Bachelor's thesis

Bachelor in Business Administration - International Business

2019

Juho Ojala

EXPLORING THE APPLICABLE USE CASES OF OPEN GOVERNMENT DATA IN AN ADVERTISING AGENCY

– Case Study TBWA\Helsinki

TURKU AMK

TURKU UNIVERSITY OF
APPLIED SCIENCES

Juho Ojala

EXPLORING THE APPLICABLE USE CASES OF OPEN GOVERNMENT DATA IN AN ADVERTISING AGENCY

- Case Study TBWA\Helsinki

TBWA\Helsinki is an international advertising agency that's a part of a global collective of over 270 offices in 95 countries. The advertising industry has been going through a fundamental shift in the 21st century due to the emergence of new disruptive technologies and companies. From TBWA\Helsinki's perspective, the biggest change in the advertising industry in recent years has been the availability and abundance of data on any subject. It's not enough to have a great idea and a great sales pitch anymore.

The purpose of this thesis was to conduct preliminary research on the applicable use cases of open government data at TBWA\Helsinki. Open government data is any data or information produced by public bodies that is in machine-readable format, can be freely used, edited and distributed by anyone for either private or commercial purposes. There's great potential in these large, open data sets, but not many companies in the advertising space know where it could be used or how to go about it.

The research found out that there are several applicable use cases of open government data in the case company, for example, optimizing outdoor advertising or as a part of a brand action. Further, the research recommends that open government data use should focus on cases where there's a clear problem and a data need, projects with large enough scope to compensate for the effort, and where the open data can be enriched with client data.

Exploratory study was chosen as the main method for both the literature review and data collection. Primary observation and informal interviews were used to asses the current activities in the case company, and semi-structured interviews with external specialist were conducted in order to gain in-depth knowledge on the subject.

Juho Ojala

KARTOITTAVA TUTKIMUS KÄYTTÖKELPOISISTA JULKISHALLINNON AVOIMEN DATAN KÄYTTÖKOHTEISTA MAINOSTOIMISTOSSA

- TBWA\Helsinki

TBWA\Helsinki on kansainvälinen mainostoimisto, joka on osa globaalia kollektiivia, johon kuuluu 270 toimistoa 95 eri maassa. Mainosala on 2000-luvulla ollut perusteellisen muutoksen kourissa: uudet mullistavat teknologiat ja yritykset ovat haastaneet sekä vanhoja toimintamalleja että perinteisiä mainostoimistoja. TBWA\Helsingin näkökulmasta viime vuosien suurin muutos mainosalalla on ollut datan saatavuus ja runsaus mistä tahansa aiheesta. Asiakkaita ei enää voiteta vain erinomaisella hissipuheella.

Tämän opinnäytetyön tarkoituksena oli suorittaa alustava tutkimus käyttökelpoisista julkishallinnon avoimen datan käyttökohteista TBWA\Helsingin toiminnoissa. Julkishallinnon avoin data on julkisen organisaation tuottamaa tai hallinnoimaa tietoa, joka on konekielisessä muodossa ja maksutta kenen tahansa käytettävissä, muokattavissa ja jaettavissa sekä yksityisiin että kaupallisiin tarkoituksiin. Suurissa, avoimissa tietovarannoissa piilee valtava potentiaali kaupalliseen hyödyntämiseen, mutta vain harvoilla yrityksillä on käsitystä siitä, missä avointa dataa voisi hyödyntää tai mistä tulisi lähteä liikkeelle.

Tutkimus löysi useita käyttökelpoisia käyttökohteita julkishallinnon avoimelle datalle TBWA\Helsingin toiminnoissa, kuten digitaalisen ulkomainonnan optimoinnissa tai osana bränditekoa. Lisäksi, tutkimus suosittelee keskittymään käyttökohteisiin, joissa on selkeä ongelma sekä tietotarve, riittävä laajuus verrattuna vaivaan, ja mahdollisuus rikastaa avointa dataa asiakkaan omalla datalla.

Päätutkimusmenetelmä sekä kirjallisuuskatsauksessa että aineistonkeruussa oli kartoittava tutkimus. Tutkittavan yrityksen aktiviteettien ja prosessien määrittämiseen käytettiin epävirallisia haastatteluja ja observointia. Syväluotaavaa tietoa aiheesta hankittiin puolistrukturoiduilla asiantuntijahaastatteluilla.

CONTENT

APPENDICES

FIGURES

PICTURES

TABLES

LIST OF ABBREVIATIONS (OR) SYMBOLS

KPI   Key Performance Indicator

OGD   Open Government Data

API   Application Programming Interface

PSI   Public Sector Information

HRI   Helsinki Region Infoshare

1 INTRODUCTION

TBWA\Helsinki is an international advertising agency that's a part of a global collective of over 270 offices in 95 countries. This thesis is written for the Helsinki office, but best practices can also be extended to partner offices if seen beneficial. TBWA\Helsinki employs over 150 professionals from various crafts: graphic design, copywriting, photography, strategy, data & analytics, media buying, social media, innovation, product design, and others. It works with both domestic and international clients from various size categories with emphasis on large companies.

During the new millenia, the advertising industry has been going through a fundamental shift due to the emergence of new disrupting technologies and companies - mainly Google and Facebook. Traditional mass media advertising agencies are being challenged by companies utilizing new doctrines, for example, Performance Marketing, Content Marketing, and Inbound Marketing. Thus, TBWA\Helsinki and its worldwide collective has had to adapt. Because embracing change is written into the company culture, the company is constantly on lookout for new, creative ways to create better advertising.

Today, one of the key initiatives for the global collective is to increase the data capabilities in its offices. From TBWA\Helsinki's perspective, the biggest change in the advertising industry in recent years has been the availability and abundance of data on any subject. This had led to brands (= clients) re-evaluating all their marketing activities to limit risk on the outset. It's not enough to have a great idea and a great sales pitch anymore. Clients want to know who you are targeting, why you have chosen a creative direction, how you aim to reach the target audience, what you aim to achieve through this activity, and how do you justify the financial investment. In conclusion, data is seen both as a tool for greater creative work and a necessity to avoid oblivion.

The purpose of this thesis is to conduct preliminary research on the applicable use cases of open government data at TBWA\Helsinki. To narrow down the research scope, this thesis addresses only Finnish open government data. The research aim is to review literature to gain knowledge on open government data and its core attributes, open data value creation, and techniques to make data usable. In addition, the research presents secondary data on the identified use cases of Finnish open government data. That is then reflected on the activities inside TBWA\Helsinki to determine where open

government data could be used. The analysis is complemented with specialist interviews about the strengths and weaknesses of open government data, and the common challenges that companies face when trying to utilize open government data. In this thesis, the term *incorporation* is used to group the use cases of open data and the techniques to make data usable.

The researcher has set the following research objectives:

- To identify relevant open government data sources and types
- To identify notable use cases of open government data
- To identify business activities in TBWA\Helsinki where open government data could be utilized

The topic will be explored through the following questions:

- What are the strengths and weaknesses of open government data?
- How can open government data be used in businesses?
- How are data and insights currently used in TBWA\Helsinki?
- In which business activities could open government data be used in TBWA\Helsinki?

The researcher has both personal and organizational motivations to pursue the chosen topic. The researcher is passionate about data-driven marketing, statistics, and data analytics, and aims to increase his competencies in the field. He has been working in the case company for seven months. As established above, both TBWA\Helsinki and the global collective are keen to develop new ways to integrate the use of data into its operations. In addition, Finland's national Open Data portal, avoindata.fi, that was developed and launched in 2014 by the Finnish Ministry of Finance (Avoindata.fi, 2018) is upgrading its offering during 2018-2019 (Avoindata.fi, 2018). This means easier access to more data from the domestic market, which in turn makes systematic incorporation of open government data as a whole a more lucrative endeavor for TBWA\Helsinki.

The thesis begins with an introduction which contains background information on TBWA\Helsinki, as well as the research purpose, objectives and motivation. The literature review of this thesis spans across chapters 2 and 3. Chapter 2 starts with defining open government data and describing different types and sources of open

government data. Following, chapter 3 introduces different use cases for data in general and then specifically for open data, and the techniques for making data usable.

After the literature review, the research methodology is presented with the specification of research design, data collection methods, and comments regarding the research's reliability, validity, and limitations. Next, a summary of the interviews is provided. Then, two benchmark case examples of the use of open government data in Finnish businesses, Vainu.io and Reaktor, are presented. Following, the case analysis chapter covers TBWA\Helsinki's activities and analysis based on all the previous information. The thesis is concluded with the research findings accompanied by recommendations for both TBWA\Helsinki and further research.

2 OPEN GOVERNMENT DATA

Open data is machine-readable, digital information containing characters and symbols. According to Open Knowledge International (2018) open data is characterized by three attributes:

- ***Availability and Access:*** *the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form.*

- ***Re-use and Redistribution:*** *the data must be provided under terms that permit re-use and redistribution including the intermixing with other datasets.*

- ***Universal Participation:*** *everyone must be able to use, re-use and redistribute - there should be no discrimination against fields of endeavour or against persons or groups.*

This definition of open data is based on the *Open Definition*, which can be accessed in full at *https://opendefinition.org/od/2.1/en/* . The common definition of open data tries to tackle the problem of interoperability: the ability to interoperate and mix various datasets to achieve a more complete view on issues (Open Knowledge International, 2018). There are various types of open data, which can all be collected and stored in different formats, and are most often found in so-called open data portals. Other common sources of open data are catalogs by single organizations, singular reports and data sets, webpage content that can be "scraped" (gathered automatically with a script), and crowdsourced questionnaires (when the data doesn't exist yet) (Toikkanen, et al., 2017).

2.1 Open Government Data

Open government data (OGD) as a term can also refer to the philosophic notion that governments should promote transparency, accountability, and value creation by making their data available for all, or the policies that make it a reality, but in this thesis open government data is used to refer to the actual data published by different government entities (OECD, 2018). A common definition of OGD is a combination of the open data definition mentioned above, and the definition of government data. Finland's Ministry of

Finance defines open government data as any data or information produced by public bodies that is in machine-readable format, can be freely used, edited and distributed by anyone for either private or commercial purposes (Valtiovarainministerö, 2017). Figure 1 illustrates the relationship between public sector information, open government data, open data, and big data.
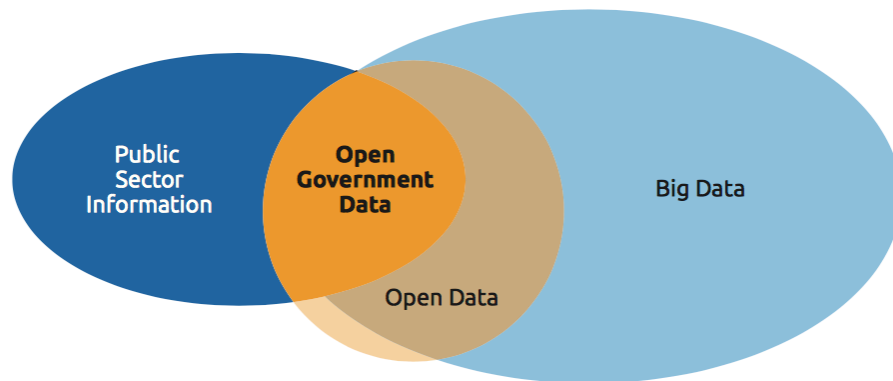
Figure 1. Boundaries of Open Government Data (Carrara, et al., 2015)

2.2 Finnish Open Government Data Providers

Finland's national open data portal avoindata.fi has 1640 data sets as of 27.12.2018, which makes it the largest open data portal in Finland. The portal aggregates data from various organizations and offers a platform to upload open data sets to. Of the 1640 data sets 853 are from regional bodies & municipalities, and 666 from government administration bodies. (avoindata.fi, 2018) Together these two categories comprise 93% of the data sets in the portal. Figure 2 illustrates the organizations within the categories and the number of data sets they have made available in the avoindata.fi portal. We can see that inside the regional bodies category, Uusimaa (and more specifically, city of Helsinki) provides the majority of the data sets. On the other hand, Ministries of Finance, Agriculture and Forestry, and Transport and Communication are the key providers in the government administration side. From the sub-agencies, Transport Agency and Environmental Institute are the largest providers.
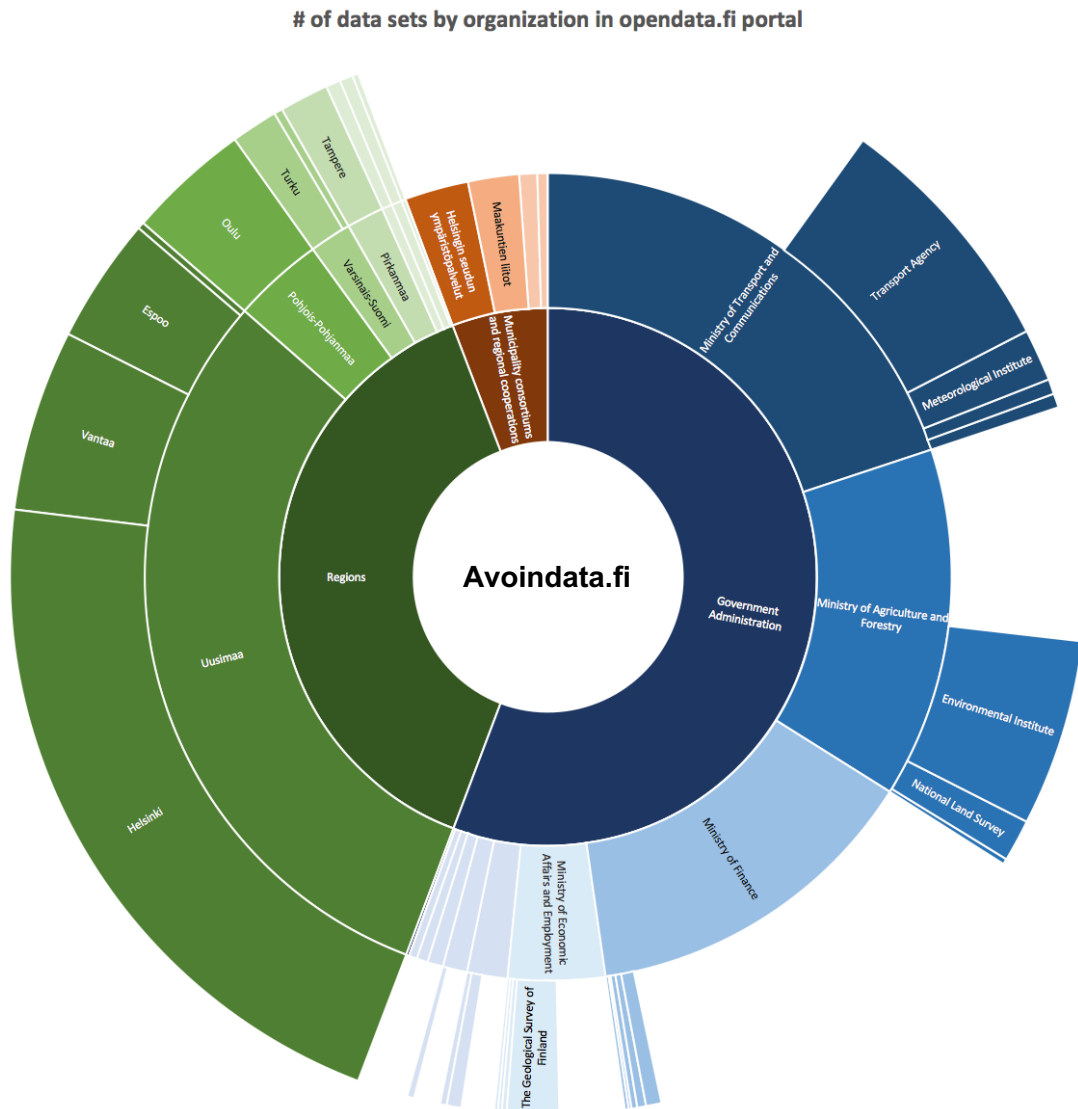
Figure 2. # of data sets by organization in avoindata.fi portal (avoindata.fi, 2018)

In addition to the national portal, there are regional open data portals created in the 6aika initiative. 6aika is a collaborative sustainable development strategy by the six largest cities in Finland (Helsinki, Espoo, Vantaa, Turku, Tampere, and Oulu), which aims to create new knowledge, business and employment. The strategy is based on three spearhead projects: open innovation platforms, open data and interfaces, and open participation and customerships. (6aika, 2018) One of the results of the open data and interfaces spearhead project was a data portal for the cities', which compiles all the available data, documents and metadata (6aika, 2018). In addition to the static data sets used in the earlier analysis, the portals also publish open government data through

Application Programming Interfaces (APIs). An API is "a set of commands, functions, protocols, and objects that programmers can use to create software or interact with an external system" (Techterms, 2019). In this case, the APIs allow users to access data from external systems by writing queries in programming languages, such as Javascript or Python. The use of APIs minimizes the amount of unnecessary data traffic by allowing the user to specify what data he wants to pull from the external system. The portals are split as seen in Table 1. The relative sizes follow the same pattern as in the avoindata.fi data set analysis with Helsinki Region Infoshare being the largest portal by far.

Table 1. 6aika data portals

Name	URL	Data sets	APIs
Helsinki Region Infoshare	https://hri.fi	647	122
Tampere Region Open Data	https://data.tampere.fi	114	75
City of Oulu Open Data	https://data.ouka.fi	83	26
Southwest Finland Open Data	https://data.lounaistieto.fi	172	26

2.3 Types of Open Government Data

Finland's national open data portal, avoindata.fi (2018), divides its open data into 13 categories (Table 2).

Table 2. avoindata.fi categories

Avoindata.fi categories	
Regions, cities	Agriculture, fisheries, forestry, food
Government, public sector	Science and technology
Energy	International issues
Justice, legal system, public safety	Transport
Education, culture, and sport	Economy and finance
Health	Population and social conditions
Environment	

These are the same categories as in the European Data Portal, which automatically aggregates data from all the national open data portals of member countries (European

Data Portal, 2018). In turn, the cities' data portals divide the data into 12 categories (Table 3).

Table 3. Cities' data portal categories

Cities' data portal categories	
Housing	Local Government
Maps	Culture and recreation
Traffic and tourism	Education
Constructed environment	Economy and taxation
Health and Social Services	Environment and nature
Jobs and Industries	Population

After this chapter gave an overview on what open government data is, what types of open government data the Finnish government bodies are producing, and where to access it, the next chapter moves to research the general ways to use of data in business activities, and more specifically focusing on value creation from open government data.

3 INCORPORATION

Data in itself is not useful, rather it can only have utility if meaning and value can be extracted from it (Kitchin, 2014). Therefore, this chapter discusses the possible uses of data in general as well as the uses of open government data in businesses, and techniques to make data usable. In the later stages these are then analysed and compared to the activities at TBWA\Helsinki to gain knowledge of the activities in which open data could be used successfully. In this thesis, the term *incorporation* is used to group the use cases of open data and the techniques to make data usable.

3.1 General Uses of Data

In *Thinking with Data* (2014), Max Shron introduces a general list of things that can be improved with data. While this list is extremely general, it delivers a basic understanding of the plethora of things where data can be useful:

- Answering a factual question
- Exploring a relationship
- Discovering a pattern
- Making a case for a decision
- Automating a process
- Judging an experiment
- Telling a story

Answering a factual question can be as simple as looking at website traffic data to see whether it had more visitors in January than in February, or something more complex, such as analyzing Customer Relationship Management system data for which salesperson has been the most effective in 2018.

Exploring a relationship between two or more variables and discovering patterns in data are common statistical problems. The relationship between variables, for example, between sleep and reaction time, can be explored to find correlations and to assess the strength of the relationship. A pattern can be defined as a set of items that occur together frequently, or are strongly correlated. For example, a grocery store chain might discover

that customers often buy Oreos and milk together. Thus, pattern discovery means uncovering these sets of items from large data sets. (Han, 2018)

Data-driven decision-making has been a buzzword for a while now. Data can help make a case for a decision in various situations: pitching a new idea to a client, A/B testing a website feature, or favouring one way of doing things over another. It can be a key element in the case such as in the A/B test, or a side-kick, as in, for example, a creative concept pitch. Data aids in judging an experiment by possibly offering a unbiased viewpoint of the success of the experiment.

Data is essential to automating processes because it gives the input to the entity handling the process, which can then be used to perform, and in some cases, optimize it. Various data sources can be combined to give a more complete outlook and free up resources in, for example, ad buying. (Forbes, 2018)

Telling a story with data in a structured way is called *data storytelling*. Data storytelling aims to communicate insights through three core elements: data, visuals, and narrative. Narrative is used to explain the data to the audience: what is happening and why is this insight important. Data visualizations enlighten the audience making the patterns and outliers pop up instead of blending them into endless rows of numbers. (Dykes, 2016)

3.2 Open Data Value Network

This section presents the value network for open data to highlight the different roles in creating value from open data. Databusiness.fi is a web site created and operated by the 6aika initiative that highlights data-based business and gives ideas on how to apply the data into business operations (Databusiness.fi, 2018). In the whitepaper "Liiketoimintaa avoimesta datasta" they elaborate on the article "Business Roles in the Emerging Open-Data Ecosystem" by Lindman et al. (2016), and present an *open-data value network* (Figure 3), which depicts the enrichment of raw data into valuable content. The network has four core roles that are linked together, and a supporting role. The first role, *Open Data Publisher* is the entity who publishes the data. *Data Extractor & Transformer* handles the importing, cleaning, and transforming of the data, while *Data Analyser* handles the visualization and modelling tasks. These tasks will be defined and discussed more in detail in the section 3.4 Techniques for making data useful.
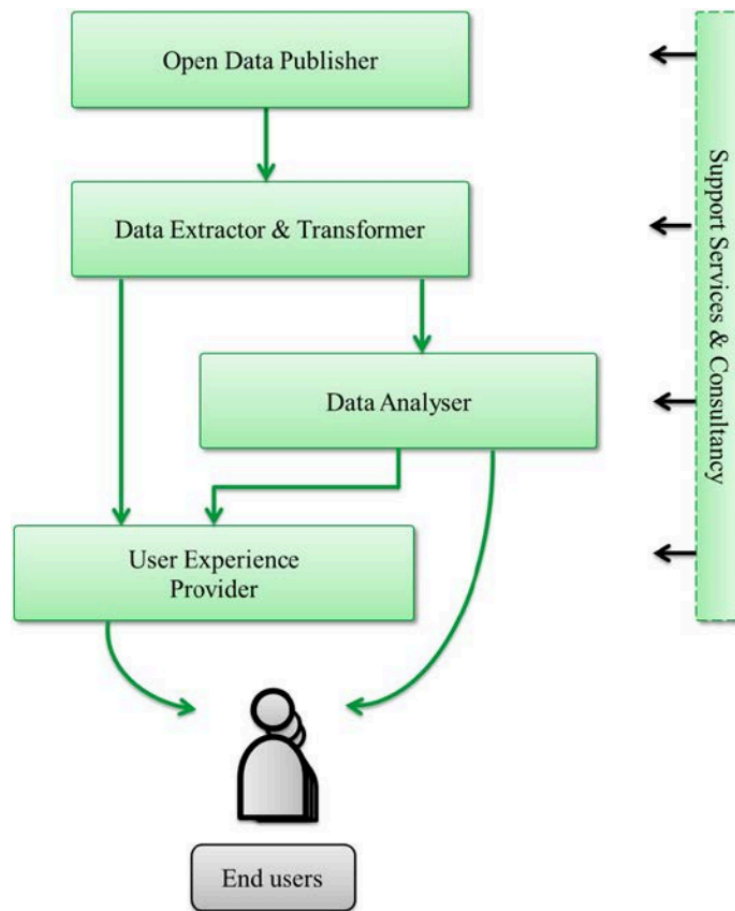
Figure 3. The open-data value network (Lindman, et al., 2016)

Figure 3 shows the relationship between different roles. We can see that all data goes through Data Extractor & Transformer, but only arrows coming from Data Analyser and User Experience Provider reach the end customer. What this means is that all open data (and therefore, open government data) has to be extracted by someone, and most often requires at least some cleaning and transforming. Businesses in the User Experience Provider role use open data as material for interesting web or mobile applications, tools, or user interfaces. In practice this can be anything where the open data is implemented in the product so that it benefits the end-user experience. For example, *Tuup,* a mobile application that offers a service that combines different modes of transport, uses an open API to incorporate public transport routes and timetables into the app (Databusiness.fi, 2017). The supporting role, Support Services & Consultancy, aids the other roles.

3.3 Open Data Business Possibilities

To understand the applicable use cases of open government data for the case company, it's essential to understand the already identified use cases. The most extensive research on the use cases of open government data in Finland is a product of the multi-year spearhead project Open data and APIs that was funded by the Europen Union. The report *Together towards Open Data Business* (Eskola & Paavilainen, 2017) by the six cities participating in the 6aika initiative outlines the use categories for open data in businesses that were identified during the three-year project:

Applications

Open data is in an essential role in many applications, with transport, map, housing search, weather, and event applications being the most prominent ones to the consumers.

Marketing

Combining different data sources aids in targeting marketing and sales, in both Business-to-Business and Business-to-Customer, and makes offering supporting solutions easier.

Products and Services

Products and services can be enriched by the use of open data. User experience can be enhanced by adding information about the environment surrounding the actual product or service, for example, a navigation app can be taught to avoid rush hours or to find a free parking spot.

Information enrichment and distribution

Business solutions such as API monitoring and data enrichment have already been sold both domestically and internationally.

Background material

Easy access to background material, for example map data, speeds up more advanced work.

Business Intelligence

Open data boosts businesses across industries: on countryside weather data improves production by optimizing procedure timings, and in cities transport data can aid in making traffic more fluent and ecological.

Data journalism

We live amongst massive piles of data. Data journalism combines these piles, and then filters, analyses, and visualizes them to create new knowledge, perspectives, and stories.

Research & Development

Open data can be used for background material for product development, identifying weak signals, and enabling collaboration.

Education & Consultancy

Due to the increasing importance of open data for the society, the demand for expert consultants and trainers will increase. In addition, open data can be used both as background and educational material.

Visualizations

Visualization transforms raw data into an easy-to-comprehend format, and makes things such as the organization's financial status more accessible for all.

The report also notes that the most prominent use cases are various applications and digital products due to the machine-readable format of the data (Eskola & Paavilainen, 2017). All of these use categories fall into either the Data Analyser or the User Experience Provider roles in the Open Data Value Network presented in the section above, which is to be expected as those are the roles that provide value to the end customer, and thus possible monetary returns to the business. It's obvious from this that there are numerous possibilities to use open data in businesses. The next section presents research on high value open data categories to further narrow down the focus for this preliminary research.

3.4 High Value Open Data for Commercial Use

European Commission's report *Creating value through open data* (Carrara, et al., 2015) presents two rankings of open data categories based on their commercial attractiveness. The OECD research from 2006 indicates the commercialization potential of different open data categories, while the Deloitte research from 2013 used the number of sectors where open data from that category could be applied as a proxy for commercial value. In both researchs' the most valuable open data categories were (in order):

1. Geographic Information
2. Meteorological and Environmental Information
3. Economic and Business Information
4. Social / Demographic Information
5. Traffic and Transport Information

The most used data of Helsinki Region Infoshare in 2018 also supports the findings of these researches: the top 5 has four sources of Geographic Information and one source of Traffic and Transport Information (Helsinki Region Infoshare, 2019). In addition, the entities that provide these types of data are the most significant open data publishers in the national open data portal (see: section 2.2).

There's no identical research of the Finnish open government data, but the 6aika open data initiative analyzed 100 different business use cases of open data and divided them into contextual categories based on the companies' fields and topics. The report notes that the categorization can't show the true division of open data business, but it identifies the top categories under which open data businesses have already been born. The categories that spurred most use cases, business, were *Economy and Taxation*, *Environment and Nature*, and *Population*. The Economy and Taxation category was used as a sort of umbrella term to group various services from both B2B and B2C fields. The high relevance of the Environment and Nature category can be seen as a manifestation of a larger environmental trend. Muurinen (2017) notes that they identified energy efficiency businesses that are using open data regarding use of different energy sources, prices, and the effects on the environment. The Population category offers lucrative sources for marketers and product & service designers to personalize and tailor experiences to the end customer. (Muurinen, 2017) Based on these findings, it is

justifiable to assume the presented research of high value categories is viable for the Finnish market as well.

3.5 Techniques For Making Data Useful

After understanding what can be done with data on a general level, one should look at the actual techniques to make the data useful. A secondary goal of this sub-chapter is to open and define these key terms. As open government data does not fundamentally differ from other machine-readable data, the same techniques can be applied to it.

A variety of field specialists' books and articles were researched to gain exhaustive knowledge of the basic techniques of making data usable. The researcher chose to review literature sources written by specialist with different perspectives (e.g. big data, programming tool guide, statistics, data mining, data journalism) to gain a more comprehensive view. Based on the researchers review, the basic steps in making data usable are almost universal and interdisciplinary.

Wickham & Grolemund's (2017) Data Science Project Workflow (Figure 4) was chosen as the framework for the standard steps in working with data due to its simple and clear structure. In addition, information regarding each step is complemented with insights from the other authors that were reviewed for this chapter. As the Enrichment isn't part of the chosen workflow but is prevalent in other texts, it is added as its own section in the end.
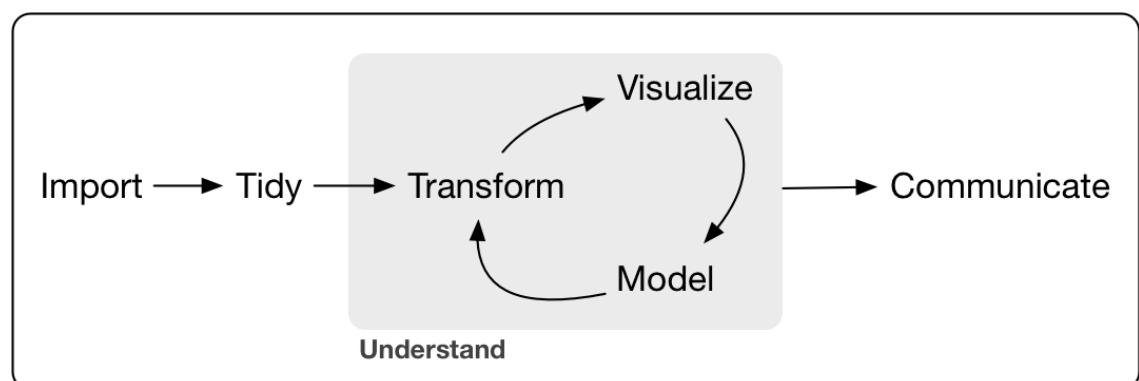
Figure 4. Data Science Project Workflow (Wickham & Grolemund, 2017)

First, the data needs to be *imported*. There are two main options for importing open government data: downloading a published data set or using an API (Application Programming Interface). Downloading a data set from catalogs (e.g. the abovementioned open data portals) is a simple, one-click task. In some data sets there's a possibility to choose the format of the file, but most often it's predefined. An API is "a set of commands, functions, protocols, and objects that programmers can use to create software or interact with an external system" (Techterms, 2019). In this case, the APIs allow users to access data from external systems by writing queries in programming languages, such as Javascript or Python. When the data being accessed is either extremely large or updates frequently, an API is the go-to-choice. (Toikkanen, et al., 2017) The data imported via either direct download or API is then loaded into the tool of choice, where the user can start performing techniques on the data (Wickham & Grolemund, 2017).

Cleaning, or tidying, the data is a part of *preprocessing* the data for the actual analysis or use. It aims to ensure a consistent form, remove noise, biases, errors, and to manage missing values (Kitchin, 2014). It's vital when dealing with large scale, real-world databases, which are highly susceptible to these inconsistencies due to their size and heterogeneous sources. Han et. al (2012) warns that low quality data leads to low quality results. For example, for analysis purposes you'd want each column to be a variable and each row to be an observation (Wickham & Grolemund, 2017). Another example is to look at the values inside the variables, are they expressed in a consistent form? Inside the variable "Gender", "Male" can be expressed with several different values, e.g. M, 1, MALE. Gray et. all (2012) add that when dealing with open government data, it has often been collected for bureaucratic reasons, which is then reflected in the accuracy of the data: if a certain variable isn't vital to the bureaucratic function it's more likely to be subject to misspellings or inaccurate content.

Kitchin (2014) defines *transforming* as reducing the dimensionality of the data. According to Wickham & Grolemund (2017) transforming includes narrowing down the amount of observations or variables (e.g. a specific country, or a variable like population count), creating new variables (e.g. population growth speed from annual population records), and calculating basic summary statistics (e.g. the average population in Finnish cities). In conclusion, transforming the data allows the researcher to cut out data that isn't needed, to condense the leftover data, and to add new, useful variables for easier visualization and modelling.

Two main methods for generating knowledge from data are *visualization* and *modelling*. Data visualizations aim to communicate data in a clear and effective way through a graphical representation (Han, et al., 2012). They aid in generating questions from the data and in getting a quick summary of the data (e.g. distribution of observations) but don't scale well because they require a human to interpret the results (Wickham & Grolemund, 2017). Figure 3 in this thesis is an example of a data visualization, more specifically a sunburst chart. Today, data visualizations are also used to create fun and intriguing graphics. *Models* are computational and mathematical tools which complement visualizations. Models scale well for larger data sets due to their mathematical nature, and certain models can be used for prediction. (Wickham & Grolemund, 2017) Testing out different models to find a good fit on the data, extracting data and patterns from large data sets is called *data mining* (Chatfield, 1995). There are numerous data mining techniques for different situations. The goal of transforming, visualizing and modelling the data is to understand the data (Wickham & Grolemund, 2017). After understanding the data, one can begin to share the found insights with others.

The final vital part of making data usable is *communication*. It doesn't matter how beneficial and ground-breaking the analysis is if it cannot be communicated to others. The role and definition of communication varies a lot between the fields, for example, The Data Journalism Handbook by Gray et al. (2012) emphasizes the story that the data conveys through words and visualizations. Communicating the useful insights is often done through the aforementioned data storytelling, which aims to communicate insights through three core elements: *data, visuals,* and *narrative*. Narrative is used to explain the data to the audience: what is happening and why is this insight important. Here data visualizations are used to enlighten the audience by making the patterns and outliers pop up instead of blending them into endless rows of numbers. (Dykes, 2016)

Enrichment in this context means combining the data with additional information, for example, market data, for deeper insights (Kitchin, 2014). A common method in data journalism is to complement the publicly available open government data with further data requested from the relevant authorities via a Freedom of Information (FOI) request (Gray, et al., 2012). On the other hand, enrichment poses the challenge of matching structures and variables from various sources. For example, two sets of data might both contain a variable called customer number with the same numbers pointing at different customers.

4 RESEARCH METHODOLOGY

The objectives of this thesis are to identify relevant open government data types and sources, to identify different use cases of open government data, and to identify business activities in TBWA\Helsinki where open government data could be utilized. As the researcher sets to explore the possibilities for TBWA\Helsinki, case study was chosen as the strategy for the research. Due to the preliminary nature of the research, a combined study was chosen as the research design, with exploratory approach chosen to fulfill the first two objectives, and an evaluative approach to apply the knowledge from the previous sections and specialist interviews to evaluate the usability of open data in different business activities of the case company. (Saunders, et al., 2016, pp. 162-176)

The primary data collection consists of a multi-method qualitative study. The chosen methods are primary observation, unstructured interviews, and semi-structured interviews. Primary observation was chosen because the researcher's position in the case company allows comprehensive observation of current methods and activities to establish base knowledge and complement other information. Unstructured interviews were used in interviewing case company employees due to the method's scheduling flexibility, informality, and the exploratory nature of the knowledge acquisition. The acquired information is used to complement the observations and internal documents to create a complete picture of the case company's activities.

Semi-structured interviews were chosen as the method for interviews with specialists. The researcher interviewed four specialist who are working directly with open government data, as well as in a position that enables them to aggregate information from numerous entities that utilize open government data. The semi-structured format allows the researcher to focus on core areas of interest, namely research questions 1 and 2, while simultaneously retaining the possibility to adapt the interview structure based on the flow of discussion, and the discovery of new, unexpected angles, which didn't come up while reviewing the literature. (Saunders, et al., 2016, pp. 388-404)

The researcher recorded, transcribed, and analysed the interviews with specialists. In addition, own notes were used to write up events that were not recognizable from the recording, such as body language and facial impressions. It was agreed with the interviewees that the transcripts of the interviews wouldn't be published, and that the researcher would ask approval for any direct quotes and comments that disclose the

interviewee to allow a more open discussion. The list of themes and key questions for the specialist interviews are available in the Appendix (Appendix 2). The researcher will only summarize the contents of the informal interviews for two reasons: first, transcribing is extremely labour intensive and would significantly limit knowledge acquirement via this method, and second, the knowledge of the fact that they are being recorded might affect the informal mood of the interview, and thus, discourage them from speaking freely (Saunders, et al., 2016, pp. 411-413). The following paragraphs comment on the research and its methods' reliability, validity, and limitations.

Reliability is concerned with the replication and consistency of the research: if an another researcher can use the same research design and end up with the same findings, the research is deemed reliable. Reliability can be divided further into *internal reliability* and *external reliability*. Internal reliability deals with the consistency within the research project and external reliability deals with the replicability of the research. In turn, *validity* is is concerned with the use of appropriate measures, accuracy of the analysis, and generalizability of the results. Validity, also, can be divided into *internal validity* and *external validity*. Internal validity is associated with statistical concepts and quantitative research, and not applicable to exploratory research, and therefore not relevant to this thesis. External validity deals with the generalizability of the results: whether the research's findings can be generalized to other groups or relevant settings. (Saunders, et al., 2016, pp. 202-207)

There are four main threats to reliability: *participant error, participant bias, researcher error, and researcher bias.* Participant error and bias consist of the factors that might affect the participant's performance in data collection, for example, an earl wake up or the proximity of a superior. Reseacher error and bias consists of the factors that might affect the researcher's performance in data collection, for example, insufficient preparation or inherent bias toward the subject. (Saunders, et al., 2016, pp. 202-207) To avoid these threats to reliability in the formal interviews, the researcher has:

- Asked the interviewees to select a time that's most suitable for them, and ensured that the interviews are conducted in a comfortable and private space.
- Shared comprehensive information about the confidentiality of the interviews and personal opinions disclosed in them, and the possibility for anonymity.
- Rigorously prepared for the interviews to eliminate the researcher error.

- Paid special attention to center on the information rather than opinions in both the interview and observation by, for example, avoiding personal remarks and writing notes in an objective manner.

Because the research has been conducted as a case study for TBWA\Helsinki, it has an inherently lower external validity, as it hasn't been designed to be generalizable, but rather specific to TBWA\Helsinki. In addition, the researcher uses multiple frameworks that are based on the Finnish open government data ecosystem and businesses, which may be radically different to the ecosystems in other countries. That said, due to the exploratory and preliminary study, the findings have moderate generalizability to other large advertising agencies operating in the Finnish market.

There were several limitations to the research. Firstly, due to the base knowledge of open data inside the company and the limited availability of previous relevant research, TBWA\Helsinki representatives requested research that would start from the fundamentals and cover different identified business use case categories to expand their knowledge of what is possible with open data. This combined with the volume limitations of the thesis led to the exclusion of cost-effectiviness, profitability, and detailed technical execution from the analysis. Therefore, this thesis can't provide answers on whether TBWA\Helsinki *should* use open government data in its business activities, but rather suggestions, ideas and hypothetical examples on where it *could* be used. Secondly, only a limited amount of experts were interviewed, and further interviews would allow for a more detailed analysis and findings. Thirdly, there is a myriad of different business activities in large advertising agencies such as TBWA\Helsinki, and as many different ideas that could be generated from thousands of open government data points. Therefore, the suggestions and ideas identified in the findings are certainly not the only possible ones.

5 SUMMARY OF THE INTERVIEWS

The researcher chose to conduct qualitative interviews to complement the literature review and secondary data with information from specialists who have concrete experience of dealing with Finnish open government data, and are in a position to aggreagate knowledge from various open data utilizers. The interviews were semi-structured with a focus on finding answers to research questions 1 and 2:

- What are the strengths and weaknesses of open government data?
- How can open government data be used in businesses?

The researcher interviewed Mika Honkanen, the owner of the Finland's national open data portal and MyData specialist, Hami Kekkonen and Tanja Lahti, specialists from Helsinki Region Infoshare, and Teemu Ropponen, the executive director of Open Knowledge Finland, which is the Finnish branch of the global organization. Together these interviewees have over 20 years of experience of the Finnish open data ecosystem. The recorded length of the interviews ranged from 30 to 35 minutes, with the total time including the pleasantries and off-the-record chats being around one hour.

Strengths of Open Government Data for the Utilizer

All four interviewees mentioned reliability as a core strength of open government data. Honkanen (2019) adds that public officials have a legal liability of acts while in office. Therefore, there's higher incentive to ensure the reliability of the data, when compared to, for example, data provided by consultancies or other commercial entities. Honkanen (2019) mentions training artificial as one use case where the reliability of the data is key to success.

Another identified core strength is the requirement for the data to be easily-accessible for everyone. There's an abundance of data and it's often in easily accessible forms. Lahti (2019) mentions that one aspect of open government data, and open government philosophy, is that the data should be available for everyone to enable democracy and transparency of government functions. This is in line with the Universal Participation attribute of open data that was presented in section 2: "everyone must be able to use,

re-use and redistribute - there should be no discrimination against fields of endeavour or against persons or groups."

Due to strict personal data protection laws, open government data never contains any personal data. This was mentioned as a benefit for companies looking to utilize open government data, as they don't have to worry about GDPR compliance.

Weaknesses of Open Government Data for the Utilizer

There were five common identified weaknesses: interoperability, findability, ease-of-use, broadness, and lack of granularity. Interoperability is the ability of a system to work with, or use the parts, of another system (Merriam-Webster, 2019). In this case, it's used to refer to the ability to combine or compare data from different sources. For example, Lahti (2019) mentions that the purchasing data of Helsinki and Espoo aren't fully interoperable as the cities have different action models and systems for accounting and purchasing process. Kekkonen (2019) says that this is an identified problem that they're trying to fix by, for example, adding common identifiers, such as Y-tunnus, in the data sets.

Findability is concerned with how easy it's to find the data or information one is looking for. All interviewees mentioned the abundance of open government data as both a strength and a weakness. A key problem for findability is the shattered nature of Finnish open government data. There are numerous entities that publish their data in numerous web sites: cities, ministries, institutes, museums. The problem has been tackled with data portals that aggregate data sets and APIs from multiple providers, with the largest being avoindata.fi. These portals then pose another challenge to findability. For example, in avoindata.fi there are over 1900 data sets and the search tools are extremely basic.

Even though ease-of-use was mentioned as a strength of open government data, it was also mentioned as a weakness due to the large variance between providers and data set quality. Some data sources are very clean and high quality, and others less so. Honkanen (2019) gives an example that in some cases the published APIs aren't user friendly and require a quite experienced developer to properly utilize it. Kekkonen (2019) notes that in instances where the data is fairly small (e.g. statistics compiled and updated in Excel), and is in such a form that transforming it into a machine-readable form would require laborious efforts, then the better option is to publish the data in a sub-optimal format, rather than not publish it at all.

Because open government data has been gathered for government processes, and not for third parties to utilize, most of the time it's not optimal from the utilizer perspective. The data sets are often vast and might contain a lot of unnecessary columns. Honkanen (2019) notes that there has been discussion regarding the service level of avoindata.fi portal. It has been decided that if there's a problem with the service or its materials, it will be fixed during the office hours (therefore, excluding weekends). Some of the companies looking to utilize the portal's data have been requesting data cleaning and transforming services (see: section 3.5), which would require large effort from the public officials.

Open data policy is implemented according to the strict personal data protection laws that were mentioned above. Due to these laws the information has to remain at a fairly toplevel, instead of having rich details. Lahti (2019) notes that the rich details are often what the companies are looking for, but the data privacy laws don't bend. This is particularly significant in the healthcare and social sectors of the government, for example, National Institute for Health and Welfare.

Other Challenges for Open Government Data Utilizers

Ropponen (2019) states that based on his experience, awareness is the number one issue for open government data utilization in Finland. Companies don't know what kind of data is available, or where and how to get it. In addition, one key challenge is the lack of competencies required for utilizing the data (see: section 3.5). Honkanen (2019) notes that the abundance of the required competencies is a key reason for why tech & software development companies are the most active utilizers of open government data.

Identifying Activities for Open Government Data Use

All interviewees agreed that open government data use should start with a problem, instead of taking certain data and trying to find uses for it. Ropponen (2019) emphasizes that the benefits of a certain data set or API aren't always obvious from the data. He gives an example that an application that showed the location of the city's plow trucks was used by disabled people to optimize their travel routes after heavy snowfalls, which certainly wasn't the original idea when thinking about opening the data.

After identifying the problem, one should refine the need for data and think if the government has any relation to the topic, and then which government entity could collect such data. Is the function under municipal responsibilities or the government, does it consider a specific ministry or institute? For example, if one needs data about environmental impact of climate change in Finland, the Environmental Institute would be the go-to choice. From there, one can find the right portal that holds the data. As seen in section 2.2, Ministry of Agriculture and Forestry's, and therefore Finnish Environmental Institute's, open data is available in avoindata.fi portal.

All the interviewees recognized that finding the right data set or API from the portal can be troublesome due to the weaknesses defined above. Kekkonen (2019) mentioned that currently one of the best ways to get an idea of the available data is to test relevant queries on both data and utilization galleries, and then manually browse the various data sets, which can be very tedious and ineffective at times. These galleries of existing utilizations of open data can be found at, for example, hri.fi and databusiness.fi. According to Kekkonen, these utilizations can give ideas of what can be done with the particular data set. Additionally, she recommends following, for example, HRI's communications to stay up-to-date on new open data and major changes.

The interviewees also reminded about the possibility of contacting public officials working with open data portals (e.g. HRI) about data needs for a problem. Then, they can point at the right direction, or make queries about the possibility of opening new government data to the public.

The Most Valuable Open Government Data

To sum up, the most valuable open government data depends on what kind business on is operating and what kind of data one needs. There's no single source of truth. On a general level, the interviewees mentioned transport, geography, and social / demographic information. These results from the interviews are in line with the identified most valuable categories presented in section 3.4. Regarding specific data sets or sources, the interviewees mentioned the Statistics Finland's data per postal code area (Paavo), the Finnish Business Register, and 3D maps of the cities.

6 CASE EXAMPLES

Two Finnish companies were mentioned by multiple interviewees as benchmarks for utilizing open government data in business: Vainu and Reaktor. Vainu.io uses open government data as a building block for its B2B prospecting tool that combines numerous data sources and connects them to companies. Reaktor builds positive brand image by using open government data in actions and applications that gather public interest. These benchmark case examples may provide valuable information on utilizing open government data in the case company, and can also be used in internal presentations to raise awareness of current applications.

6.1 Vainu

Vainu is a Finnish B2B software company founded in 2013 by three Finns. Currently, the company has offices in Helsinki, Oslo, New York, Amsterdam, and Stockholm with a total of 180 employees. Vainu is currently on a mission to "collect, read and understand all the information ever written about every company in the world, and then make this information comprehensible to everyone". (Vainu, 2019) The company extracts information from open APIs and data sets, RSS-feeds, csv-files, social media, web site source code, and web portals. As last section highlighted, data interoperability is one of the key challenges for using open government data. Here Vainu transforms the data from various (open) sources into an interoperable form, analyses it, and connects it to the correct companies. Vainu then offers this knowledge as a tool for prospecting (and, therefore, selling) better. Instead of cold calling to all the possible companies, one can find prospects based on different criteria to focus on the ones that are most likely to need your product or service. For example, Autoklinikka, a car repair chain, is using Vainu to find companies with a car fleet of a certain size. Vainu uses open government data published by Trafi to supply this information. (Vainu, 2019)

In addition to providing interoperability, a vital part of Vainu's business model is the *enrichment of data* (see: section 3.5). The Trafi data in itself doesn't provide any immediate business value, but when it's extracted, transformed, analysed, enriched, and then presented in an understandable way (e.g. web application) it has a clear value

proposition to businesses like Autoklinikka. The Trafi data can be enriched with, for example, the yearly revenue of the company, company location, profit growth.

Vainu operates in three roles in the open data value network presented in section 3.2: Data Extractor and Transformer (extracting the data from different sources and transforming it into an interoperable form), Data Analyser (providing actionable insights from the enriched data, using machine learning to predict future events and company needs, e.g. change of office), and User Experience Provider (the web application, SaaS product).

6.2 Reaktor

Reaktor is a Finnish company established in 2000 that offers consultancy and agency services that are built on exceptional technological competences. They combine human, cultural, and technological understanding to creative thinking and practical skills. The company's core competencies are in IT expert services, software development, digital design, and communications. It has in total over 500 employees in five locations, with the headquarters being located in Helsinki. Reaktor uses open government data to build a positive brand image, which is also a core task in TBWA\Helsinki. Reaktor has built and launched two web applications that use open government data to provide free services to users, Kannattaakokauppa.fi and Kaupunkifillarit.fi. In the open data value network presented in section 3.2, Reaktor's roles here would be Data Extractor & Transformer, Data Analyser, and User Experience Provider.

These web applications can be seen as brand actions (see: section 7.1), as the brand (Reaktor) is creating something concrete to connect their brand to something positive. They are tackling a problem (housing prices, city bike use) through a solution that gives them free visibility (articles in, for example, Iltalehti.fi, hs.fi and yle.fi) and highlights their core competence (web development). It's brilliant marketing in a package that doesn't feel like marketing.

Kannattaakokauppa.fi

Kannattaakokauppa.fi presents an interactive map of Finland with a prediction of housing price development by area codes. The service uses open government data, statistical

models, machine learning, and visualization to offer a tool to follow housing prices and their predicted development (see: section 3.5). The data has been imported via an API from Statistics Finland's apartment price statistics for 2010-2017 and population density statistics (Särelä, 2015). Statistics Finland publishes the averages for apartment prices on different areas, but if the average is calculated from less than six deals, it's censored. Reaktor solved this problem by building a model that takes into account the values from neighbouring areas, population density, and the correlations between the development and historical price point, which allowed them to provide prediction on areas with low total number of deals.
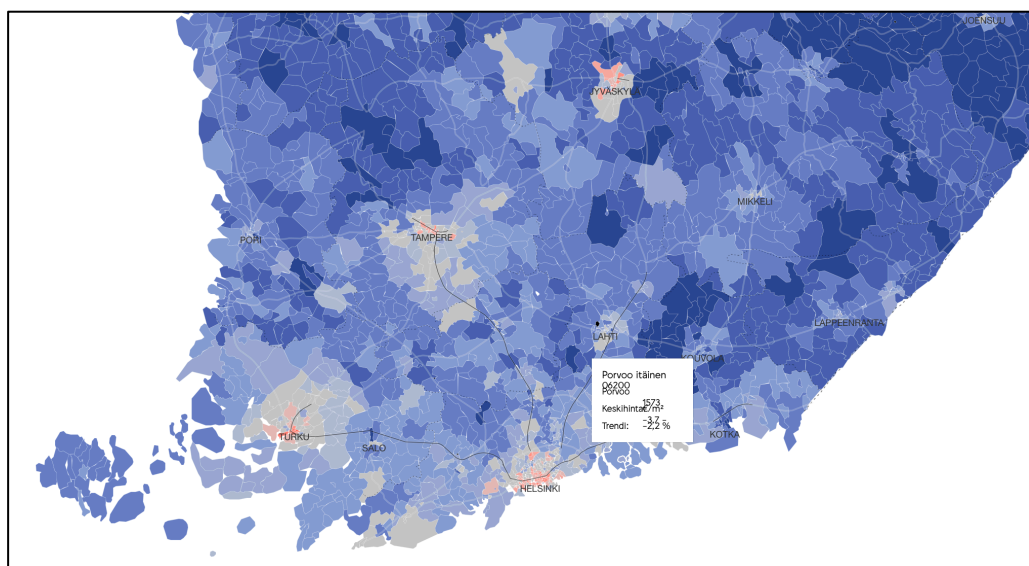
Picture 1. Kannattaakokauppa.fi

Kaupunkifillarit.fi

Kaupunkifillarit.fi web application that visualizes the availability and exact location of city of Helsinki's rental bikes. It taps into Helsinki Region Infoshare's Kaupunkifillari-API, which then supplies the locations of the bike stations and the number of available bikes. It was built by five biking enthusiasts at Reaktor. The site had over 40,000 sessions during its first three months of existence (Partanen, 2016).

7 CASE ANALYSIS

The case analysis chapter first introduces the creative executions as well as the data and insight use at TBWA\Helsinki to understand the current activities inside the case company. These activities are then compared to the open data value network and open data business possibilities presented in the literature review section to identify use cases in existing activities. The identified use cases are further analysed based on the insights from specialist interviews and the techniques to make data usable to provide recommendations on most likely activities where open government data could be used. Finally, the analysis is concluded with concrete hypothetical examples on how open government data could be utilized at TBWA/Helsinki.

7.1 Short Introduction to Creative Executions at TBWA\Helsinki

Due to the size and variety of work at TBWA\Helsinki, it's impossible to exhaustively map out all the activities that relate to creative executions. Additionally, there's no official listing and the terminology varies inside the company between different client teams. Therefore, this short introduction aims to present the core categories to give the reader an understanding of the key activities based on informal interviews with the case company employees and observation by the researcher during the past 6 months.

- The company plans, creates, and executes traditional advertising campaigns. These campaigns come in various sizes and forms: tactical social media campaign, a nation-wide television commercial, print media, or a combination.
- Brand actions are a way for brands to embed themselves to the culture: doing something concrete, instead of just advertising. Brand actions also come in various sizes, but they aim to gain visibility through making it to the news organically. For example, in the #LIFEINHEL campaign, a Chinese celebrity named Ryan Zhu moved to a small cabin inside Helsinki Airport for a month.
- Content production for websites, social media, and articles. TBWA\Helsinki manages social media accounts and aids in corporate communications.
- Creation of brand stories and visual identities.
- Product design for consumer and B2B markets.

7.2 Description of Data and Insight Use at TBWA\Helsinki

Data and Insight Use at TBWA\Helsinki can be divided into four core steps in an iterative process, which is visualized in Figure 7. The definition of data in this context is digital information containing characters and symbols, often numerical. The definition of Insight in this context is presented in the start of the next paragraph. This process depicts the complete and optimal process that is not always the case in real scenarios. For example, TBWA\Helsinki might not be in charge of all four steps, or it's executing only certain tasks inside the steps. The information in this section is based on TBWA\Helsinki's internal documents, informal interviews with the employees, and observation by the researcher, unless stated otherwise.
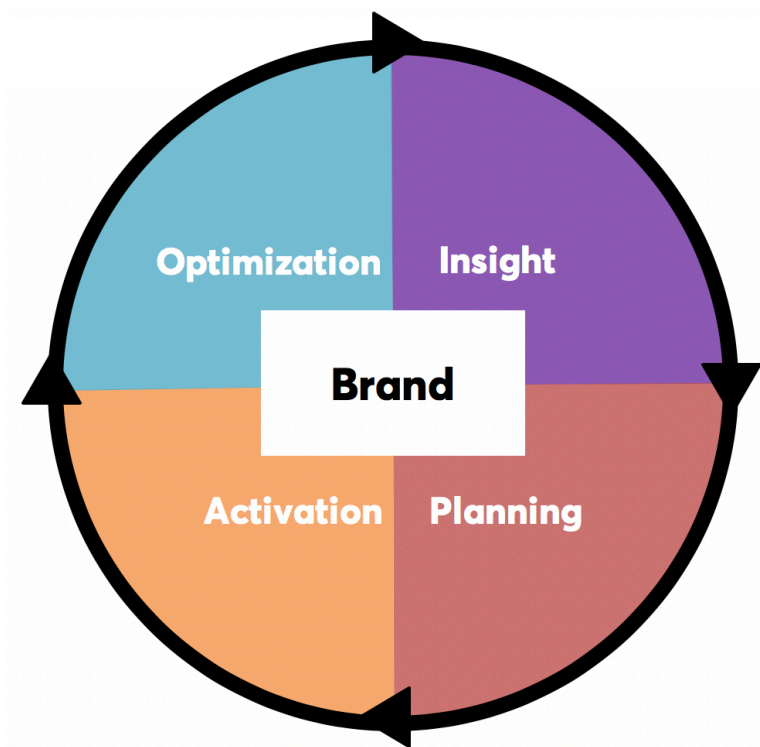

Figure 5. Data & Insight Use Workflow in TBWA\Helsinki

Insight

The Cambridge dictionary defines insight as "*a clear, deep, and sometimes sudden understanding of a complicated problem or situation*" (Cambridge Dictionary, 2019). At TBWA\Helsinki, the word insight is used in the same vein, but in a looser manner. For example, an insight could be that young women donate blood twice as often as young men, or that the amount of money Finnish engineering companies use in employer branding has been rising steadily for the past four years. These insights don't necessarily provide a deep understanding of a complicated problem, but rather provide knowledge on a cultural, statistical, or other occurrence that can then be used as an either starting point or proof point for further activities. Therefore, the most common use cases of data here are exploring a relationship, discovering a pattern, and making a case for a decision (see: section 3.1).

The process begins with insights: researching both the client brand and its marketplace. TBWA\Helsinki conducts research to identify the conventions in the client brand's field(s) and market(s), which is then supplemented with insights regarding cultural trends happening both globally and locally. In the case of new business, the global megatrends are mapped more in detail in its own research. For example, we could find a convention that all key competitors use the term 'smart home' in their communications, which could lead us to use the term 'intelligent housing' instead, to differentiate. Or, we could spot a trend that 44% of Finnish recruiters are predicting an increase in employer branding investment (Sanoma Media Finland, 2018).

In addition, the brand's digital presence is evaluated and compared to that of competitors to understand its quality. The evaluation includes the brand's social media channels, web pages, and other visibility (e.g. earned media) in digital. The objective is to establish a clear starting point and identifying the key areas of improvement. Here, in accordance to section 3.1, data is used to answer a factual question, for example, which competitor is having most engagement on Facebook. After the abovementioned, basic Key Performance Indicators (KPIs) are set for the campaign/project.

Planning

The planning stage combines the insights from the first step with knowledge about the target audience, events & other relevant occurences, and the right metrics for measurement to create the creative concept and its executions. The aim is to understand the target audience for the client's campaign: who are they, where are they and what are the important characteristics. For example, what's the age distribution in our target group, which websites do they use, or what kind of content they consume. These insights are then used to guide the creative concept, as well as the targeting in the Activation step. In the planning stage, the most relevant use cases for data are: making a case for a decision (e.g. what should be the tone of voice when talking to our target group), exploring a relationship (e.g. what are the similarities and differences between two target groups), and discovering a pattern (e.g. all of our target groups list Kauppalehti as their favourite newspaper).

In addition to understanding the audience, the company looks for ways to tie its work into culture. We look for occurences that we could leverage to be relevant: is there going to be a large event that we could tie our work into, can we expect something happening in the near future that would aid us in our task, are there special days coming up.

Activation

After gathering insights on the target audience(s) and relevant cultural occurences, we move to plan the ways to activate the audience. Activating audiences requires distribution. The core actions in the Activation step include media planning, influencer hunt & engagement, and performance marketing. Media planning contains the selection and sourcing of optimal media platforms and placements for the case. These selections are based on the audience insights from the previous step. The aim of media planning is to reach the right audience at the right time with the right message.

At TBWA\Helsinki, many campaigns make use of influencers. In this context the definition of influencer contains all the subgroups from social media micro-influencers to worldwide superstars. Hunting for influencers entails collecting people who fit the requirements for the campaign's purposes and sometimes also evaluating and

demonstrating their viability. These findings are then discussed together with the team responsible for the strategic and creative execution to ensure best fit and activation.

Performance marketing is an umbrella term that refers to (mainly online) marketing & advertising activities where the advertiser pays the publisher for specific actions (Performance Marketing Association, 2019). For example, in Facebook, the advertiser can choose to pay Facebook (the publisher) based on clicks (action). Based on the average cost for click across different services, ad variations or targeting, the performance-marketer can then start optimizing his efforts. Notable performance marketing activities include social media advertising, display advertising and search-engine marketing. In performance marketing activities, the use of data is often already integrated into these platforms. For example, Facebook Ads Manager has all the aspects mentioned in section 3.1.

Optimization

Optimization is in the core of modern data-driven advertising. In addition to marketers optimizing their spend across different channels & services to ensure best results for their chosen KPIs, almost all modern digital advertising platforms have built-in optimization capabilities. Still, to be able to do any kind of optimization, one first has to measure. In the case of TBWA\Helsinki, data set-up is focused on acquiring and tidying data from various digital platforms: social media, media tracking, website, ad accounts, and so forth. Then, in accordance with the data science workflow presented in section 3.5, the data is transformed, visualized, modelled, and communicated internally, to the client, or both. These outputs are then used to optimize activities either within the current project or similar projects in the future. The most common visualization and communication methods in TBWA\Helsinki are slideshow reports and live dashboards that combine data from multiple sources and update automatically to give and overview on a campaign's performance. Data use is in the core of optimization, therefore, it's not a surprise that in the optimization part we use data to answer factual questions, explore relationships, discover patterns, making decisions, and judging experiments (see: section 3.1).

7.3 TBWA\Helsinki's Roles in the Open Data Value Network

The open data value network presented in section 3.2 outlines four core roles and one supporting role in value creation from open data. As concluded earlier, all open government data has to go through Data Extractor & Transformer. It was also highlighted in the interviews that extracting and cleaning the data is a prerequisite for any successful use case, which means that TBWA\Helsinki would need to take that role, or outsource it. Nevertheless, the company isn't interested in creating direct monetary value from that role, by, for example, extracting and transforming the data and then selling it to another company, similar to Vainu in section 6.1. As an advertising agency, TBWA\Helsinki aims to influence the end customer. The only roles that are in contact with the end customer are the Data Analyser and User Experience Provider roles. The company doesn't provide any analytical products for the end customer (for example, a consulting company, such as Gartner, would). That said, as described in section 7.2, we do conduct analysis in various parts of our internal processes, for example, in planning and activation stages. In addition, providing user experiences is in the core of creative executions at TBWA\Helsinki.

To conclude, the roles where TBWA\Helsinki could create commercial value from open government data, based on the open data value network are Data Analyser (for internal data & insight processes, section 7.2) and User Experience Provider (for creative executions, section 7.1). The next section looks at the open data business possibilities from section 3.3 that fall under these two identified roles, and are similar to the activities of TBWA\Helsinki presented in sections 7.1 and 7.2 to narrow down to the most relevant.

7.4 Open data business possibilities at TBWA\Helsinki

Table 4 compares the open data business possibilities from section 3.3 and compares them to the activities at TBWA\Helsinki that were introduced in sections 7.1 and 7.2. Furthermore, the analysis is limited to the high value open data categories for commercial use presented in section 3.4 and specialist interviews.

Table 4. Open Data Business Possibilities at TBWA\Helsinki

Open Data Business Possibilities	Use case in TBWA\Helsinki
Applications	Brand actions
Marketing	Targeted advertising, Performance Marketing
Products and Services	Brand actions
Information enrichment and distribution	
Background material	Market analysis, Audience & Cultural Insights, Content Production
Business intelligence	Competitive intelligence, Audience Insights
Data journalism	Content Production, Insights (proof-points and starting points)
Research & Development	
Education & Consultancy	
Visualizations	Content Production, Brand actions

There are identified use cases in six different categories: *applications*, *marketing*, *background material*, *business intelligence*, *data journalism*, *products and services,* and *visualizations*. The data journalism possibilities can be questioned, as TBWA\Helsinki doesn't exactly do journalism. The researcher has included in here, because the company could use similar techniques for marketing purposes, for example, as a basis for content production, or as a starting point insight for a larger campaign. In the next

section, these identified use cases inside the case company are analysed further with the information from the interviews and the techniques for making data usable (see: section 3.5).

7.5 Recommendations for Use Cases at TBWA\Helsinki

The following preliminary viability analysis of the identified use cases is based on primary observation by the researcher, specialist interviews, and theory on the techniques to make data usable (see: section 3.5).

Both the interviews and theory highlight the significant effort required to pre-process (import, clean, and transform) the data (see: sections 3.5 & 5). In addition, the interviewees highlighted the challenges regarding interoperability, quality and vastness of open government data (see: section 5). Working with data isn't a core competence at TBWA\Helsinki, and is done mainly by the internal data & insight team. The employees working directly with creative executions often don't have competences to pre-process or visualize the data for smaller content production tasks, such as sporadic blog or social media content. In addition, the monetization structure is different for marketing content production than, for example, medias that get ad revenue from data journalism articles. Therefore, TBWA\Helsinki is most likely to benefit from open government data in large executions, such as long-term brand actions, large campaigns or brand platforms/applications.

Enrichment is one of the key components in creating value from open government data. Therefore, it would be advisable to focus on clients who already have considerable amounts of data, which can then be combined with the open government data to gain deeper insights and more compelling user experiences. Another key benefit of enriching open data with client data is that it adds in something that's *not* open for everyone, and thus, not replicable by competitors. Vainu accomplishes this by combining a vast amount of sources and using proprietary methods to transform the data into interoperable from and to automatically draw insights from the massibe amounts of raw data. All interviewees mentioned enrichment as an essential aspect for generating business value from open government data.

It's clear that TBWA\Helsinki shouldn't try to use open government data for explorative purposes in the beginning. There's a possibility of looking into different data sets and

apply different transformations, visualizations, or models, and see if anything interesting comes up. That said, these techniques might not provide anything of use and be extremely time-consuming. Instead, the company should follow the process identified in the interviews (see: section 5).

In conclusion, open government data use at TBWA\Helsinki should, in the beginning, focus on large scale projects where there's a clear data need related to a government area and it's possible to enrich the open data with client's data. The next section provides concrete hypothetical examples on how open government data could be applied, which can later be used to demonstrate the possibilities for the company employees and raise awareness of the issue.

7.6 Hypothetical Use Case Examples

These two hypothetical examples combine the findings from alll the analysis presented earlier: they are based on TBWA\Helsinki activities (sections 7.1 & 7.2) that are within the Data Analyser or User Experience provider roles of the open data value network (section 7.3) and match the identified open data business possibilities (Table 4). These options are then constricted to high value open data categories (sections 3.4 & 5), and by recommendations for use cases at TBWA\Helsinki (section 7.5). In addition, the examples follow the action flow based on the specialist interviews presented in section 5.

Weather Optimized Performance Marketing

TBWA\Helsinki has a nationwide indoor park chain as a client, who wants to advertise its indoor leisure activities on social media during the summer 2019. As mentioned in section 7.2, the company wants to reach the right person with the right message at the right time. The client has informed TBWA\Helsinki that one of their key problems during summer is that people don't visit the indoor parks when it's sunny outside.

Currently, TBWA\Helsinki uses tools that enable automatic creation of numerous different versions of social media advertisements. For example, a campaign that's aims to gain people to visit churches in five different cities could use five different headlines based on the city and five different images of local churches. In total, this would result in

25 different ad variations. Then, a person living in Turku would get an ad that mentions Turku in the headline and has Turku Cathedral as the key image. The goal of this personalization is to reach the right person with the right message, as outlined in the section 7.2, and improve the ads' effectiveness. The tool uses an API to query data from a Google Sheets worksheet with all the possible variations (e.g. Turku, Tampere, Helsinki, Vantaa, Oulu and the church images). The tool also allows automatic dynamic changes to social media advertisements based on certain conditions, triggers. For example, if a value in a specific cell in a worksheet is >0, it will show the picture of St. Michael's church instead of Turku Cathedral. The team working on the case identifies that the Finnish Meteorological Institute could have data that would allow the tool to switch the text based on weather conditions.

If one were to connect the live weather data from the Finnish Meteorological Institute, one could create a trigger that would change the copytext or other components of the ad based on the weather conditions in the location where the ad is distributed. The team decides to create an ad variation that has a headline "Raining? Take shelter and enjoy our indoor activities!" with a -10% discount offer attached that's active only when it's raining in the target location.

Utilizing the open data from the Meteorological Institute would require locating the correct API, connecting the API to a Google Sheets worksheet, and then ensuring interoperability of the data with other inputs.The setup would require a considerable amount of effort, which means that it most likely wouldn't be worth it in small scale campaigns, as analysed in the previous section. The campaign would preferably contain numerous locations around the country, with a clear value proposition for the use of varying copytexts based on weather.

Targeted Marketing Based on Foot, Bike and Car Traffic

TBWA\Helsinki uses a notable amount of digital outdoor advertising to distribute its campaigns. The team is drafting a media plan for an outdoor campaign in Helsinki city area. As mentioned in 'Planning' in section 7.2, media planning contains the selection and sourcing of optimal media platforms and placements. The team recognizes that the city of Helsinki might gather data about different means of traffic inside the city center. The data could tell unbiased information on the amount of potential reach for each outdoor screen. For example, if the data tells us that 1000 people pass by screen A on

weekdays, but 1500 people pass by screen B, it would make sense to rent space on screen B, if other factors are equal.

In the interview conducted for this thesis, Kekkonen (2019) said that Helsinki Region Infoshare is looking into publishing open data on foot, bike and car traffic on the capital area during spring 2019. Depending on the final form of the released open data, it could be used to optimize both the timing and placement of digital outdoor advertising to maximise reach. This approach has a clear value proposition, but it needs to be noted that the companies responsible for selling the advertising space could also use the data in their valuations.

8 CONCLUSION

The objectives of this thesis were to identify relevant open government data sources and types, notable use cases of open government data, and business activities in TBWA\Helsinki where open government data could be utilized. Additionally, due to a request from the case company, the researcher provided two hypothetical examples of open government data use cases to illustrate the analysis results. The researcher succeeded in realizing these objectives by exhaustive review of Finnish open government data and its use cases, field specialist interviews about the characteristics, challenges, and methods of using open government data, and thorough observation and analysis of the case company's activities.

The Strategy Director at TBWA who commissioned the research stated that it yielded valuable insight into open government data and its use, and appreciated the easy-to-understand concrete examples. He noted that the research provides an applicable framework on open government data use consideration, which is essential when trying to implement new methods. In addition, the researcher has agreed to present the research findings in the company's weekly meeting to all current employees to raise awareness of the identified possibilities.

The research found out that majority of Finnish open government data is located in open data portals, of which the most notable one's are the ones maintained by Population Registre Center (avoindata.fi) and the six largest cities in Finland, especially the capital area (6aika data portals, hri.fi).

Based on the completed research, it's clear that there are several applicable use cases for open government data in TBWA\Helsinki. The research also found that a vital part of introducing open government data use is to raise the awareness inside the company. The amount of available data is massive, and it's impossible for a person to exhaustively map all the possible use cases of each data source. Therefore, it's essential to find ways to categorize the activities that are *most likely* to benefit from open government data.

The research identified that inside TBWA\Helsinki, open government data could be used to optimize targeted advertising and performance marketing activities, as part of a brand action, in competitive and market analysis, audience & media planning, content production, and as a background material for other activities with information needs. Both

the literature review and interviews indicate that the focus should be put to the high value open data categories: geographic, meteorological and environmental, economic and business, social / demographic, traffic and transport information. In addition, grounded on both theory and interviews, the researcher recommends to begin with projects that have large enough scope to cover for the effort required to preprocess and analyze the data, where the open data can be enriched with client data, and where there's a clear problem that requires information and relates to a government responsibility area.

The researcher has the following recommendations for further research regarding the topic at the case company. As mentioned in section 4.3 as a limitation of the research, it cannot exhaustively answer whether the case company *should* use open government data. Based on the findings of thesis, TBWA\Helsinki can begin scanning for its first projects that utilize open government data. When such a project goes live, the company should conduct deeper research on the actual business value and profitability of the open government data use.

In addition, to better understand what's the current awareness and knowledge level of open government data inside the company, a quantitative research mapping out the familiarity of the key concepts and players could be conducted. This information could then be used to educate at least one person from each client team to know the basic concepts and situation with open government data, so that it would be taken into consideration in the planning stages for activities that were identified as most likely to benefit from open government data.

Through the thesis research, I, the researcher, learned to embrace change, resilience, and the importance of having a vision. The final thesis was nothing like I imagined it to be in the beginning. Due to the exploratory nature of the research, I had to adapt my research along the way to fit the newly realized realities and narrower focus. I had to remove large portions of work to make room for more relevant data and analysis. Because of these constant, and sometimes frustrating, set backs, it required a great deal of resilience to stay motivated to strive for excellence. That's where having a clear vision helped me: I knew the problem I wanted to solve, even if the way to get there wasn't clear yet.

9 REFERENCES

6aika, 2018. *Älykaupungit tehdään yhdessä.* [Online]
Available at: https://6aika.fi/6aika-avoimia-ja-alykkaita-palveluja/
[Haettu 27 December 2018].

6aika, 2018. *Open Data and Interfaces.* [Online]
Available at: https://6aika.fi/open-data-and-interfaces/
[Haettu 27 December 2018].

Avoindata.fi, 2018. *About.* [Online]
Available at: https://www.avoindata.fi/fi/about
[Haettu 20 October 2018].

Avoindata.fi, 2018. *Avoindata.fi -palvelu uudistuu vuonna 2018.* [Online]
Available at: https://www.avoindata.fi/fi/content/avoindatafi-palvelu-uudistuu-vuonna-2018
[Haettu 20 October 2018].

avoindata.fi, 2018. *Front Page.* [Online]
Available at: https://www.avoindata.fi/fi
[Haettu 27 December 2018].

avoindata.fi, 2018. *Open data categories.* [Online]
Available at: https://www.avoindata.fi/en
[Haettu 16 December 2018].

avoindata.fi, 2018. *Organisaatiot.* [Online]
Available at: https://www.avoindata.fi/data/fi/organization
[Haettu 27 December 2018].

Cambridge Dictionary, 2019. *Meaning of "insight" in the English Dictionary.* [Online]
Available at: https://dictionary.cambridge.org/dictionary/english/insight
[Haettu 2 February 2019].

Carrara, W., Chan, W. S., Fischer, S. & van Steenbergen, E., 2015. *Creating value through open data,* s.l.: European Commission.

Chalabi, M., 2019. *Following Prince Philip's car crash, how many elderly drivers are on the road?.* [Online]
Available at: https://www.theguardian.com/news/datablog/2019/jan/27/prince-philip-car-crash-elderly-drivers-data-uk-us
[Haettu 29 January 2019].

Chatfield, C., 1995. *Problem solving: A statistician's guide.* 2nd Edition toim. s.l.:Chapman & Hall/CRC.

Databusiness.fi, 2017. *Showcases: Tuup.* [Online]
Available at: https://www.databusiness.fi/fi/showcase/tuup/
[Haettu 20 January 2019].

Databusiness.fi, 2018. *Info.* [Online]
Available at: https://www.databusiness.fi/fi/info/
[Haettu 20 January 2019].

Davies, C., 2019. *Prince Philip involved in car crash near Sandringham estate.* [Online]
Available at: https://www.theguardian.com/uk-news/2019/jan/17/prince-philip-involved-in-car-accident-near-sandringham
[Haettu 29 January 2019].

Dykes, B., 2016. *Data Storytelling: The Essential Data Science Skill Everyone Needs.* [Online]
Available at: https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/
[Haettu 22 December 2018].

Eskola, K. & Paavilainen, U.-M., 2017. *Yhdessä kohti avoimen datan liiketoimintaa,* s.l.: 6Aika avoin data ja rajapinnat -hanke .

European Data Portal, 2018. *Homepage.* [Online]
Available at: https://www.europeandataportal.eu/en/homepage
[Haettu 16 December 2018].

Forbes, 2018. *10 Strategic Ways To Automate Your Internal Business Workflows.* [Online]
Available at: https://www.forbes.com/sites/forbestechcouncil/2018/05/29/10-strategic-

ways-to-automate-your-internal-business-workflows/#4402a82f6972
[Haettu 28 December 2018].

Gray, J., Bounegru, L. & Chambers, L., 2012. *The Data Journalism Handbook.* 1st toim. Sebastopol: O'Reilly Media, Inc.

Han, J., 2018. *What Is Pattern Discovery? Why Is It Important?.* [Online]
Available at: https://www.coursera.org/lecture/data-patterns/1-1-what-is-pattern-discovery-why-is-it-important-CEzHJ
[Haettu 28 December 2018].

Han, J., Kamber, M. & Pei, J., 2012. *Data mining: Concepts and techniques.* 3rd Edition toim. Waltham: Morgan Kaufmann Publishers.

Helsinki Region Infoshare, 2019. *Analytiikka.* [Online]
Available at: https://hri.fi/data/statistics
[Haettu 7 February 2019].

Honkanen, M., 2019. *Utilizing open government data* [Haastattelu] (11 February 2019).

Kekkonen, H., 2019. *Utilizing open government data* [Haastattelu] (11 February 2019).

Kitchin, R., 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences.* London: SAGE Publications.

Lahti, T., 2019. *Utilizing open government data* [Haastattelu] (11 February 2019).

Lindman, J., Kinnari, T. & Rossi, M., 2016. Business Roles in the Emerging Open- Data Ecosystem. *IEEE Software,* 33(5), pp. 54-59.

Merriam-Webster, 2019. *interoperability.* [Online]
Available at: https://www.merriam-webster.com/dictionary/interoperability
[Haettu 16 February 2019].

Muurinen, R., 2017. *Liiketoimintaa avoimesta datasta: 100 tapauksen ajankohtaiskatsaus,* Helsinki: 6aika avoin data ja rajapinnat -hanke.

OECD, 2018. *Open Government Data.* [Online]
Available at: http://www.oecd.org/gov/digital-government/open-government-data.htm
[Haettu 10 December 2018].

Open Knowledge International, 2018. *Open Data Handbook: What is Open Data?.*
[Online]
Available at: http://opendatahandbook.org/guide/en/what-is-open-data/
[Haettu 14 October 2018].

Partanen, P., 2016. *Kaupunkifillari-API herätti koodausinnon.* [Online]
Available at: https://hri.fi/fi/kaupunkifillari-api-heratti-koodausinnon/
[Haettu 29 January 2019].

Performance Marketing Association, 2019. *WHAT YOU SHOULD KNOW ABOUT PERFORMANCE MARKETING.* [Online]
Available at: https://thepma.org/our-work/the-performance-marketing-industry/
[Haettu 3 February 2019].

Ropponen, T., 2019. *Utilizing open government data* [Haastattelu] (13 February 2019).

Särelä, J., 2015. *Asuntojen trendit ja miten niistä tehdään luotettavia ennusteita.*
[Online]
Available at: https://www.reaktor.com/blog/asuntojen-trendit-ja-miten-niista-tehdaan-luotettavia-ennusteita/
[Haettu 26 January 2019].

Sanoma Media Finland, 2018. *Tutkimus: Yritykset panostavat työnantajakuvaan, tekoäly vähentää rutiinitöitä.* [Online]
Available at: https://sanoma.fi/tiedote/tutkimus-yritykset-panostavat-tyonantajakuvaan-tekoaly-vahentaa-rutiinitoita/
[Haettu 4 February 2019].

Saunders, M., Lewis, P. & Thornhill, A., 2016. *Research Methods for Business Students.* 7th toim. Harlow: Pearson Education.

Shron, M., 2014. *Thinking With Data.* 1st Edition toim. Sebastopol: O'Reilly Media.

Techterms, 2019. *API definition.* [Online]
Available at: https://techterms.com/definition/api
[Haettu 2 January 2019].

Toikkanen, T., Kalliala, E., Poikola, A. & Sillanpää, L., 2017. *Avoimen datan hyödyntäjä, johdanto.* [Online]

Available at: https://courses.p2pu.org/en/courses/2797/avoimen-datan-hyödyntäjä-johdanto/
[Haettu 2 January 2019].

Vainu, 2019. *About us.* [Online]
Available at: https://company.vainu.io
[Haettu 24 February 2019].

Vainu, 2019. *Asiakastarina: Autoklinikka.* [Online]
Available at: https://customers.vainu.io/fi/asiakastarina/autoklinikka/
[Haettu 24 February 2019].

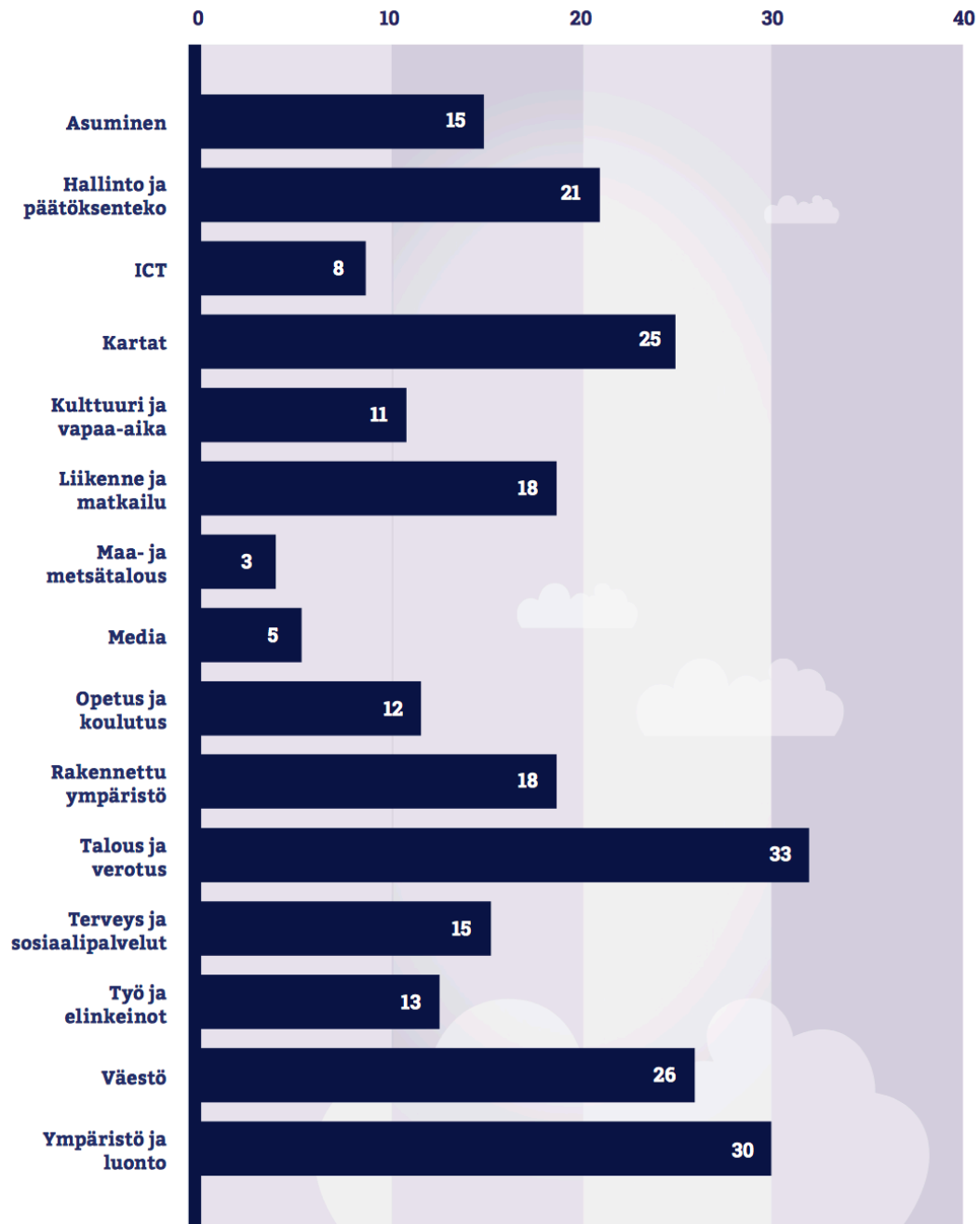Valtiovarainministerö, 2017. *Avoin tieto.* [Online]
Available at: https://vm.fi/avoin-tieto
[Haettu 16 December 2018].

Wickham, H. & Grolemund, G., 2017. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data.* 2nd Edition toim. Sebastopol: O'Reilly Media.

Miller, H.J. and Han, J. (2009) 'Geographic data mining and knowledge discovery: an overview', in H.J. Miller and J. Han (eds), Geographic Data Mining and Knowledge Discovery. Taylor and Francis, London, pp. 3–32

Open Government Data Use Cases Split Into Contextual Categories

Category	Value
Asuminen	15
Hallinto ja päätöksenteko	21
ICT	8
Kartat	25
Kulttuuri ja vapaa-aika	11
Liikenne ja matkailu	18
Maa- ja metsätalous	3
Media	5
Opetus ja koulutus	12
Rakennettu ympäristö	18
Talous ja verotus	33
Terveys ja sosiaalipalvelut	15
Työ ja elinkeinot	13
Väestö	26
Ympäristö ja luonto	30

(Eskola & Paavilainen, 2017)

Interview Questions Chart

[Introduction of the research, case company, and the purpose of the interview]

General

1. **Can you shortly describe your role in the [organization] and your tasks related to open data?**

Open Government Data

2. **In your opinion, what kind of positive characteristics does open government data have (for the utilizer)?**

3. **In your opinion, what kind of negative characteristics does open government data have (for the utilizer)?**

4. **Currently, what are the most valuable open government data sources for businesses?**
   a. Categories
   b. Specific data sources

Utilizing Open Government Data

5. **What are the core challenges regarding open government data utilization?**

6. **How should a company approach utilizing open government data?**
   a. Roadmap
   b. Thought-process

7. **How can one recognize an activity where open government data could be utilized?**

Use cases

8. **Significant use cases of Finnish open government data?**