should be interviewed in this Thesis?

REFERENCES

- Afuah, A. (2004). *Business Models: A Strategic Management Approach*. McGraw-Hill Higher Education. ISBN 978-0-07-288364-0
- Aitamurto, T., Sirkkunen, E., & Lehtonen, P. (2011). *Trends in Data Journalism*. Next Media, a Tivit programme.
- Aitamurto, T., & Lewis, S. C. (2012). Open Innovation in Digital Journalism: Examining the Impact of Open APIs at Four News Organizations, Forthcoming in *New Media & Society*
- Albarran, A. B. (2010). *The media economy*. New York: Routledge
- Andersson, C. (2009). *Free: The future of a radical price*. New York: Hyperion.
- Berners-Lee, T. (2006). Linked data - Design issues.
<http://www.w3.org/DesignIssues/LinkedData.html>
- Berners-Lee, T. (2012). Introduction. In J. Gray, L. Bounegru, & L. Chambers (Eds.), *The Data Journalism Handbook* (pp. 1-21). O'Reilly Media, Inc.
- Chesbrough, H., & Rosenbloom, R.S. (2002). The role of the business model in capturing value from innovation: evidence from Xerox Corporation's technology spin-off companies. *Industrial and Corporate Change*, Volume 11, Number 3, pp. 529-555.
- COSMOS Corporation. (1983) (cited in Yin 2003) Case studies and organizational innovations: Strengthening the connection. Bethesda, MD: Author
- Creative commons website. (n.d.). Retrieved on September 9, 2012 from <http://creativecommons.org/licenses/by-nc-sa/3.0/>
- Creswell, J. W. (2009). *Research design: qualitative, quantitative, and mixed methods approaches*. Sage Publications Inc. ISBN 978-1-4129-6557-6
- Finlex Money Collection Act 31.3.2006/255,
<http://www.finlex.fi/en/laki/kaannokset/2006/en20060255.pdf>
- Fitzgerald, B. (2006). The Transformation of Open Source Software. *MIS Quarterly*, 30:4, 587-598.
- Google Trends. (n.d.). Web Search Interest: "open data". Retrieved on October 2, 2012 from <http://www.google.com/trends/explore#q=%22open%20data%22&cmpt=q>

- Hagiu, A., & Wright, J. (2011). Multi-Sided Platforms (Working Paper No. 12-024). Harvard Business School.
- Helsingin Sanomat. (2012, August 23). (Press release) Helsingin Sanomat improves its online service and increases the amount of paid content. Retrieved on November 6, 2012 from <http://sanoma.com/about-us/sanoma-news/news/helsingin-sanomat-improves-its-online-service-and-increases-the-amount-of-paid-content>
- HSL Mobiilikisa. (n.d.). Mobiilikisa. Retrieved on May 28, 2012 from <http://hslmobiilikisa.blogspot.com/>
- HSL Palvelut muissa kanavissa . (n.d.). Palvelut muissa kanavissa. Retrieved on May 28, 2012 from <http://www.hsl.fi/FI/aikataulutjareitit/avoimentiedonpalvelut/Sivut/default.aspx>
- HSL website news. (2011). HSL:n Reittiopas 10 vuotta: Keskimäärin 150 000 kävijää päivässä. Retrieved on May 7, 2012 from http://www.hsl.fi/fi/mikaonhsl/uutiset/2011/Sivut/Page_20111107082427.aspx
- KAMU Ry background. (n.d.) Mikä Kamu on?. Retrieved on May 25, 2012 from <http://www.kansanmuisti.fi/about/background/>
- KAMU Ry rules. (n.d.). Kansan muisti KAMU ry:n säännöt. Retrieved on May 25, 2012 from <http://www.kansanmuisti.fi/about/rules/>
- Kayser-Bril, N. (2011) (Cited in Aitamurto 2011) Presentation on data journalism. Retrieved on May 29, 2012 from <http://prezi.com/e7tfgnu2zpuarepublica-xi-110413/>
- Kidder, L., & Judd, C. M. (1986) (cited in Yin 2003) Research methods in social relations (5th ed.). New York: Holt, Rinehart & Winston
- Kuk, G., & Davies, T. (2011). The Roles of Agency and Artifacts in Assembling Open Data Complementarities. In Proceedings of Thirty Second International Conference on Information Systems, Shanghai, China.
- Latif, A., Saeed, A., Hoefler, P., & Stocker, A. (2009). The linked data value chain: A lightweight model for business engineers. *International Conference on Semantic Systems*, 568-575. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.181.950&rep=rep1&type=pdf>
- Lehtonen, P. (2011). Open data in Finland - Public sector perspectives on open data. Next Media, a Tivit programme.
- Louhos website. (n.d.). Datawiki: ohjeet R-laskentaympäristölle. Retrieved on October 11, 2012 from <https://github.com/louhos/sorvi/wiki/Data>
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*, Sage Publications Inc. ISBN 0-8039-2431-3

- Lorenz, M., N., Kayser-Bril, N., & McGhee, G. (2011). Media Companies Must Become Trusted Data Hubs. Retrieved on May 29, 2012 from <http://owni.eu/2011/02/28/media-companies-must-become-trusted-data-hubs-catering-to-the-trust-market/>
- McCandless, D. (2012). Introduction. In J. Gray, L. Bounegru, & L. Chambers (Eds.), *The Data Journalism Handbook* (pp. 1-21). O'Reilly Media, Inc.
- McQuail, D. (2010). *McQuail's Mass Communication Theory*. Sage Publications Ltd. ISBN 978-1-84920-292-3
- Mäkinen, E. (2011, March 16). Tuloksia HS Openista #1: Budjettivajepeli ja ehdokkaiden ammatteja. Retrieved on May 15, 2012 from <http://blogit.hs.fi/hsnext/tuloksia-hs-openista-1-budjettivajepeli-ja-ehdokkaiden-ammatteja>
- Mäkinen, E. (2011, March 28). HS julkaisee vaalikoneensa avoimena tietona ennen vaaleja. Retrieved on October 10, 2012 from <http://blogit.hs.fi/hsnext/hs-julkaisee-vaalikoneensa-avoimena-tietona-ennen-vaaleja>
- Mäkinen, E. (2011, April 12). 15 uusiokäyttöä HS:n vaalikonedatalle – viikossa. Retrieved on May 25, 2012 from <http://blogit.hs.fi/hsnext/15-uusiokaytto-hsn-vaalikonedatalle-viikossa>
- Mäkinen, E. (2012, January 3). Helsingin Sanomat julkaisee vaalikoneen tiedot avoimena rajapintana. Retrieved on May 25, 2012 from <http://blogit.hs.fi/hsnext/helsingin-sanomat-julkaisee-vaalikoneen-tiedot-avoimena-rajapintana>
- Mäkinen, E. (2012, February 10). HS Open -tapahtumissa luodaan tietojournalismia. Retrieved on May 25, 2012 from <http://blogit.hs.fi/hsnext/hs-open-tapahtumissa-luodaan-tietojournalismia>
- Mäkinen, E. (2012, February 28). Ehdotus datajournalismin bisnesmalliksi: X euroa per kertajulkaisu. Retrieved on May 25, 2012 from <http://blogit.hs.fi/hsnext/ehdotus-datajournalismin-bisnesmalliksi-x-euroa-per-kertajulkaisu>
- O'Murchu, C. (2012). Introduction. In J. Gray, L. Bounegru, & L. Chambers (Eds.), *The Data Journalism Handbook* (pp. 1-21). O'Reilly Media, Inc.
- Open Definition. (n.d.). Defining the Open in Open Data, Open Content and Open Services. Retrieved on May 16, 2012 from <http://opendefinition.org/okd/>
- Open Government Data Dashboard. (n.d.). Data Catalogs. Retrieved on October 5, 2012 from <http://dashboard.opengovernmentdata.org/catalogs/>
- Osterwalder, A. (2004). *The business model ontology - A proposition in a design science approach*. Doctoral dissertation, l'Ecole des Hautes Etudes Commerciales de l'Université de Lausanne
- Oxford English Dictionary. (n.d.). Word: media. Retrieved on October 31, 2012 from <http://www.oed.com/view/Entry/115635?rskey=8tJJxX&result=2#eid>

- Parker, G.G., & Van Alstyne, M. W. (2005). Two-sided Network Effects: A Theory of Information Product Design. *Management Science*, Vol. 51, No. 10, October 2005, pp. 1494–1504.
- Parpola, A., & Kiljala, J. (2005). *Hyvä vai paha tieto?*. asiakastieto. ISBN 952-9708-13-0
- Philliber, S. G., Schwab, M. R., & Samsloss, G. (1980) (cited in Yin 2003) Social research: Guides to a decision-making process. Itasca, IL: Peacock.
- Picard, R. G. (1989) (Cited in Albarran 2010) *Media economics*. Newbury Park, CA: Sage
- Poikola, A., Kola, P., & Hintikka, K. A. (2010). *Julkinen data - johdatus tietoverkkojen avaamiseen*. Helsinki: Liikenne- ja viestintäministeriö. ISBN 978-952-243-146-2.
- Rajala, R., Westerlund, M. (2007). Business models - a new perspective on firms' assets and capabilities. *International Journal of Entrepreneurship and Innovation*, Vol 8, No 2, pp. 115-125.
- Rajala, R. (2009). Determinants of Business Model Performance in Software Firms. Retrieved from <http://hsepubl.lib.hkkk.fi/pdf/diss/a357.pdf>
- Rochet, J.-C., & Tirole, J. (2003). Platform Competition in Two-Sided Markets. *Journal of the European Economic Association*, Vol. 1 (2003), pp. 990–1029.
- soRvi GitHub website. (n.d.). Retrieved on October 11, 2012 from <https://github.com/louhos/sorvi>
- Tammisto, Y., and Lindman, J. (2011). Open Data Business Models. The 34th Information Systems Seminar in Scandinavia, Turku, Finland.
- The Guardian Open Platform website. (n.d.). Frequently asked questions. Retrieved on May 24, 2012 from <http://www.guardian.co.uk/open-platform/faq>
- The New York Times Company. (2011, March 17). (Press release) The New York Times Launches Digital Subscriptions. Retrieved on November 6, 2012 from <http://phx.corporate-ir.net/phoenix.zhtml?c=105317&p=irol-newsArticle&ID=1540299&highlight=>
- Yin, R. K. (2003). *Case study research: design and methods*. Sage Publications Inc. (Applied social research methods series; vol. 5). ISBN 0-7619-2553-8
- Yoo, Y. Temple University Philadelphia professor, speech at Aalto University School of Economics 11.5.2012
- Web of Knowledge. (n.d.). Search. Retrieved on October 2, 2012 from <http://wokinfo.com/>