

# MUNICIPAL OPEN DATA AND REGIONAL DEVELOPMENT

Identification of Open Datasets of City of  
Lahti

LAHTI UNIVERSITY OF APPLIED  
SCIENCES  
Degree programme in  
Business Information Technology  
Bachelor's Thesis  
Spring 2017  
Lenka Zatkova

Lahti University of Applied Sciences  
Degree Programme in Business Information Technology

ZATKOVA, LENKA:

Municipal Open Data and Regional  
Development  
Identification of Open Datasets of  
City of Lahti

Bachelor's Thesis in Business Information Technology

63 pages, 13 pages of appendices

Spring 2017

ABSTRACT

---

Open data is a term that resonates more and more in municipalities all over the world. The definition of datasets with the greatest potential for economic and technological development is behind the stories of the most successful open cities.

The purpose of this thesis was to identify the datasets of the city of Lahti that could be the most relevant for regional development of Lahti area. This thesis also studied the awareness of open data among the local companies. Additionally, the work investigated the best ways of communication between the city and local community.

The data used in this study was collected in interviews with experts and in a survey focused on local companies. Consequently, the data was analysed using qualitative and quantitative research approach.

The results pointed out some datasets that could be potentially used for existing business ideas of local companies. But generally speaking, the data requirements for the most part did not come from existing business ideas. This thesis revealed a low awareness of open data. The effective ways of communication were detected.

The definition of open datasets for the city of Lahti in this thesis was just one of the first steps of a big project. The active cooperation of the city with the local community was identified as the only way that would make city of Lahti another successful open city.

Key words: open data, datasets, municipal data, regional development, city of Lahti, Lahti region, Ladec, public sector

## CONTENTS

1	INTRODUCTION	1
1.1	Theoretical Background	1
1.2	Pragmatic Background	2
1.3	City of Lahti	3
2	RESEARCH APPROACH	4
2.1	Research Question and Objectives of the Study	4
2.2	Research Methods	5
2.2.1	Data Collection Methods	7
2.2.2	Data Analysis Methods	8
2.3	Research Framework and Limitations	9
2.4	Research Process	10
3	AN INTRODUCTION TO MUNICIPAL OPEN DATA	13
3.1	The Open Data Phenomenon	13
3.1.1	Definition of Data	13
3.1.2	How and Why Open Data Appeared	15
3.1.3	Definition of Open Data	16
3.1.4	How to Open up Data	17
3.2	The Role of Municipal Open Data in Local Entrepreneurship Development	18
3.2.1	Why to Open Up Municipal Data for Local Businesses	18
3.2.2	Municipal Open Data Categorisation	21
3.2.3	Municipal Datasets in Finland	22
4	RESEARCH DATA	27
4.1	Data Collected from the Interview with the Ladec Development Manager	27
4.2	Data Collected from the Interview with the Project Coordinator of the Six City Strategy Project	29
4.3	Data Collected from the Survey	31
5	DATA ANALYSIS	35
5.1	Analysis of the Interviews	35
5.1.1	Open Coding	35
5.1.2	Axial and Selective Coding	38
5.2	Analysis of the Survey	39

5.2.1	Business Location	39
5.2.2	Awareness of Open Data	40
5.2.3	Innovations and Business Development Plans	41
5.2.4	Datasets Requirements	45
6	CONCLUSIONS	50
6.1	Answers to the Research Questions	50
6.2	Summary	54
6.3	Generalisability of the Results	55
7	DISCUSSION	57
	REFERENCES	58
	APPENDICES	64

## LIST OF FIGURES

Figure 1 - Research Process Phases	10
Figure 2 - Open Datasets of Cities in Finland Local Open Data Census, part 1	25
Figure 3 - Open Datasets of Cities in Finland Local Open Data Census, part 2	26
Figure 4 - Awareness of Open Data	40
Figure 5 - Business Development Plans of the Companies within Two Years	43
Figure 6 - Effect of Municipal Data on Business Development Plans by Categories	44
Figure 7 - Business Development Plans of the Knowing Group within Two Years	45
Figure 8 - The Most Required Municipal Datasets	46
Figure 9 - The Most Required Municipal Datasets by the Knowing Group	48

## LIST OF TABLES

Table 1 - Comparison of Quantitative and Qualitative Research Approach	6
Table 2. Municipal Data Categorisation in Finland Local Open Data Census	22
Table 3 - Answers to the Survey - Question 1	31
Table 4 - Answers to the Survey - Question 2	32
Table 5 - Answers to the Survey - Question 3	33
Table 6 - Answers to the Survey - Question 4	33
Table 7 - Open Codes for Q1 – Q3	36
Table 8 - Axial and Selective Coding for Q1 – Q3	38
Table 9 - Recommended Datasets based on the Experience of Other Cities	50
Table 10 - Recommended Datasets based on the Survey	52
Table 11 - Categories of Data in Survey, Finnish version	68
Table 12 - Complete List of Answers to the Survey - Question 5	70
Table 13 - Datasets Required by the Knowing Group	74

## 1 INTRODUCTION

Organisations in both the private and public sector collect extensive amounts of data. Data that is relevant to the public sector is not necessarily relevant to the private sector, and vice versa.

The role of government is to manage population as well as to deliver services to citizens. In order to accomplish these tasks, governments need to have knowledge about the society. Such knowledge is based on information and data. All the citizens and entities are identified and their activities are tracked by the government. Consequently, statistical agencies generate and analyse data to provide input for decision making. (Kitchin 2014, 114-115.)

The use of computers and the implementation of information systems affect how data is gathered and stored. However, accelerating the process and making it more accurate and secure is not the main asset of computerisation. The biggest benefit comes from generating and processing big data. Big data is data in volume of terabytes or petabytes. It is created in real-time. Big data varies in type, structure and other specifics. (Kitchin 2014, 68.) In other words, information technology helps make large amounts of data available for different purposes in the public or private sector with minimum effort and costs.

### 1.1 Theoretical Background

The right to information together with existing public sector data has resulted in a great demand for public sector open data. This increased demand for open data has occurred for several reasons such as desired transparency in the public sector, the participation of citizens in public administration and the perceived value of data. (Conradie & Choenni 2014, 3.)

Public sector data is collected primarily on local level, and therefore the success of public sector open data is determined by municipal data. There

is a variety of municipal data that can be open. The nature of data influences its release. In general, it is easier and faster to release

- data about public objects than data intended, for example, for the creation of municipal regulations
- data that is used and stored centrally
- data that is already in suitable digital format
- data without privacy sensitivity or copyright restrictions. (Conradie & Choenni 2014, 4, 7 – 9.)

However, the technical or legislative suitability of data does not say anything about its potential. The comparison of investment in release of data with the rating of re-use of datasets must show that the time and effort was spent on the right datasets. The most reliable method how to identify the datasets with big potential is a demand-based approach which means a close cooperation with potential end users. (Conradie & Choenni 2014, 7 – 9.)

## 1.2 Pragmatic Background

One of the major goals of cities all over the world is opening up their data towards citizens and businesses (Veljkovic, Bogdanovic-Dinic & Stoimenov 2011, 196). The identification of datasets that should be published is the most important part of the process. The assessment of eight European cities including Amsterdam, Barcelona, Berlin, Copenhagen, London, Paris, Stockholm and Vienna suggests the best practice for opening up municipal data. In 2016, the numbers of datasets published by these cities varied from 935 on Berlin portal to 175 in Paris. (European Data Portal 2016, 3.)

The best practices recommendations that should help to identify the relevant datasets are the following:

- A city has to identify municipal data that is available for publishing.
- The benefits of open data must be discussed with citizens and organisations.



- The most required datasets must be identified and released first.
- It is better to start with easily publishable and simple datasets as pilot project.
- The effect of dataset releases should be monitored and evaluated.  
(European Data Portal 2016, 19.)

### 1.3 City of Lahti

Lahti is the capital of Päijänne, Tavastia, a region of 200,000 residents, situated in Southern Finland. It is one of the most important business and industrial centers in Finland. (Lahti 2017a.) In recent years, Lahti, like many other cities, also experienced the decrease of economic growth as a consequence of world economic recession. In addition, technological development in the region is falling behind the development in other regions, for example Helsinki. The lack of development is followed by low number of new business opportunities and a higher unemployment rate. (European Commission 2017.)

The city of Lahti decided to use the promising potential of municipal open data to support and revive its regional, especially technological, development. In spring 2017, the open data project was at its very beginning. Public discussion related to municipal open data had not started yet. The main focus of this initiative was on local companies.

The goal of the city was to start active collaboration and discussion with local companies. The most important part of this phase of the project was to specify relevant datasets in cooperation with local companies.

## 2 RESEARCH APPROACH

### 2.1 Research Question and Objectives of the Study

This thesis aims to answer the following research question:

What municipal datasets should the city of Lahti open up to support regional development and local businesses?

Regional development is the ability of a region to achieve economic, social and environmental growth and prosperity. Material prosperity can be measured with economic indicators such as gross domestic product (GDP). Companies contribute to GDP by producing products and services and selling them. Customers need money to buy products and services. Customers have money if they have income, for example a salary from their employers or income from their own businesses. Selling and buying local products and services supports local companies and employment. Companies and customers pay taxes, one of the biggest sources of public finances. With public finances cities can improve public services and the quality of life also for the socially weaker groups of citizens. Companies with good economic results are pillars of regional development. Companies need to offer products and services that are competitive at the market and required by customers. (Eurostat 2017.)

Opening up its municipal data is one of the ways in which the city of Lahti wants to support the competitiveness and technological strength of local companies. The municipal data have great potential to be used for the development of new products and services. Moreover, it should arouse interest in new technologies and positively affect the industry structure in the Lahti region, improve the situation on the labour market, help regional demographic development and attract investments in the region.

The resources of the city are created publicly so they have to be spent wisely. For this reason, there is a big need for clarification of local companies' demand for municipal data. Currently, the city of Lahti is not able to state which datasets are of the greatest importance for local

companies. This thesis should help to solve this question. Considering this, the main research question is supported with the following subsidiary research questions:

- Is the current understanding of open data among local companies sufficient enough to identify what municipal datasets they would find useful?
- Which communication channels should the city of Lahti use to increase the companies' understanding and use of open data?

The answer to the first subsidiary question aims to reveal how well the concept of open data and its benefits are known to companies and to what extent the companies are currently interested in municipal data. The second subsidiary research question is supposed to specify the forms of cooperation between the city and the companies in defining municipal datasets.

The objectives of the study are formulated as follows:

1. to identify the municipal datasets needed by local companies to support the development of their products or services;
2. to evaluate how familiar local companies are with the concept of open municipal data;
3. to specify the communication channels that can be used to support cooperation between the city and local companies.

## 2.2 Research Methods

The answers to the research questions depend on the collected data. Based on the analysis of data, suggestions for needed datasets and the evaluation of companies' needs will be formulated.

The research approach can be inductive or deductive.

The logical ordering of induction is the reverse of deduction as it involves moving from the 'plane' of observation of the empirical world to the construction of explanations and theories about what has been observed (Gill & Johnson 2002, 40).

Therefore, the inductive research approach will be applied in this study. Induction can be quantitative or qualitative.

Quantitative research is based on a big sample of data of numerical, percentage or monetary character. In quantitative research, the research data is usually collected by a survey or through observation. Quantitative research, then again, applies statistical tools such as arithmetic mean, median, mode and others. (Krishnaswami & Satyaprasad 2010, 6.) Generalisation is based on statistical data.

On the other hand, qualitative research applies a subjective assessment of the observed reality. Generalisation is based on discussions or interviews. (Krishnaswami & Satyaprasad 2010, 7.) Qualitative research focuses on opinions, beliefs or the behavior of individuals. It helps to understand the complex reality in context. (Mack, Woodsong, MacQueen, Guest & Namey 2005, 1.) The comparison of quantitative and qualitative research approach is summarised below in Table 1.

**Table 1 - Comparison of Quantitative and Qualitative Research Approach (Mack et al. 2005, 3)**

	<b>Quantitative</b>	<b>Qualitative</b>
<b>General framework</b>	Structured methods (questionnaires, surveys)	Semi-structured methods (in-depth interviews)
<b>Analytical objectives</b>	Quantification and prediction	Explanation
<b>Question format</b>	Close-ended	Open-ended

<b>Data format</b>	Number-based	Text-based
<b>Flexibility</b>	Flexible	Inflexible

### 2.2.1 Data Collection Methods

For the purposes of this thesis two data collection methods will be used: a survey and in-depth interviews.

A survey is a research method that will be used here to collect the data from local companies. It is an efficient means of gathering large volumes of data in a required form and structure. Surveys can be implemented in a form of pen-and-paper questionnaires or questionnaires distributed by e-mail or on the Web. Questionnaires can be simple or more structured, long or short but should be easy to understand. The purpose of the research has to be explained. The content of questionnaires is not the only important part. It is also necessary to select a relevant sample of participants to obtain relevant data and guarantee the protection of privacy. (Ruel, Wagner & Gillespie 2016, 2-8.)

In this study, data from companies will be collected through an online questionnaire. The research questions require a collection of data from a bigger sample because the needs can be very different in different companies. If the sample size is too small or too specific it may not reflect the overall reality. At the same time, the study has to be done in a short time period because the project has already started.

An in-depth interview is a qualitative research method that usually involves two people – an interviewer and a participant. The interviewer is considered to be a researcher and the participant is an expert. The interviewer asks questions that encourage the participant to share his or her point of view on the topic at hand. The interviewer has to listen carefully and ask follow-up questions. The aim of an in-depth interview is

to obtain the perspective of an individual including opinions and experiences. (Mack et al. 2005, 29.) It is assumed that the research topic is a part of the participant's everyday life. However, an in-depth interview is not an ordinary conversation. An in-depth interview is semi-structured which means that it has suggested questions and a particular focus. At the same time, the interviewer has to be open to new and unexpected topics that may appear during the interview. The participant can also change the attitude or discover new aspects during the interview. (Brinkmann & Kvale 2015, 31-35.)

In this study, in-depth interviews will be conducted with experts on regional development or municipal open data. The in-depth interviews are very important for this study because the expertise, overview and previous experience help to see the project of municipal open data coherently. Municipal open data links together two kinds of subjects. On the one side, there is a city opening up the data. On the other side, there are potential users who have needs for the data. It is important to think about the research from both perspectives.

### 2.2.2 Data Analysis Methods

The analysis of data collected in the survey will be done by using a statistical approach. The collected data will be inserted into a table as variables. Each question will be represented by a variable with a unique identifier. In case of multiple-response questions, the options will have separate variables. (Ruel et al. 2016, 198, 200.) The variables will be categorical with no ordering. The answers will be evaluated in both ways:

- individually for each variable
- in the context of other variables.

The first question will discard the companies that do not represent the target sample. The result of the analysis will reveal the interests of the companies in particular categories of municipal open data as well as their level of awareness on the potential of open data.

The interviews will be analysed using open, axial and selective coding. Open coding means repeated reading of the data and finding the most relevant concepts. The concepts come from the meaning of the data. Open code has properties that characterise the concept. Finally, there are examples of words from the collected data. Axial coding finds the relationships between open codes. Selective coding means finding a variable that is related to all the data. (Gallicano 2013.) The analysis will reveal the following:

- business ideas of local companies in Lahti area
- companies' existing or potential needs for municipal data for their business ideas
- the ways to identify those needs
- and the datasets with the biggest potential usability for regional development.

### 2.3 Research Framework and Limitations

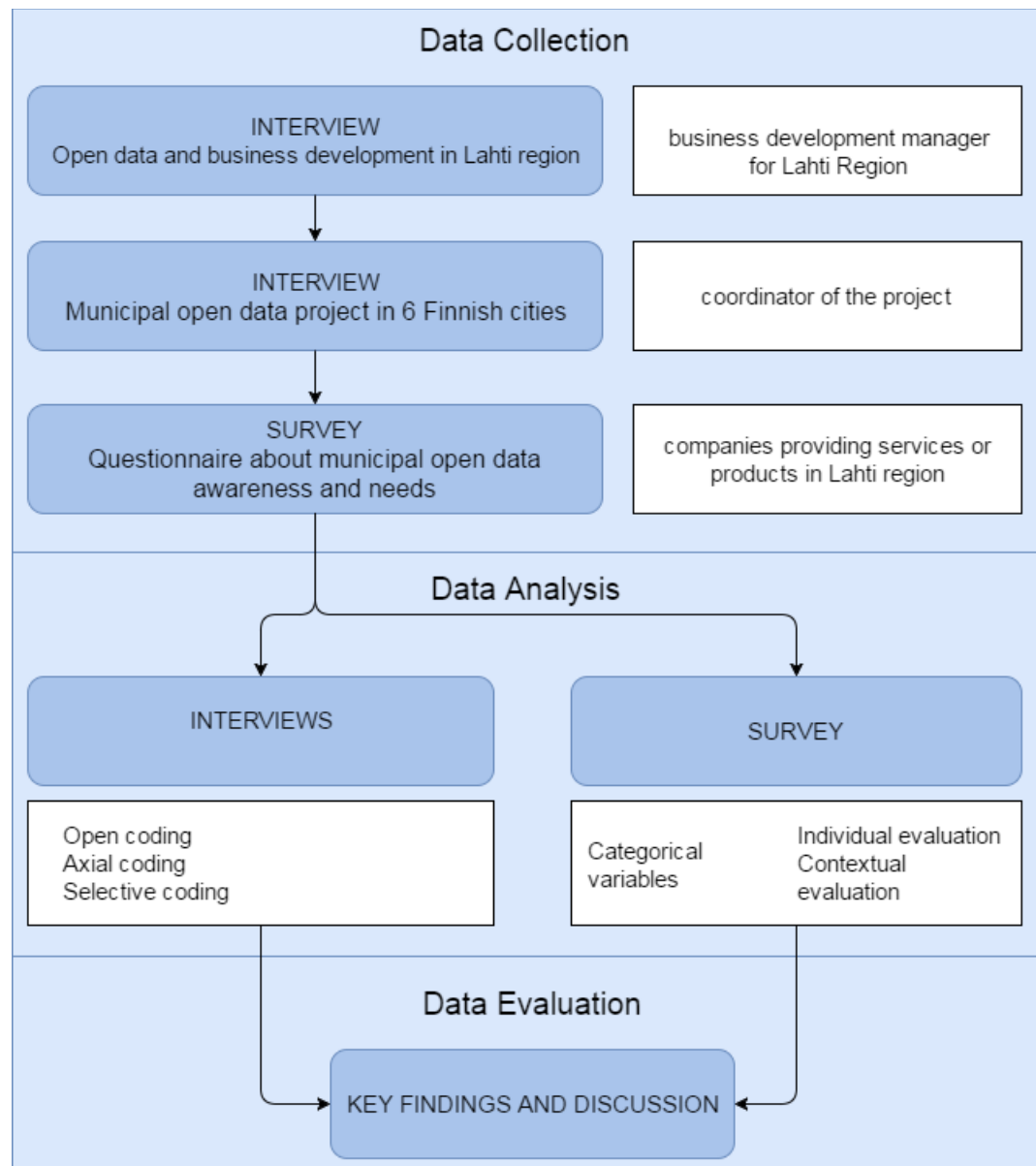
The study focuses on private sector companies that concentrate their business activities in the Lahti region. The list of included companies is generated by Ladec, a regional development organisation, without any specific limitations such as industry or the size of the company.

Since there is no detailed analysis of structure or form of data owned by the city of Lahti, the categorisation of municipal data is based on existing statistical categorisations. The categorisation of data in the survey may not reflect the reality in detail. This fact can affect negatively the integrity of the survey. However, a careful review of existing statistics was made in order to reduce this risk.

The concept of open data and its potential benefits have not been discussed with the companies before. Therefore, there is a risk that the concept and the survey will be misunderstood or ignored. To prevent this, an explanation will be sent together with the invitation to the survey.

## 2.4 Research Process

The phases of the research process are illustrated below in Figure 1.



**Figure 1 - Research Process Phases**



The research process is divided in three main phases:

- data collection
- data analysis
- data evaluation.

Data collection includes the survey and interviews. The interviews will be conducted in English. The survey will be in Finnish. The first interview will be executed with a business development manager from Ladec. Ladec is a Lahti region development organisation that helps local entrepreneurs with starting a new business, growth and innovations (Ladec 2017b). Their everyday communication and close contact with local entrepreneurs provides them with an overview of current trends in local business ideas and the focus of local entrepreneurs.

The data from this interview will help

- identify the potential areas that can be supported with specific categories of municipal open data
- and consider the possible regional activities supporting the awareness of open data potential.

The second interview will be executed with a development manager from Tredea Ltd. in Tampere. He is also a project coordinator of the Six City Strategy (6Aika). Tredea Ltd. is Tampere Region Economic Development Agency. 6Aika is a project including the six largest cities in Finland, Helsinki, Espoo, Vantaa, Tampere, Turku and Oulu. One part of the project is related to municipal open data. The leading partner of the project is the city of Tampere. The main focus of this part of the project is on opening municipal data for local companies. The companies share their data needs and use open data for developing new services. The aim is also that the companies open their own data. The project started in 2014.

The interview will help

- understand the best practice for cooperation models that support the best utilisation and usability of data in business

- and identify the datasets that the city of Lahti should focus on at the beginning. (6Aika 2017.)

The survey will be done with a sample of companies that provide their services or products in the Lahti area. The questionnaire will be implemented online and the invitation letter and link to the form will be sent via e-mail by Ladec. The aim of the survey is to identify

- the companies' awareness of open data
- and the needs for municipal open data for development or innovation of products and services of local companies.

Data analysis will be done after data collection. The methods that will be used are described in chapter 2.2.2 of this document. Afterwards, the data will be evaluated and the key findings will be formulated.

### 3 AN INTRODUCTION TO MUNICIPAL OPEN DATA

Municipal open data is a phenomenon which can, if its potential is truly understood and revealed, make a big difference in regional development. Each city or region has its own specific attributes, such as location, size or local concerns, which impact decisions regarding which municipal datasets can stimulate regional development the most.

The question about the benefits of open data is still open. It has been stated by experts that the full range of benefits has not yet been discovered. However, each city can contribute to the investigation by starting its own project on open data.

#### 3.1 The Open Data Phenomenon

To better understand the open data phenomenon, it is helpful to explain first the definition of its basic element. The basic element is data. The explanation of the concept of data will help to clarify how the role of data became important in our society.

##### 3.1.1 Definition of Data

The meaning and context of data was evolved concurrently with the development of science and humanity through the centuries. The term data has been a part of the English language since the seventeenth century. It was adopted from Latin and used mainly in the area of historiographical records, records of astronomical phenomena or other sciences of those times. Originally the term referred to facts that were indisputable or facts revealed through experimentation. (Rosenberg 2013, 17-18, 20.)

The current concept of data can be further clarified from the following perspectives:

- A knowledge-oriented approach characterises data as a collection of facts.

- An informational approach defines data as information.
- A computational interpretation describes data as a collection of binary elements in the electronic environment.
- From diaphoric perspective data is understood to be without interpretation. (Floridi 2008, 2-6.)

Data can be classified as

- primary data in a database
- secondary data constituted by absence of data
- metadata describing format or other properties of data
- operational data describing system operations
- derivative data from existing data. (Floridi 2008, 7-8.)

Data can have the following forms:

- Quantitative data includes numeric facts that describe physical or non-physical properties.
- Qualitative data includes non-numeric content in the form of text, pictures, sound recordings or videos.

The transformation of qualitative data into quantitative data is possible using data mining or some other techniques. (Kitchin 2014, 4-5.)

Depending on the structure data can be

- structured in the form of a data model
- semi-structured or irregular structured
- unstructured.

An example of unstructured data is qualitative data. (Kitchin 2014, 5-6.)

The ways in which data is generated are as follows:

- Captured data is data collected with the intention of creating a database.
- Exhaust data is generated as a side product of other activities that include data acquisition. In both cases the data is considered raw.

Structured raw data can be analysed and processed and the result is called derived data. (Kitchin 2014, 6-7.)

Depending on the origin there are the following categories of data:

- Primary data is a product of a researcher.
- If primary data is used by other users the data is secondary.
- Tertiary data is usually a statistical outcome produced by institutes of statistics.

By definition, data is just a collection of existing facts without any added value. However, data became a precious article. Data is the base of the pyramid of human knowledge. It can have many different forms, languages or means of storage but the value of collected data remains. Data is a rich source with unlimited potential in generating new information. The information is obtained by putting the data into a structure and interpreting it within a context. However, information is not the final stage of data conversion. By applying cognitive processes, theoretical and practical understanding, learning and experience people are able to transform the information into knowledge. (Kitchin 2014, 7-12.)

### 3.1.2 How and Why Open Data Appeared

A reputable sociologist Robert King Merton already in 1942 declared that scientific data should be freely accessible and scientists should share their knowledge. The concept of open data was mentioned for the first time in 1995 when the authors of an environmental report called for openness of scientific data between countries. This openness was called for in order to help understand the global geophysical phenomena. With the development of information technologies, publicly available information commons started to be seen almost as a public good. Everybody can use it without decreasing the value for the others. On the contrary, the value of data even grows. In 2007, a group of activists met in Sebastopol near San Francisco and declared that the concept of open data should be built on openness, participation and collaboration. (Chignard 2013.)

The evolution of data reflects the cultural changes and development of society. Open data has the potential to facilitate innovation, transparency and solve political and economic issues. (Chignard 2013.)

### 3.1.3 Definition of Open Data

“Open data is governmental data of public interest that is available without any restrictions and can be easily found and accessed” (Veljkovic et al. 2011, 196). Open data is data that is freely accessible by anyone for any purpose. The openness of data is technological, non-proprietary and legal. (Halonen 2012, 18.)

In European legislation, the ownership of databases is a part of intellectual property rights. If facts are collected, verified and added to a database, the owner of the database automatically gains intellectual property rights over them because it is considered to be a result of creative work. Copyright is related to the database, not to the facts that were collected. The owner has the right to decide about the rules of using it. It is realised in the form of a licence. (Open Data Institute 2017.) The license must allow free use, redistribution, modification, separation, compilation, non-discrimination and use for any purpose without any charge. The license may require attribution of creators, integrity of versions, share-alike copies, retention of copyright notes, form of modified works that makes possible further modification, technical restriction prohibition and public additional premissions grant. (Open Knowledge International 2017b.)

Licensing is regulated by licensing law. Each country has its own licensing law. For European Union countries the recommended standard is the Creative Commons Licence. The process of making data open must respect all relevant rules and laws, and must take in account all relevant security and ethical aspects. (Open Data Institute 2017.)

### 3.1.4 How to Open up Data

Opening up municipal data has some rules that help to achieve better results. The definition of municipal datasets is an iterative process that should be done in four steps. The first step is the analyses of best practices around the world. Open datasets of other cities provide an overview of the most common open datasets. The second step is the analyses of local government e-services which help to reveal the current situation of the project. The third step is the analyses of internal databases. A city has to know what data they can provide, as well as what the current form of the data is. The fourth step is the most important and it is the analyses of the public interest which is defined by potential users. (Veljkovic et al. 2011, 199-201.) A very important group of potential users are infomediaries. Infomediaries are those users that will take the data and transform it into a more usable form. For example, they can make maps from GPS coordinates data. (Open Knowledge International 2017c.) The interaction with end users should be regular and it should begin before the first datasets are created. The strategy of municipal open data policy has to start with rise of awareness of open data. The most effective way to increase the understanding of the potential of open data is by organising events such as hackatons for the public. In-person events establish the communication and collaboration between municipalities and citizens. The concept of open data should be explained to people with the focus on practical applications. The most important topics should be identified and solved at first. The policy should include partnerships with other cities and organisations. The strategy requires also legislative and financial support. (Moneo 2016.)

The process of opening data should start with creating relatively small and simple datasets, and then continuing with more datasets as a learning process. It is crucial that the public is well informed about the project and all the questions and fears are discussed and explained. (Open Knowledge International 2017c.)

For the purpose of making data publicly available the data should have a standardised form and structure to ensure their direct usability by software tools (Gurstein 2011). Open data should be available as a set. The datasets should be in a machine-readable format such as, for example, CSV, XML, HDF5 or JSON (World Wide Web Consortium 2016). There are the following online methods how to make the open data accessible:

- The data can be provided in the form of files downloadable from existing websites.
- The platforms of third parties store data and provide the infrastructure for accessing that data.
- File Transfer Protocol (FTP) is a method that is suitable for software developers or scientists.
- BitTorrent is an easy way how to share large volumes of data with other users.
- Application Programming Interface (API) allows users to access up to date databases and select required data. (Open Knowledge International 2017c.)

### 3.2 The Role of Municipal Open Data in Local Entrepreneurship Development

Data that seems to have no development potential from the perspective of the originator can have big potential from the perspective of companies. Usually the companies analyse and use the data or make the data usable in the form of applications or online services. (Halonen 2012, 87.)

#### 3.2.1 Why to Open Up Municipal Data for Local Businesses

Finland joined the Open Government Partnership with the decision made by the Minister for Public Administration and Local Government in 2012 (Open Government Partnership 2017).

Data openness in Finland is supported also by the Act on the Openness of Government activities from December 1, 1999.



## Section 21 — Production of sets of data on request

(1) When requested to do so, an authority may compile and deliver a set of data formed from signs contained in one or more computerised information management systems and maintained for various purposes, if such delivery is not contrary to the provisions on document secrecy and the protection of personal data owing to the search criteria used, the volume or quality of the data or the intended use of the set of data.

(2) Where permitted by the authorities concerned and subject to the provisions in subsection 1, the set of data may be compiled also from information management systems maintained by different authorities. (Act on the Openness of Government Activities 621/1999, Chapter 4, Section 21.)

Finland's Action Plan on Open Government contains a commitment to open and publish new data and change the form of existing open data into a machine readable form that will enable the citizens to access the data efficiently. (Open Government Partnership 2013, 7-8.)

There are undoubtedly economic benefits of accessing municipal data. The benefits are explained by the following examples:

- Citizens have the right to access the data but they should also benefit from the development of local services and products. Businesses' ability to identify new business opportunities, make decisions about innovations, build an efficient supply chain or evaluate the effectiveness of their activities strongly depends on information. The public sector has a key role as a source of data especially in some areas, such as transport, geography or weather, where other sources of data can be hardly found (Pollock 2008, 2).

It is expected that companies will use public sector information to create new products and services. As a result, the innovation will invoke economic growth, increase of income and employment.

- Another positive aspect is the cost-efficiency which will reduce the costs of companies provided that the licences of open data allow the companies to use it for commercial purposes. Small and medium-sized businesses can especially benefit from open data since if there were charges for access and use of this data, they would probably not use it. (Halonen 2012, 28-29.)
- If the data is open the companies can access more timely and accurate information (Yiu 2012, 20).
- In addition to previous benefits, open data will encourage partnership whether between companies or between companies and the city (Kitchin 2014, 56).

The economic effect of governmental open data is proven by the following examples from countries all over the world:

- The economies of the United States (US) and Europe are approximately the same size but the US weather industry is ten times larger which is explained by the fact that US weather data is publicly available (Pollock 2006, 14).
- Porch.com is a Seattle-based company that created a database with information that helps to estimate the costs of house building and reconstruction projects. There were three times more new businesses created in Seattle in 2011 than in the rest of the country.
- Opening global positioning data led to development of a new expanding industry producing GPS products and services.
- Opening public weather data in the US increased the growth of consumer websites with weather information which consequently significantly increased the number of newly created weather related products and services. (Cashman 2014.)
- Open data in Helsinki enabled the creation of several applications that are being sold in Finland and also in other countries. The most

significant ones use mostly public transport and geographical data. (Sulopuisto 2014.)

- The Open Data Institute analysed 270 companies in the United Kingdom (UK). These companies have open data as a part of their business. 70% of the companies use data provided by the government. The most used data is geospatial data that is used by 57% of companies. Transport data including live traffic and CO2 emissions is used by 43% of companies. Environmental data is used by 42% of companies and demographic data such as census, postcode or claimant count data is used by 18% of companies. The companies found a lack of open government data to be a big difficulty. (Open Data Institute 2015.)

### 3.2.2 Municipal Open Data Categorisation

Clarke (2010) identified that public datasets can be generally categorised as historical, planning, infrastructural or operational data. This categorisation supports the sector-specific research. (Halonen 2012, 21.) Cities can provide the datasets in categories like transportation, demographics, education, city administration and other categories. Other cities can make the datasets available in alphabetical order, for example schools, parking, street lighting and so on. (Veljkovic et al. 2011, 200.)

The analyses of municipal open data portals all over the world indicate the most common municipal data categories. Based on the analyses the categorisation should involve finance and economy, environment, health, energy, education, employment, transportation, infrastructure and population data which each city should have and provide to the public. The description and content of categories is supposed to be different in different cities depending on regional specific features and city interests. The categories can be extended, for example, by geographic information system (GIS), sport, events or others. Each city has to make an effort to identify the categories that are relevant for the region. (Veljkovic et al. 2011, 202.)

### 3.2.3 Municipal Datasets in Finland

The current status of access to municipal datasets in Finnish cities is monitored by the Finland Local Open Data Census. The census shows the progress and categories of open datasets in different cities. It is updated periodically by volunteers and municipalities and it is run by Open Knowledge International. (Open Knowledge International 2017a.)

The list of cities that participate in the census currently includes Espoo, Hämeenlinna, Helsinki, Jyväskylä, Kauniainen, Kuopio, Mikkeli, Oulu, Pori, Rovaniemi, Tampere, Turku and Vantaa. The census is updated for the years 2013 – 2016 by now. Table 2 contains the definitions of municipal datasets that are monitored by the census. (Open Knowledge International 2017a.)

**Table 2. Municipal Data Categorisation in Finland Local Open Data Census (Open Knowledge International 2017a)**

<b>Data category</b>	<b>Definition</b>
<b>Public Facilities</b>	Location information about various Public facilities such as schools, parks, hospitals, daycare etc
<b>Real-Time Transit</b>	Real-time information about major municipal-run or commissioned transit services (buses, subway, rail, tram etc). Real-time transit information means things like the location of actual services (individual buses and trains, etc)
<b>Election Areas</b>	Boundaries of electoral districts in this municipality
<b>Transport Timetables</b>	Timetables (schedules) of all municipally run or commissioned transit services (buses, subway, rail tram etc). Locations of stops would also be good (as geodata) but is not required to answer this question affirmatively.
<b>Business Listings</b>	Key information for businesses in the municipal area, such as name, address, contact information, business type.
<b>Air Quality</b>	Data on air quality (e.g. levels of major pollutants) on a

	granular basis - that is at least broken down by month (preferably by day). Geographic breakdown (e.g. by grid point) would be nice but is not required.
<b>Procurement Contracts</b>	Per contract information on municipal contracts including amount, awardee (name, address), data awarded etc
<b>Expenditure (detailed)</b>	Records of actual (past) municipal spending at a detailed transactional level, for example, at the level of month to month expenditure on specific items (usually this means individual records of spending amounts at a fairly granular level - e.g. \$5-50k rather than at the \$1m+ level). (Note: a database of contracts awarded or similar is <i>not</i> considered sufficient. This data category refers to detailed ongoing data on <i>actual</i> expenditure)
<b>Annual Budget</b>	Municipal budget at a high level (e.g. spending by sector, department etc). This category is about budgets which are plans for expenditure (not actual expenditure in the past).
<b>Crime Statistics</b>	Data on municipal crime, preferably at a reasonably disaggregated level (best would be exact date, location and type but per day per street or post/zip code would be acceptable)
<b>Agendas and Decisions</b>	Agendas, meeting minutes and decisions of the municipal board (hallitus), council (valtuusto) and committees (lautakunnat).
<b>Food Safety Inspections</b>	Outcomes of food safety inspections of restaurants and other similar providers of food to the public
<b>Traffic Accidents</b>	Statistics on road traffic accidents including time and location
<b>Service Requests</b>	Non-Emergency service requests to municipal authorities, for example regarding potholes, graffiti, etc (311, D115 etc). Data should be at granular (per request) level.
<b>Building Permits</b>	Building / construction permits

The tables below in Figure 2 and Figure 3 show the datasets that have been open by Finnish cities. The rows of the tables include a complete list

of the categories of municipal datasets. The columns show the score of data openness of each category by city together with the year of the last update. The percentage says how much the datasets meet the following conditions:

- the data exists
- the data is in digital form
- the data is openly licensed
- the data is publicly available
- the data is available online
- the data is available for free
- the data is available in bulk
- the data is machine readable and up-to-date.

Public facilities, real-time transit, election areas and transport timetables are the categories that were open by the largest number of municipalities in 2016. Business listings, air quality, procurement contracts, expenditure, annual budget and crime statistics datasets were open largely. (Open Knowledge International 2017a.)

Dataset	Score by City									
	Helsinki		Vantaa		Espoo		Tampere		Kauniainen	
	%	Last Update	%	Last Update	%	Last Update	%	Last Update	%	Last Update
Election areas	100	2016	100	2016	100	2016	100	2016	100	2016
Public facilities	100	2017	100	2016	100	2016	100	2016	100	2016
Transport timetables	100	2016	100	2016	100	2016	100	2016	100	2016
Annual budget	90	2016	90	2016	90	2016	90	2016	45	2016
Business listings	90	2016	90	2016	90	2016	90	2016	90	2016
Real-time transit	100	2016	100	2016	100	2016	100	2016	100	2016
Procurement contracts	100	2017	90	2016	90	2016	90	2016	90	2016
Air quality	100	2016	100	2016	100	2016	45	2016	100	2016
Agendas and decisions	100	2016	45	2016	45	2016	45	2016	55	2016
Expenditure (detailed)	90	2016	90	2016	90	2016	90	2016	90	2016
Crime statistics	70	2016	70	2016	70	2016	70	2016	70	2016
Food safety inspections	45	2016	45	2016	45	2016	45	2016	45	2016
Traffic accidents	90	2016	35	2016	35	2016	10	2016	5	2016
Service requests	100	2016	0	2016	0	2016	20	2016	0	2016
Building permits	30	2016	45	2016	10	2016	25	2016	0	2016
Business permits	10	2016	10	2016	10	2016	10	2016	10	2016

Figure 2 – Open Datasets of Cities in Finland Local Open Data Census, part 1 (Open Knowledge International 2017a)

Dataset	Score by City									
	Turku		Oulu		Hämeenlinna		Jyväskylä		Mikkeli	
	%	Last Update	%	Last Update	%	Last Update	%	Last Update	%	Last Update
Election areas	100	2016	100	2016	75	2014	65	2014	-	
Public facilities	100	2016	100	2016	65	2014	35	2014	-	
Transport timetables	100	2016	100	2016	55	2014	45	2014	-	
Annual budget	55	2016	45	2016	55	2014	100	2014	85	2014
Business listings	90	2016	90	2016	100	2014	%	2014	-	
Real-time transit	100	2016	100	2016	0	2014	10	2014	-	
Procurement contracts	0	2016	90	2016	5	2014	100	2015	-	
Air quality	90	2016	45	2016	75	2014	-		-	
Agendas and decisions	45	2016	75	2016	75	2014	45	2014	45	2014
Expenditure (detailed)	20	2016	45	2016	10	2014	5	2014	-	
Crime statistics	55	2016	50	2016	-		-		-	
Food safety inspections	45	2016	30	2016	85	2014	-		-	
Traffic accidents	90	2016	5	2016	-		20	2014	-	
Service requests	45	2016	45	2016	75	2014	-		-	
Building permits	35	2016	5	2016	75	2014	-		-	
Business permits	5	2016	5	2016	-		-		-	

Figure 3 – Open Datasets of Cities in Finland Local Open Data Census, part 2 (Open Knowledge International 2017a)



## 4 RESEARCH DATA

This chapter contains the data collected in two interviews and one survey. The data is organised into three subchapters. Each interview has its own subchapter because the questions were different for each interviewee.

### 4.1 Data Collected from the Interview with the Ladec Development Manager

The interviewee's specialisation is helping new entrepreneurs with business ideas and building connections. He provides the clients with guidance on how to make sure that the business idea is feasible. He helps with validation whether the clients have something that they can sell and how much they understand about who their customers are. The interviewee was selected for this study because of his experience and knowledge of local business development and his close cooperation with local companies.

The interview was done in person on the 22nd February 2017 on the premises of Ladec in Lahti. The interviewee was asked 5 questions. The list of questions for this interview can be found in the Appendices (Appendix 1). The interview was recorded and transcribed afterwards. The collected data came from several categories of questions.

*Category 1: The products and services offered by new start-ups in the Lahti region and their relation to data*

There are some business ideas dealing with data. One is an advertisement office gathering data about products in the form of public products reviews made by users who tried the products.

Another business idea based on data is a mobile application for municipalities where citizens give feedback. Data collected with this application could be also open to the public. (Future Dialog 2017.)

There are a few cases of start-ups in the energy field. There is an ambition to optimise the use of energy and avoid the waste of energy. If these kinds of companies have information about the use of electricity in buildings it could give them a picture about energy spending and what they are able to do about it. Another company came up with technology that can get energy out of waste water or other sources such as polluted air, steam or solar energy (Wasenco 2017).

There is around 40 companies established through Ladec start-up services. The list of business ideas can be found in the Appendices (Appendix 2).

Every company needs data to make decisions. They need information about potential clients. For example, if they want to open a cafeteria they need to know how many people walk by to decide how big the cafeteria should be. If there is not enough people they need to decide how to bring more people there. This kind of open data can be interesting for the private service sector.

#### *Category 2: Ladec involvement in municipal open data in Lahti region*

Ladec intends to organise hackathon events in the future. As an example, there is a hackathon event that will be held in Oulu in autumn 2017. Participating teams can develop new services based on open data. The hackathon will open up data including municipal data. (Arctic Startup 2016.) The idea is about creative people being able to develop new services. It is important to get people, developers, students enthusiastic.

Protomo is a team-based business development service provided by Ladec. Working in teams should generate new business concepts and growth of new companies. (Ladec 2017a.) More than 500 people are registered to become members. Hackathon events in Lahti should help to get as many new ideas as possible. To enable the companies to decide what municipal data they need, it is necessary to provide them with the information which categories of municipal data are available.

### *Category 3: Municipal data business potential*

There is a lot of information that may have the potential. When there are innovative people, if they see what is available, they can get ideas. There is potential of lowering the costs which is always interesting for municipalities. In some areas the public sector would prefer not to provide services. Where public services are going to be private services there is a lot of new business opportunities. In healthcare there are private organisations taking care of medical services. Companies are buying cleaning services from other companies, they are not employing people. Municipalities pay for family houses for children with family issues that can not live at home.

#### 4.2 Data Collected from the Interview with the Project Coordinator of the Six City Strategy Project

The interviewee is a coordinator of the Six City Strategy project. The activities of Tredea Ltd. include a data portal for the six cities and a DataBusiness.fi website that promotes the business opportunities and real cases that use open data in business. They founded more open data projects as a part of the Open Tampere project which supports new businesses and the growth of companies (AvoiTampere 2017). The interviewee was selected for this interview because of his experience with opening municipal data in six Finnish cities.

The interview was done in the form of e-mail communication in the days from the 3rd of March to the 9th March of 2017. The answers were obtained as a text file. The list of questions for this interview can be found in the Appendices (Appendix 3). There are several categories of data.

### *Category 1: Promotion of awareness of municipal open data*

The communication must be active and direct. The communication channels can be the local newspapers or other media, participation in already existing channels on Facebook like Finnish Open Data Ecosystem (over 4000 members), Kuntien avoin data (over 300 members) or

{API:Suomi} (over 300 members), local company clusters and student groups. On Twitter it is #avoindata. Organisation of open data hackathons for producers and users of data will encourage new ideas and organisation of competitions for the public will increase the use of the data.

### *Category 2: Defining the first datasets*

The city should organise a public open data kick off meeting for local companies and citizens including students and researchers. The city must show a real commitment to cooperation with the local community. The ideas must be collected and responded to by the city. Published datasets must be modified according to the feedback and kept updated. The city itself can come up with datasets and not just wait for the requirements. The positive results can appear after a long time.

The city should start with datasets that are easy to open. The usage of data should have simple and effective use cases to demonstrate the value of the data, for example transit applications. Analysis of other cities' datasets or applications will help to understand which datasets could bring a value to Lahti. The most open datasets in other cities could be useful also for city of Lahti. In addition, the data should have a proper metadata and documentation. The data model and format is important, as well as the license for the data, portal where the data will be published and how the data will be updated.

### *Category 3: The most efficient ways of identification of companies' needs for municipal data*

The communication must be focused on identification of problems and ideas of companies and searching for datasets that could solve the issues. Surveys are not the best way of communication with companies. Invitations to events will show the real determination of the city to open datasets for the companies. Experimenting in cooperation with the companies is the right choice.

*Category 4: The most used datasets in the 6 biggest Finnish cities*

Mobility data (transportation, traffic, parking information), geospatial data, city services and events are the datasets of the biggest usage in the 6 biggest Finnish cities. It is possible that city services or geospatial data could be widely beneficial.

#### 4.3 Data Collected from the Survey

The survey was done in the time period from the 14th of March to the 24th of March 2017. The first invitation was sent the 14th of March and a reminder was sent on the 21st of March. The invitation was sent in a form of an e-mail containing the Ladec's newsletter. The e-mail was sent automatically by Ladec's software to 1800 subscribed companies. The newsletter contained the invitation to the survey only. This database of companies was selected for this survey because it contains an extensive list of local businesses who are the target group of this study. 40 companies participated.

The list of questions for this survey can be found in English and Finnish language version in the Appendices (Appendix 4 and Appendix 5). The answers are presented in this chapter in the form of tables.

Answers to the first question specify the location (Table 3). For the survey it is required that only companies with business in Lahti area can participate. The participants had to choose one of two possible answers.

**Table 3 - Answers to the Survey - Question 1**

<b>1. Does your company produce products or services in Lahti area?</b>	
<b>Answer</b>	<b>Number of answers</b>
Yes	40
No	0

Municipal data can be used by any kind of business. There are no limitations and no rules related to size or industry. For this reason, no further general information about the companies was required.

Answers to the second question describe how the participants evaluate their familiarity with the concept of open data (Table 4). The participants had to choose one option that describes their understanding of open data most accurately. The answers show how confident the companies are regarding open data questions.

**Table 4 - Answers to the Survey - Question 2**

<b>2. Please, evaluate how well do you understand what open data is:</b>	
<b>Answer</b>	<b>Number of answers</b>
I understand what open data is and I have an idea how our company could use municipal open data	10
I understand what open data is but I do not know how our company could use municipal open data	11
I partly understand what open data is	8
I do not understand what open data is	11

The third question asks about the development intentions of the companies (Table 5). The participants had to evaluate each answer. The evaluation of the answers shows how likely it is that the company will develop a new service or product within the next 2 years.

Table 5 - Answers to the Survey - Question 3

<b>3. Within next 2 years, how likely it is that you will do:</b>				
<b>Answer / Number of answers</b>	<b>Very likely</b>	<b>Mostly likely</b>	<b>Mostly unlikely</b>	<b>Very unlikely</b>
Innovate at least one of your existing products	25	7	1	7
Innovate your current services	29	8	2	1
Come up with new product	19	9	5	7
Come up with new services	20	11	7	2
Try to reach out to new target group of customers	17	12	8	3

The fourth question asks about the business development intentions in a situation when the company can access the municipal data (Table 6). The participants evaluated each answer. The evaluation of the answers shows how open data can influence the business development plans of the companies.

Table 6 - Answers to the Survey - Question 4

<b>4. If you have an access to municipal data of city of Lahti, how likely it is that you will do within next 2 years:</b>				
<b>Answer / Number of answers</b>	<b>Very likely</b>	<b>Mostly likely</b>	<b>Mostly unlikely</b>	<b>Very unlikely</b>
Innovate at least one of your existing products	19	11	3	7
Innovate your current services	22	13	3	2
Come up with new product	16	10	5	9

Come up with new services	18	13	7	2
Try to reach out to new target group of customers	16	13	8	3

Question number five asked: *In case that you answered "Very likely" or "Mostly likely" at least in one of questions 3. or 4., what kind of municipal data would you need/find helpful to make the plans real?*

This question aims to gather the most specific data requirements from the companies. The question contains a list of categories of datasets. Companies could vote for an unlimited number of categories that should be open, and they could add their own categories. The number of votes for categories varies from 0 to 20. There were 11 categories with 10 or more votes. A complete list of the answers can be found in the Appendices (Appendix 6).

Moreover, the companies added these categories:

- repairs and construction of public buildings (schedule information, developers, contractors, designers)
- traffic in the city center (including pedestrians)
- annual plan and budget of technical board
- open data also helps to sell anywhere if there is awareness of potential
- energy consumption information and the energy strategy of city of Lahti and the will of the city to implement measures focused on energy efficiency
- energy savings targets for various construction projects, the introduction of new innovations in the field of energy. (translated from Finnish by Zatkova)



## 5 DATA ANALYSIS

The data is analysed in two sections. One section contains the analysis of the two interviews. The other section analyses the data collected in the survey.

### 5.1 Analysis of the Interviews

The interviews are analysed using open, axial and selective coding as described in chapter 2.2.2 of this document. The coding is done using data collected from both interviews combined.

The analysis is structured with respect to the research questions. There is one main research question and two subsidiary research questions. For the analytics part the main research question is named Q1, the subsidiary questions are Q2 and Q3.

- *Q1: What municipal datasets should the city of Lahti open up to support regional development and local businesses?*
- *Q2: Is the current knowledge of open data among companies sufficient to identify what municipal datasets they find useful?*
- *Q3: Which communication channels should the city of Lahti use to increase the understanding and use of open data by companies?*

#### 5.1.1 Open Coding

The key concepts from the interviews are listed below in Table 7. Each key word was characterised individually by properties and examples. The properties describe the meaning and content of concepts. The examples are instances of the concepts that come from the collected data.

The codes of research question Q1 identify potentially the most relevant business areas for municipal datasets usage. The list of codes also contains suggestions of datasets. The examples of businesses and datasets are particularly important for the research question.

There is only one open code for research question Q2. The code says that there are no current business ideas based on open data.

The open codes of research question Q3 are the most relevant concepts describing the recommended role of the city. As in the research question Q1, the examples of communication channels are of great importance for the research question.

**Table 7 - Open Codes for Q1 – Q3**

<b>Open code</b>	<b>Properties</b>	<b>Examples</b>
<b>Q1</b>		
Business ideas in energy saving and pollution reduction	Trying to solve environmental issues Looking for ecological use of waste Searching for solutions for efficient heating of buildings	Devices that monitor the temperature in buildings Saving energy is a key area Getting energy out of waste water and other polluted sources
Business ideas in software and ICT companies	Bringing new software solutions Open to new software ideas	Building information management software Open sources software and services Special games
Business ideas in healthcare and social services	Providing high level of healthcare and social services Providing services more effectively than public sector Identifying new opportunities Lowering municipality costs	Private organisations taking care of medical services Family houses for children with family issues
Mobility, geospatial data, services and events	Showing big usage of data Providing large scale of usability	Transportation Traffic Parking information City services Events
Datasets suggested by a city	Discovering the potential of data Trying to meet the needs	City can learn from other cities City can come up with

	Analysing the datasets of other cities	datasets
Datasets suitable to be open in the beginning	Easy to open	Simple and effective use cases (transit applications)
Datasets solving a problem	Focusing on real issues Searching for ideas of companies	Datasets that could solve the problem
Data attributes	Containing proper documentation Having suitable format	Metadata Documentation Data model and format License
<b>Q2</b>		
Data gathering in business ideas	Gathering data as a business Using data for the owner's purposes Keeping data with no intention to open it	Gathering marketing data about products Mobile application for municipalities where citizens give feedback
<b>Q3</b>		
Organisation of hackathon events in future	Supporting active cooperation Addressing local community Providing with information what municipal data are available	New services based on open data Creative and enthusiastic people, developers, students As much new ideas as possible Enable companies to decide what data they need
Active and direct communication channels	Encouraging new ideas Increasing the use of data	Local media Facebook groups Twitter Hackathons
City's real commitment	Inviting to the meetings Collecting the ideas Responding to the ideas Experimenting in cooperation with the companies	Public open data kick off meeting for local companies and people including students and researchers Invitations to events Discussion and modification of datasets Data updates

### 5.1.2 Axial and Selective Coding

This chapter contains a table and description that identify the relationships between open codes as axial codes. Finally, the selective code for each category of axial codes is defined. The coding is done separately for each research question below in Table 8.

The open codes were divided into categories. The categorisation depended on the context and meaning of each open code. Open codes representing one idea were included in one category. The formulation of the idea/axial code had to cover all the open codes in one category. The general statement /selective code covers all the axial codes.

**Table 8 - Axial and Selective Coding for Q1 – Q3**

<b>Open Codes</b>	<b>Axial Codes</b>	<b>Selective Code</b>
<b>Q1</b>		
Business ideas about energy saving and pollution reduction Business ideas about software and ICT companies Business ideas in healthcare and social services Mobility, geospatial data, services and events	Business ideas of local companies must be identified and supported	City must be able to provide datasets that reflect real needs of local companies
Datasets suggested by a city Datasets solving a problem	Datasets must solve issues with assistance of the companies	
Datasets suitable to be open in the beginning Data attributes	Datasets must be known and available	
<b>Q2</b>		
Data gathering in business ideas		
<b>Q3</b>		
Organisation of hackathon events in future	Communication with companies must support	The city must be very initiative and responsive in

Active and direct communication channels	the creation of new business ideas	communication with companies
City's real commitment	City must take into account the results of communication with the companies	

There are three axial codes for question Q1. The first one identifies the business ideas as the main source of datasets needs and usability. The second axial code points out the importance of the asset of datasets. The last axial code refers to the availability of data. The selective code states that the city must be able to provide datasets that cover real needs of local companies.

There is only one open code for question Q2. There can not be any relationships identified.

There are two axial codes for Q3. The first axial code says that the city must support business ideas based on municipal datasets. The second axial code emphasises the importance of feedback. The selective code states that the communication must be interactive and initialised by the city.

## 5.2 Analysis of the Survey

The data collected from the survey is categorised. The categorisation is based on the survey questions. The responses to some of the questions affect the analysis of other categories of data.

### 5.2.1 Business Location

The first question of the survey asked if the company produces the products or services in Lahti area. The answers should ensure that only relevant data is analysed.

As expected, 100% of the companies run their business in Lahti area. Therefore, the data from all the participating companies can be included in the analysis.

### 5.2.2 Awareness of Open Data

The analysis of the second category data reveals the level of participants' understanding of open data. The percentage of answers is shown below in Figure 4.

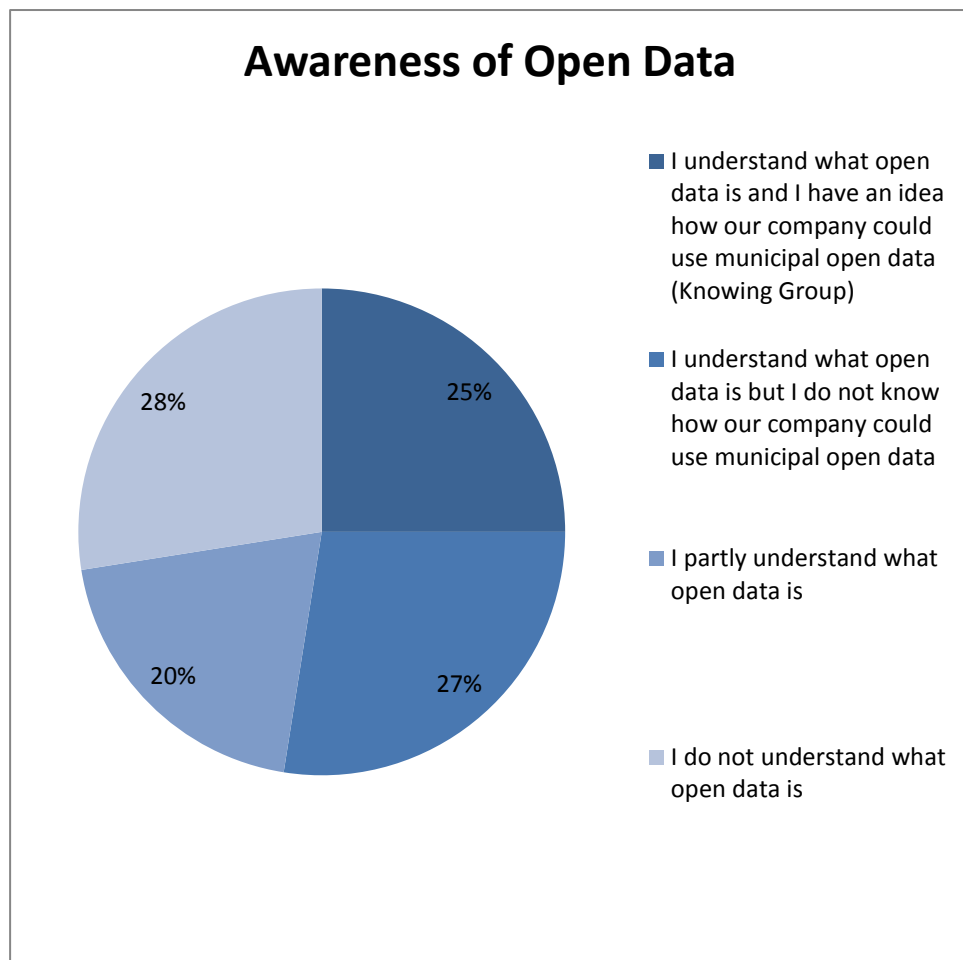


Figure 4 - Awareness of Open Data

There are 25% of companies who declared that they not only understand the concept of open data but they also have an idea how they could use municipal open data. For future reference this group of companies will be referred to as the Knowing Group.

The concept of open data is clear to 52% of participants. This fact does not remove the rest of the companies from the survey. Even if the company does not have any particular idea about open data, they have their plans regarding products and services. The company may assume what kind of information could support their innovations.

However, the level of awareness will be taken into account when analysing the following categories of data. The reason is that the relevancy of data requirements can vary between the four categories of answers to this question.

### 5.2.3 Innovations and Business Development Plans

The data of this category comes from answers to two questions. At first, the companies were asked about their business development plans within two years. Then they were asked about their business development plans assuming they can freely access municipal data. It is generally expected that the business development intentions increase when companies can access the municipal data. Despite this fact, twelve participants surprisingly decreased the business development intentions if they had access to municipal data. The possible explanation is that the question was understood incorrectly by those companies. The answers were removed from this category of data. The analysis was done for the rest of the companies including the companies from the Knowing Group.

Each company evaluated individually five business development categories. They evaluated how likely it is that their company will

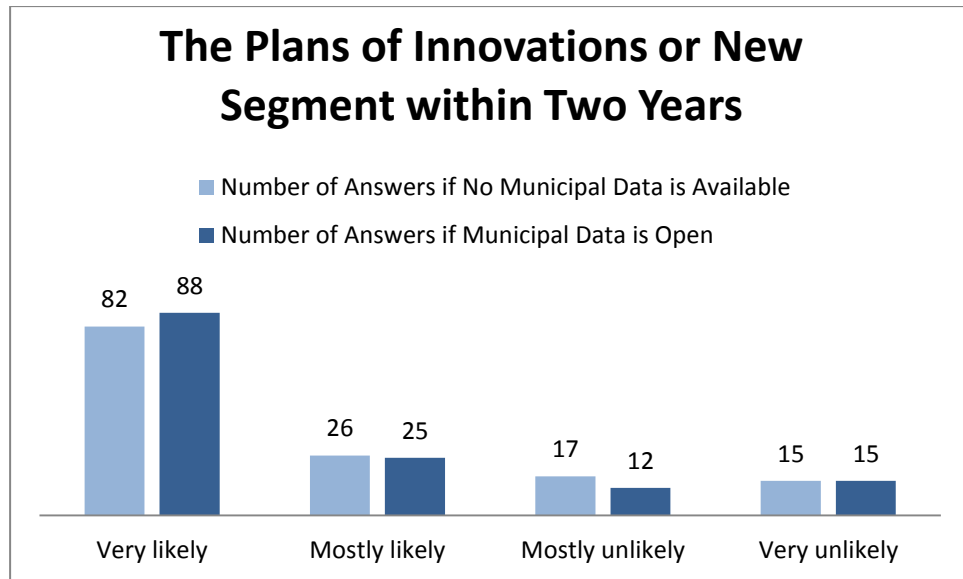
- innovate existing products

- innovate existing services
- develop new product
- develop new services
- reach out to new target group of customers.

There were four levels that could be used as answers. Answers in all the categories were counted together for each level. The answers show the overall business development plans. The number of answers in the chart below in Figure 5 shows how many times the companies answered that they would innovate their products or services or reach out to a new segment within two years. The exact number of answers is shown above the corresponding column. For each level of likelihood there is a comparison of two cases. The first set of columns (light blue) shows the decisions made assuming that the companies have no access to municipal data. In the second set of columns (dark blue) there is a number of answers about the business development in a situation when the companies could freely use the municipal data.

The chart shows a slight increase of innovation or expansion plans in case that the municipal data is open. However, the difference is very small.

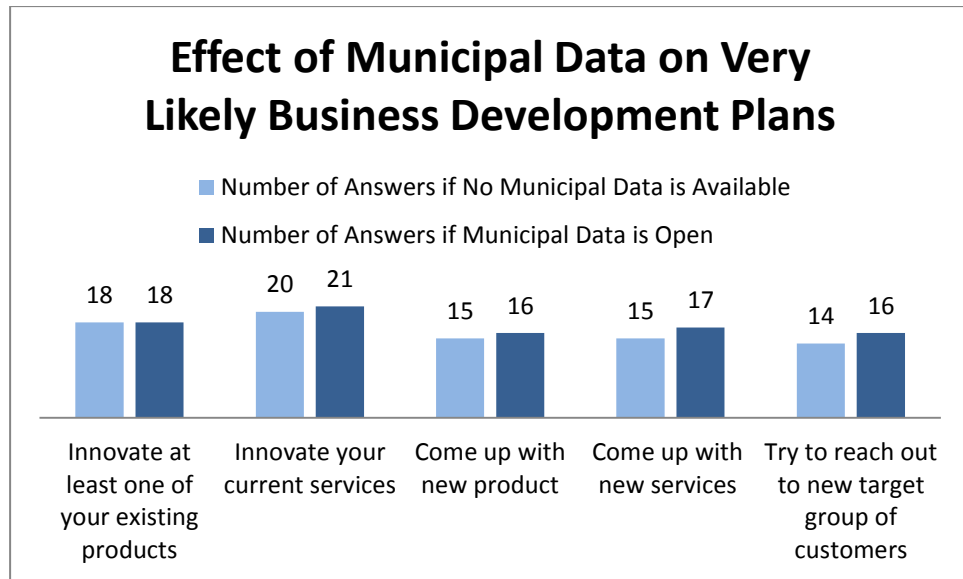




**Figure 5 – Business Development Plans of the Companies within Two Years**

The effect of municipal data on companies' decisions about individual categories of business development is illustrated below in Figure 6. The chart includes only the answers of the companies that said that the business development was very likely in the category. The exact number of answers is visible above the columns. There are two sets of columns for each business development category. The first one (light blue) shows the number of answers provided that the companies can not access the municipal data. In the second set (dark blue) there is a number of answers in a situation when municipal open data is available.

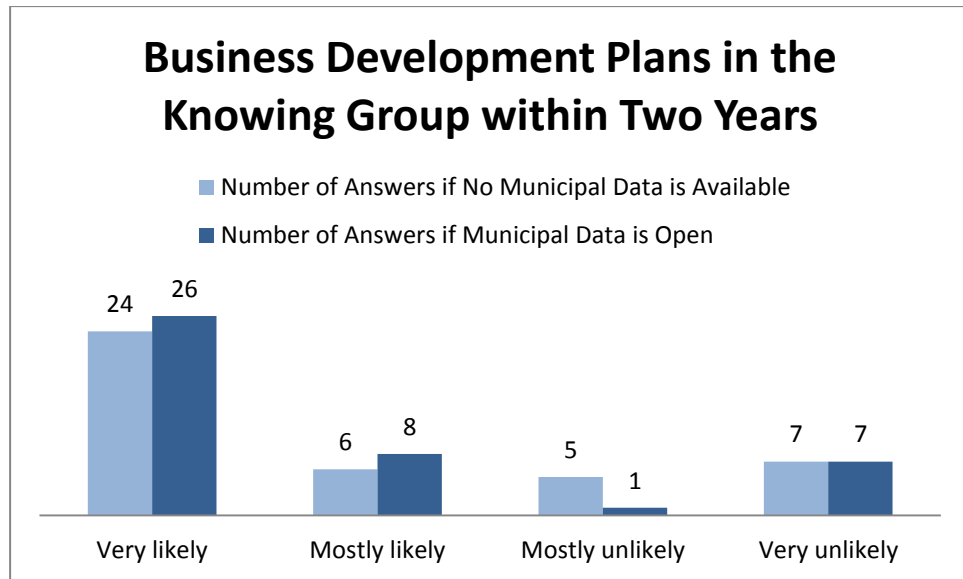
There is a small increase of business development plans. The chart shows that the municipal data affects the decision to introduce new services or reach out to new segment the most.



**Figure 6 - Effect of Municipal Data on Business Development Plans by Categories**

As an additional analysis of this category of data, the analysis of the Knowing Group's answers only was done. The chart below in Figure 7 contains the numbers of answers about all the business development categories together. The exact number of answers of the companies is shown above the columns. For each level of likelihood there are two sets of columns. The decisions in the situation when the companies have no access to municipal data (light blue) are compared with the situation when the municipal data is open (dark blue).

The most relevant information from this chart is that the mostly unlikely business development plans were almost completely transformed into mostly likely or very likely business development.



**Figure 7 - Business Development Plans of the Knowing Group within Two Years**

#### 5.2.4 Datasets Requirements

The data of this category comes from the selection of municipal datasets required by the companies including the companies from the Knowing Group. The chart below in Figure 8 shows the results.

## The Most Required Municipal Datasets

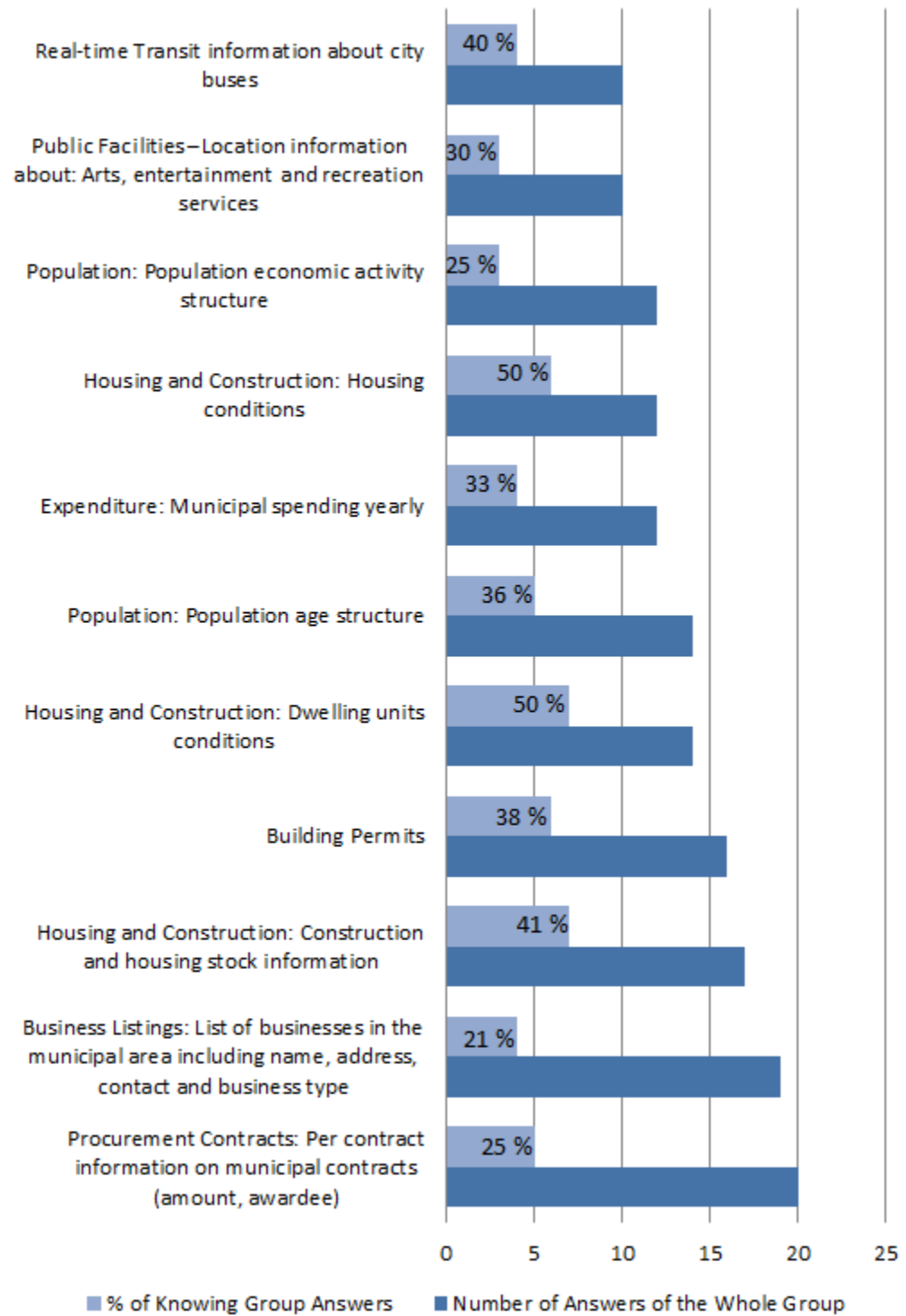


Figure 8 - The Most Required Municipal Datasets

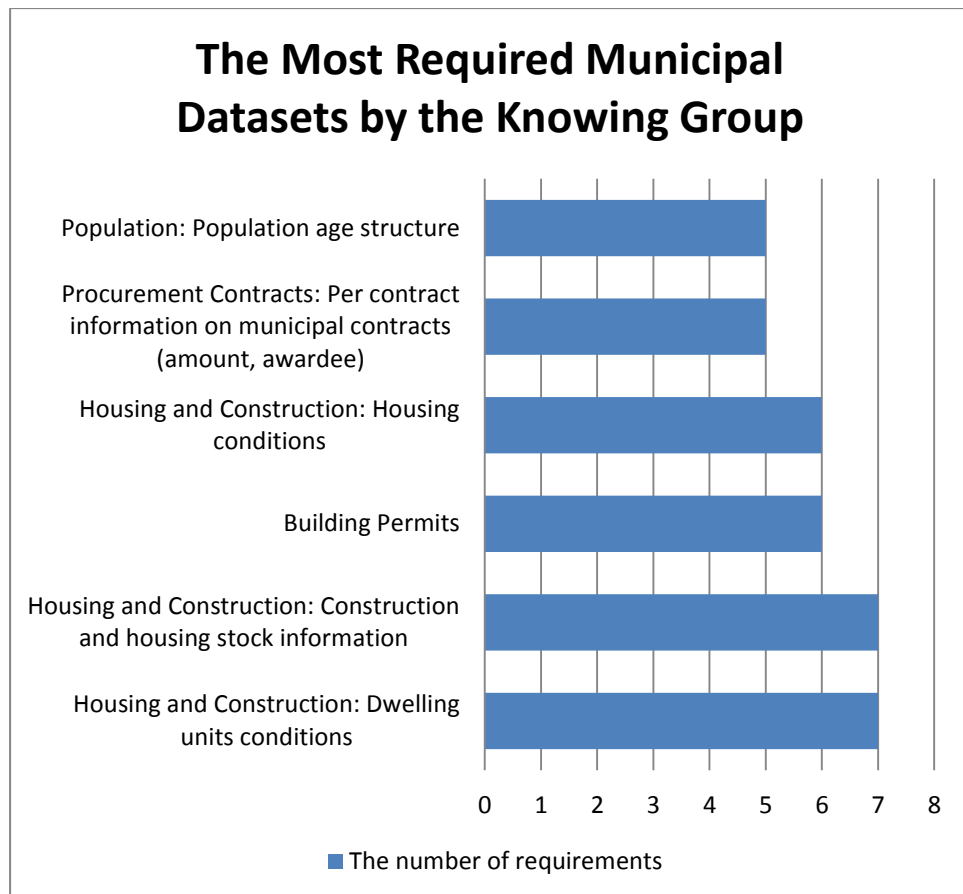
The chart shows the number of companies who require a particular dataset. Only datasets required by ten or more companies are included. The percentage shows the share of Knowing Group companies in each dataset.

The preferences of datasets are quite significant considering that there were 48 datasets categories to select from. The procurement contracts, business listings, construction and housing stock information and building permits are required by more than 15 companies participating in the survey which is 37,5 %.

Additionally, the analysis of the Knowing Group's answers only was done. A complete list of datasets required by this group of companies can be found in the Appendices (Appendix 7).

The chart below in Figure 9 shows the number of companies from the Knowing Group who require a particular category of datasets. The chart shows only the datasets required by five or more companies.

The four most required categories of datasets are related to housing and construction and building permits. Also in this group of companies the procurement contracts is a category that is significantly desired.



**Figure 9 - The Most Required Municipal Datasets by the Knowing Group**

There are these datasets suggested and required directly by the companies:

- repairs and construction of public buildings (schedule information, developers, contractors, designers)
- traffic in the city center (including pedestrians)
- annual plan and budget of technical board.

These are the datasets suggested and required by the companies from the Knowing Group:

- energy consumption information and the energy strategy of city of Lahti and the will of the city to implement measures focused on energy efficiency

- energy savings targets for various construction projects, the introduction of new innovations in the field of energy.

## 6 CONCLUSIONS

This chapter contains the most important findings of the work. It provides the answers to the research questions and a summary and assessment of the work.

### 6.1 Answers to the Research Questions

*What municipal datasets should city of Lahti open up to support regional development through support of local businesses?*

This is the main research question. The question can be answered from three different perspectives. The first one is based on the experience and best practice of other cities. This approach is applied because it is not only recommended by the literature but it also results from the analysis of this work. The recommended datasets based on the experience of other cities are listed below in Table 9.

**Table 9 – Recommended Datasets based on the Experience of Other Cities**

<b>Rate</b>	<b>Dataset</b>
<b>1</b>	Election areas Public facilities Transport timetables Real-time transit
<b>2</b>	Air quality
<b>3</b>	Annual budget Business listings Procurement contracts Agendas and decisions Service requests



The list of datasets comes from the experience of cities in Finland. As it was stated earlier in this document, each city has its own specifics that should be taken into account when defining the content of datasets. The list of most open datasets of other cities does not contain any details about the usage of data. However, it is expected that the datasets were open after the analysis of local needs and potential was done. If some kind of datasets works for many other cities, there is a big chance that it will work for city of Lahti, too.

The second approach to the main research question comes from the project of the 6 biggest Finnish cities. This perspective is more specific and includes also the real use of data. The list of recommended datasets contains

- public transportation
- traffic
- parking information
- geospatial data
- city services
- events.

There are other aspects important to the decision about the first datasets that city of Lahti should open. The datasets and their metadata must be well documented. The licence must allow publishing. The form of the data must be easy to publish. The use cases of data should be simple.

The last but not the least important approach to the main research question takes into consideration the communication with local companies. The companies from Lahti area expressed their preferences of datasets in the survey. The list of the most important datasets is shown below in Table 10.

**Table 10 - Recommended Datasets based on the Survey**

<b>Dataset</b>	<b>Rating by the Whole Group</b>	<b>Rating by the Knowing Group</b>
Procurement contracts: Per contract information on municipal contracts (amount, awardee)	<b>1</b>	<b>5</b>
Business listings: List of businesses in the municipal area including name, address, contact and business type	<b>2</b>	<b>9</b>
Housing and construction: Construction and housing stock information	<b>3</b>	<b>2</b>
Building permits	<b>4</b>	<b>3</b>
Housing and construction: Dwelling units conditions	<b>5</b>	<b>1</b>
Population age structure	<b>6</b>	<b>5</b>
Expenditure: Municipal spending yearly	<b>7</b>	<b>11</b>
Housing and construction: Housing conditions	<b>8</b>	<b>4</b>
Population economic activity structure	<b>9</b>	<b>20</b>
Public Facilities – Location information about: Arts, entertainment and recreation services	<b>10</b>	<b>14</b>
Real-time transit information about city buses	<b>11</b>	<b>7</b>

The list includes requirements of all the participants. However, there was a couple of companies who answered in the survey that they already knew how they would use the municipal data for their business ideas. This group was named Knowing Group. It is worth considering whether their requirements are more relevant for decision-making.

Two different data sources in this study pointed out the field of energy savings. One source was the interview about business ideas of local start-ups. Another source was the survey. The attention of the city should be focused on this field. Moreover, these datasets were directly required by the companies who already know how they would use the municipal data. The data that this group of companies requires is

- the energy consumption information and the energy strategy of city of Lahti and the will of the city to implement measures focused on energy efficiency
- the energy savings targets for various construction projects, the introduction of new innovations in the field of energy.

The other specific requirements of other companies that should not be missed are

- the repairs and construction of public buildings (schedule information, developers, contractors, designers)
- the traffic in the city center (including pedestrians)
- the annual plan and budget of technical board.

All three approaches should be considered when deciding about the first datasets to be published. The answers to the subsidiary research questions should support the main research question with additional insight.

*Is the current knowledge of open data among companies sufficient to identify what municipal datasets they find useful?*

The answer to the second research question is that the companies in Lahti region have a very low level of knowledge of open data. The evaluation of their needs for municipal data is mostly theoretical and hypothetical. This can negatively affect the relevancy of their requirements. There is a risk that more potential datasets will not be revealed and published because the companies do not know how they could use the data, or even what open data is. The companies need to find business ideas.

Most of the companies do not really think about their business development in relation to open data. As the analysis in this work indicates, opening municipal data does not affect their development decisions too much. Only a few companies have a clear idea about the use of municipal data for their business development. On the other hand, these companies showed evidently that municipal datasets would make a

big difference and they would make their business development plans very likely.

The number of companies with a real need for municipal data is low. However, it can be much higher as the awareness of municipal open data increases.

*Which communication channels should city of Lahti use to increase the understanding and use of open data by companies?*

The analysis showed that the communication with local community is the core of identification of datasets. The experience of other cities confirmed that the following ways of communication are the most effective:

- kick-off meeting at the beginning of the project for local community including companies, researchers, students and developers
- hackathon events and competitions for the same community
- active participation in existing communities on social media
- local media.

The city should do the work in cooperation with the regional development organisation. Ladec is already considering activities in the field of municipal open data.

## 6.2 Summary

The purpose of this work was to identify the datasets that city of Lahti should open for the local community with the intention to support regional economic and technological development. The process of searching for datasets is more complex. The context of identification of the datasets is important. Thus, this thesis investigated the needs for municipal datasets together with the open data awareness and the ways in which it can be improved and how to develop effective communication.

The most important part of the work was the survey. Inviting the companies to the survey was the first contact of the city of Lahti and Ladec

with the companies in the field of municipal open data. The feedback showed that the topic must be explained to the companies. The companies are interested in many datasets but they are not sure about the use.

Nonetheless, there are findings that point out some datasets as more promising than the others. The datasets were identified by the companies who have ideas about business development using municipal data. The energy savings and building and construction appear to be quite interesting areas for those companies.

The best practice of other cities is a valuable source of facts about the use of municipal open data. It indicates that the requirements of the companies in Lahti area are not so far from the real needs of other cities.

The thesis gives importance to direct and open communication initiated by city of Lahti and Ladec. The identification of datasets is an open process where response to the feedback and continual modification and updates are obligatory. If a dataset is open, the work does not end there. In addition, the ideas and the issues that have to be solved are the generators of municipal data requirements. It is possible to find them only if there is an active cooperation between the city, regional development organisation and local community of companies, developers, students, researchers and all the enthusiastic people.

In summary, the findings of this thesis provide a picture of the initial phase of a bigger project. It can help the city of Lahti to determine what datasets to focus on in the beginning. There is no doubt that the further cooperation of the city with the local community will reinforce and develop the findings.

### 6.3 Generalisability of the Results

The sample of this study includes 40 companies from Lahti region. There were 1800 companies invited to the survey. The number of companies that provided data for this study is very low in comparison to the number of potential users of municipal data of city of Lahti. Thus, the findings may not

be entirely accurate or reliable. On the other hand, all the companies were given the chance to formulate their needs. It was their choice not to do so. That can be also a signal indicator for this study.

The survey does not provide any identification of the persons who answered the questions on behalf of the companies. There was no specification of the person who should fill in the questionnaire either. Therefore, the awareness of the respondents may not entirely reflect the awareness of the companies.

This thesis uses three different sources of data. The data comes from the local companies, Lahti regional development expert and a coordinator of a big project of municipal open data of six cities. The findings reflect the range of factors which affects the generalisability of the results positively.

In conclusion, the generalisation of results of this work is very limited. It is possible to make a general statement that the awareness is low and the cooperation is crucial for future success. It is not possible to generalise the findings about datasets to all companies in Lahti area.

## 7 DISCUSSION

The findings of this study correspond with the recommendations coming from earlier research and experience of other cities. The city of Lahti can use this document as one of the sources of information in the decision-making process about municipal datasets. This study does not solve the task of the city completely but it provides a starting point and recommendations for further development.

In addition, the study reveals also a number of questions that need to be investigated in early stages of the project. First of all, the identification of availability and form of municipal data is crucial for datasets definition. The analysis of the data should be done together with the description of the ways of opening and keeping the data updated. Companies need to know what municipal data they can use before they develop new business ideas.

Another key point is the identification of responsibility for the data. The data must be ready for publishing. The published data must be kept updated. This may require the definition of new internal processes of city of Lahti.

## LIST OF REFERENCES

### Printed sources:

Brinkmann, S. & Kvale, S. 2015. *InterViews: Learning the Craft of Qualitative Research Interviewing*. London: SAGE Publications.

Floridi, L. 2008. Data. In Darity, W. (ed.) *International Encyclopedia of the Social Sciences*. 2<sup>nd</sup> edition. Detroit: Macmillan, preprint.

Gill, J. & Johnson, P. 2002. *Research Methods for Managers*. Third Edition. London: SAGE Publications.

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

Krishnaswami, O. & Satyaprasad, B. 2010. *Business Research Methods*. Mumbai: Himalaya Publishing House.

Mack, N., Woodsong, C., MacQueen, K., Guest, G. & Namey, E. 2005. *Qualitative Research Methods: A Data Collector's Field Guide*. North Carolina, USA: Family Health International.

Rosenberg, D. 2013. Data Before the Fact. In Gitelman, L. (ed.) "Raw Data" is an oxymoron. Cambridge, Massachusetts: MIT Press, 15-40.

Ruel, E., Wagner, W. & Gillespie, B. 2016. *The Practice of Survey Research: Theory and Applications*. London: SAGE Publications.

Veljkovic, N., Bogdanovic-Dinic, S. & Stoimenov, L. 2011. Municipal Open Data Catalogues, chapter in Danube University Krems CeDEM11: *Proceedings of the International Conference for E-Democracy and Open Government*. Krems: Edition Donau-Universität Krems, 195-207.

### Digital sources:

6Aika. 2017. Spearhead projects. 6Aika [accessed 18 March 2017]. Available at: <https://6aika.fi/spearhead-projects/>



Arctic Startup. 2016. OutHack: The Search For Truth About Finns. Arctic Startup [accessed 24 March 2017]. Available at:

<http://arcticstartup.com/article/outhack-search-truth-finns/>

AvoinTampere. 2017. Open Tampere – Minds Wide Open. AvoinTampere [accessed 25 March 2017]. Available at:

<http://www.avointampere.fi/en/the-program/>

Cashman, T. 2014. The Economic Impact of Open Data. Socrata. Blog [accessed 5 February 2017]. Available at:

<https://socrata.com/blog/economic-impact-open-data/>

Chignard, S. 2013. A Brief History of Open Data. ParisTech Review [accessed 4 February 2017]. Available at:

<http://www.paristechreview.com/2013/03/29/brief-history-open-data/>

Conradie, P. & Choenni, S. 2014. On the Barriers for Local Government Releasing Open Data. Government Information Quarterly. Vol. 31, Supplement 1 June 2014 [accessed 17 February 2017]. Available at: [https://www.researchgate.net/publication/261989071\\_On\\_the\\_barriers\\_for\\_local\\_government\\_releasing\\_open\\_data](https://www.researchgate.net/publication/261989071_On_the_barriers_for_local_government_releasing_open_data)

European Commission. 2017. South Finland. European Commission [accessed 21 March 2017]. Available at: <https://ec.europa.eu/growth/tools-databases/regional-innovation-monitor/base-profile/south-finland-0>

European Data Portal. 2016. Analytical Report n4: Open Data in Cities. European Data Portal [accessed 17 February 2017]. Available at: [https://www.europeandataportal.eu/sites/default/files/edp\\_analytical\\_report\\_n4\\_-\\_open\\_data\\_in\\_cities\\_v1.0\\_final.pdf](https://www.europeandataportal.eu/sites/default/files/edp_analytical_report_n4_-_open_data_in_cities_v1.0_final.pdf)

Eurostat. 2017. GDP at regional level. Eurostat [accessed 27 February 2017]. Available at: [http://ec.europa.eu/eurostat/statistics-explained/index.php/GDP\\_at\\_regional\\_level](http://ec.europa.eu/eurostat/statistics-explained/index.php/GDP_at_regional_level)

Future Dialog. 2017. Case: City of Lahti / Porukka. Future Dialog [accessed 24 March 2017]. Available at: <https://www.futuredialog.fi/client/lahti-porukka>

Gallicano, T. 2013. An Example of How to Perform Open Coding, Axial Coding and Selective Coding. The PR Post [accessed 19 March 2017]. Available at: <https://prpost.wordpress.com/2013/07/22/an-example-of-how-to-perform-open-coding-axial-coding-and-selective-coding/>

Gurstein, M. 2011. Open data: Empowering the Empowered or Effective Data Use for Everyone? First Monday. Nro 16/2011 [accessed 30 January 2017]. Available at: <http://firstmonday.org/ojs/index.php/fm/article/viewArticle/3316/2764>&quot;%3b&gt;%3b

Halonen, A. 2012. Being Open About Data. Analysis of the UK open data policies and applicability of open data. Finnish Institute in London [accessed 5 February 2017]. Available at: <http://finnishinstitute.cdn.coucouapp.com/media/W1siZiIsIjIwMTYvMDgvMjUvMDkvMzgvMjEvZGQyOGQ5MTUtY2Q4YS00N2U0LWEyZTAzM2FiOWUxN2M0Njc2L2JlaW5nIG9wZW4gYWJvdXQgZGF0YS5wZGYiXV0/being%20open%20about%20data.pdf?sha=cad600352a6ffb0b>

Ladec. 2017a. Protomo – Älä yritä yksin! Ladec [accessed 24 March 2017]. Available at: <https://www.ladec.fi/protomo>

Ladec. 2017b. We Are Here for Your Company. Ladec [accessed 27 February 2017]. Available at: <https://www.ladec.fi/en/our-services>

Lahti. 2017a. Lahti Info. Lahti [accessed 21 March 2017]. Available at: <https://www.lahti.fi/en/information-about-lahti/lahti-info>

Lahti. 2017b. Tilastot ja Tilastojulkaisut. Lahti [accessed 13 February 2017]. Available at: <https://www.lahti.fi/tietoa-lahdesta/tilastot-ja-tilastojulkaisut>

Lahti. 2017b. TILDA – tilastotietoa Lahdesta. Lahti [accessed 13 February 2017]. Available at: <http://www4.lahti.fi/verkkotilastointi/>

Ministry of Justice, Finland. 1999. Act on the Openness of Government Activities. Ministry of Justice, Finland. [accessed 5 February 2017]. Available at: <http://www.finlex.fi/en/laki/kaannokset/1999/en19990621.pdf>

Moneo, A. 2016. Five Strategies to Enact a Municipal Open Data Policy. Blog [accessed 18 March 2017]. Available at: <http://opendatacon.org/five-strategies-enact-municipal-open-data-policy/>

Open Data Institute. 2017. Publisher's Guide to Open Data Licensing. Open Data Institute [accessed 30 January 2017]. Available at: <http://theodi.org/guides/publishers-guide-open-data-licensing>

Open Data Institute. 2015. Open Data Means Business: UK Innovation Across Sectors and Regions. Open Data Institute [accessed 11 February 2017]. Available at: <http://theodi.org/open-data-means-business>

Open Government Partnership. 2013. Finland's Action Plan on Open Government. Open Government Partnership [accessed 5 February 2017]. Available at: [http://www.opengovpartnership.org/sites/default/files/legacy\\_files/country\\_action\\_plans/20130314%20OGP%20Action%20Plan%20Finland.pdf](http://www.opengovpartnership.org/sites/default/files/legacy_files/country_action_plans/20130314%20OGP%20Action%20Plan%20Finland.pdf)

Open Government Partnership. 2017. Finland: Introduction. Open Government Partnership [accessed 7 April 2017]. <https://www.opengovpartnership.org/country/finland>

Open Knowledge International. 2017a. Finland Local Open Data Census. Open Knowledge International [accessed 11 February 2017]. Available at: <http://fi-city.census.okfn.org/>

Open Knowledge International. 2017b. Open Definition 2.1. Open Knowledge International [accessed 10 February 2017]. Available at: <http://opendefinition.org/od/2.1/en/>

Open Knowledge International. 2017c. The Open Data Handbook [accessed 10 February 2017]. Available at:  
<http://opendatahandbook.org/guide/en/how-to-open-up-data/>

Pollock, R. 2008. The Economics of Public Sector Information. University of Cambridge [accessed 5 February 2017]. Available at:  
<http://www.econ.cam.ac.uk/research/repec/cam/pdf/cwpe0920.pdf>

Pollock, R. 2006. The Value of the Public Domain. Institute for Public Policy Research [accessed 5 February 2017]. Available at:  
[http://rufuspollock.org/papers/value\\_of\\_public\\_domain.ippr.pdf](http://rufuspollock.org/papers/value_of_public_domain.ippr.pdf)

Statistics Finland. 2017. Description of Statistics. Statistics Finland [accessed 14 February 2017]. Available at:  
[http://www.stat.fi/meta/til/index\\_en.html](http://www.stat.fi/meta/til/index_en.html)

Statistics Finland. 2008. Standard Industrial Classification TOL 2008. Statistics Finland [accessed 12 February 2017]. Available at:  
[http://tilastokeskus.fi/meta/luokitukset/toimiala/001-2008/index\\_en.html](http://tilastokeskus.fi/meta/luokitukset/toimiala/001-2008/index_en.html)

Sulopuisto, O. 2014. How Helsinki Became the Most Successful Open-Data City in the World. Citylab [accessed 11 February 2017]. Available at:  
<http://www.citylab.com/tech/2014/04/how-helsinki-mashed-open-data-regionalism/8994/>

Wasenco. 2017. Don't pour energy down the drain – recover it and reuse it!. Video. Wasenco [accessed 24 March 2017]. Available at:  
<http://wasenco.com/en/>

World Wide Web Consortium. 2016. Data on the Web Best Practices. W3C [accessed 27 January 2017]. Available at:  
<https://www.w3.org/TR/2016/PR-dwbp-20161215/>

Yiu, C. 2012. A Right to Data: Fulfilling the Promise of Open Public Data in the UK. Policy Exchange [accessed 5 February 2017]. Available at:  
<https://policyexchange.org.uk/wp-content/uploads/2016/09/a-right-to-data-mar-12.pdf>

Translation:

Monni, M. 2017. Information Technology Manager. City of Lahti.

## APPENDICES

### APPENDIX 1. Questions for the Interview with the Ladec Development Manager

1. From your current cooperation with existing or potential companies in Lahti region can you identify that they want to develop new products or services based on data? If so, is it also municipal data?
2. If companies intend to come up with new products or services based on municipal data, can you identify areas of their interest? (areas of life, e.g. education, health care, traffic, air pollution etc.)
3. Do you have information about companies that were struggling with their products or services because of a lack of municipal data? If so, which data it was?
4. From your experience – what is your opinion, what municipal data have the biggest business development potential?
5. Can you see promising business ideas in products or services directly connected to public products or services? If so, in which areas it could be the most?

### APPENDIX 2. The List of Business Ideas of Start-Up Companies Established with the Help of Ladec Start-Up Services

- optimisation of cleaning services
- building information management software
- website for municipalities requesting offers for smaller purchases
- floor cooling system
- job seeking
- new television graphics
- programmable gaming device
- heating solutions for restaurant food
- 3D advertisement for web shops
- clothing
- sport coaching

- open sources software and services
- jewelery
- game company producing special games, for example for children in hospitals
- piper construction concept
- appealing shelves for stores
- advertisement company
- drone taking photos
- mobile application for municipalities
- solar heating solutions
- devices for energy saving in buildings
- new building using metal structure inside the buildings
- bus company
- chemicals
- tar paper recycling
- ICT company
- selling company
- flight airplane technology
- turbine solution for energy use
- public responsibility consultation
- 3D printing company
- diving equipments
- diving navigator
- ice hockey concept
- metal 3D printing
- international trade to Southern America
- gas turbine
- cleaning of dirty water.

### APPENDIX 3. Questions for the Interview with the Project Coordinator of the Six City Strategy Project

1. How would you describe the role and main activities of Tredea in municipal open data project in Tampere?

2. What do you consider to be the most effective forms, ways, tools of promoting the awareness of municipal open data among companies and citizens?
3. The city of Lahti needs to identify what municipal datasets should be opened as first. The main goal is to support the local businesses/end users in the development of new services, products and technologies. What is crucial in communication with companies/end users about the initial datasets that should be opened?
4. What aspects, other than communication with end users, a city should take into account when defining the first datasets that should be open?
5. When you look back at the beginning of your municipal open data project, what would you do differently in the process of definition of the first datasets for business development? (Maybe some advice for city of Lahti how to avoid the mistakes?)
6. Which ways of requirements collection or identification of companies' needs for municipal data appeared as the most efficient in your project?
7. From the cooperation of the 6 biggest Finnish cities – can you say that some types of datasets are more required and used in general for development of new services or products? If so, could you specify them? If not, you do not have to answer this question.
8. Would you say that there are datasets that a city can open anyhow and it will be beneficial for local companies? (Even without asking the companies?) If so, could you specify them? If not, you do not have to answer this question.
9. Is there anything that you would like to add as a recommendation for a city that is starting the identification of municipal datasets that should be open?



#### APPENDIX 4. Questions for the Survey in English

1. Does your company produce products or services in Lahti area?
2. Please, evaluate how well do you understand what open data is.
3. Within next 2 years, how likely it is that you will innovate at least one of your existing products, innovate your current services, come up with new product, come up with new services, try to reach out to new target group of customers?
4. If you have an access to municipal data of city of Lahti, how likely it is that you will do within next 2 years innovate at least one of your existing products, innovate your current services, come up with new product, come up with new services, try to reach out to new target group of customers?
5. In case that you answered "Very likely" or "Mostly likely" at least in one of questions 3. or 4., what kind of municipal data would you need/find helpful to make the plans real?

The English language version of dataset categories can be found in Appendix 6 in Table 12.

6. Is there anything you would like to add?

#### APPENDIX 5. Questions for the Survey in Finnish

(translated from English by Monni)

1. Tuottaako yrityksenne tavaroita tai palveluita Lahden alueella?
2. Olkaa hyvä ja arvioikaa kuinka hyvin tunnette avoimen datan käsitteen ja merkityksen.
3. Kuinka todennäköisesti seuraavan kahden vuoden aikana: kehitämme vähintään yhtä olemassa olevaa tuotettamme, kehitämme olemassa

olevaa palveluamme, julkaisemme uuden tuotteen, julkaisemme uuden palvelun, yritämme löytää uuden asiakassegmentin?

4. Jos yrityksellänne olisi käytettävissä pääsy Lahden kaupungin avoimeen dataan, kuinka todennäköisesti seuraavan kahden vuoden aikana: kehitämme vähintään yhtä olemassa olevaa tuotettamme, kehitämme olemassa olevaa palveluamme, julkaisemme uuden tuotteen, julkaisemme uuden palvelun, yritämme löytää uuden asiakassegmentin?

5. Jos vastasit “Hyvin todennäköisesti” tai “Melko todennäköisesti” ainakin yhteen kysymysten 3 tai 4, niin mikä kunnan tarjoama avoin data olisi hyödyllistä tai tarpeellista suunnitelmien toteuttamiseksi?

**Table 11 - Categories of Data in Survey, Finnish version (Translated from English by Monni)**

1	Public Facilities - Sijaintitieto: Koulut ja oppilaitokset
2	Public Facilities - Sijaintitieto: Terveystieteiden ja sosiaalitoimen palvelut
3	Public Facilities - Sijaintitieto: Taidelaitokset, vapaa-aika- ja virkistyspalvelut
4	Reaaliaikainen joukkoliikennetieto (Reaaliaikainen tieto julkisesta liikenteestä, linja-auto, lähi- ja kaukoliikenne, juna. Tieto sisältää mm reaaliaikaisen sijaintitiedon): Reaaliaikainen tieto paikallisliikenteestä
5	Reaaliaikainen joukkoliikennetieto: Reaaliaikainen tieto kaukoliikenteestä
6	Reaaliaikainen joukkoliikennetieto: Reaaliaikainen tieto junista
7	Äänestysalueet, Kunnan sisäinen äänestysaluejako, Äänestysalueiden rajat
8	Aikataulut: Paikallisliikenteen aikataulut
9	Aikataulut: Kaukoliikenteen aikataulut
10	Aikataulut: Junien aikataulut
11	Aikataulut: Paikallisliikenteen pysäkkien sijaintitiedot
12	Aikataulut: Kaukoliikenteen pysäkkien sijaintitiedot
13	Liiketoimintatieto: Lista yrityksistä sisältäen yrityksen nimi, osoite ja kontaktitiedon sekä liiketoiminta-alueen
14	Ilmanlaatu: Ilmanlaatu kuukausittain
15	Ilmanlaatu: Ilmanlaatu päivittäin
16	Ilmanlaatu: Ilmanlaadun maantieteellinen jakauma
17	Hankintasopimukset: Sopimusten tiedot sopimuksittain: esimerkiksi sopimuksen

	kohteet, sopimuksen volyyymi, sopimusikumpanit
18	Kunnat menot (Kunnan menotiedot riittävällä tarkkuustasolla: vuosi- kuukausi- ja viikkotasolla. Kirjanpidon tilitapahtumat tapahtumatasolla (=tarkkustaso)): Kunnan menot vuositasolla
19	Kunnat menot: Kunnan menot kuukausitasolla
20	Kunnat menot: Kunnan menot viikkotasolla
21	Kunnat menot: Kunnan menot päivätasolla
22	Kunnan tulot: Tietoa kunnan verotuloista
23	Kunnan tulot: Tietoa kunnan muista taksoista ja maksuista
24	Kunnan vuosibudjetti: Kunnan vuosibudjetti toimialoittain
25	Kunnan vuosibudjetti: Kunnan vuosibudjetti osastoittain
26	Esityslistat ja pöytäkirjat: Kaupunginhallituksen esityslistat, kokousaikataulut, pöytäkirjat ja tehdyt päätökset
27	Esityslistat ja pöytäkirjat: Valtuuston ja lautakuntien esityslistat, kokousaikataulut, pöytäkirjat ja tehdyt päätökset
28	Esityslistat ja pöytäkirjat: Lautakuntien esityslistat, kokousaikataulut, pöytäkirjat ja tehdyt päätökset
29	Palvelupyynnöt: Ei kiireelliset palvelupyynnöt kunnalla, esimerkiksi kuopat tiestössä, ympäristön epäsiisteudet yms. Mukana paikkatieto.
30	Rakennusluvut
31	Asuminen ja rakennuskanta: Asunto-olot
32	Asuminen ja rakennuskanta: Asumismuodot
33	Asuminen ja rakennuskanta: Vapaa-ajan asunnot
34	Asuminen ja rakennuskanta: Rakentaminen ja asuntokanta
35	Asuminen ja rakennuskanta: Asuntojen hinnat
36	Asuminen ja rakennuskanta: Osakehuoneistojen hinnat
37	Asuminen ja rakennuskanta: Vapaa-ajan asuntojen hinnat
38	Väestö: Väestön ikärakenne
39	Väestö: Väestön sukupuolijakauma
40	Väestö: Väestön kansallisuus
41	Väestö: Väestön kielijakauma
42	Väestö: Väestön siviilisääty
43	Väestö: Väestön uskontojakauma
44	Väestö: Väestön koulutustaso
45	Väestö: Väestön taloudellinen asema
46	Väestö: Elämäntapahtumien määrä, esim. avioliitot, avioerot
47	Väestö: Koulutus rakenne
48	Väestö: Muuttotilastot
49	Väestö: Perherakenne
50	Väestö: Väestönkasvu

51	Työttömyys: Työttömyysaste
52	Työttömyys: Työttömyyden kehittyminen
53	Työttömyys: Työpaikkojen määrä
54	Koulutus, Koulutusorganisaatioiden tiedot: Aloittavien opiskelijoiden määrä (so. määrä ja jakaumat)
55	Kulttuuri ja vapaa-aika: Kirjastojen yleistiedot (kävijämäärät, lainamäärät)
56	Kulttuuri ja vapaa-aika: Kaupunginorkesterin konserteissa käyneet
57	Kulttuuri ja vapaa-aika: Kaupunginteatterin näytöksissä käyneet
58	Kulttuuri ja vapaa-aika: Museoiden vierailijamäärät
59	Emme tarvitse kunnan tarjoamaa avointa data toteuttaaksemme ideamme
60	Emme tiedä
61	Muu tieto

6. Haluatko lisätä jotain?

## APPENDIX 6. – Complete List of Answers to the Survey – Question 5

Table 12 – Complete List of Answers to the Survey <sup>1</sup> - Question 5

5. In case that you answered "Very likely" or "Mostly likely" at least in one of questions 3. or 4., what kind of municipal data would you need/find helpful to make the plans real?		
Category		Number of positive answers
1	Public Facilities – Location information about: Education services	8
2	Public Facilities – Location information about: Human health and social work services	9

<sup>1</sup> The definitions of categories are according to Standard Industrial Classification TOL 2008 (Statistics Finland 2008), Finland Local Open Data Census (Open Knowledge International 2017a), survey research practice (Ruel et al. 2016,33), statistical information from Lahti (Lahti 2017c), statistics and statistical publications (Lahti 2017b) and description of statistics (Statistics Finland 2017)

3	Public Facilities – Location information about: Arts, entertainment and recreation services	10
4	Real-Time Transit (Real-time information about major municipal-run or commissioned transit services (buses, subway, rail, tram etc). Real-time transit information means things like the location of actual services (individual buses and trains, etc)): Real-time information about city buses	10
5	Real-Time Transit: Real-time information about long distance buses	6
6	Real-Time Transit: Real-time information about trains	6
7	Election Areas, Boundaries of electoral districts in this municipality	1
8	Transport Timetables: Timetables of city buses	6
9	Transport Timetables: Timetables of long distance buses	3
10	Transport Timetables: Timetables of trains	4
11	Transport Timetables: Location of city buses stops	6
12	Transport Timetables: Location of long distance buses stops	3
13	Business Listings: List of businesses in the municipal area including name, address, contact and business type	19
14	Air Quality: Air quality data by months	2
15	Air Quality: Air quality data by days	7
16	Air Quality: Geographic breakdown of air quality data	4
17	Procurement Contracts: Per contract information on municipal contracts (amount, awardee)	20
18	Expenditure (Records of actual (past) municipal spending at a detailed transactional level, for example, at the level of month to month expenditure on specific items (usually this means individual records of spending amounts at a fairly granular level - e.g. \$5-50k rather than at the \$1m+ level). (Note: a database of contracts awarded or similar is <i>not</i> considered sufficient. This data category refers to detailed ongoing data on <i>actual</i> expenditure)): Municipal spending yearly	12

19	Expenditure: Municipal spending monthly	3
20	Expenditure: Municipal spending weekly	2
21	Expenditure: Municipal spending daily	2
22	Municipal revenue: Information of municipal tax revenue	4
23	Municipal revenue: Information of municipal fees revenue	6
24	Annual Budget: Municipal budget by sectors	8
25	Annual Budget: Municipal budget by departments	7
26	Agendas and Decisions: Agendas, meeting minutes and decisions of the municipal board	7
27	Agendas and Decisions: Agendas, meeting minutes and decisions of council	8
28	Agendas and Decisions: Agendas, meeting minutes, decisions of committees	8
29	Service Requests: Data on non-emergency services requests per request	8
30	Building Permits	16
31	Housing and Construction: Housing conditions	12
32	Housing and Construction: Dwelling units conditions	14
33	Housing and Construction: Free-time buildings conditions	8
34	Housing and Construction: Construction and housing stock information	17
35	Housing and Construction: Prices of houses	6
36	Housing and Construction: Prices of dwelling units	4
37	Housing and Construction: Prices of free-time buildings	3
38	Population: Population age structure	14
39	Population: Population sex structure	8
40	Population: Population nationality structure	5
41	Population: Population language structure	4
42	Population: Population citizenship structure	4

43	Population: Population religion structure	2
44	Population: Level of education	8
45	Population: Population economic activity structure	12
46	Population: Information about life events, specify (e.g. marriage, divorce etc.)	1
47	Population: Education structure	7
48	Population: Migration information	7
49	Population: Family information	4
50	Population: Population growth information	7
51	Unemployment: Unemployment over the population	8
52	Unemployment: Unemployment from time perspective	5
53	Unemployment: Vacancies over all the labor market	5
54	Education, Education institutions structure: Enrollment to educational institutions (e.g. structure and number of students)	6
55	Culture and leisure: City library overall information about customers (visits, borrowers etc.)	1
56	Culture and leisure: City orchestra overall information about visitors	1
57	Culture and leisure: City theatre overall information about visitors	1
58	Culture and leisure: Museums overall information about visitors	1
59	We do not find municipal data important to make our ideas real	2
60	We do not know	1
61	Other: <ul style="list-style-type: none"> <li>- Repairs and construction of public buildings (schedule information, developers, contractors, designers)</li> <li>- Traffic in the city center (including pedestrians)</li> <li>- Annual plan and budget of technical board</li> </ul>	5

	<ul style="list-style-type: none"> <li>- Open data also helps to sell anywhere if there is awareness of potential</li> <li>- Energy consumption information and the energy strategy of city of Lahti and the will of the city to implement measures focused on energy efficiency</li> <li>- Energy savings targets for various construction projects, the introduction of new innovations in the field of energy (Translated from Finnish by Zatkova)</li> </ul>	
--	--	--

## APPENDIX 7. – Complete List of Datasets Required by the Knowing Group

**Table 13 - Datasets Required by the Knowing Group**

<b>Category</b>	<b>Number of positive answers</b>
Housing and Construction: Dwelling units conditions	7
Housing and Construction: Construction and housing stock information	7
Building Permits	6
Housing and Construction: Housing conditions	6
Procurement Contracts: Per contract information on municipal contracts (amount, awardee)	5
Population: Population age structure	5
Real-Time Transit: Real-time information about city buses	4
Real-Time Transit: Real-time information about trains	4
Business Listings: List of businesses in the municipal area including name, address, contact and business type	4
Air Quality: Air quality data by days	4
Expenditure: Municipal spending yearly	4
Housing and Construction: Free-time buildings conditions	4
Public Facilities – Location information about: Human health and social work services	3
Public Facilities – Location information about: Arts,	3



entertainment and recreation services	
Real-Time Transit: Real-time information about long distance buses	3
Transport Timetables: Timetables of city buses	3
Transport Timetables: Timetables of trains	3
Transport Timetables: Location of city buses stops	3
Annual Budget: Municipal budget by sectors	3
Population: Population economic activity structure	3
Population: Population growth information	3
Public Facilities – Location information about: Education services	2
Transport Timetables: Timetables of long distance buses	2
Transport Timetables: Location of long distance buses stops	2
Air Quality: Geographic breakdown of air quality data	2
Expenditure: Municipal spending monthly	2
Service Requests: Data on non-emergency services requests per request	2
Housing and Construction: Prices of houses	2
Housing and Construction: Prices of dwelling units	2
Housing and Construction: Prices of free-time buildings	2
Population: Population sex structure	2
Population: Population language structure	2
Population: Population citizenship structure	2
Population: Level of education	2
Population: Migration information	2
Unemployment: Unemployment over the population	2
Unemployment: Unemployment from time perspective	2
Unemployment: Vacancies over all the labor market	2
Education, Education institutions structure: Enrollment to educational institutions (e.g. structure and number of students)	2
Air Quality: Air quality data by months	1
Expenditure: Municipal spending weekly	1
Expenditure: Municipal spending daily	1
Municipal revenue: Information of municipal fees revenue	1
Annual Budget: Municipal budget by departments	1
Agendas and Decisions: Agendas, meeting minutes, decisions of committees	1
Population: Population nationality structure	1
Population: Population religion structure	1

Population: Information about life events, specify (e.g. marriage, divorce etc.)	1
Population: Education structure	1
Population: Family information	1
Culture and leisure: City library overall information about customers (visits, borrowers etc.)	1