

Master thesis

In this thesis a research into success factors of open data is performed. Within the research, the motivations, benefits, and barriers of open data are also examined. To do so, a literature study is performed, 15 interviews are conducted, and an evaluation is executed.

Keywords: Open data, success factors, open data projects

A.K.G. Super

14-8-2015

Version 1.0

П

NAME: Anniek Super

STUDENT NUMBER: 3542319

CONTACT: anniek.super@gmail.com

TITLE THESIS: Success factors of open data projects RESEARCH PERIOD: December 1st, 2014 – August 21st, 2015 DATE OF DEFENCE: August 21st, 2015

INSTITUTION: Utrecht University DEPARTMENT: Information and Computing sciences GROUP: Organisation and Information MASTER: Business Informatics SUPERVISORS: prof.dr. ir. R.W. Helms & prof. dr. S. Brinkkemper

HOST INSTITUTION: Utrecht Data School SUPERVISOR: Dr. Mirko Tobias Schäfer

ACKNOWLEDGEMENTS

This is it! After nine months of working on this thesis, I can finally present it to you. I can definitely say that I have learned a lot during these months, as I had almost no knowledge of open data. This project showed me what open data comprehends, can bring, and how it is intertwined in people their lives.

Therefore, I would like to start with thanking all fifteen interviewees for their cooperation, new insights and the nice conversations. It is safe to state that it would not have been possible to complete this thesis without them. Furthermore, I would like to thank my first supervisor Remko, for his guidance during the process and providing me with his insights and feedback. I would also like to thank Sjaak, my second supervisor, for his input in the early phases of my thesis and the feedback at the end of the process.

Besides my supervisor from the university, I would also like to thank Mirko for his time and guidance during the project. Moreover, I am grateful for all the help and insights I got from people involved in the GOOD DATA project. Hopefully, this research can be of value for you.

I would like to thank my father for helping me tackle the database and my reviewers for reading this thesis and providing me with useful feedback. Finally, I would like to thank my friends and family for all their support and mainly their understanding during this hectic period.

Hopefully, you enjoy reading this thesis.

ABSTRACT

Open data is a timely topic. However, a clear definition is lacking, just as an overview of open data projects. Moreover, it is unknown what the benefits, barriers and motivations are concerning open data. It is also not clear when an open data project is a success and which factors contribute to that success. These subjects are covered in this thesis, resulting in the following main research question:

'What are success factors of open data projects?'

In order to answer this question a literature study has been performed. To validate the findings of the literature study, fifteen interviews have been conducted. To make sure it is not only a theoretical contribution, the success factors found in the interviews and literature are evaluated using existing open data projects.

First of all it is important to know what open data is, therefore, a definition is created. The definition is derived from literature. More importantly the definition is based on the idea that open data is not a type of data, it is about the way the data is made available. The definition is as follows:

Open data is data that is publicly available, easy accessible and has no restrictions for the users.

According to both literature and the interviews, open data is gaining momentum. The first initiatives are up and running and the first lessons have been learned from these examples. Therefore, this is the moment for organizations to start getting involved with open data. Moreover, the focus is shifting from opening as much data as possible towards realizing the benefits of open data projects. This implicates that it is more beneficial to join the movement now, as more attention will be paid to its benefits.

These benefits are promised to be huge. Both literature and the interviewees are confident that disclosing data will result in an increase in innovation. Furthermore, the relation between an organization and its customers will intensify, which might result in better products and services. The largest benefit is that of transparency, as the data is open and available for everyone, people will be able to see what happens within an organization. On the other hand, there are still some barriers for organizations to open their data, these include accountability and cultural issues. The added value of open data entails more than just the economic value that can be achieved, however, for organizations it is unclear what they can gain from open data.

During hurricane Katrina in 2005, the collaboration between organizations was very poor. This led to a huge amount of deaths, as all organizations came to the same place to provide help. If they would have made their data publicly available, they could have used that data to coordinate their help more efficiently. This is one example where open data could have been beneficial.

An open data project differs from other projects in that they cannot exist without the open data. Keeping that in mind, the analysis of the interviews and literature resulted in a list of 42 success factors. To create an overview and to make it more directly applicable, the success factors are mapped to the Open Data Process. The success factors provide organizations with a first indication of what they have to think of when getting involved with open data. Figure 9 (p. 74) provides an overview of all 42 success factors, mapped to the Open Data Process.

TABLE OF CONTENTS

Acknowledgements IV		
AbstractV		
List of Figures IX		
List of TablesX		
1. Introduction		
1.1 Problem Statement		
1.2 Research Questions		
1.3 Scoping		
1.4 Relevance		
1.5 Thesis Structure		
2. Research Approach		
2.1 Literature Study		
2.2 Interviews		
2.3 Evaluation		
2.4 Validity Threats		
3. Definition Open Data		
3.1 Constructing a Definition		
3.2 Concluding Remarks		
4. Theoretical Background		
4.1 Why now?		
4.2 What do we know so far?		
4.3 Promises of open data 12		
4.4 Motivations for open data14		
4.5 Barriers for open data reuse15		
4.6 Success and open data projects16		
4.7 Economic value		
4.8 Summary		

5.	Interviews	25
	5.1 Interviewees	25
	5.2 Results of interviews	25
	5.3 Summary	39
6.	Literature vs. Interviews	40
	6.1 Why now?	40
	6.2 What do we know so far?	40
	6.3 Promises of open data	41
	6.4 Motivations for open data	42
	6.5 Barriers for open data reuse	43
	6.6 Success and open data projects	45
	6.7 Economic value	45
	6.8 Summary	46
7.	Processing the Success Factors	47
	7.1 Prioritization Success Factors	47
	7.2 Mapping the Success Factors to the Open Data Process	50
	7.3 Prioritization Fail Factors	55
	7.4 Summary	56
8.	Open Data Projects vs. Other Types of Projects	57
	8.1 Comparison	57
	8.2 Summary	61
9.	Evaluation Success Factors	62
	9.1 Creating a Workable List	62
	9.2 Selecting Projects from the Database	62
	9.3 Selected Projects for Evaluation	67
	9.4 Conclusion of Evaluation	67
	9.5 Summary	69
10	. Conclusion & Discussion	71
	10.1 Conclusion	71

10.2 Discussion	74
References	77
Appendices	83
Appendix A: Interview Questions	83
Appendix B: Interview Summaries	84
Appendix C: Mapping on Success Factors	108
Appendix D: Evaluations	111

LIST OF FIGURES

Figure 1 Research Approach	4
Figure 2 Systematic Literature Process	5
Figure 3 Open Data Process (adopted from (Zuiderwijk et al., 2012))	11
Figure 4 Open Data Value Chain (adopted from (Vickery, 2011))	21
Figure 5 Open Data Process – overview (comments by the author)	23
Figure 6 Success factors mapped to the Open Data Process	56
Figure 7 General success/failure factors (adopted from (Belassi & Tukel, 1996))	58
Figure 8 Result random sampling generator	64
Figure 9 Success factors mapped to the Open Data Process	74

LIST OF TABLES

Table 1 Benefits per stakeholder	
Table 2 Success factors from literature	19
Table 3 Motivations for open data	26
Table 4 Elements of open data definition	27
Table 5 Success Factors mentioned by interviewees	
Table 6 Fail Factors mentioned by interviewees	37
Table 7 Prioritized list of success factors	50
Table 8 Success factors in the Open Data Process	51
Table 9 Comparison of success factors with Atsu et al. (2009)	59
Table 10 Projects with multiple occurrences	63
Table 11 Overview fields GOOD DATA database	66
Table 12 Overview of success factors found per project	68
Table 13 Evaluated success factors	72
Table 14 Mapping to the prioritized list	109
Table 15 Paper and number matching	109
Table 16 Mapping to the phases	110
Table 17 InstaPLAY as in GOOD DATA	111
Table 18 Vienna City Bike as in GOOD DATA	115
Table 19 Hochwasser Krems as in GOOD DATA	118
Table 19 Hochwasser Krems as in GOOD DATA Table 20 An analytical view on data as in GOOD DATA	118
Table 19 Hochwasser Krems as in GOOD DATA Table 20 An analytical view on data as in GOOD DATA Table 21 Taxi Wien as in GOOD DATA	118 121 124
Table 19 Hochwasser Krems as in GOOD DATA Table 20 An analytical view on data as in GOOD DATA Table 21 Taxi Wien as in GOOD DATA Table 22 Wanderwege Wien as in GOOD DATA	118 121 124 127
Table 19 Hochwasser Krems as in GOOD DATA Table 20 An analytical view on data as in GOOD DATA Table 21 Taxi Wien as in GOOD DATA Table 22 Wanderwege Wien as in GOOD DATA Table 23 Jeco Guides as in GOOD DATA	
Table 19 Hochwasser Krems as in GOOD DATA Table 20 An analytical view on data as in GOOD DATA Table 21 Taxi Wien as in GOOD DATA Table 22 Wanderwege Wien as in GOOD DATA Table 23 Jeco Guides as in GOOD DATA Table 24 Sozialmärkte Wien as in GOOD DATA	118 121 124 127 130 133
Table 19 Hochwasser Krems as in GOOD DATA Table 20 An analytical view on data as in GOOD DATA Table 21 Taxi Wien as in GOOD DATA Table 22 Wanderwege Wien as in GOOD DATA Table 23 Jeco Guides as in GOOD DATA Table 24 Sozialmärkte Wien as in GOOD DATA Table 25 Dog Toilet Map Salzburg as in GOOD DATA	

1. INTRODUCTION

In this chapter the problem statement, research question, scope, scientific and societal relevance, and the structure of the remainder of the thesis are discussed.

1.1 PROBLEM STATEMENT

In 2013 President Obama signed an executive order making open data the default for all governmental information (Obama, 2013). For many countries this was the driver to start thinking about open data themselves. In 2015, all member states of the European Union are obligated to actively open their data (European Commision, 2011; European Parliament & European Council, 2013). However, this is still limited to open data policies.

Not much is known about the potential of open data for organizations. Most research regarding open data is performed with regard to the public sector, as most initiatives come from governmental bodies (Hall, Shadbolt, Tiropanis, O'Hara, & Davies, 2012). Moreover, there is no clear overview containing open data projects, as most databases are focused on the data sources.

An initiative to collect information about existing open data projects is executed by the Utrecht Data School. They are in the process of creating a database containing information about open data projects, such as data sources, funding, and geographic location (i.e. the GOOD DATA database). Only projects that are based in Europe are selected for the database. The focus of the project is to create an interactive database which will be machine-readable. The goal is to enable comparative analyses.

"Weather apps, it is remarkable that the government has nothing to do with those apps, they only provide the data. However, it is unfortunate that we still rely on these examples."

Interviewee 15

Still, an overview and knowledge about success factors for open data projects is missing (Möller, Hausenblas, Cyganiak, & Handschuh, 2010). For instance, Zuiderwijk and Janssen (2014) looked into the relation between open data and governmental policies. They found that the policy at various levels of government need to be improved to increase collaboration with other organizations (Zuiderwijk & Janssen, 2014). However, this is only one aspect that might contribute to the success rate of an open data project.

To get organizations more interested in open data projects, governmental organizations are promoting

Public Sector Information (PSI) reuse. PSI is linked to open data as it is also freely available, however PSI is about governmental documents (The European Thematic Network on Legal Aspects of Public Sector Information, 2010), whilst open data goes beyond those documents. Therefore, it does not convince organizations to look into open data and possible projects that led from that open data.

So far, it is unclear what success factors are for organizations, but also what organizations can gain from open data projects (Bonina, 2013; Vickery, 2011). Although some research is performed into what the potential economical benefits of open data could be (Chui, Farrell, & Kuiken, 2013; Janssen, Charalabidis, & Zuiderwijk, 2012; Jetzek, Avital, & Bjørn-Andersen, 2013), this is also focused on governmental organizations and it is important to look further than just governmental data (Vickery, 2011). The impact of public information is broader than just those institutions (Jetzek et al., 2013; Vickery, 2011). However, governments themselves are, still, mostly focused on the transparency acquired by opening data and not yet on the economics of the open

data (Neuroni, Riedl, & Brugger, 2013). Since data and information are (potential) sources for competitive advantage, it is interesting to know their potential value (Vickery, 2011). This could also function as a driver for organizations to look into (the possibilities of) open data.

1.2 RESEARCH QUESTIONS

This thesis project will focus on success factors of open data projects for organizations. Therefore, the main research question is:

'What are success factors of open data projects?'

To answer the main research question, the following sub questions are drafted:

1. What is open data?

To answer this research question a definition of open data will be established. This will form the background and foundation on which the rest of the thesis is build.

- 2. Why should open data be of interest for organizations? The answer to this question will be obtained by looking at open data from different perspectives. For instance, what is the added value of open data and what can be learned from previous experiences with regard to open data. The purpose of this question is to find the reasons for organizations to get involved with open data.
- 3. What defines a successful project and which success factors can be distinguished? As this sub question resembles the core of this research, multiple approaches will be combined to answer this question. It will result in a list of success factors that will be validated by looking at existing projects to see whether the factors can be found in practice as well.
- How does an open data project differ from other projects?
 The answer to this question will indicate what characteristics are specific for open data projects.
 Moreover, it will indicate whether the success factors found are specific for open data projects.

1.3 SCOPING

Although most open data originates from public organizations, the focus within this thesis will be on all open data, regardless of the original source of the dataset. However, as it is impossible to incorporate all open data projects, the main focus will be on European projects. This can be explained by the fact that the database of the GOOD DATA project is aimed at European projects. The database will be used to select projects for the evaluation of the success factors found in previous stages.

1.4 RELEVANCE

The discussion of the relevance of this research is divided into the scientific and societal contribution.

1.4.1 SCIENTIFIC RELEVANCE

To the knowledge of the author, no research has been conducted towards success factors of open data projects. However, some research that is related to success factors of open data has been performed and published. These researches are focused on the public sector and on the relation between open data and policies (e.g., Conradie & Choenni, 2014; Hogge, 2010; Huijboom & Broek, 2011; Janssen et al., 2012; Janssen & Zuiderwijk, 2012; Zuiderwijk & Janssen, 2014). Some research is found on success and fail factors of projects in general. This will be discussed in chapter 8 on page 57.

1.4.2 SOCIETAL RELEVANCE

Increasingly more data is generated (Bauer & Kaltenböck, 2011). By making this data publicly available, it can be used to generate new information and knowledge. However, for most organizations it is unclear what they can gain from open data (Hogge, 2010; Huijboom & Broek, 2011). This research will provide an overview of success factors of open data projects to provide more focus for organizations when they are considering open data.

Current initiatives, such as PSI, are completely focused on governmental data and documents. Most research, as discussed above, is also focused on the public sector. However, open data can be beneficial for (commercial) organizations as well (Socrata, 2014). Unfortunately, it is still unclear how organizations can be successful in open data projects and what the potential benefits are for them.

An exemplary case of how open data can add value, can be found in the United States, where organizations did not work together for hurricane Katrina, leading to chaotic situations and many deaths, because of the miscommunication and lack of collaboration between those organizations (Peled, 2011). If their data would have been open, the coordination of the help could have benefited.

1.5 THESIS STRUCTURE

This thesis is structured as follows: firstly, the research approach is discussed, in which the activities for the creation of this thesis are described. Validity threats are discussed in this chapter as well. In chapter 3, the working definition of open data as used within this thesis is created. Subsequently, chapter 4 discusses the theoretical background on which this thesis is based. Within this theoretical background, multiple aspects of open data will be touched upon. Then, the findings of the interviews are covered in chapter 5, followed by a comparison of the findings from literature and the interviews in chapter 6. As the interviews and literature both provide input for the final list of success factors, they are processed in chapter 7. In chapter 8 open data projects are compared to other type of projects to find out what differences exist between open data projects and other projects. The success factors will be evaluated in chapter 9. The thesis will be concluded in chapter 10, with a conclusion and discussion in which future research is covered as well.

Throughout this thesis examples of open data projects can be found, which are placed in separate boxes.

2. RESEARCH APPROACH

In this chapter the research approach within this thesis project is described. Three main activities are performed for this thesis: a literature study, interviews, and an evaluation of the findings in the previous activities. In conclusion, the validity threats are also discussed.



Figure 1 Research Approach

Figure 1 depicts the research approach of this thesis. In the first phase a literature study has been performed. To validate and complement the findings of the literature study, interviews were conducted. As the interviews led to new information (i.e. new success factors) an evaluation is performed to find the success factors in existing projects. In the remainder of this chapter, each phase will be described.

2.1 LITERATURE STUDY

The literature study will form the foundation for this thesis project. By conducting a (systematic) literature review, existing research is examined. This is done to acquire an indication of how success of a project can be determined.

The literature study, performed during this research, is undertaken in two phases. In the first phase, the guidelines for a structured literature were followed (Moher, Liberati, Tetzlaff, & Altman, 2009). The first step in a structured literature review (SLR) is to determine the keywords. The following keywords were used: "open data", "open data projects", and "open data success". The combination of these keywords cover the main research question and give an overview of available literature concerning the topic of this research. Subsequently, these keywords were entered into one electronic source, namely Google Scholar. Google Scholar is solely used because it has indexed a lot of journals and a quick search at other sources (such as Scopus) led to similar results. Furthermore, a first elimination of search results was made based on two criteria: both patents and citations were excluded, as the goal was to find full papers.

Figure 2 depicts the process of the SLR. The search for papers was executed on 24th of November 2014, resulting in 148.223 search results. Due to the amount of search results, it was decided to merely include the first hundred search hits for a detailed evaluation. These papers were assessed by the researcher for eligibility by reading the title and abstract. Papers which were not available in full text were eliminated from the results. In total, 32 papers were included in this first phase of the literature study.

The second phase consisted out of snowballing. This means that from the literature found in the first phase, the relevant references are noted. Subsequently, these references were retrieved. For each of the found papers, the relevancy was determined by reading the full article. From these papers the relevant references were also noted, resulting in a new list of papers. Due to time constraints, the literature process concluded with that round.





2.2 INTERVIEWS

Moreover, interviews are conducted with so called experts of open data projects. The experts are chosen using the purposive sampling technique (Flick, 2009; Tongco, 2007). This technique was used as the desired information is held by members of the open data community (Tongco, 2007). Although a bias arises by using purposive sampling, this bias contributes to the efficiency and proved to be as robust as random probability sampling (Tongco, 2007). The first interviewees were selected based on their expertise with open data, as they all presented at a conference on open data. Subsequently, they were asked for new interviewees during the interviews.

These interviews provided more insight into success factors of open data projects. The interviewees were asked what they expect to be success factors for open data projects. By doing so, an indication of success factors is obtained, which will be merged with the success factors from literature and tested with the GOOD DATA database. Based on information found in literature, the questions for the interviews are drafted.

The interviews during this research are performed as semi-structured interviews, as this allows for dialogues and more in-depth questions while the established questions make sure the interview keeps its focus (Flick, 2009). The interviews are all conducted by the researcher. They took place either face-to-face or via Skype. To make sure the interviewees were comfortable, the interviewees were asked to pick the location (and media) for the interview. Due to the differences in talkativeness of the interviewees, the interviews varied between 45 minutes and 90 minutes.

After each interview an interview report was sent to the interviewees to make sure the researcher correctly interpreted their sayings. By doing so, an extra form of validation was built into the interviews (Flick, 2009). These interview reports were made by using the recordings of the interviews and additional notes made by the

researcher. The interviews were conducted in Dutch, the interview reports are sent in English. In this way, the interviewee could check the translation of his statements as well as the content of the statements.

As the interviews were semi-structured, most questions were posed as open questions and not all questions were thought of beforehand (Flick, 2009). This enables the interviewee to answer the question based on his or her knowledge at hand, although this leads to what is called subjective theory (Flick, 2009). Moreover, this approach allows the interviewees to inspire the interviewer to ask more or different questions at the next interview. By adapting the questions after each interview, the approach became less generic (Flick, 2009).

Each interview was adapted to the interviewee, to make sure that each interviewee was able to transfer his or her knowledge about his or her specific domain. An example of the questions can be found in Appendix A: Interview questions on page 83. By choosing experts from different disciplines, a more complete overview is obtained.

Although it is recommended to pose theory-driven questions (Flick, 2009), it is chosen not to do so, as the purpose of the interviews is to complement the literature (Flick, 2009) and to find out what associations the interviewees have with open data. The main goal of the interviews was to develop a theory about success factors of open data projects.

To process the input provided by the interviewees, their input is analyzed by looking at the elements that they mentioned. Their answers were first translated to English, after which the interviewees checked whether the translation was correct. For each subject that is discussed in the interviews the elements mentioned by the interviewees are compared. For instance, their definitions are divided into elements, these elements are mapped to the definition as created in this thesis.

2.3 EVALUATION

To validate the list of success factors of open data projects, the database of the GOOD DATA project will be used. The data from the database is used to perform an analysis to look for the success factors as found in the literature and interviews. The goal of this evaluation is to find out whether the success factors found in theory can be found in practice.

The evaluation has been done with projects from the GOOD DATA database. After a project was selected, the project was looked up by the researcher. The researcher looked for information about the project on the website of the project, or in the application itself. This included looking into the terms and conditions, searching for news items about the project, and taking a look at the project itself. For each success factor, it is indicated whether it is found and whether it was of influence on the project. Unfortunately, it was not possible to contact the project initiators directly due to time constraints.

The evaluation ended when there was no new information found, so when all success factors were found in the projects, the evaluation came to an end. This might lead to a survival bias, because only the factors that are found in literature and the interviews are used in the evaluation. However, as the subject of the evaluation is the factors themselves and not the success rate of the projects, the risk on survival bias is limited.

2.4 VALIDITY THREATS

In this section the threats to the validity of this thesis are discussed. Various forms of validity threats can arise, a division of four types is made (Wohlin et al., 2012):

• **Construct validity** is concerned with whether the research is measuring what the researcher wants to measure. A threat to construct validity arises in the interviews. It is important to make sure that the researcher and the interviewee agree about, in this case, the definition of open data. To make sure both the

interviewee and the researcher were talking about open data as defined in this thesis, the interviewees were asked for their definition and for their reaction on the definition, as defined in this thesis. The success factors are derived from the interviews, however, to find out what the interviewee thought of as success and fail factors, they are explicitly asked to provide those factors at the end of the interview. Furthermore, next to the success factors, the fail factors are also taken into account during the interviews.

- Internal validity is concerned with causal relations. In this thesis this implicates that there might be other factors influencing the success rate of open data projects or (one of) the success factors. Within this thesis there are no assumptions made based on the presence of success factors in a project. So, there is no causal relation proposed between the success rate of a project and the success factors found in this research.
- External validity is concerned with the generalizability of the research and whether it is of interest of people outside those who are involved in the research. The success factors can be applied to all open data projects, not just to the ones used in the evaluation. Moreover, as stated in section 1.4.2 Societal relevance as well, the success factors, but also the other findings of this research, can be of interest for organizations who are considering to move towards open data. To increase the generalizability the interviewees are selected from diverse working fields, this entails that not only governmental organizations are approached, but also commercial ones.
- **Reliability** is concerned with independency. It should be possible for another researcher to find the same results by following the research approach of this research. All activities of the research are described, enabling other researchers to perform the research as well. However, as semi-structured interviews were held, the findings will be different. Moreover,

since the second part of the literature study consists out of using the snowballing technique, this is also dependent on the interpretations of the researcher. It might be that the author thought a certain paper was important and looked into that paper, whilst other researchers might neglect that paper as they do not see its importance, or vice versa. In general, the research can be performed by another researcher, however, the used techniques leave some room for interpretation, which probably result in a difference in findings. To increase the level of reliability, the interviews are transcribed and all input for the literature study is stored.

"In logistics they use on board systems and planning software to calculate the most efficient route for trucks. Nowadays, the planning software is based on road data (the amount of traffic present on a specific road) and weather forecasts (as the probability of rain influences the duration of a trip). By using both those open data sources, the software is able to adapt the routes dynamically. This is of strategic importance for logistical service providers as they will be able to work more efficiently."

Interviewee 1

3. DEFINITION OPEN DATA

In this chapter open data is defined.

3.1 CONSTRUCTING A DEFINITION

Before a definition of open data can be constructed, it should be clear what is meant with data in general. A widely accepted definition is provided by Davenport and Prusak, who state that "[d]ata is a set of discrete, objective facts about events" (1998, p. 2). Every organization deals with data, some organizations are even dependent on data for their daily business (Davenport & Prusak, 1998).

Currently, increasingly more data is generated, resulting in a bulk of data (Chui et al., 2013; STT Netherlands, 2002). Many researchers are therefore supporting the idea to make more data available to the public. However, there is no clear overview of what data is available and in which way it is made available. This has led to the creation of multiple open data initiatives, to create an overview of what data is available. However, to be able to do this, a clear definition of open data is needed. As there is no standard definition (Davies, 2014; Vickery, 2011), the definition used in this work originates from several sources (Chernoff, 2010; Janssen et al., 2012; Open Knowledge Foundation, 2014; Rijksoverheid, 2014). From these definitions of open data, core aspects of the definition are adopted and included in this definition. In this work, open data will be defined as follows:

Open data is data that is publicly available, easy accessible and has no restrictions for the users.

This definition is based on the idea that open data is not a type of data. Open data is about the way that the data is made available. It is released in such a way that the public can access it without setting any kind of charges. The data is easy accessible, as it is easy to find and is offered in a format that can be processed with an open-source software tool. The creation of open data is done for a public purpose, in such a way that it is machine-readable, released within open standards and available under an open license. The open license ensures that there are no restrictions posed on the data.

Since open data should always be accompanied with an open license, a definition of open license is also necessary. The definition used in this work is derived from Open Knowledge Foundation (2014) and Janssen et al. (2012).

The open license that is related to open data is about the distribution and use of the data, more specifically it entails "the legal conditions under which the work is made available" (Open Knowledge Foundation, 2014). The license enables free use and redistribution of the data for everyone and should therefore not restrict the usage of the data to specific purposes. Moreover, it should also be possible to reuse and modify parts of the data. When (a part of) the data is reused, the new work should also be created under the open license. However, the license might require the new or modified data to have a different name to be able to cope with different versions.

3.2 CONCLUDING REMARKS

So far there has been no consensus about the definition of open data. Therefore, a definition is created in this thesis, which is stated above. Even though the definition is based on the use of open standards, in practice there are no standards used yet, as they are for now focusing on disclosing the data. An open data project is a project that is based on open data. This entails that the project cannot exist without the open data.

4. THEORETICAL BACKGROUND

In this chapter the findings from the literature study are discussed. This chapter is divided in seven subsections, followed by a conclusion. The division of the sections is based on the content found in literature.

4.1 WHY NOW?

Already in the nineties, some research into openness was performed (Streeter, Kraut, & Caby, 1996), resulting in the finding that the error rates within organizations decrease when organizations are part of an open network, such as an open ecosystem. Shortly afterwards, the connection between open data and crowd sourcing was made, as open data is also updated by the crowd (Hall et al., 2012).

At the moment, open data is gathering momentum, as the rise of open data started in the late 2000s (Davies, 2014). Since then, the amount of data that is made open and the ability to cope with such amounts of data is improved (Chui et al., 2013; Shadbolt et al., 2012). Moreover, the increased use of the Internet and the rise of technologies for government transparency create new opportunities for, and unlock the value of, open data (Huijboom & Broek, 2011; Jetzek et al., 2013; Ubaldi, 2013). Therefore, it can be stated that open data is IT-enabled, as IT allows the publication of data by governments and citizens use IT to be able to use the data (Davies & Edwards, 2012; Janssen & Zuiderwijk, 2012).

Currently, the aim is shifting from opening as much data as possible, to realizing its benefits. The realization that adding data to an existing database increases the informativity of the dataset (Attila, Michel, & Blegind, 2013) is settling in. Since 2009, the focus is shifted from open data as a new phenomenon towards the potential economic value of open data, as national policies in both the United States as the United Kingdom are developed (Bates, 2014; Davies, 2014; Weiss, 2002). However, the effect of the global economic crisis should be taken into account. Due to that crisis, determining the economic value of open data was impeded (Davies, 2014). Currently, open data has reached the point where stakeholders are curious towards the impacts (Kaschesky & Selmi, 2014).

Increasingly more organizations are collecting and analyzing open data, furthermore, open data is, after syndicated data, the most important type of data that executives perceive as valuable for deciding on strategy changes (Bonina, 2013). Governments and public organizations are the largest contributors of open data (Bonina, 2013; Ding, Peristeras, & Hausenblas, 2012; Janssen & Zuiderwijk, 2012). In addition, the public sector is of crucial importance in the overall sustainability of open data (Kassen, 2013). However, in practice it appears to be very difficult to decide how to actually open data (Gurstein, 2011; Janssen & Zuiderwijk, 2012). Moreover, the focus of governmental organizations should be changed to how to capture the value of open data (Evans & Campos, 2013; Neuroni et al., 2013).

For public organizations a lot of research has been done regarding Public Sector Information (PSI). The goal of PSI is to facilitate the reuse of public sector information. By doing so, economic activity, innovation, and competition are stimulated (Lee, Cyganiak, & Decker, 2014).

Often, PSI is mentioned as open data within public organizations. However, as PSI and open data cannot be used interchangeably, a clear distinction is needed (Bonina, 2013). PSI can be seen as a subset of open data. Parts of PSI can be open, but not all PSI is open by definition. PSI refers to all documents held by public organizations. The PSI Directive of the European Union encourages the reuse of the information, however, it also puts some restrictions on its access and use (Lapsi Network, 2010).

In addition, open data is often associated with linked data (Bizer, Heath, & Berners-Lee, 2009), as it facilitates the open data movement. However, linked data is not limited to open data. Linked data is a method to create a

collection of datasets, which are related within the context of the semantic web. With linked data these datasets are connected to each other (Linked Data Community, 2014). To be able to create linked data, a standard format needs to be used, RDF is often suggested (W3C, 2013). By using RDF users are able to navigate through data from multiple data sources (Auer, Bizer, Kobilarov, & Lehmann, 2007). Moreover, linked data also includes exposing and sharing data, information and knowledge (Linked Data Community, 2014).

Various initiatives (Gurstein, 2011) try to stimulate organizations to share their data and make it more open. One such initiative is the Linking Open Data Community project of W3C. The aim of this project is to connect data of various open data sets from different data sources (W3C, 2014). The result can be imagined as a cloud of interconnected data sets. By doing so, their objective is to create a web in which one can go from a data item in one source to another data item from another data source. However, it is important to accompany this data with metadata, describing the dataset (Bizer et al., 2009), which is not done at the moment (Oliver, Palacin, Domingo, & Valls, 2012).

4.2 WHAT DO WE KNOW SO FAR?

The opening of data is accompanied by some learning points. For instance, the first open data initiatives showed that releasing all data at once has a negative effect, as users are confused and overwhelmed by the amount of data that is suddenly available (Conradie & Choenni, 2012; Davies, 2011; Ding et al., 2012; Harrell, 2013; Janssen et al., 2012). The same applies to the difficulty of the data, if the data is only readable for trained people or aimed at one specific program, the amount of users and therefore the reusability will be low (Bauer & Kaltenböck, 2011; Conradie & Choenni, 2012; Davies & Edwards, 2012; Davies, 2011; Gurstein, 2011; Hall et al., 2012; Harrell, 2013; Vosough, 2013). Therefore, the, sometimes subtle, difference between access to data and the ability to use data should be kept in mind (Davies & Edwards, 2012; Davies, 2011; Gurstein, 2011).

Moreover, it has been taught that there should be a balance in place between at one end proprietary and openness and at the other end control and empowerment in order to reach the full potential of open data (Roy, 2014). From Roy (2014) and Zuiderwijk and Janssen (2014), the lesson learned is that in order to let local initiatives blossom, each municipality should have its own policy. However, for the overall adoption of open data it is advised to have one overarching strategy (Zuiderwijk & Janssen, 2014). Combining these two insights, an overall strategy in which (public) organizations can adjust details seems to be the best solution, as it allows organizations to learn from each other (Zuiderwijk & Janssen, 2014). This implies using an overarching policy from which diverse open data initiatives are coordinated (Roy, 2014; Zuiderwijk & Janssen, 2014). An example can be found within the European Union, as the European Commission has set an open data policy. Member states are then able to modify the policy for local use.

In practice, open data policies are often not driven by user needs (Janssen et al., 2012; Ubaldi, 2013; Zuiderwijk, Janssen, Choenni, Meijer, & Alibaks, 2012; Zuiderwijk & Janssen, 2014), creating a gap between the offered data and the data that is asked for by the users. This can even result in citizens rejecting the idea of open data as they see the new transparency as a marketing action (Jetzek et al., 2013). Although in the UK it led to an increase of trust in the government (Bates, 2014). In general public agencies think they are in fact meeting the citizens demands and wishes and contributing to more active citizen participation (Evans & Campos, 2013).

Next to the driver user needs, public organizations should be highly motivated to use open data, as they are setting the example (Jetzek et al., 2013). In the Netherlands, it was found that governmental organizations were not publishing open data to the national portal, leading to only a few new datasets (Leeuw & Lemmer, 2011). Furthermore, no sound evidence of the impact of open data policy is found by Huijboom and Broek (2011). This leads to a difference in adoption speed, some organizations are more motivated to adopt an open data policy than other organizations, that experience open data policies as an obligation, making it harder to create a culture in which sharing data is common use (Zuiderwijk & Janssen, 2014).

In order to facilitate the opening and reusing of data, the Open Data Process is created (Zuiderwijk et al., 2012). The intention of the open data process is to facilitate the process of opening data by providing five steps. By presenting the open data process, Zuiderwijk et al. (2012) introduced more standardization in creating open data. As indicated by Davies (2011), the need for more standardization for open data exists.

Figure 3 depicts the Open Data Process. The first phase is the creation of data, this includes collecting of data, storing of data as well as paying for data. Phase two is opening the data, which can be done via a national platform or on the website of the organization itself. In phase three, users find the open data, followed by phase four where they actually use the open data for their own projects. In phase five feedback is provided on the open data, so the quality of the open data can be improved. By including a feedback loop in the process, two-way communication is ensured (Janssen et al., 2012).



Figure 3 Open Data Process (adopted from (Zuiderwijk et al., 2012))

Open data has a very broad impact, involving divers stakeholders, with the most prominent being governments, which are taking the role of publishers of the data and can therefore be found at the top of Figure 3. The organizations responsible for the data that is being opened, are already generating this data as part of their function to fulfill their public task (Vickery, 2011). Governments are still exploring how open data can be used in the development process of a country and other potential benefits of open data (Davies & Edwards, 2012).

In addition to publishers, customers and developers are also involved (Kaschesky & Selmi, 2014; Vosough, 2013) and can be seen as the reusers in Figure 3. The roles can be fulfilled by various parties, e.g. organizations can take on the role of data publishers, but can also act as customers and use the data produced by others. The idea that sharing data is an added value is growing, resulting in both large and small organizations that want to 'do something' with open data (Murray-Rust, 2013). Moreover, public sector organizations can be seen as an intermediary between the source and end-users of open data (Ubaldi, 2013; Vickery, 2011). This is not directly reflected in the figure, but between opening the data and finding the data, these organizations could facilitate the findability of the open data. As open data is a relatively new field, a more complete stakeholder analysis is needed to understand the needs and capabilities of the divers group of users of open data (Neuroni et al., 2013).

4.3 PROMISES OF OPEN DATA

Currently, a gap exists between the promised benefits of open data and the realized benefits (Conradie & Choenni, 2014; Davies, 2011, 2014). The majority of the promised benefits of open data are described below. After which the way to reach those benefits is touched upon.

4.3.1 PROMISED BENEFITS

Firstly, open data will facilitate the reuse of data (Ding et al., 2012; Hall et al., 2012; Zuiderwijk & Janssen, 2014), leading to a more transparent society (Attila et al., 2013; Bauer & Kaltenböck, 2011; Bonina, 2013; Cucciniello, Nasi, & Valotti, 2011; Hall et al., 2012; Heimstädt, 2014; Janssen & Zuiderwijk, 2012; Kassen, 2013; Lee et al., 2014; Open Data Research Network, 2013; Ubaldi, 2013; Vosough, 2013; Wright, Prakash, Abraham, & Shah, 2010). Data released by public organizations is assumed to be reliable and sustainable (Nilsen, 2010), and will therefore be very useful. Since more data is available, more is possible, leading to innovative products and services (Attila et al., 2013; Bates, 2014; Conradie & Choenni, 2012; Dekkers, Polman, te Velde, & de Vries, 2006; Hall et al., 2012; Heimstädt, 2014; Iemma, 2012; Janssen et al., 2012; Lee et al., 2014; Manyika et al., 2013; Nilsen, 2010; Pollock, 2006; Ubaldi, 2013; Verhulst, Noveck, Caplan, Brown, & Paz, 2014; Vickery, 2011; Vosough, 2013; Vries et al., 2011; Zuiderwijk & Janssen, 2014). This includes unexpected innovation, because more (diverse) people are able to do something with the data, unexpected benefits might be achieved (Attila et al., 2012; Oliver et al., 2012; Vickery, 2011). In addition, it will lead to standardization of metrics (Hall et al., 2012). Moreover, it will contribute to the marketing of public service provision (Bates, 2014; Neuroni et al., 2013).

By opening data to citizens, the interaction between stakeholders will intensify (Bauer & Kaltenböck, 2011; Chui et al., 2013; Hall et al., 2012; Neuroni et al., 2013; Nilsen, 2010; Ubaldi, 2013; Vosough, 2013), making it easier for governments to fulfill the wishes and needs from their citizens, leading to more engaged citizens (Ding et al., 2012; Hall et al., 2012; Kassen, 2013; Lee et al., 2014; Neuroni et al., 2013; Ubaldi, 2013; Zuiderwijk & Janssen, 2014). Moreover, a side-effect is that the trust between government and their citizens will be restored as accountability will increase (Hall et al., 2012; Heimstädt, 2014; Janssen & Zuiderwijk, 2012; Lee et al., 2014; Manyika et al., 2013; Neuroni et al., 2013; Open Data Research Network, 2013; Ubaldi, 2013; Vosough, 2013; Wright et al., 2010). Furthermore, the communication between stakeholders, as well as the communication between different levels within an organization will be improved (Hall et al., 2012; Iemma, 2012; Kassen, 2013; Wright et al., 2010).

Open data promises to stimulate economic growth, lower costs of data processes, and lead to new economic activities (Bates, 2014; Bauer & Kaltenböck, 2011; Bonina, 2013; Ding et al., 2012; Hogge, 2010; Iemma, 2012; Janssen & Zuiderwijk, 2012; Lee et al., 2014; Manyika et al., 2013; Open Data Research Network, 2013; Roy, 2014; Ubaldi, 2013; Verhulst et al., 2014; Zuiderwijk & Janssen, 2014). The economic benefits might differ per stakeholder, for instance, customers (or citizens) might benefit from a better service; researchers have access to more data that they can use in their research; organizations could see new opportunities based on the newly available data (Bonina, 2013; Fioretti, 2010; Hall et al., 2012; Janssen & Zuiderwijk, 2012; Lee et al., 2014; Vickery, 2011; Vosough, 2013). Moreover, organizations will be able to reach more (potential) customers, as third parties are also allowed to use their open data (Hall et al., 2012; Vries et al., 2011). In addition, this provides the opportunity to better align businesses (Hall et al., 2012).

Another promised benefit of opening data by organizations is that customers (and citizens) have easy access to the data, as open data is a non-rivalrous and non-excludable good (Nilsen, 2010), resulting in new applications and services, which in its turn can lead to job creation (Vosough, 2013). This also leads to more citizen initiatives, improved citizen participation and empowerment, and improved democratic processes (Evans & Campos, 2013; Heimstädt, 2014; Huijboom & Broek, 2011; Neuroni et al., 2013; Vickery, 2011). This is one type

of value that can be created by open data, as the core idea of open data is to create new value through the reuse of existing data (Manyika et al., 2013; Neuroni et al., 2013; Ubaldi, 2013; Vickery, 2011). Different types of value might require different types of data (Ubaldi, 2013).

Stakeholder	Example	Promised benefits
Publishers	Governments, organizations	Facilitate reuse of data More transparent society Improved communication Economic growth Unexpected innovation Standardization of metrics Lower data processing costs More (potential) customers Improved participation Improved democratic processes
Intermediary	Public sector organizations	Facilitate reuse of data More transparent society Improved communication Economic growth Improved participation Improved democratic processes Marketing of public service provision Fulfilled wishes of citizens Increased trust in government
Reusers	Organizations, developers	Facilitate reuse of data More transparent society Improved communication Economic growth Improved democratic processes Reliable and sustainable data Easy access to data

Table 1 Benefits per stakeholder

These (promised) benefits lead to the conclusion that each stakeholder has something to gain from open data. However, the benefits differ per stakeholder (Manyika et al., 2013). Therefore, it is prudent to identify the stakeholders and what their benefits are for each individual project (Ubaldi, 2013). A general overview of the (promised) benefits mentioned in this section and the accompanying stakeholder can be found in Table 1.

4.3.2 HOW TO REACH THESE PROMISED BENEFITS?

To be able to reach those potentials governments need to know their citizens wishes, behaviors and mental models (Chui et al., 2013). Furthermore, (public) organizations will require adaptations in core processes (technological, organizational, and institutional) to be able to reach those potentials (Conradie & Choenni, 2012; Hall et al., 2012; Janssen & Zuiderwijk, 2012; Ubaldi, 2013), as they will go from publishing information to publishing raw facts (Lee et al., 2014). Especially when public participation is one of the goals, public organizations then need to shift their current, somewhat closed, mindset to a mindset of openness (Davies, 2011; Jetzek et al., 2013; Oliver et al., 2012; Roy, 2014). Moreover, it is expected from public organizations to embrace open data policies (Janssen & Zuiderwijk, 2012), whereas other organizations might experience less pressure resulting from legislations.

An efficient open data policy can lead to multiple benefits, as it enables the creation of new services based on government data, lowers barriers for new open data adopters, and it enables the exploration of new technologies, whose adoption might lead to better performances (lemma, 2012; Jetzek et al., 2013; Ubaldi,

2013). Furthermore, an efficient open data policy stimulates the democratic participation about significant issues and emerging needs (Iemma, 2012). Moreover, the public administration itself benefits from the open data policy, as they will be able to work more efficiently as the internal processes will be renewed in order to adopt the policy (Iemma, 2012). According to Neuroni et al. (2013), open data policies should focus on different stages of the implementation of open data. At first, the focus should be on increasing data transparency, then it should be on improving open participation, leading to improved citizen participation as not only the data itself is considered (Evans & Campos, 2013; Neuroni et al., 2013).

Data will always be needed to create innovative products, and open data enables innovation to come from multiple sources, as more people have access to the data (Chui et al., 2013; Jetzek et al., 2013). However, open data will only reach its potential when it is presented in the right way. Therefore, the way and time of presenting the data is essential (Harrell, 2013). Furthermore, it is important that the open data is presented in a consistent manner (Böhm et al., 2012). Therefore, Harrell (2013) suggests to learn from the lessons learned in user experience to create accessible and usable data. To be able to reach its full potential, the data should be easy to find (Attila et al., 2013; Hall et al., 2012; Ubaldi, 2013).

4.4 MOTIVATIONS FOR OPEN DATA

Just as there are promises of open data, diverse motivations to adopt open data incentives exist. A lot of the motivations **come from outside the organization** (Huijboom & Broek, 2011). The foremost motivation is to **increase the transparency** of the organization (Chui et al., 2013; Conradie & Choenni, 2012; Huijboom & Broek, 2011; Kaschesky & Selmi, 2014; Roy, 2014). At the same time, the position of the government will be reinforced (Chui et al., 2013; Huijboom & Broek, 2011). The Cabinet Minister of the United Kingdom stated that "We don't just want to lead the world in releasing government data — our aim is to make the UK an international role model in exploiting the potential of Open Data to generate new businesses and stimulate growth." (Capgemini

"I want to create the same effect as when you are at primary school. You see some kids playing football and you think, hey I also want to play, because it seems fun."

Interviewee 10

Consulting, 2013, p. 3). This indicates that governments are willing to create political leadership and stimulate innovation and knowledge transfer by adopting an open data policy (Huijboom & Broek, 2011; Leeuw & Lemmer, 2011; Manyika et al., 2013; Neuroni et al., 2013). Sometimes local governments already adopted open data, resulting in an incentive for a national strategy (Huijboom & Broek, 2011). The other way around, also provides a motivation for governments, the EU legislation evolving open data provides a push for governments to open their data (Huijboom & Broek, 2011).

Another popular motivation is to increase the customers' engagement and create more enthusiasm on a more local level (Chui et al., 2013; Conradie & Choenni, 2012; Huijboom & Broek, 2011; Roy, 2014). By doing so, new products and services will be created (Huijboom & Broek, 2011), which might lead to economic benefits (Huijboom & Broek, 2011; Kaschesky & Selmi, 2014). Moreover, in the United Kingdom, this was one of the motivations because "a more informed citizen is a more empowered citizen" (Huijboom & Broek, 2011, p. 4). By sharing the data with more people, the decision-making processes, collaborations, and knowledge sharing can be improved (Hall et al., 2012; Janssen et al., 2012). This provides the opportunity to involve citizens or customers in tough decisions, such as budget costs (Huijboom & Broek, 2011).

Current policies are limiting the public good as they restrict the access to the data, making data sharing a costly and inefficient process (Roy, 2014) is another motivation. Moreover, noticing that other countries are doing well is a trigger for governments to adopt an open data policy (Huijboom & Broek, 2011). Examples from other

countries can be inspiring, leading to market initiatives which pressure the government to open its data (Huijboom & Broek, 2011).

The rise of new technologies is also a motivation to adopt an open data strategy (Huijboom & Broek, 2011; Ubaldi, 2013). The opportunities arising from these new technologies making it possible for users to go directly to the data source (Davies & Edwards, 2012). By using (new) technologies, organizations are able to collaborate. It is this collaboration that is another motivation for opening data or adhering to an open data policy (Neuroni et al., 2013).

Economic growth is a motivation as well, although there are no estimations available of the size of it (Neuroni et al., 2013). At first, organizations were worried about other organizations making profit from their data, however, they realize that opening data will lead to an increase in taxes, which will make up for the extra costs for opening the data (Neuroni et al., 2013).

4.5 BARRIERS FOR OPEN DATA REUSE

In contrast to the motivations, barriers can be found within the organization, as they are often aimed at the data publishers (Huijboom & Broek, 2011; Vosough, 2013). This might be due to the fact that open data policies are currently inward-looking and not yet outward-looking (Jetzek et al., 2013; Zuiderwijk & Janssen, 2014). Furthermore, barriers differ per culture (Huijboom & Broek, 2011; Vickery, 2011; Wright et al., 2010), per stakeholder (Manyika et al., 2013), and might also differ per type of data (Conradie & Choenni, 2012; Manyika et al., 2013). Zuiderwijk et al. (2012) provide an overview of barriers found in literature. They conclude that most barriers are related to the actual use of open data. Evans and Campos (2013) and Dekkers et al. (2006) add to this notion that the timing and wording of the directive, created by the European Commission, might have acted as a barrier, preventing new ways to rise to reap the (potential) benefits of opening data.

A major barrier for both opening data and the actual reuse of data is the **lack of knowledge** about how the open data market works, about market transparency, and about establishing fair prices (Neuroni et al., 2013; Vickery, 2011). This lack of knowledge often results in unclear reuse conditions, complex procedures, and barriers to the development of international markets (European Commission, 2010; Jetzek et al., 2013; Vickery, 2011). In addition, it is very hard to distinguish fake, or incomplete, information from true information (Borges, De Faria Cordeiro, Campos, & Marino, 2011; Jetzek et al., 2013). Furthermore, the current community centered on open data is too fragmented and divergent (Jetzek et al., 2013).

The culture within an organization can be a major barrier. In some organizations, employees are rewarded for secrecy and not for openness, creating a gap for the transition towards an open data policy (Böhm et al., 2012; Conradie & Choenni, 2012; Huijboom & Broek, 2011; Jetzek et al., 2013; Vickery, 2011). This same effect can be caused by the **privacy legislation** which is applied within the organization (Manyika et al., 2013). Opening the data can lead to tensions between the organization and its customers (Conradie & Choenni, 2012; Huijboom & Broek, 2011). In addition, opening the data can lead to **security threats** for the organization (Huijboom & Broek, 2011). Therefore, the organization should learn how to cope with open data (Neuroni et al., 2013).

Next to the culture within the organization, the **branch** of the organization is also a factor. In some branches the profit results from selling data, opening data would therefore lead to a loss in sales, making those organization reluctant to open their data (Conradie & Choenni, 2012; Huijboom & Broek, 2011; Janssen et al., 2012; Jetzek et al., 2013). The Open Government Program initiated by President Obama proposes a solution by ordering all public organizations to open non classified data. Therefore, none of those organizations are harmed as they all cannot benefit from selling data anymore (Peled, 2011).

Another barrier is the **quality of the data**, some data is of too low quality to be released to the public (Böhm et al., 2012; Conradie & Choenni, 2012; Huijboom & Broek, 2011; Jetzek et al., 2013). A low quality of the data

leads to a drop in the amount of users of the data, leading to a decrease of value (Heimstädt, 2014; Peled, 2011). In addition, **imperfect information** leads to (project) failure (Nilsen, 2010). This results in an inclusion problem, as it is unclear what data should be included and actually are included (Peled, 2011; Vickery, 2011). Although it is worth mentioning that it is hard to determine the quality of data, this is even amplified when the data is imperfect (Conradie & Choenni, 2014; Nilsen, 2010). Furthermore, it is also hard to determine and measure **the impact** of opening the data (Conradie & Choenni, 2012; Davies, 2014; Fioretti, 2010; Huijboom & Broek, 2011; Open Data Research Network, 2013; Verhulst et al., 2014).

Existing databases need to be adapted in order to be suited to function as an open database, this entails making the database more user friendly, to make it also usable for non-experts and ensure equal access to the data (Huijboom & Broek, 2011; Janssen et al., 2012; Vickery, 2011). However, as there is **no standardization** in place, this is a difficult process (Böhm et al., 2012; Borges et al., 2011; European Commission, 2000; Huijboom & Broek, 2011; Jetzek et al., 2013). Especially as public organizations often **lack the required technical knowledge** (European Commission, 2010; Jetzek et al., 2013; Ubaldi, 2013; Vickery, 2011). Furthermore, the capacity of existing networks should be expanded in order to be able to coop with the new amount of data traffic (Huijboom & Broek, 2011). All these **extra measures** to open data entail extra costs and a request for a proper licensing system (Vickery, 2011), creating new barriers. This often results in data that is being opened without the metadata, necessary for the users, in order for them to determine the relevance, reliability, and value of the dataset (Borges et al., 2011; European Commission, 2010; Evans & Campos, 2013).

Opening data entails the problem of accountability, it is unsure who is responsible for the data and for actions that result from the, perhaps erroneous, data (Conradie & Choenni, 2012; Janssen et al., 2012; Kulk, Loenen, & Ploeger, 2012; Manyika et al., 2013; Ubaldi, 2013). Since it is harder to rectify incorrect data once it is published than to resolve an incorrect data upstream (Peled, 2011), it is important to find the one responsible.

Nevertheless, open data architects support the idea that once something is posted it cannot be withdrawn or adjusted later on (Peled, 2011). However, the common belief within the open data community is that "the invisible hand of the market will take care of" (p. 2) the data quality (Ferro & Osella, 2012).

As stated in the introduction, increasingly more data is being released, however this also entails that people do **not** have **an overview** any more of what information or data is available (Bates, 2012; European Commission, 2000; Ubaldi, 2013; Vickery, 2011). This results in decreasing motivations to actually use open data (Vickery, 2011). Moreover, the data that is available **does not interoperate** well, as multiple formats are being used and it is still unclear how to collaborate properly (Borges et al., 2011; European Commission, 2000; Jetzek et al., 2013; Neuroni et al., 2013; Vickery, 2011). This use of **multiple formats** and the lack of a standard format caused major implications in dealing with emergency situations, as depicted by (Borges et al., 2011).

"A laborious project was a project that was initiated to allow elderly people to live longer in their own house instead of moving to elderly homes. The issue has so many aspects and involves a delicate group in difficult times that solutions to make it work were hard to find. I am sure it will need a multi discipline approach to tackle such an issue."

Interviewee 9

All these barriers lead to new risks for organizations (Manyika et al., 2013).

4.6 SUCCESS AND OPEN DATA PROJECTS

Having discussed the motivations of and barriers towards open data, this section is focused on what is known about success and open data projects. Not much has been written about this subject, therefore most of the information as discussed in this section is derived from the papers and is not explicitly stated by the authors of those papers.

First of all, it is useful to have a definition of what is meant by a success factor. Success factors are defined as a limited number of characteristics, conditions, or variables that have a direct and serious impact on the effectiveness, efficiency, and viability of an organization, program, or project. It is important to distinguish between success criteria, which are measures by which the success rate of a project is judged, and success factors, which are inputs to the project that lead (in)directly to success (Cooke-Davies, 2002).

To assess whether the open data project is a success, Zuiderwijk and Janssen (2014) propose to look at the usage of the data, the possible negative impacts and the potential positive outcomes. Harrell (2013) mentions an open data project which, according to her, is successful because it approaches the available data with a different perspective. Leeuw and Lemmer (2011) add to this that the data should be utilized in an innovative way, having both an economic as a societal relevance. In addition, Cucciniello et al. (2011) and Weiss (2002) indicate that having access to more information, which will happen by opening data, is significant for economic and social prosperity.

A lot of open data is collected at a local level, therefore, the support of the local level is a factor for the successful adoption of open data (Conradie & Choenni, 2012). When the focus is on a more local level, the focus is more narrow, making success more likely (Conradie & Choenni, 2014). Moreover, in the analysis of social data, the location is seen as a success factor (Kalampokis, Hausenblas, & Tarabanis, 2011).

Planning and coordination are also recognized as being relevant for the success of open data (European Commission, 2010). The attitude within an organization towards open data influences the open data policy. When organizations are mainly focused on the risks that the policy entails, the policy will be more closed as compared to organizations that develop their policy because they want to become more open (Zuiderwijk & Janssen, 2014). Moreover, it is important that the public sector remains an active player in the market, as this enhances the reuse of open data and trust of citizens (Nilsen, 2010; Ubaldi, 2013).

Chui et al. (2013) propose the following six elements that are needed for a successful open data initiative.

1. Consideration of potential value

When releasing open data the potential value has to be considered. At the moment, one of the few considerations when releasing the data is the ease of releasing it. This might lead to lost opportunities, as the easiest releasable data will not always be the data with the most economic value potential.

2. Presence of an ecosystem

To enable the potentials of open data, contributors have to be able to work together on the dataset, an ecosystem might be such a solution. Ecosystems exist of loosely connected participants that interact with each other (Heimstädt, 2014). By collaborating the data can be enriched (Auer et al., 2007; Capgemini Consulting, 2013; Jetzek et al., 2013; Leeuw & Lemmer, 2011; Neuroni et al., 2013) and the error rates within an organization will decrease (Streeter et al., 1996). Moreover, users are able to contribute information and collective intelligence to the data (Kassen, 2013; Neuroni et al., 2013; Ubaldi, 2013; Vries et al., 2011; Weiss, 2002) and the existence of such an ecosystem encourages reuse as the wishes of the users can be taken into account (European Commission, 2010; Lee et al., 2014; Neuroni et al., 2013). An example of such an ecosystem is the phenomenon of Linked Open Government Data (LOGD), as it brings data publishers, data processors, and data consumers together (Ding et al., 2012). For each field, the ecosystem will be different, because other factors affect the choice for most suited ecosystem (Hall et al., 2012). And even within the overarching open data ecosystem there will originate different categories of ecosystems, as the publisher of the data, the organizations in between, and the users are likely to create their own ecosystems (Ubaldi, 2013).

3. Presence of an infrastructure

An infrastructure to be able to manage the data is necessary. The infrastructure provides metadata about the data and facilitates the findability of the data (Davies, 2011). This will enable monitoring of the data, which could be used to improve the data and the data usage (Janssen & Zuiderwijk, 2012). In

turn, the infrastructure can enable the integration of multiple datasets (European Commission, 2010; Neuroni et al., 2013).

4. Choose appropriate channels

The appropriate channels to release the data have to be chosen per dataset to be able to take full advantage of the data. This will lead to improved access to the data, which has a positive outcome on the number of users and development of new uses (Vickery, 2011).

5. Protect your data properly

Data that needs to be protected, should be properly protected. Therefore, the policy to make open data the standard within an organization might not always be the best policy. In addition, some marketing is necessary to raise the awareness of the availability of the open data (Chui et al., 2013; European Commission, 2010).

6. Presence of a clear leader

For a successful open data project a clear leader is essential. In some cases the benefits of open data might indicate risks that managers are not willing to take. In those cases a leader needs to step in to ensure the open data project is not shut down. Moreover, a leader will also be responsible for responding to concerns and suggestions of the data consumers. This person should promote open data to the remainder of the organization, as this is a critical factor to achieve the positive effects of open data (Zuiderwijk & Janssen, 2014). This is effective as the value increases when more people are using it (i.e. the open data) (Nilsen, 2010).

"Basic registration large-scale topography (BGT), a lot of municipalities are in the middle of that process, releasing datasets. Everything is released in the same way and the same format, and is collected in a national receptacle, from which it is published as a web service. It is useful for citizens, but also for internal use within municipalities as it simplifies our architecture, since all data is available at that receptacle. Open Streetmap is also a nice project, resulting in very detailed locations, however the BGT is more precise."

Interviewee 3

Understanding the potential economic value of open data is necessary in order to generate this value (Chui et al., 2013; European Commission, 2010; Neuroni et al., 2013; Nilsen, 2010; Ubaldi, 2013). The ecosystem and leader make it easier for users to contact the data provider, therefore, more feedback could be collected, resulting in an improvement of the data quality (Janssen & Zuiderwijk, 2012).

There is not one strategy for a successful open data policy, for instance, in Tanzania the push from top level was the key for a successful adoption of open data (Hogge, 2010). However, the same study indicates that other countries in Africa might benefit from a push from international aid organizations. Therefore, Hogge (2010) created a checklist for open data policy in general, making it easier to create an adequate strategy for each situation (Capgemini Consulting, 2013; Pollock, 2009). Since there is not one strategy for a successful open data policy, Huijboom and Broek (2011) wonder whether open data strategies are really strengthen citizen engagement and yielding new innovative businesses.

According to literature, a couple of things are essential

for the success of a project. First of all, there is a need for a motivated group of users that are willing to use the open data to create projects, as well as, access to resources and the skills to use them (European Commission, 2000; Hogge, 2010). The working practices, organizational contexts, incentives and appropriate technologies also matter (Hall et al., 2012). The presence of a high-level mandate is also essential for potential success. For this mandate, it might be useful to adopt Freedom of Information laws, which enable free access to

information (Hogge, 2010; Ubaldi, 2013). Furthermore, it is of importance to define organizational processes clearly (European Commission, 2010; Neuroni et al., 2013).

As defined in project management courses, a project is successful as the stakeholders of the project are satisfied, and the project team achieved what the stakeholders had asked for (Cooke-Davies, 2002). Therefore, it is important to realize that project success can only be measured when the project is completed, during the project only its' performance can be measured (Cooke-Davies, 2002). The ultimate measure of success is whether the project results in lasting, long-term, value creation (Cooke-Davies, 2002). On top of that, the benefits must outweigh the costs (Nilsen, 2010).

In summary, the following list of 25 factors that influence success, as mentioned in literature, can be composed. They can be found in Table 2, which also indicates the sources where these factors can be found.

Success factor	Source(s)	
Usage of the disclosed data.	(Nilsen, 2010; Zuiderwijk & Janssen, 2014)	
Possible negative impacts, when adopting open data.	(Zuiderwijk & Janssen, 2014)	
Possible positive outcomes, when adopting open data.	(Zuiderwijk & Janssen, 2014)	
Different perspective on (existing) datasets.	(Harrell, 2013)	
Innovative project.	(Leeuw & Lemmer, 2011)	
Project is initiated at a local level.	(Conradie & Choenni, 2012, 2014)	
Location of the project.	(Kalampokis et al., 2011)	
Planning and coordination of the project.	(European Commission, 2010)	
Attitude of the organization.	(Zuiderwijk & Janssen, 2014)	
Active presence of the public sector in the market.	(Nilsen, 2010; Ubaldi, 2013)	
Adequate strategy (or policy), as set by governments.	(Capgemini Consulting, 2013; Hogge, 2010; Pollock, 2009; Zuiderwijk & Janssen, 2014)	
Consideration of potential value that can be reached by opening data, both for the organization itself as for reusers.	(Chui et al., 2013; European Commission, 2010; Neuroni et al., 2013; Nilsen, 2010; Ubaldi, 2013)	
Presence of an ecosystem, surrounding the open data.	(Chui et al., 2013; European Commission, 2010; Kassen, 2013; Lee et al., 2014; Streeter et al., 1996; Ubaldi, 2013; Vries et al., 2011; Weiss, 2002)	
Collaboration between parties to exchange knowledge, both on how to open data as on the datasets themselves.	(Auer et al., 2007; Capgemini Consulting, 2013; Jetzek et al., 2013; Leeuw & Lemmer, 2011; Neuroni et al. 2013)	
Presence of an infrastructure to spread the open data.	(Chui et al., 2013; European Commission, 2010; Neuroni et al., 2013)	
Availability of metadata accompanying the open dataset.	(Davies, 2011; Janssen & Zuiderwijk, 2012)	
Appropriate release channels for the distribution of the open data.	(Chui et al., 2013; Vickery, 2011)	
Proper security of the data.	(Chui et al., 2013)	
Raise awareness of the existence of the project.	(Chui et al., 2013; European Commission, 2010)	
ear leadership to overcome barriers. (Chui et al., 2013; Zuiderwijk & Janssen, 2014)		
Motivated group of users to ensure reuse of the data.	(European Commission, 2000; Hogge, 2010)	
High-level mandate.	(Hogge, 2010; Ubaldi, 2013)	
Clear working practices and organizational processes within the project.	(European Commission, 2010; Hall et al., 2012; Neuroni et al., 2013)	
Satisfied stakeholders.	(Cooke-Davies, 2002)	
Long-term value creation by the project.	(Cooke-Davies, 2002; Nilsen, 2010)	

Table 2 Success factors from literature

4.7 ECONOMIC VALUE

As stated above, the adoption of open data promises an economic value (Chui et al., 2013). Moreover, successful projects lead to a maximized return on investment, resulting in the creation of new economic value (Cooke-Davies, 2002). Furthermore, research indicates that charging marginal costs¹ for open data reuse will lead to optimal economic growth in society, as cost inefficiency reduced the productivity impact significantly (European Commission, 2000; Pollock, 2009; Vickery, 2011; Weiss, 2002). According to Weiss (2002), this is achieved by 1) increased indirect taxes resulting from higher sales of products that incorporate open data, and 2) by increased income taxes resulting from a growth of the number of jobs. On top of that, Nilsen (2010) states that there are no negative economic impacts from public organizations releasing their data. However, this seems a little bit naïve, as organizations will have to put in time and effort to open their data.

Although not much has been written and researched about open data and its economic value (Dekkers et al., 2006; Heimstädt, 2014; Nilsen, 2010; Open Data Research Network, 2013), some research is performed into the potential economic impact of open data. This research evolves mostly around the European Commission as they have performed a targeted research at the economic value of open data (i.e. PIRA and MEPSIR). They found that most value can be unlocked in the geographical sector (European Commission, 2000). PIRA provided the first estimations of the economic value of open data, however, this was mainly based on estimates and only some fields were taken into account (Dekkers et al., 2006; Vickery, 2011; Weiss, 2002). Therefore, their analysis is not complete and their estimations are not accurate. The MEPSIR study based its predictions of the economic value of open data all estimations regarding the economic value of open data can be traced back to PIRA, as there is not enough empirical data available (Dekkers et al., 2006). Lee et al. (2014) state that there is clear evidence of the economic value of open government data, however, this information is not available. Fortunately, Vickery (2011) looked into the pitfalls of the PIRA and MEPSIR studies, and thought of a formula to calculate the marginal cost pricing of digital open data:

Gains = $2/5F\lambda\varepsilon$

in which F represents the revenue under average cost pricing (i.e. when the price is set to balance the costs), λ represents the multiplier, and ϵ the elasticity of the demand. However, the formula is not explained in detail, so it is still unclear how it is composed. Moreover, this still does not provide an indication of the economic value of open data. Even though it is a starting point to be able to calculate the size of that value, unfortunately, this is beyond the scope of this research. In this research, only the conclusion of Vickery (2011) is taken into account, which is that improved access to and use of open data has a major contribution to the economy.

Traditionally, the value of information is determined based on its scarcity (Hall et al., 2012; Nilsen, 2010), even though it is very hard to establish this value (Pollock, 2006). However, this will not apply to open data, as it will be available for everyone. The value lies in the potential to result in better decisions, new products and services, and the possibility to combine resources (Nilsen, 2010). Therefore, the relation between "Buienradar results in a lot of money. That is only possible thanks to the Royal Netherlands Meteorological Institute (KNMI). At first they thought, opening the data will cost us a lot on yearly basis, but now Buienradar and all the other apps built on that data result in so much money, that it multiplies the tax money it takes for KNMI and the resulting tax money (because of those new businesses) with a factor 100."

Interviewee 2

¹ Note that for digital data, the marginal cost will, often, be zero (Pollock, 2009).

open data and value generation is not necessarily a direct one (Dekkers et al., 2006; Jetzek et al., 2013; Pollock, 2006; Ubaldi, 2013; Verhulst et al., 2014). This is also depicted in

Figure 4; the added value is realized after several adaptations to the data are made. However, it is difficult to ascertain the exact difference between raw data and the final, sellable, product (Nilsen, 2010). According to Heimstädt (2014), organizations in between the data publisher and the consumer are responsible for the second and third step in Figure 4, with the support of an open data ecosystem.

In essence, Figure 4 focuses on phase 1 and 2 of Figure 3, which are the phases that need to be performed, before the reuser can find the data. In addition, the value chain ends with a final product, whereas the Open Data Process is based on the idea that feedback from the reusers can be used by publishers to improve the dataset. Therefore, the value chain of Figure 4 is part of the process depicted in Figure 3.

Opening data saves money that otherwise would have been spent on using that same data, therefore it is possible to cut public expenses (Fioretti, 2010; Vickery, 2011). Moreover, on the demand side, the effects for the end-user are presumable greater than just the reduction of costs (Vickery, 2011). Lee et al. (2014) indicate that only high-value datasets are able to generate economic value. They define high-value datasets as datasets where there is active engagement with potential users and the release is demand-driven. In order to achieve such a high-value dataset, the satisfaction of the users and the data usage metrics need to be evaluated on a regular basis (Lee et al., 2014).



Huijboom and Broek (2011) observed that, in addition to scientists, policy makers also acknowledge that the precise economic impact of open data policy remains unclear. In addition, the fact that organizations publish a lot of data does not automatically imply that those organizations understand the value and benefits of open data (Bonina, 2013). The findings of lemma (2012) illustrate this as he found that the revenue from open data initiatives within public organization is often less than one percent.

It appears to be easier to calculate the costs accompanying open data, than its (potential) economic impact (Bonina, 2013; Nilsen, 2010; Vickery, 2011). This is reinforced by the divers use of open data (Nilsen, 2010). Nonetheless, Conradie and Choenni (2012) indicate that the costs for open data are also difficult to establish. Bonina (2013) indicates the relevance of business models for this matter, as "business models are about explaining how value is created as to how value is captured" (p. 10). In addition, Hall et al. (2012) and Kassen (2013) acknowledge the value of open data that arises when organizations can involve third parties to enrich their data. According to them, "[d]ata becomes more valuable when it is possible to reuse it in unanticipated contexts – and hence it must be represented in such as a way as to facilitate this" (Hall et al., 2012, p. 34). Kaschesky and Selmi (2014) add to this notion that the view that sharing actually creates value is just starting to rise.

Chui et al. (2013) distinguish five ways in which open data can unlock its value.

1. <u>Transparency</u>; being able to access the right information in a decision-making process could save money.

- 2. <u>Exposing variability and enabling experimentation</u>; based on the open data variability between, for instance, school results can be made public, then, experiments can be performed based on this data.
- 3. <u>Dividing the population</u>; by dividing the population actions can be tailored to specific target audiences, making it more likely for an action to reach its desired effect.
- 4. <u>Automating human decision-making</u>; open data can enrich the data that is used to improve or even automate decision-making.
- 5. <u>Create new products, services, and business models</u>; as described before, by opening the data more people will take a look at it, resulting in new products, services, and business models.

Currently, the European policy is focused on releasing data, which are only phase one and two as depicted in Figure 3, less attention is paid to how these releases should create impact and value (Kaschesky & Selmi, 2014; Oliver et al., 2012; Zuiderwijk & Janssen, 2014). Therefore, the focus of their policy should be on use and benefit, instead of access and potential for reuse (Ubaldi, 2013; Wright et al., 2010). Pollock (2009) and Capgemini Consulting (2013) indicate that the content of the policy should be adequate for the situation to be truly effective. However, is remains unclear how open data creates economic value (Bonina, 2013; Jetzek et al., 2013), which should be discovered first, before a policy focused on the impact of open data can be formulated.

lemma (2012) proposes two charging schemes for public sector information; one for short term earnings, the other for long term financials. Bonina (2013) proposes three models to generate money from open data. However, both authors based their models on the intention to charge people for using the data. In addition, Vries et al. (2011) propose to use a freemium business model, which entails that the basic application can be used for free, users only have to pay for premium functions. This is in conflict with the definition of open data as provided in chapter 3 on page 8.

Although the fact that increasingly more data is available, can be seen as a barrier, it also provides the opportunity to create value from combining datasets (European Commission, 2000). By creating this added value, the project owners are able to let customers pay for their product, resulting in economic gains (Vries et al., 2011).

According to Davies (2014) and Fioretti (2010), the economic benefits of open data consist of economic growth, job creation, innovation, national competitiveness and stimulating organizations. Fioretti (2010) also acknowledges the savings within an organization, as opening data provides opportunities to remove activities or make them more efficient. In addition, Leeuw and Lemmer (2011) propose three ways in which open data can create economic value: 1) the released data meets a need as the interest in open data increases; 2) innovative products will unlock the economic potential; 3) participation and monitoring by citizens.

According to Janssen et al. (2012) and Vickery (2011), it is not possible to calculate or predict the return of investment for opening data. Nilsen (2010) and Dekkers et al. (2006) add to this that the value of open data will differ for each user. Most important to realize is the fact that opening up data does not create value (Chui et al., 2013; Janssen et al., 2012; Manyika et al., 2013). It is the use of that open data that creates value (Chui et al., 2013; European Commission, 2000; Lee et al., 2014; Neuroni et al., 2013; Vickery, 2011; Zuiderwijk et al., 2012). Hall et al. (2012) added the notion that the value of open data arises from the ability to use and process the data with diverse tools. The analysis of the surroundings of the data are an important factor in determining the economic value (Bates, 2014; Kassen, 2013). However, because not all data is open, the potential value of open data remains restricted (Kaschesky & Selmi, 2014). According to Zuiderwijk et al. (2012) and Ubaldi (2013), it is too ambitious to create public value from the use of open data, as too many factors are not optimal yet. In addition, Jetzek et al. (2013) and Nilsen (2010) state that there is still no consensus reached on how to measure the different aspects of open data.

4.8 SUMMARY

This chapter discussed the findings from literature. The literature turned out to be homogeneous, resulting in statements that are supported by multiple sources. To summarize these findings, the Open Data Process will be used. In Figure 5, the Open Data Process and some summarizing statements are depicted.

The reason that open data is a timely topic can be seen by the fact that increasingly more data is created and used by organizations. By opening data, even more data will become available. So it is important to prepare an organization to be able to cope with such an amount of data. On the other hand, it is already known that not all data should be released at once, as users might feel overwhelmed with the amount of data. Therefore, the creation of data is already being done.

To go to phase 2 (i.e. Opening data), publishers need to know the (potential) benefits to create motivations to actually open their data. As presented in Table 1, the benefits might differ per stakeholder, however since the organizations will be publishing the data, their benefits are most important for them to actually open their data. Some of the benefits from their perspective include: improved communication, economic growth, lower data processing costs, more (potential) customers, and improved participation. The actual motivations to open data come mostly from outside the organization. Legislations and examples of other countries or organizations are just two of the motivations found in practice.



Figure 5 Open Data Process – overview (comments by the author)

When the data is opened according to the definition of open data, reusers should not encounter any problems finding the data. For the actual use of the open data some barriers exist. One of the barriers is lack of knowledge of the reuser, resulting in erroneous interpretations and misuse of the data. Another barrier is about the format of the data. As long as there is no standard format, different formats will be used, which do not interoperate well, making it hard to combine datasets and actually reap the benefits of open data.

When the barriers are put aside, or overcome, the data will be used by reusers. This is in itself already a success for the publishers. For the reusers the success can be indicated by economic value, although the economic benefits of open data are very hard to measure. For both sides, the success is dependent on various factors, as mentioned in Table 2.

5. INTERVIEWS

In this chapter, the findings of the interviews are presented. First, the interviewees are shortly introduced, after which their answers are presented.

5.1 INTERVIEWEES

As stated in the method section (p. 5), the purposive sampling technique (Flick, 2009) was used, resulting in the following interviewees:

- 1. Data journalist
- 2. Project manager Open State Foundation
- 3. Information advisor municipality
- 4. Initiator open data projects, such as openOV
- 5. GIS consultant
- 6. Social entrepreneur (in small and medium enterprises)
- 7. Founder Buienradar & Weerslag
- 8. Employee Belastingdienst
- 9. Part of management of ministry of Infrastructure and the Environment (including spatial planning), involved in coordination center open data
- 10. Open data coordinator municipality
- 11. Information analyst (on a freelance basis)
- 12. Technical program manager
- 13. Policy officer, focused on real time weather data on national and international level
- 14. Service manager at NDW (National Data Warehouse for Traffic Information)
- 15. Part of Leer- en Expertisepunt Open Overheid (to help public professionals with Open Government), focused on Open Data

The summaries of all interviews can be found in Appendix B: Interview summaries. Quotes used within this thesis are derived from the summaries and can be found in Appendix B.[number of interviewee].

5.2 RESULTS OF INTERVIEWS

The interviews consisted of three parts: an introduction, focused questions adapted to the interviewee, concluding questions related to success and failure of projects. An example of the questions can be found in Appendix A: Interview questions.

5.2.1 PART 1: INTRODUCTION

The introduction of the interview is aimed at getting the interviewee to talk. It includes questions as "Why are you interested in open data?" and "What is your definition of open data?". Based on the answers of the interviewees more in depth questions were posed. At the end of this section, the definition as used in this thesis was given, the interviewees were then asked what they thought about that definition in comparison to their own definition.

PERSONAL MOTIVATIONS FOR OPEN DATA

In the first part of the introduction phase the focus was on finding out how the interviewee got involved with open data and how open data is interwoven with his daily work. Table 3 provides an overview of the reasons for the interviewees to get involved in open data.

Motivation	Mentioned by interviewee
Hobby	1, 15
Interested in all data	1, 5, 7
Promote democratic transparency	2, 12
Promote participation	2, 12
Transparency	3
Access to data	4
Stimulating innovation	4
Intrigued by possibilities	7, 8
More than just data	10
Policy	13, 14

Table 3 Motivations for open data

For interviewee 15 open data started as a hobby, as he learnt from open data out of the United States. It was an interesting phenomenon, which is now the focus of his daily work. For interviewee 1, open data is a hobby. As interviewee 1 indicates, he is interested in all data. By opening data the job of a data journalist becomes easier, as more data and sources are available. It will be easier for such a journalist to look critically at the statements made by politicians. The same applies for interviewee 5, he does not make a difference between open data and data that is easy accessible, even though it would be nice if all data would be open. Moreover, interviewee 7 created a successful data project (Buienradar) without making use of open data. As addition to interviewee 1, he also acknowledge the advantage of open data as being easy accessible and ready to use, however, he also mentioned the challenge that it might lead to misinterpretation.

The political view on open data is the motivation for interviewee 2, 3 and 12, as they are mainly interested in open data because of the transparency it can provide for democracy. As interviewee 2 states "a decision will be better when the people which are subject of the decision know what it is about". Interviewee 3 connects transparency with trust in government, "as the citizen can observe and check what the government is doing". In addition, interviewee 2 and 12 look at open data as a means to increase participation.

Interviewee 4 started looking into open data, because he wanted to access data that was not open for customers or he needed to pay a lot of money for getting the data. OpenOV as well as openKvK arose from the fact that the government holds data which it does not want to open. Both are a copy of the data of the government, acquired by volunteers, by which those datasets are now open. As an effect he noticed that it led to innovation, as the market was acting upon his projects. Currently, he is using open data to stimulate innovation in some markets.

Both interviewee 7 and 8 are intrigued by the possibilities of data. According to interviewee 8 the value of combining datasets is just being discovered. As can be read in the following quote, combining datasets can be very helpful.

"It helped us to reveal an enormous scheme that was going on in the Netherlands in which people pretend to have a right to subsidy, which they did not. This costs us a lot of money, and by combining datasets we were able to reveal this scheme."

Interviewee 8

For interviewee 10 it is about more than just data. For him it is about making people aware of open data, changing their mindset so that they understand what opening data can do for them, and also about organizational change.
Interviewee 13 and 14 got in touch with open data because it became an issue in the European Commission, followed by a directive from that commission. Both interviewees work at organizations that are influenced by this directive.

Interviewee 1, 6 and 11 acknowledge that there is a lot of data available and that it is hard for people to find the data that is interesting for them. Interviewee 6 also indicates that there is a need for people who act as a kind of broker to connect people to the right dataset. According to interviewee 1 it takes special skills to be able to disclose the data that you are looking for, which are mostly acquired by experience.

DEFINITION

The second part of the introductory phase of the interview consisted out of defining open data. In this section, the elements mentioned by the interviewees are used to see whether they are covered in the definition used in this thesis. First the elements mentioned by the interviewees are described, after which a comparison is made with the definition used in this thesis.

Element	Mentioned by interviewee	Captured in definition?
In public domain	1, 6, 9, 14, 15	\checkmark
In a structured way	1, 9	
Without restrictions regarding government or businesses	1	
Financed and gathered using public means	2	
No restrictions on reuse	2, 3, 7, 12, 15	\checkmark
Preferably in an open standard	2	\checkmark
Machine readable	2, 3, 9, 12, 13	\checkmark
Taking away as much barriers as possible for reuse	3, 4	\checkmark
Easy to find	3	\checkmark
If it is possible to make it free, it should be free	4	
The attribution clause, is the only limitation that is allowable	4, 9	
Accessible for everyone	5	\checkmark
Freely accessible	5, 7, 9	\checkmark
Open to society	6, 9, 10, 11, 12	\checkmark
Availability	8, 13	\checkmark
Possible to acquire at once	9	
Goal: to give meaning to the city	10	
General data	11	
Raw data	13	\checkmark
Making it available as soon as possible	14	

Table 4 Elements of open data definition

ELEMENTS OF INTERVIEWEES' DEFINITIONS

Table 4 depicts the elements mentioned by the interviewees, it also indicates who has mentioned the element, and whether the element is incorporated in the definition used in this thesis. Four elements are named by five interviewees, which implicates that those elements are important. These elements are:

- In public domain

Most of the interviewees that have mentioned the public domain are working for a governmental organization. From their perspective, open data is data that is opened by public organizations, or organizations in the public domain, and therefore financed with public means. For instance, interviewee 15 has based his definition solely on the law WOB and the law on reuse that is being

implemented at the moment. Other interviewees indicated that it is about data that is currently available within organizations.

- No restrictions on reuse

This element is mentioned by interviewees who think about how people are going to use open data, but are not actively using the data themselves. Interviewees that are more involved in using open data are more careful by stating that as much barriers for reuse as possible should be eliminated. Interviewee 2 specifies some of the restrictions that can be put on data, but are not allowed within the definition of open data: obligations to pay, login screen.

- Machine readable

Interesting to see is that interviewees that are on the side of opening data incorporate machine readable in their definition, whereas interviewees who are using the data and indicate that it is one of the success factors, do not incorporate it in their definition. So, apparently in order for open data projects to be successful it is important that the data is machine readable, but it is not part of the definition according to those interviewees, who are working with open data themselves. Interviewee 9 adds to this notion that it should be possible to acquire all data at once. This does not imply that it should be machine readable, but when it is machine readable it is often easier to acquire the data at once.

- Open to society

The data should be opened to a larger audience than it reaches currently, in such a way that it is publicly released. This is highly related to accessible for everyone as posed by interviewee 5 and the availability as added by interviewee 8 and 13. As can be seen below, interviewee 8 only looks at the availability of data for himself.

"For me it is open if I can reach it, so if this is within the government, it is not open towards everyone, but for me it is."

Interviewee 8

The other elements that are mentioned by the interviewees are also important aspects, but are less frequently mentioned. For example, interviewee 14 indicates that the data should be released as soon as possible, but he also states that doing so is something what they do, without getting that as an assignment from the ministry. So they are not obligated to do so.

In addition, some elements are about the way the data is opened. Interviewees 1 and 9 mentioned that the data should be opened in a structured way to enable the findability of the data. Interviewee 3 also mentioned this, as he states that the data should be easy to find. This is also related to the addition of interviewee 2, that the data should preferably be opened using an open standard.

There are also some extra notions about the restrictions, interviewee 1 adds that it should not pose restrictions on government and organizations. Moreover, the attribution clause is seen as the only restriction that might be applied to the data. However, interviewee 4 indicates that it is not free by definition, it should be free if it is possible, but if it is not possible it is only fair to pay a compensation. Interviewee 5, 7, and 9 state that it should be freely available at all times. Interviewee 7 adds that it is allowed to pay for the delivery or infrastructure of the data, but the data itself should be free.

Some remarks are made about the data. For instance, the data should always be general data. Interviewee 11 specifies the data that can be opened, as this cannot be critical business data, data that might disrupt the market or data that brings privacy at stake. Interviewee 13 adds to this that it should be raw data.

"We define raw data as digital information which form the basis for a subsequent adding value process. That might imply that we already perform some actions on the data, but we don't create an end product."

Interviewee 13

Lastly, interviewee 10 incorporated the goal of open data into his definition. For him it is important that the city benefits from the open data and that the data provides meaning for the city.

COMPARISON WITH DEFINITION

The definition used in this work, established in chapter 3, is as follows:

Open data is data that is publicly available, easy accessible and has no restrictions for the users.

As is shown in Table 4, about half of the elements mentioned by the interviewees are incorporated in this definition. Below, for all elements it will be explained why they are or are not included in the definition.

- In public domain The definition includes that the data should serve a public purpose and is therefore in the public domain. However, the definition is not limited to data that is opened by public organizations, data from private organizations is also included, as open data is about the way that the data is made available and not about the source.
- In a structured way It is not explicitly incorporated in the definition of open data that the data should be opened in a structured way, however, the data should be easy accessible. As there is no standard defined yet for opening data, it is very hard to do so in a structured way. When there will be a standard, this should be incorporated in the definition.
- Without restrictions regarding government or businesses The definition is aimed at the reuse of data, so the potential restrictions for publishing organizations are out of scope of the definition.
- **Financed and gathered using public means** As with the previous element, the definition is aimed at the reuse of data. However, as the data should be available without setting any charges, it is most likely that it is financed using public means.
- **No restrictions on reuse** This is explicitly incorporated in the definition, as it is one of the key elements of open data.
- **Preferably in an open standard** This is also included in the definition. However, as stated above, there are no standards defined yet. Therefore, it is not explicitly stated in the definition, but included in the description of the definition in chapter 3.
- **Machine readable** It is incorporated in the explanation of the definition in chapter 3 that the data should be machine readable. As indicated by some interviewees, at this moment it is more important to open the data than to make it machine readable as well.
- **Taking away as much barriers as possible for reuse** The definition states that there are no restrictions for reuse, which implies that as much barriers as possible need to be eliminated.
- **Easy to find** The data being easy to find is also incorporated in the explanation of the definition.
- If it is possible to make it free, it should be free This is not included in the definition, as the goal of open data is to make the data freely available for users.
- The attribution clause, is the only limitation that is allowable The attribution clause is part of the license that accompanies the open data. However, in the definition this is also aimed at the reuse of the data. It is more important that the data can be used than that an attribution clause is or is not included. As long as the reuse is not impeded, it is fine within this definition.
- Accessible for everyone This is implied within the definition, as the data should be publicly available.

- **Freely accessible** In the explanation of the definition it is stated that the data should be available without any charges being put on the reuser.
- **Open to society** This is implied within the definition, as the data should be publicly available.
- **Availability** In order to be able to use the data it should be available, this stated in the definition by the data being publicly available.
- **Possible to acquire at once** The definition incorporates that the data should be easy accessible, if it is possible to acquire the data at once, it is very accessible, however, this is a gradation that is not included in the definition.
- Goal: to give meaning to the city The goals of open data are left out of the definition.
- **General data** The definition clearly states that open data is not about the type of data, but about the way it is opened. Therefore, the restriction of privacy-sensitive data is left out. Moreover, this is often arranged in legislation and regulations.
- **Raw data** A separate definition of data is provided in chapter 3. However, in the definition of open data, it does not state whether the data should be raw data. The definition is aimed at the reuse of data, the type of data is out of scope for this definition. As interviewee 11 indicates, he would like to see a combination of raw data and already processed data.
- Making it available as soon as possible It would be nice if the data is made available as soon as possible, and for real-time applications it is a necessity, but it is not part of the definition of open data. It could be considered as a success factor.

Interviewee 4 had an addition for the definition as he would like to see some guarantee towards the user. The reuser of the data should be able to trust that the data that is being opened is reliable. However it is not possible to give such a guarantee, as interviewee 9 stated, when the data is usable within an organization it can be opened towards the public.

CONCLUSION PART 1

In the introductory part of the interviews, the motivations and definitions of the interviewees are discussed. The motivation that is most often mentioned is that the interviewee is not particular interested in open data, but he is interested in all data. Having open data is for them just easy as it is better accessible. Another frequently mentioned motivation is transparency. Some of the interviewees had a political focus, which resulted in them wanting to use open data to increase (democratic) transparency.

Regarding the definition of open data, most interviewees agree with the definition as used in this thesis. However, they all provided their own definition before that one was shared with them. All elements mentioned by the interviewees are mapped on the used definition, and for each of their elements an explanation is provided as for why it is (not) incorporated in the definition used in this work.

5.2.2 PART 2: FOCUSED QUESTIONS

In the second part of the interview multiple subjects were discussed. Based on the background of the interviewee and the answers in the first part a subject was chosen. The various subjects will be discussed in this section.

CURRENT POLICY

One of the focused questions was about the current policy regarding open data. The question was posed as an open question, resulting in answers directed both at the European level, as well as Dutch legislation. This subject was discussed with interviewee 2, 3, 9, and 10.

All four interviewees agree that the policies that are in place at the moment function as guidelines. They all indicate that the policies leave room for interpretation. However, only interviewee 2 states that the policy should be a bit severe by forcing public organizations to supply a data register. The other interviewees do not feel the need to make the policies stricter, as they think that for the first years the current policies are capable of providing a guideline for organizations to open their data. As interviewee 9 indicates, with the amount of member states it is likely that the European guideline on open data is the result of a compromise. Therefore, it is not striking that it "leaves some room for interpretation". Interviewee 2 adds to this notion that the focus of the European directive is too much on linked (open) data, while he states that is a step too far ahead.

"Let us first focus on opening the data, linking the data is a next step."

Interviewee 2

Even though all interviewees agree that opening data by public organizations should be more proactive instead of just reactive as it is now, they do not see a need for a stricter policy. As indicated by interviewee 2 and 3, public organizations are obliged to provide data, but only when someone request that data. Without adapting the current Dutch policy, interviewees 2, 3, and 10 see that the number of organizations that is proactively opening data is increasing.

According to interviewee 2 and 10, the Dutch government should focus at their national open data portal. They both find a lot of broken links in this portal, which depreciates the goal of the portal, which is to provide an easy accessible overview of open datasets.

ADDED VALUE OF OPEN DATA

Another subject discussed in the interviews is the (potential) value of open data. This subject is discussed with interviewees 1, 2, 3, 4, 5, 7, 8, 12, and 15.

The economic value of open data is promised to be billions, however, it is very difficult to show how this will be achieved. According to interviewee 2, the economic value of open data is realized by the fact that data is disclosed, making it available for others, resulting in new products and services. For interviewee 3 the economic value is already reached when the data is available for others and when the data is easy to find.

"Economic value is something that originates from a successful project and the success factors."

Interviewee 4

Interviewee 15 indicates that the billions of Euros that are promised to arise from open data might be found in costs that are saved. He also indicates that it is very hard to measure the value of open data as traditional methods are used to measure a new phenomenon. An example of how open data can contribute to somewhat hidden economic value can be found below.

"When there would be no weather apps, more people might be having a cold. This might be due to the fact that they do not longer have an overview of when they will be out in the rain, therefore, they might be outside in the rain all the time. By using weather apps this number is reduced, as people will be able to take appropriate measures. In its turn this will save on medical costs."

Interviewee 15

Most of the interviewees acknowledge that the added value of open data goes beyond just the economic value. By just making the data available in a usable way, enabling reusers to combine datasets, some added value is already achieved, as it contributes to transparency, according to interviewees 1, 3, 4, 8, and 12. Interviewee 7 adds that the added value of open data is in combining data in unique possibilities and the interpretation of the data. It is not in the data itself, it is in the possibilities of the data. As highlighted by interviewee 5, the added value of open data is much clearer in the context of developing countries in comparison with more developed countries that are less dependent of open data.

Interviewee 1 suggests that the most interesting open data projects are the ones with a high impact on society. This is especially true if the project was not expected to have a high impact, but the realized potential turned out to be unexpectedly large. This is affirmed by interviewee 4, who achieved an unexpected impact with one of his projects (i.e. openOV). Up front they considered the project a success if they could find 30 participators, instead they found 92. For him it is important that "the establishment realizes that there is a force that is there to stay", so in essence, he likes to challenge existing parties.

NEW PROJECT

All interviewees are asked what project they would start and which data they would use if they would start a project at this moment. It is interesting to see that both interviewee 1 as interviewee 2 would like to see the data of city councils as open data. They both indicate that this would be beneficial for local democracy.

Another topic that was mentioned multiple times is regarding sustainability. Interviewee 3, 4, and 14 would start a project related to sustainability. Interviewee 3 would compare energy usage of various municipalities, interviewee 4 would like to see smart devices to control the energy usage in a household, and interviewee 14 would make an analysis of the amount of emissions in a particular area over a period of time.

Other interviewees would like to start a project with a societal impact. Interviewee 6 would like to see a project with disease progression, interviewee 8 would create a project regarding poverty alleviation, and interviewee 9 would like to visualize roads and buildings so that firemen can use those to prepare for firefighting. Some other interviewees are more focused on small societal impact, for instance interviewee 10 would create an app in which all buildings of Utrecht are mapped including their opening hours to reinforce consumer behavior.

The conclusion for this question is that every interviewee mentioned a project which is in line with his current work. They all named projects which would be relevant for them at the moment. One of the barriers mentioned was that there is so much data that it is impossible to create a useful project. However, the answers to this question indicate that people will stay within certain boundaries of what is already familiar to them. Therefore, the amount of data in total has less influence on losing sight of what can be done, as people will mostly reason from their own experiences.

GOOD DATA DATABASE

It is hard to stay up to date with open data, as also indicated by interviewee 4, this was one of the questions that all of the interviewees got. However, this resulted in the fact that people make use of their own networks and hear about new events regarding open data via those networks. Therefore, interviewees were also asked whether they see a need for or would use the GOOD DATA database themselves.

In general, the interviewees definitely saw a use for the database; all from their own perspective. They all expressed their own whishes for the final version of the database. Interviewee 1 and 2 see a possibility to present the top stories of open data, whereas interviewee 3, 10, 11, and 14 would use the database to see who is using their data and what products are developed based on their data.

The fact that such a database would provide an overview of (active) projects, which data is used, what products are created, and who is working on it already is most of the added value of the database, as this is mentioned by ten of the fifteen interviewees. Moreover, interviewees 2, 3, 10, and 11 indicate that they would use the database to get in touch with others, either to collaborate or to find out whether their data is up to the expectations of the reuser.

As interviewee 4 indicates the database could also function as a pr tool, as others can see what you have done. Furthermore, the datasets that are used might get a boost out of it. When people see that a successful project has used dataset X, they might start looking into that dataset to see what other things can be done with it.

The database should be very usable, as current initiatives to create such an overview often lack the ability to perform a directed search. Therefore it is important that it is easy to search within the database, so if someone is looking for a particular project, he or she can actually find it by combining multiple search terms or tags. By linking from the database to the original source of the project (Interviewee 5), the database can function as a central point for open data projects (Interviewee 4).

There are also some barriers raised by the interviewees. For instance, interviewee 10 emphasizes that it is not up to public organizations to facilitate and create the database. This should be done by the community. In its turn, this is what interviewee 7 thinks that will break the project, as "I do not think companies are going to tell you their detailed plan", resulting in general descriptions of the projects. Interviewee 6 contributes to this by stating that "a lot of innovative organizations are unwilling to share their ideas with others". So, for the GOOD DATA database to become a useful source, it is important to ascertain the possibility of an active community contributing to the database in a meaningful way.

CONCLUSION PART 2

In the second part of the interview, multiple topics are discussed as the topic was dependent on the interviewee. One of the topics concerned the current policies regarding open data. Both the policy on international level (i.e. as stated by the European Union) as the Dutch national level are discussed. Even though all interviewees think the current policy is not strict enough, they do not feel like it is necessary to make the policy stricter.

The added value of open data is also discussed. Most of the interviewees acknowledge that the added value of open data goes beyond economic value. They also indicate that this might be where the promised billions of Euros come from. When they were asked which project they would start themselves, they often mentioned a project with a societal impact. This also indicates that there is more than just the economic value.

Lastly, the need for a database such as the GOOD DATA database is discussed. In general, they all see a use for the database, although not everyone would actually use it. Moreover, they all indicated how they would like the database to look like.

5.2.3 PART 3: SUCCESS AND FAILURE OF PROJECTS

In the last part, the interviewees were asked about their definition of project success, what they think of as success factors, and whether they could think of a way to measure success of open data projects. Since it is important to also capture the other side of success, to avoid a bias, the interviewees were also asked about project failure and potential failure factors. In addition they were asked to name a successful open data project as well as a failed open data project.

SUCCESS AND SUCCESS FACTORS

This section will tell about the definitions of success, how to measure it, and the success factors mentioned by the interviewees. Table 5 provides an overview of all success factors that were mentioned by the interviewees. The table also indicates by who the factor was mentioned.

Category	Success Factor	Mentioned by interviewee
General		
	Satisfied reusers and customers	4, 5, 6
	Market acceptation	6
	Knowledge of data	7
	Knowledge sharing	10
	Learning during the process	10
	Being able to sell your idea	14
Opening data		
	Usefulness	1
	Cooperation of the publisher	1
	Continuity, up to date data	1, 7, 11, 12
	Access to data	2
	Reuse of data	2, 3, 4, 13, 14
	Combining data	4, 7
	Availability of data	8, 11, 12
	People and their networks	9
	Mapping your data landscape	10
	One central datacenter	10
	Timeliness, accuracy, and completeness	11
	Legal status of information	12
	Legislation	12
	Availability of documentation accompanying the data	12
	Publicity	2, 13
Effect of open data		
	Support for transparency	2
	Increase of trust in government	3
	Citizen participation	3, 10
	Impact of reuse	4
	Ecosystem surrounding the project	4
	Embedded in standard practice	14
	Able to show effects	15

Table 5 Success Factors mentioned by interviewees

Within the listed success factors of Table 5 different perspectives can be found. First of all, some success factors are generally applicable to projects.

- Satisfied reusers and customers; the goal of each project should be to satisfy the customer. With open data projects the customers are often the reusers of the data. If they are not satisfied with the data, they will stop using it. So it is important to listen to their feedback.
- *Market acceptation*; for a project to have any chance to become a successful project, it is important that it is accepted by the market.
- *Knowledge of data*; for each project it is important that the content is familiar for the team members. Otherwise, the data can be misused or misinterpreted. However, it is not a necessity that team members are already familiar with the data, as they can also learn about it during the process. The importance of this factor strongly depends on the project.

- *Knowledge sharing*; to get the most out of a project, it is wise to share acquired knowledge with other parties, and learn from each other.
- *Learning during the process*; when new insights are gained during the project, it is important to be "brave enough to act upon the things that you have learnt" (Interviewee 10).
- Being able to sell your idea; for the individual who is working with the data and gets inspired, seeing a lot of possibilities. However, those are out of reach for the individual self, to accomplish them more money or manpower is needed. If a large organization steps in and buys the idea of the individual, the idea can still be executed.

Moreover, some factors are related to the process of opening datasets.

- Usefulness; whether the data that is being opened is useful to society. This could be measured using factors that a dataset can score on a scale from 0 to 10, for example whether it is beneficial for democracy. It will then serve as an indication of the usefulness and added value of the dataset.
- *Cooperation of the publisher*; the publisher has to be cooperative to ensure that customers will be able to use their data.
- *Continuity*; it is important that reusers can rely on the data that it is up to date, consistent and relevant over time. This is important as a lot of projects contain a real-time component.
- Access to data; having access to data itself can already be seen as a success factor.
- *Reuse*; when the data is actually reused in other projects it can be considered a success. Organizations can stimulate this reuse themselves, however, this can also be arranged by the ecosystem. When the open data is used in several applications, the success is even bigger. Also when the data is used by increasingly more customers, who are spread amongst society.

"We try to stimulate the actual reuse of data by organizing hackatons, app challenges and so forth. We do this also to show public organizations that their data is really used and that it is applied in innovative ways."

Interviewee 2

- *Combining data*; by combining datasets, it is possible to filter the data and create added value.
- Availability of data; in addition to being able to access the data, the data should be available. If the data is not available, it is not possible to use it at all. Although interviewee 8 states that the format of the data is not important, interviewee 11 and 12 indicate that it is important to look at how much effort does it take to make the data usable. Interviewee 11 does state that when someone really wants to use a dataset, the format is not decisive, but when he is looking for interesting data PDF files are not opened, so format is a factor to consider, regarding availability. As the number of open datasets increases, the usage of standard formats will increase.

"We often forget that if we would close all datasets again, a lot of applications will stop working."

Interviewee 15

- *People and their networks*; it all starts with enthusiastic people that are initiating a project and using their networks to bring organizations together.
- *Mapping your data landscape*; knowing which data is available within an organization, to be able to say which datasets can be opened.
- One central datacenter; it is confusing and time consuming for customers to go to several separate portals to collect the data that they want to use.
- *Timeliness, accuracy, and completeness*; three aspects of a dataset that are of importance for the success rate of open data projects based on a dataset.

• *Legal status of information*; without an accompanying license it is hard for reusers to determine whether they can reuse the dataset. Therefore, it is better to use a license that is as open as possible.

"For instance, OpenStreetMap does not allow the restrictions on data of having to use references to the original source. If they would accept that, their product becomes unusable as the complete right side will consist out of references to original sources."

Interviewee 13

- *Legislation*; as there are differences in legislation, the legislation can make a difference for the success rate of a project. In the European Union governments are allowed to put restrictions on the reuse of their open data, as in the United States all data is open by definition.
- Availability of documentation; when there is no documentation accompanying the data, it is very hard to use the data. So this documentation should be complete, describing what the dataset is about and how to interpret particular parameters.
- *Publicity*; sometimes it is not clear that particular data is open. So by advertising with an open data portal or dataset, other parties are able to see what data is available. Interviewee 13 indicated that organizations are often surprised about the data that is available. Moreover, the publicity could also serve as a first guide of how to use the data, to reduce the chance of misinterpretation and misuse.

In addition to those factors, the following factors are related to the effect of opening the data.

- *Support for transparency*; by using open data reusers are contributing to the transparency of the publishing organization, as the data of the publisher is further processed and spread amongst a larger population.
- *Increase of trust*; by being more transparent and open, the expectation is that the trust in government will increase. This can be measured using a survey.
- *Citizen participation*; opening data supports citizen participation, as citizens are able to track all decisions made. Interviewee 10 adds to this notion that it is important to involve citizens to create the basis for the ecosystem and ensuring that the data is appealing to (potential) reusers.
- *Impact of reuse*; when it is not only beneficial for the publishing organization, but also for its environment. When the effect of the data can be measured from both a societal perspective as well as an economic perspective. For example, that it is noticed when there is a failure within the data, or as interviewee 14 said: "When the portal is down, you get a lot of reactions".
- *Ecosystem surrounding the project*; when the project is initiated by enthusiastic people (see previous category), an ecosystem should arise. This ecosystem can provide feedback, but also money.
- *Embedded in standard practice;* it is assumed that it is always there, so when it is not, it is being missed.
- Being able to show effects; even though it is not possible to show all effects, the ability to show some effects is already a success. For instance, when an open data event is organized, the number of registrations for today's events are much higher than those from a few years back. That is one of the effects that can be shown.

Interviewee 6 proposes to define open data as a KPI in a project to measure its added value within that particular project.

FAILURE AND FAIL FACTORS

This section will tell about when an open data project fails and which factors contribute to that, according to the interviewees. Table 6 provides an overview of all fail factors that were mentioned by the interviewees. The table also indicates by who the factor was mentioned. As mentioned by interviewee 12, the fail factors can be

seen as an "inverse of the success factors". Other interviewees indicated that the adoption rate of open data will differ per country and cultural differences will determine what will work at different places.

According to interviewee 1, a project fails if the effort put into the project is disproportionate to the outcome of the project. However, it might be that this can only be determined based on a comparison with similar projects. In addition, the failure of a project can be determined at the end of a project, when there are no results coming from the project, yet it can also be determined halfway down. Interviewee 13 and 14 also refer to the efforts put into the project. They indicate that when the data is not used at all, the project of setting up an open data platform has failed.

Interviewee 4 is more focused on the data itself, as he mentioned that it fails if the "data is sold as open data, but is not really open". If that happens, it is no longer an open data project, and the efforts to create one have failed. This is more directly related to the factors mentioned by interviewees 5 and 7, which are also often aimed at the data itself. As they indicate that erroneous and incomplete data and quality of the delivery of the data as fail factors. Remarkable is that interviewee 5 and 9 also state that open data projects do not fail because of the data, but because of action performed on the data. Interviewee 8 indentifies just one fail factor, as he indicates that only the database law prohibits one "from using the data and creating something new".

Category	Fail Factor	Mentioned by interviewee
Project Management		
	Mismanagement	1, 2, 6
	Differences in expectation	1, 11
	Knowledge drain	1
	Not thinking about your target group	9, 10, 15
	Resistance of the business	10
	No longer able to work with each other (other municipalities and organizations)	10
(Open) Data Projects		
	Lack of (domain) knowledge	1, 11, 13
	Changes in legislation and regulations	1
	Momentum	2
	Copyright issues	2
	Fear towards opening data	2, 9, 12
	Not putting words into action, only writing an advice, not acting upon the advice	2
	Thinking too much as a public organization	2
	No usage of the data	3, 10, 13, 14
	When the data is not ready to use	4
	Erroneous data	5
	Reliability of the delivery of the data	7
	Database law	8
	Lack of documentation accompanying the dataset	11
	Not seeing the added value of open data	12

Table 6 Fail Factors mentioned by interviewees

As can be seen in Table 6, diverse fail factors are mentioned. On the one hand, factors are mentioned that have to do with project management in general. These factors include:

• *Mismanagement*; this includes putting together the wrong people, erroneous communication, setting unrealistic expectations, and being too focused. As it is not possible to think of everything beforehand it is important to have an open mind during the project.

- *Differences in expectation*; for a project to succeed it is important that the goal of the project is clear to all members, with differences in expectation it is hard to achieve the goal.
- *Knowledge drain*; it is important to think about what happens if one of the team members leaves. If that person happens to be the only one with knowledge about a particular part, the knowledge will be gone together with that person.
- Forgetting to think about your target group; it is important to always keep your target group in mind during the process. As interviewee 9 pointed out: "if your target group consists of people over 60, how likely is it that they will be able to use a smartphone or a tablet". According to interviewee 10, it is therefore, important to keep communicating with your end-users. As interviewee 15 pointed out, at hackatons a lot of intelligent applications are created, "however there is no practical use for it, so it disappears again".
- *Resistance of the business*; as with all projects it is important to also have commitment of the business to be able to reach the goal. If the business offers resistance to opening data (or proceeding with a project), it is very hard to complete the project.
- *Issues in the collaboration with other parties*; as with the business commitment, it is important to keep working together. If there is an issue in the collaboration, the focus will shift from the project and it will be very hard to reach success.

On the other hand, some fail factors are more specific for (open) data projects, including:

- Lack of (domain) knowledge; with the amount of data that is currently available and will be available in the near future, it is tempting to start working with data about which one has no knowledge. At this point it is an easy mistake to misinterpret the data, leading to erroneous analyses.
- *Changes in legislation and regulations*; opening data is sometimes arranged using legislation and regulations, if these change, a dataset might no longer be up to date.
- *Momentum*; the right moment needs to be there to start an open data project, if the subject is not appealing at the moment, the project will probably fail, as no ecosystem will be formed.
- *Copyright issues*; when the data is published without a license, it is hard to determine which (part of the) data is available for reuse.
- *Fear towards opening data*; when an organization is afraid of the consequences of opening data, such as that people will sue them, or that people will see what goes really on in the organization, and thereby also see all the things that an organization might prefer to keep from their customers, prevents data from being opened.
- *Thinking too much as a public organization*; a lot of IT projects within the public sector fail because they miss the mentality of a private sector organization.
- No usage of the data; if nobody uses the opened data, it is pointless to keep putting in the effort. As interviewee 14 illustrated: "Is it worth the effort if no one even notices that the server is down for a weekend". Interviewee 10 indicated that if no one is using the data, the project did not fail from the perspective of the publisher, but the project did fail as a whole.
- *Not ready to use data*; when it requires another party to prepare the data before it can be used, it does indicate a certain economic value, however, it also indicates the failure of the data publisher.
- *Erroneous data*; this also includes incomplete data. It is always important to check the validity of the data before it is opened to the world.

"In Sierra Leone villages have different names; one village can have five different names due to language differences. If this is not complete, it is a challenge for aid organizations to coordinate their actions, especially in crisis situations"

Interviewee 5

- *Reliability of delivery of the data*; if the data has a lot of down time, the product loses quality.
- *Database law*; the database law might prohibit one from using a particular dataset, which entails that it cannot be used to create something new.

- *Lack of documentation*; without proper documentation accompanying the dataset, it is extra difficult to interpreted and use the data in the right manner.
- Not seeing the added value of open data; the most valuable data is not opened yet, this is often because the organization holding the data does not see the added value for themselves to open the data.

Some of these factors emanate from personal struggles. As interviewee 2 indicates, sometimes an advice regarding open data is drafted, however, it is immediately put into a drawer. So, nobody is acting up on the advice. That is a shame, as the plan is already there, they just forget to execute it. Other factors were thought of as they might be of influence, such as the change in legislation and regulations. Interviewee 1 does not have experience with such a situation.

As can be seen in Table 6 the most frequent mentioned fail factor is that the data is not used at all. However, all four interviewees do not expect this to happen. They expect that there always will be someone looking at and using their data. At the moment they can monitor their portal (this applies to interviewees 3, 10, 13 and 14), and all of them see their portal being visited, not only by Dutch visitors, but also by foreign people. However, all of them would like an overview to see what is actually being done with their data.

CONCLUSION PART 3

In this last part of the interview, the success and fail factors are discussed. In total 28 success factors were mentioned by the interviewees. The reuse of data and the continuity of the data were mostly mentioned by the interviewees. In addition, twenty fail factors are named. Some of them can be linked to a success factor, as one of the interviewees already indicated, the fail factors are basically the "inverse of success factors" (Interviewee 12).

A more in depth comparison of the success factors can be found in chapter 9, where the success factors will be prioritized for their evaluation, based on the number of mentions in literature and in the interviews.

5.3 SUMMARY

In conclusion, the interviews were divided into three separate parts. In the first part, the motivations of the interviewees to get involved with open data are discussed, resulting in the most mentioned motivation that they are interested in all data and that they want to increase (democratic) transparency. The interviewees were also asked for their definition of open data, the elements mentioned in those definition (see Table 4) are held against the definition used in this thesis. For each element it is explained whether or not it is incorporated in the definition as created in this work. Overall, the interviewees agreed with the definition created in this thesis.

Regarding the current policies concerning open data, the interviewees feel like the policies should be stricter. However, they do not see the urge to make that happen, as they think organizations should focus first on opening the data. Furthermore, they foresee added value on different aspects. At the moment, they focus mostly on the societal impact of open data projects.

The success and failure factors of the third part of the interviews will be discussed and combined with the ones found in literature in chapter 7.

6. LITERATURE VS. INTERVIEWS

In this chapter the findings from literature and the interviews are compared. For this comparison, the same division is made as when the literature was discussed in chapter 4.

6.1 WHY NOW?

As indicated in section 4.1 (p. 9), open data is gathering momentum (Davies, 2014). This is also reflected in the interviews, as especially interviewee 10 and 15 are actively involved in getting open data at the agenda (for respectively a municipality and the Dutch government). This is nicely put by interviewee 15 as can be seen in the following quote.

"The conversation has started, organizations are aware of open data and are trying to get their data ready to open."

Interviewee 15

However, although the focus of literature is shifting from opening as much data as possible towards realizing the benefits, organizations are still focused on opening datasets. Even though interviewee 11 indicates that just opening datasets is not enough, the dataset really have to mean something for the reuser. Interviewees 10 and 13 indicate that their organizations are just in the process of opening data and making open data the standard within their organizations. Interviewee 13 even indicates that even though the ministry states that they should make their data more usable, their organizations sees this as something for the future. However, the involvement of the ministry is also highlighted in literature, as the public sector is crucial for the survival of open data (Kassen, 2013).

Even though public sector information (PSI) is often confused with open data in literature, the interviewees never mentioned PSI. Nonetheless, interviewee 8 gave a definition of open data which was close to PSI, as he stated that for him data is open if it is available for himself. Data that is only available within the public sector or within governments but which is still available to him, is also open data according to his definition.

6.2 WHAT DO WE KNOW SO FAR?

For large organizations it is easier to adapt to open data as those organizations are often already used to deal with large amounts of data (Interviewee 1). This is one of the reasons that organizations are encouraged to open their data as is, in this way it does not take a lot more effort to start with open data. As interviewee 15 indicates, the discussion about open formats et cetera is for a later stage.

"When you look at the United Kingdom, they say they are very mature in opening the data, but the data that is opened, is not the data that you want to use, as it is not up to date for real time services."

Interviewee 7

In literature a lot can be found about open data policies. This is probably due to the fact that an open data policy is one of the more tangible points regarding open data, as it is all in writing. However, not much is written about the implications of the policies. Interviewee 7 stated that the current legislations provided by the European Commission are "a joke", as the compliance with those legislations is very low. Of all data that should

have been opened by now, according to legislations, only a fraction is actually open. In addition, the data that has been opened is not always truly open, as can be read in the quote of Interviewee 7 above.

However, as interviewee 5 and 8 indicate, at the moment people are already happy that they can access the data. Whether the data is polluted, or it is misinterpreted, is of minor importance at the moment. Despite the fact that interpretation is key for successful combining (open) datasets (Interviewee 8) and data is needed to achieve a business model (Interviewee 6).

"The rationale behind the system of key registers is very nice, linking all data using a single source. However, they forget to incorporate the principle of multireality: what is true for one, is not per definition true for another. Therefore, the context sensitivity of data should be incorporated in the system of key registers."

Interviewee 8

As mentioned by Leeuw and Lemmer (2011) the government in the Netherlands is not publishing data to its national open data portal. This was also mentioned by interviewees 2 and 10 as they found a lot of broken links at the national open data portal. As seen from the other sides, owners of open data portals are not capable of monitoring what is being done with the data that is on the portal. Interviewee 14 only indicates this problem, while interviewee 10 is acting upon it. Interviewee 10 acknowledges the importance of a community and is working with divers parties to be able to monitor what is being done with the data.

"We see that a lot of data is being downloaded, but we don't know who is doing that and what products are created from the data."

Interviewee 14

The variety of stakeholders is both indicated in literature as by the interviewees. In section 4.2, this is described using the open data process. Interviewee 5 emphasizes this by stating that all stakeholders have their own interests. He adds to this that stakeholders will therefore not drown in the available data, as long as they have it clear for themselves what they want to find. This is opposite to what is stated in literature, as multiple papers mention that users get overwhelmed by the amount of data that is suddenly available (Conradie & Choenni, 2012; Davies, 2011; Ding et al., 2012; Harrell, 2013; Janssen et al., 2012).

6.3 PROMISES OF OPEN DATA

A lot of papers found in the literature study mention an increase in transparency (e.g. (Attila et al., 2013; Hall et al., 2012; Heimstädt, 2014; Lee et al., 2014; Vosough, 2013; Wright et al., 2010). In the interviews, transparency was also mentioned, however, the interviewees also mentioned (potential) effects of the increased transparency. Interviewee 2 mentions a decrease in venation as a result of increased transparency, while interviewee 5 relates an increased transparency to governments becoming more auditable, which will then benefit the democracy, which is also mentioned by interviewee 15. Just as a lot of authors (e.g. (Heimstädt, 2014; Janssen & Zuiderwijk, 2012; Manyika et al., 2013; Neuroni et al., 2013; Ubaldi, 2013), interviewee 10 sees an increased transparency as a basis for renewed trust in the government, which will lead to an increase in participation. This is supported by the findings of (Evans & Campos, 2013; Heimstädt, 2014; Huijboom & Broek, 2011; Neuroni et al., 2013; Vickery, 2011).

Innovation is another promise of open data which is mentioned in a lot of papers (e.g. (Bates, 2014; Conradie & Choenni, 2012; Iemma, 2012; Manyika et al., 2013; Verhulst et al., 2014; Zuiderwijk & Janssen, 2014). Interviewee 10 and 12 agree with them, as they also see open data as an accelerator for innovative products.

Even though interviewee 10 still foresees some issues, as the current mindset needs to change to enable opening datasets. However, interviewee 7 states that open data does not lead to more innovative projects. Since data is being opened, he has seen only one innovative project, all the other projects are more of the same.

"I think that if someone is creative, he will look for the data and if he really wants to go for it, it does not matter whether it is open or not."

Interviewee 7

Interviewee 8 and 9 anticipate on the increase of innovative products. According to interviewee 8, the rise of innovative products shows that even though one does not know all applications that can come from the data, the applications will be there. Interviewee 9 indicates that the innovative products contribute to the smartness of the city and therefore that of the citizens. In addition, interviewee 11 indicates that more people will be able to look at the data and they will, therefore, see new possibilities to link datasets. This is also found in literature, as Attila et al. (2013), Oliver et al. (2012), and Vickery (2011) stated this as well.

A promise of open data that will become reality in the future, as expected by interviewee 11, is that "[t]he use of standards will increase the usability of open data". Using standards will facilitate the cooperation of organizations (Hall et al., 2012). Interviewee 15 refers to this as an increase in efficiency, as open data allows organizations to look in each other's data and use it as well.

"For instance, if you have a datasets which involves municipalities, it would be nice if you use the same abbreviations or codes as for example the CBS does."

Interviewee 11

6.4 MOTIVATIONS FOR OPEN DATA

Huijboom and Broek (2011) indicate that most motivations to actually open data come from outside the organization. Interviewee 7 indicates that internal motivations are hard to find, as he stated that when an organization opens its data it will get a lot more competition for their product. As most organizations do not want more competition, it would not be wise to open their data. therefore, motivations for opening data should be sought after outside the organization. On the other hand, interviewee 4 provides a counterargument, as he approaches opening data as a form of knowledge sharing. By opening data organizations can make easier use of other data, while they do not have to pay for all data.

Existing legislations can function as a motivation for organizations to open their data (Huijboom & Broek, 2011). This is supported by the motivations mentioned by interviewees 13 and 14. Moreover, examples from higher level function as a motivation (Huijboom & Broek, 2011). In line with this vision, interviewee 15 and his team are focusing on setting the right example at the Dutch ministries, so that lower level governments can join the movement and do not have to start themselves. In this way, it is easier to spread open data, as lower level governments can provide from the example set by the Dutch ministries, which means it will take them less effort to use open data.

As Chui et al. (2013), Conradie & Choenni (2012), Huijboom and Broek (2011), and Roy (2014) indicate, one of the motivations to open data is to increase the customers' engagement. However, not all organizations are ready to do so. Therefore, interviewee 9 indicates that it is important for organizations to realize that it is not a bad thing to receive feedback on their data, as it allows them to improve the data.

"By spreading successful examples, where organizations have dealt with feedback on their data, you create more understanding and acceptance of open data."

Interviewee 9

Another motivation to open datasets is that organizations already have the data, they only have to open it. This is what interviewee 14 mentioned, as his organization only created a new environment for their open data to protect their own production environment. Therefore, they did not experience any issues with opening their data. However, as interviewee 3 stated, a lot of organizations do not have their data available in a structured way. He proposes to use the process of opening data as a means to get more structured data, which is then easier to open.

6.5 BARRIERS FOR OPEN DATA REUSE

Barriers for opening data can be found within organizations (Huijboom & Broek, 2011; Vosough, 2013). This is also indicated by interviewees 5 and 7. They state that it will take some time before data will be opened, as commercial organizations are now making money on combining closed datasets (Interviewee 5). For these parties it will be necessary to change their business model when all data they work with is opened. Moreover, there are always organizations that are unwilling to open their data (Interviewee 7). Interviewee 13 add to this that the opening of datasets would be facilitated if there would be some financing to support it.

Zuiderwijk et al. (2012) found that most barriers are related to the actual use of open data. This is also reflected in the answers provided by the interviewees. First of all, interviewees 2 and 11 indicate that just opening the data is not enough. People still need a little boost to actually use the data. Moreover, interviewees 2 and 5 indicate that it is important that the data is well structured, as it will take reusers too much effort to use the data when it is not well structured. One solution to get more well structured data is to introduce standards to which organizations and their open data have to comply. However, interviewee 13 sees some issues with the introduction of standards, as not all reusers will be able to work with those standards, as it requires specific knowledge to do so.

Knowledge can also act as a barrier, this can be either a lack of knowledge about how to open datasets (Neuroni et al., 2013; Vickery, 2011) or what to do with the open data, or acquiring too much knowledge (Interviewee 6). When parties get too much knowledge they can become a threat for the market and for other organizations. A lack of knowledge about the data used in an application can also be a barrier, as it can be perceived as a danger as stated by interviewee 7, see also the quote below.

"It is dangerous that app users always assume that the data behind the app is correct and contains the right information."

Interviewee 7

However, according to interviewee 11 this is not an issue of open data, but in the whole internet culture, a lot of things are misinterpreted. Furthermore, interviewee 7 adds to this that human interpretation of the data will always be important. Interviewee 8 also refers to this, as he foresees problems concerning the context of the data. He indicates that the meaning of data differs per system, even though systems think they understand each other, they do not, as their interpretation of the same input differs.

As indicated by Manyika et al. (2013) and interviewee 7 privacy is one of the concerns regarding open data. Interviewee 8 predicts that privacy will always be an issue that cannot fully be covered within the revolution of open data. Legislations concerning privacy are not dealing with the combination of datasets.

"So I am allowed to see your income and I am also allowed to know the worth of someone's house, but it does not say whether you are allowed to know the combination of those two."

Interviewee 8

According to interviewee 5, open data will be mostly used by organizations and not by individual customers. Therefore, it is important to train organizations in their open data use, as suggested by interviewee 2. This training can also be used to overcome the fear towards opening data, that is influencing some organizations, mostly municipalities, in their decision whether to open their data. According to interviewee 15, the mindset should change, as it is more important to start a project from a societal issue than just starting a project because the data is available.

In addition to the barriers mentioned in literature and in the interviews, interviewee 15 provided an overview of 45 perceived barriers as found in practice. These are all reasons people have given of why they are reluctant to open data. These excuses are related to what is mentioned in the interviews and in literature and are as follows:

- 1. Costs too much money
- 2. There is no business case
- 3. Has commercial value
- 4. It is privacy sensitive
- 5. It is a secret
- 6. It is not ours and we do not know whose it is
- 7. The quality is unknown
- 8. We do not know where it is
- 9. It is not our work
- 10. It is not a usable format
- 11. I am not authorized
- 12. People will take advantage of it
- 13. Image damage for the minister
- 14. It is not done yet
- 15. The government loses its reputation
- 16. The files are too large
- 17. We do not have enough bandwidth
- 18. It is our first step
- 19. We can find it, but we do not have access to it
- 20. It is out of date
- 21. We only have it on paper
- 22. We do not know whether it is legal
- 23. Our management says no
- 24. We have not done it before
- 25. We have no idea what someone should do with it
- 26. It has no value
- 27. No time / resources

- 28. They can rob it
- 29. We open it (but adjust 90%)
- 30. It is incomplete
- 31. It is incorrect
- 32. Commercially sensitive
- 33. It is dangerous if it is combined
- 34. People will take the wrong things out of it
- 35. It leads to unnecessary discussions
- 36. We cannot confirm or deny that we have it
- 37. We know there are mistakes in the data and people will send improvements to us which we then have to process
- Our IT supplier will charge an astronomical amount to create a data extract
- 39. Our website cannot place local files
- 40. The data is not ours and we do not have permission of the owner
- 41. We already published the data (on an untraceable place in an unusable format)
- 42. People will store the data themselves and will be working with this outdated data
- 43. We do not collect the data in a structured way
- 44. Too much people will download the data, making our servers collapse
- 45. People will become angry because of what they learn

As stated in section 4.5, all these barriers result in new risks for organizations (Manyika et al., 2013). Interviewee 15 adds to this notion that each organizations handles these barriers differently.

6.6 SUCCESS AND OPEN DATA PROJECTS

Discussing success and open data from a literature point of view led to the creation of 25 success factors. The comparison of the success factors, from literature and the ones from the interviews, will be done in a separate section, as this will result in a prioritized list of success factors that will be evaluated. Therefore, this can be found in chapter 7, where the factors are processed and in chapter 9 which covers the evaluation of those success factors.

Next to the success factors, the interviewees gave indications of successful implementations of open data. Interviewee 8 stresses the importance of the ethical side by stating that "[e]ven though it is allowed by law, possible by technology, that does not mean you also have to do it". This corresponds with the findings of Zuiderwijk and Janssen (2014), who indicated that it is important to think about the consequences of open data.

"I can use it, I am allowed to use it but still, will I use it?"

Interviewee 8

Within the organization, interviewee 10 discovered that it is beneficial to have one person responsible for open data. Everyone will be able to find that person and they will know that they have to go to that person with regard to all open data issues. This refers to clear leadership, which is one of the six elements that are composed by Chui et al. (2013). This particular element is supported by Zuiderwijk and Janssen (2014) as they also acknowledge the importance of a central person who can promote open data to the remainder of the organization.

Not only inside the organization, but also the service towards their reusers is important. As interviewee 11 states, organizations should be answering the questions from their reusers. At the moment, a lot of the organizations are not doing this, interviewee 14 states that that will be done once all their data is open. Until then, they will grow towards such an intensive customer relation. On the other hand, interviewee 14 adds that the data itself is most important, "[i]f your product is usable, customers will start using it".

According to interviewee 11, it is hard to determine the success of some open data projects. Some open data projects originate just because people want to practice their skills with (open) data, they do not have a specific goal. Therefore, it is hard to indicate whether their goal is achieved.

6.7 ECONOMIC VALUE

As stated in section 4.7 (p. 20), the European Commission performed a study into the economic benefits of open data (European Commission, 2000). A lot of the interviewees also heard of this research and were also astonished about the rough estimates provided by that research. For example, interviewee 13 indicates that it is "a political statement to move towards open data with the expectation that it will bring a high economic result". He also indicates that it is more important to realize that the data that is being opened, is already paid for by taxes.

"The first order effect is the app developer that makes some money with his app, the second order effect could be someone's happiness and is therefore much bigger than the first order, but also immeasurable."

Interviewee 15

As this subject is already discussed in section 5.2.2 (p. 30), those remarks will not be repeated in this section. Additional information acquired from the interviews will be used to compare them to literature. For instance, interviewee 15 foresees two degrees of effect by opening data, as can be read in the quote above.

The relationship between open data and its value creation is not a direct one (Dekkers et al., 2006; Jetzek et al., 2013; Pollock, 2006; Ubaldi, 2013; Verhulst et al., 2014), making it hard to estimate its complete value (Nilsen, 2010). Interviewee 12 adds to this that open data is not always visible, see also the quote below. Moreover, interviewee 11 experienced the added value of open data himself as he got a new job assignment out of his open data project. His open data project functioned as a marketing piece to show potential employers what he can do with open data.

"For instance the train delays are included in Google Maps, that is based on open data, however, not much people realize that. This is also the beauty of open data."

Interviewee 12

As indicated by Bonina (2013), Nilsen (2010), and Vickery (2011) it is easier to calculate the costs of opening data than to make an indication of the benefits. Interviewee 13 reasons from the loss of revenue because of terminating the license fees. He did not perceive more revenues by opening the data, which is partly due to the satisfaction of the meteorological market. Furthermore, Bonina (2013) mentioned the importance of business models to assess the value of open data. However, interviewee 3 is not positive about the most used business model in open data projects, in which users have to pay when they want more information, the so called freemium business model (Vries et al., 2011).

"For example the first ten are free and then you have to pay for each extra data point you want access to. It is really insipid, why don't they just open up their complete database."

Interviewee 3

6.8 SUMMARY

The findings of the interviews are confirmed by the findings in the literature study. Although not all aspects that were found in literature came back in the interviews, the interviews act as an addition to the existing literature. Moreover, the interviews confirm that open data is gathering momentum. However, it should be noted that the interviewees were found based on their activities involving open data.

Both literature as well as the interviews indicate that the focus is shifting from opening as much data as possible towards realizing the benefits, however, in practice, organizations are still working on how to open their data. They are not yet busy with realizing benefits, even though this can function as a motivation to open their data. The largest contribution of the interviews is the list of success factors, which will be evaluated in the chapter 9.

7. PROCESSING THE SUCCESS FACTORS

In this chapter the success factors from literature and the interviews are prioritized and combined. To provide more overview over the complete list, the success factors are mapped onto the Open Data Process. In the last section the fail factors are prioritized.

7.1 PRIORITIZATION SUCCESS FACTORS

In this section the success factors as found in literature and in the interviews are prioritized. By doing so, a more workable list is obtained. This list will then be used in the evaluation of the success factors in chapter 9.

7.1.1 SUCCESS FACTORS FROM LITERATURE

Based on the number of occurrences of a success factor in literature as found by the author, the prioritization of the success factors is as follows:

- 1. Presence of an ecosystem, surrounding the data (mentioned 8 times)
- Collaboration (mentioned 5 times) Consideration of potential value (mentioned 5 times)
- 3. Adequate strategy (or policy) (mentioned 4 times)
- 4. Clear working practices and organizational processes (mentioned 3 times) Presence of an infrastructure (mentioned 3 times)
- 5. Active presence of the public sector in the market (mentioned 2 times) Appropriate release channels (mentioned 2 times) Availability of metadata (mentioned 2 times) Clear leadership (mentioned 2 times) High-level mandate (mentioned 2 times) Long-term value creation by the project (mentioned 2 times) Motivated group of users (mentioned 2 times) Project is initiated at a local level (mentioned 2 times) Raise awareness of the existence of the project (mentioned 2 times) Usage of the data (mentioned 2 times)
- 6. Attitude of the organization (mentioned 1 time) Different perspective on (existing) datasets (mentioned 1 time) Innovative project (mentioned 1 time) Location of the project (mentioned 1 time) Planning and coordination of the project (mentioned 1 time) Possible negative impacts (mentioned 1 time) Possible positive outcomes (mentioned 1 time) Proper security of the data (mentioned 1 time) Satisfied stakeholders (mentioned 1 time)

Within this prioritization success factors with an equal amount of mentions are put in alphabetical order.

7.1.2 SUCCESS FACTORS FROM INTERVIEWS

In addition to the success factors found in literature, the interviews also resulted in a list of success factors. These factors are also prioritized based on the number of times they are mentioned in the interviews. As with the success factors from literature, factors that have an equal amount of mentions are put in alphabetical order.

- 1. Reuse of data (mentioned 5 times)
- 2. Continuity, up to date data (mentioned 4 times)
- Availability of data (mentioned 3 times)
 Satisfied reusers and customers (mentioned 3 times)
- Citizen participation (mentioned 2 times) Combining data (mentioned 2 times) Publicity (mentioned 2 times)
- 5. Able to show effects (mentioned 1 time) Access to data (mentioned 1 time) Availability of documentation accompanying the data (mentioned 1 time) Being able to sell your idea (mentioned 1 time) Cooperation of the publisher (mentioned 1 time) Ecosystem surrounding the project (mentioned 1 time) Impact of reuse (mentioned 1 time) Increase of trust in government (mentioned 1 time) Knowledge of data (mentioned 1 time) Knowledge sharing (mentioned 1 time) Learning during the process (mentioned 1 time) Legal status of information (mentioned 1 time) Legislation (mentioned 1 time) Mapping your data landscape (mentioned 1 time) Market acceptation (mentioned 1 time) One central datacenter (mentioned 1 time) People and their networks (mentioned 1 time) Support for transparency (mentioned 1 time) Timeliness, accuracy, and completeness (mentioned 1 time) Usefulness (mentioned 1 time) Embedded in standard practice (mentioned 1 time)

7.1.3 COMBINING LITERATURE AND INTERVIEWS

When the prioritized lists of success factors from literature and interviews are combined, the following list of success factors is the result. The combination of both lists has been done based on the aggregation of the number of occurrences in literature and the number of mentions by the interviewees. Table 7 presents the final list, indicating which success factors are combined. Once again, success factors with an equal amount of mentions are ordered alphabetically.

Rank	Success Factors	As from literature	As from interviews	Total mentions
1	Ecosystem of people surrounding the data	Presence of an ecosystem, surrounding the data	Ecosystem surrounding the project People and their networks	10
2	Reuse of data	Usage of the data	Reuse of data	7
3	Collaboration between parties to exchange	Collaboration	Knowledge sharing	6

Rank	Success Factors	As from literature	As from interviews	Total mentions
	knowledge			
4	Consideration of potential value to be reached with open data	Consideration of potential value		5
	Strategy and legislation as set by the government	Adequate strategy (or policy)	Legislation	5
5	Continuity, up to date data		Continuity, up to date data	4
	Publicity for the open data project	Raise awareness of the existence of the project	Publicity	4
	Satisfied stakeholders	Satisfied stakeholders	Satisfied reusers and customers	4
6	Availability of data		Availability of data	3
	Availability of metadata accompanying the open dataset	Availability of metadata	Availability of documentation accompanying the data	3
	Clear working practices and organizational processes	Clear working practices and organizational processes		3
	Market acceptation of the open data by reusers	Motivated group of users	Market acceptation	3
	Presence of an infrastructure	Presence of an infrastructure		3
	Value creation for the publishing organization as well as for the reuser	Long-term value creation by the project	Impact of reuse	3
7	Active presence of the public sector in the market	Active presence of the public sector in the market		2
	Appropriate release channels	Appropriate release channels		2
	Citizen participation		Citizen participation	2
	Clear leadership	Clear leadership		2
	Combining data		Combining data	2
	Effects of opening datasets	Possible positive outcomes	Able to show effects	2
	High-level mandate	High-level mandate		2
	Initiated at local level	Project is initiated at a local level		2
8	Access to data		Access to data	1
	Attitude of the organization	Attitude of the organization		1
	Being able to sell your idea		Being able to sell your idea	1
	Cooperation of the publisher		Cooperation of the publisher	1
	Different perspective on (existing) datasets	Different perspective on (existing) datasets		1
	Embedded in standard practice		When it becomes a matter of course	1
	Increase of trust in government		Increase of trust in government	1
	Innovative project	Innovative project		1

Rank	Success Fa	ctors		As from literature	As from interviews	Total mentions
	Knowledge	of data			Knowledge of data	1
	Learning process	during	the		Learning during the process	1
	Legal informatio	status n	of		Legal status of information	1
	Location of	[:] the proje	ct	Location of the project		1
	Mapping landscape	your	data		Mapping your data landscape	1
	One centra	l datacent	ter		One central datacenter	1
	Planning an of the proje	nd coordii ect	nation	Planning and coordination of the project		1
	Possible ne	gative im	pacts	Possible negative impacts		1
	Proper se data	curity of	f the	Proper security of the data		1
	Support for	r transpar	ency		Support for transparency	1
	Timeliness, completen	accuracy ess	/ and		Timeliness, accuracy and completeness	1
	Usefulness				Usefulness	1

Table 7 Prioritized list of success factors

7.2 MAPPING THE SUCCESS FACTORS TO THE OPEN DATA PROCESS

As this combined list contains 42 success factors, they are mapped onto the Open Data Process, that was introduced on page 11. By doing so, the list is divided into six parts, which makes it easier to read. Moreover, by plotting the success factors to the Open Data Process it is visible which success factor is related to which phase in the process. The open data process consists of five phases, a sixth category is added for the success factors that cannot be mapped directly. Table 8 provides an overview of the success factors per phase. After the table, for each phase the success factors will be described. To create an overview, the success factors are highlighted in the descriptions in the remainder of this section.

Phase	Role	Success Factor			
Creating data	Dublisher	Continuity, up to date data			
Creating data	Publisher	Citizen participation			
		Consideration of potential value			
		Availability of data			
		Availability of metadata			
		Presence of an infrastructure			
		Appropriate release channels			
		High-level mandate			
Opening data	Publisher	Attitude of the organization			
		Legal status of information			
		Location of the project			
		Mapping your data landscape			
		Location of the project Mapping your data landscape Possible negative impacts Proper security of the data			
		Timeliness, accuracy and completeness			
		Publicity			
Finding open data	Reuser	Access to data			
		One central datacenter			
Using open data	Bousor	Reuse			
	הכשטכו	Clear working practices and			

Phase	Role	Success Factor	
		organizational processes	
		Market acceptation	
		Clear leadership	
		Combining data	
		Effects of opening datasets	
Using open data	Reuser	Different perspective on (existing)	
		datasets	
		Embedding in standard practices	
		Knowledge of data	
		Usefulness	
		Ecosystem	
		Collaboration	
	Satisfied stakeholdersValue creationPublisherActive presence of the the marketCooperation of the public Increase of trust in gove	Satisfied stakeholders	
		Value creation	
Discussing open data		Active presence of the public sector in	
Discussing open data		the market	
		Cooperation of the publisher	
		Increase of trust in government	
		Learning during the process	
		Support for transparency	
		Strategy and legislation	
		Initiated at a local level	
Other	-	Being able to sell your idea	
		Innovative project	
		Planning and coordination of the project	

Table 8 Success factors in the Open Data Process

7.2.1 CREATING DATA

This is the first phase of the Open Data Process, only two of the success factors are related to the creation of the data. The first success factor in this phase is 'Continuity, up to date data'. This factor ensures that the data that is released is up to date. This is important as the number of real-time applications is increasing and they cannot rely on obsolete data.

The other success factor related to the creation of data is 'Citizen participation'. This success factor was mentioned in the interviews and is about involving citizens in the open data project of governmental organizations. This involvement can have various interpretations, for instance it might entail that citizens will use the open data of a municipality to create open data projects, or that they work together on such a project. The factor is mapped onto this first phase as citizens can also participate in the generation of data.

7.2.2 OPENING DATA

The second phase concerns the opening of the data, twelve success factors are related to this phase. To be able to achieve a successful project, literature indicates that it is important to consider the potential value that the open data project can unlock. The success factor **'Consideration of potential value'** is placed in this phase as it is important to make a cost-benefit analysis at the beginning of the project, which starts when the data is disclosed.

Another important characteristic of the data is the 'Availability of data', as this is a requirement for the data to be reused. If the data is not available, the project is actually not an open data project, as the availability of the data is one of the key characteristics of open data.

In order to be able to use the data in a correct way, the 'Availability of metadata' is important. By including metadata, the publisher has the opportunity to explain how the dataset should be interpreted and used. By doing so, the dataset can be used in a better way, which probably result in more projects that are based on the data. It is important to consider the metadata when opening a dataset.

Moreover, by having thought of an infrastructure, the project can be easily distributed. This is captured in the next success factor: 'Presence of an infrastructure'. When an infrastructure is in place, it is also easier to select 'Appropriate release channels'. This success factor is about choosing the right channels to communicate the data to the intended public. If the intended users are not able to access the data, or they do not know about its existence, the data will not be reused.

To ensure that the project will continue, a 'High-level mandate' will help. By obtaining this, the person providing the mandate indicates that he is supporting the project. This will also influence the next success factor, which is 'Attitude of the organization'. When the general attitude within the organization is negative towards open data, the mandate ensures the process towards opening data or using open data will continue. However, it is important that the attitude in the organization is positive towards open data, to increase the success rate of the open data projects within that organization.

The next factor in this phase 'Legal status of information' is also influenced by the legislation of the government. However, it also concerns the legal status that has been put on the data by the organization that owns the data. Therefore, it has to be considered before the data can be opened.

The organization that publishes the data, influences the 'Location of the project'. The exact location of the project is also of influence on the success rate of a project, as the location often determines the language and thereby the potential users.

As indicated by one of the interviewees, it might be beneficial to map the existing data landscape before opening data, this is captured in the following success factor: 'Mapping your data landscape'. By creating an overview of what data is available within the organization and in which way it is structured, it will be easier to create an overview of what data can be opened.

In order to achieve a successful project, it is important to also incorporate the 'Possible negative impacts' at the beginning of the project. As it is always important to include a cost-benefit analysis for a project, it is also included as a success factors.

Before data is opened, it is important to think about the security of the data ('**Proper security of the data**'). As interviewee 14 indicated, they have built a new platform for the open data. Even though the datasets are identical to each other, they do not want to risk people hacking into their server park.

The last success factor describes three characteristics of the data: 'Timeliness, accuracy and completeness'. They were mentioned by one of the interviewees as general characteristics of datasets. This is confirmed in literature by Pipino, Lee, and Wang (2002) and Wang and Strong (1996), although they indicate that these are just three of many characteristics.

7.2.3 FINDING OPEN DATA

The third phase of the Open Data Process concerns the findability of open data. To enhance the findability, it is recommended to commercialize the data, which is related to the 'Publicity'. If there is no publicity evolving the data, less people will know of its existence. This will result in less reusers, making the data less interesting.

The findability of the data is related to the next success factor, which is 'Access to data'. When the (potential) reusers of the data are not able to access the data, they will not build projects with the data. Therefore, this characteristic of the dataset is key for its potential success.

For the usability of open data, and also to make it more attractive, **'One central datacenter'** will be beneficial. By creating one place where people can find all open datasets, it is easier for them to find the datasets. Moreover, it will also be an indication of the state of the art indicating which datasets are already open.

7.2.4 USING OPEN DATA

The fourth phase is about reusers that are actually going to use the open data. The first success factor in this phase is 'Reuse of data', as it is the goal and therefore added value of open data projects is to stimulate reuse. It is the foundation of the open data movement. When the data of a project is not reused, it is definitely not a success.

For each project it is recommended to have 'Clear working practices and organizational processes' in place, so this is also true for using open data. By mapping the practices and processes it is also easier to copy them to another project.

In order to reach the goal of publicity, it is important that the project is accepted by the market, which is captured in the next success factor: 'Market acceptation'. If the data or the project is not accepted by (potential) users, it will fall into oblivion. Moreover, the use of open data will lead to market acceptation.

The next success factor is 'Clear leadership'. Within the organization, but also within the project (e.g. which can be based on open data), it is important that someone is in charge. That person should be able to make decisions and act accordingly. When there is too much doubt it will harm the project.

The added value of the project is increased when datasets are combined, which is captured in the second success factor of this phase: 'Combining data'. The combination of datasets happens when the data is actually used, therefore this factor is mapped to this phase.

The use of open data will lead to several 'Effects', one effect of opening data is that more people are able to look at a dataset, this will result in 'Different perspective on (existing) databases', which is the next success factor. Another effect is that it becomes 'Embedded in standard practice'. This entails that it is such a normal application or project that it is hard to think it out of our daily lives.

Before someone, be it a person or an organization, gets involved with open data it is important to ensure that there is 'Knowledge of data'. To be able to use the open data in the right way, knowledge of the data is a prerequisite. This prevents misinterpretations of the data. This is also mentioned three times as a fail factor. When knowledge about the domain of the data is lacking, it is very hard to create a project based on this data.

Lastly, the 'Usefulness' of the data determines whether it will be used frequently by other parties. If the data is not perceived as useful, it will not be used. As one of the goals of open data is to ensure reuse of the data, the usefulness of the data contributes to the success rate.

7.2.5 DISCUSSING OPEN DATA

In the fifth phase it is all about improving the open data, this entails the publication of data as well as the characteristics of an open dataset, eight success factors are related to this phase. The first success factor in this phase is 'Ecosystem of people surrounding the data', according to the number of mentions in literature and in the interviews, this is also the most important factor. Within an ecosystem it is important that various parties work together, in this case its goal is to ensure the survival of the project. The ecosystem is mostly about the people and their networks, in such a way that they are surrounding the open data.

'Collaboration' is the second success factor in this phase. Even though it is part of the ecosystem, the ecosystem consists of more than just the collaboration. Therefore, collaboration is another success factor. Moreover, collaboration and the knowledge sharing because of that collaboration are mentioned separately in both the interviews and literature.

When a project is carried out successful, the requirements and whishes of the stakeholders are met, resulting in 'Satisfied stakeholders'. It is clearly an effect of an open data project, it is also the goal of a project to end up with satisfied stakeholders. To determine whether the wishes of the stakeholders are met, a discussion or evaluation at the end of the process is beneficial.

'Value creation' is also captured in this phase, as it can only be determined at the end. Each open data project will create some value, the kind of value depends on the specific project. Also the impact of reuse is taken into account within this factor. When a project creates much value, its impact will also be greater.

The next factor 'Active presence of the public sector in the market' is beyond the abilities of publishers and reusers, however, it can be evaluated in the last phase of the Open Data Process. Even when a public organization fulfills one of these roles, one organization cannot be held responsible for the presence of the complete public sector in the market. The active presence can be stimulated by organizations that are involved with open data, but they cannot ensure it.

To fully reap the benefits, the publisher of the data benefits from feedback from its reusers. Therefore, the **'Cooperation of the publisher'** is a success factor in order to improve (the usability of) the data.

As open data will result in an increase of transparency, it will be possible to see what happens within organizations with that data, or at least what data is processed within a particular organization. Therefore, a successful project will lead to an 'Increase of trust in government'.

In the process of opening the data or reusing open data it is important to learn from experiences. Therefore, the success factor 'Learning during the process' is incorporated in this phase, as it is an evaluation of the project.

By creating open data projects, there will be more attention for transparency and especially to the accompanying benefits. This will lead to 'Support for transparency', which can be evaluated in the discussion as well.

7.2.6 OTHER

As the success factors are not created to be mapped on the Open Data Process, some cannot be mapped to it. In this section, the other eight success factors are described. The first success factor is 'Strategy and legislation'. As both the strategy and the legislation are set out by the government, organizations or individuals that are opening or reusing data are bound to them. As it is related to both the opening of data as well as the using of the data, it cannot be placed in a phase. Another characteristic of an open data project is whether it is '**Initiated at local level**'. According to literature, it contributes to the success rate of a project when it is initiated at the level of a city or municipality instead of at a national level. Since it concerns the initiation phase of the project, it is also a characteristic of the project.

The next success factor is 'Being able to sell your idea'. When someone is able to sell his or her idea to another (commercial) party, it can be considered as a successful project.

The innovativeness of a project is also a success factor ('Innovative project') and a characteristic of an open data project. By adding new perspectives to the dataset the success rate of the project increases.

Most of the insides of the project are captured in the next success factor: 'Planning and coordination of the project'. The internal processes of the project include the planning and accompanying coordination of resources.

7.3 PRIORITIZATION FAIL FACTORS

Next to the success factors, the fail factors also need to be taken into account. The fail factors are also prioritized based on the number of mentions by the interviewees. Fail factors that have an equal amount of mentions are put in alphabetical order.

- 1. No usage of the data (mentioned 4 times)
- Fear towards opening data (mentioned 3 times) Lack of (domain) knowledge (mentioned 3 times) Mismanagement (mentioned 3 times) Not thinking about the target group (mentioned 3 times)
- 3. Differences in expectation (mentioned 2 times)
- 4. Changes in legislation and regulations (mentioned 1 time) Copyright issues (mentioned 1 time) Database law (mentioned 1 time) Erroneous data (mentioned 1 time) Knowledge drain (mentioned 1 time) Lack of documentation accompanying the dataset (mentioned 1 time) Momentum (mentioned 1 time) No longer able to work with each other (other municipalities and organizations) (mentioned 1 time) Not putting words into action, only writing an advice, not acting upon the advice (mentioned 1 time) Not seeing the added value of open data (mentioned 1 time) Reliability of the delivery of the data (mentioned 1 time) Resistance of business (mentioned 1 time) Thinking too much as a public organization (mentioned 1 time) When the data is not ready to use (mentioned 1 time)

As might be noticed from the list, a lot of the fail factors can be traced back to the success factors. For instance,

'No usage of the data' is the opposite of the success factor 'Reuse', and 'Lack of (domain) knowledge' can be mapped to 'Knowledge of data'. As the collection of the fail factors was limited to the interviewees, the fail factors are not included in the evaluation. More research is necessary to create a complete overview of these factors and to establish their relation to the success factors.

7.4 SUMMARY

In this chapter the success factors as well as the fail factors are prioritized. For the success factors, the factors from literature and interviews are combined. This resulted in a list of 42 factors, which can be found in Table 7. To keep the list manageable, the success factors are mapped onto the Open Data Process. The phases and the accompanying success factors can be found in Table 8 as well as in Figure 6 below. In this figure, the success factors are placed next to the phase they relate to. The success factors in the middle of the figure do not relate to a specific phase of the Open Data Process.



Figure 6 Success factors mapped to the Open Data Process

8. OPEN DATA PROJECTS VS. OTHER TYPES OF PROJECTS

In this chapter the relation of open data projects to other type of projects will be discussed. The focus will be on success factors found for not open data projects, because of the limited amount of literature available to compare open data projects with other projects on characteristics.

8.1 COMPARISON

In order to be able to make a comparison it is important to have a clear definition of what is meant by an open data project within this thesis. As stated in chapter 3, an open data project is a project that is based on open data. This entails that the project cannot exist without the open data. So, the open data is fundamental for the project's existence.

A lot of the characteristics of open data projects are general applicable to all projects. In literature, not much distinction is made between various types of projects. In fact, there is not much written about project characteristics in general. In 2001, Shenhar created a model in which engineering projects can be placed in various categories. However, this model only takes the level of technological innovation and system scope into account (Shenhar, 2001). In the model only specific projects can be placed as the focus of each project determines its place in the model, therefore, types of projects cannot be placed in his model.

Regarding the differences between open data projects and closed data projects, Bachmann and Bernstein (2009) found that even the data quality used within open projects is comparable to that of closed projects. Even though the data quality is perceived as one of the main dangers of open data.

8.1.1 SUCCESS FACTORS

A comparison based on the success factors is made, the overall comparison can be found in Appendix C (p. 108). In this Appendix, the success factors found in this research form the basis of the comparison, the factors as found in the comparing literature are mapped to those.

Belassi and Tukel (1996) proposed a framework in which they mapped success and failure factors for projects in general. As their framework is very elaborated it will serve as a basis for the comparison, their framework can be found in Figure 7. The factors they mention are very general. In addition, Atsu, Andoh-Baidoo, Osatuyi, and Amoako-Gyampah (2009) examined success factors for ICT projects in developing nations. They came to a list of fifteen factors, as found during their case study. The success factors of Belassi and Tukel (1996) and Atsu et al. (2009) have a lot in common. However, the factors of Atsu et al. (2009) are more focused on the internal processes of a project, whereas Belassi and Tukel (1996) also consider the environment of the project.

Compared to the list of success factors established in this research, some factors of Belassi and Tukel (1996) are overlapping. The overlapping factors are indicated in Figure 7. All overlapping factors have a number to facilitate the comparison of the factors. For each factor it will be indicated which success factor from the prioritized list is related to that factor.

- 1. Ability to delegate authority corresponds with the success factor 'Clear leadership'.
- 2. Ability to coordinate corresponds with the success factor 'Planning and coordination of the project'.
- 3. Value corresponds with the success factor 'Value creation'.

- 4. Uniqueness of project activities corresponds with the success factors 'Being able to sell your idea' and 'Innovative project'.
- 5. Top management support corresponds with the success factor 'High-level mandate'.
- 6. Customer acceptance corresponds with the success factors 'Satisfied stakeholders' and 'Market acceptation'.
- 7. Effective project management corresponds with the success factors 'Planning and coordination of the project' and 'Appropriate release channels'.
- 8. Project preliminary estimates corresponds with the success factor 'Consideration of potential value'.
- 9. Availability of resources corresponds with the success factor 'Availability of data'.
- 10. Political environment corresponds with the success factor 'Strategy and legislation'.



Figure 7 General success/failure factors (adopted from (Belassi & Tukel, 1996))

As can be concluded from Figure 7, not all success factors are found in the interviews and vice versa. Three phases are covered by the factors of Belassi and Tukel (1996). They have not defined a success factor for the phases 'Creating data' and 'Finding open data'. This can be explained by the fact that these phases are very specific for open data projects.

Table 9 depicts the factors of Atsu et al. (2009) and the corresponding factor of the list of success factors created in this thesis. First of all, the factors of Atsu et al. (2009) are spread over the phases. They only do not mention a success factor related to the phase of 'Finding open data'. This is also visible in the factors for which there is no corresponding factor. Those factors have to do with starting a project, for example, the availability

of funds refers to the starting capital. Moreover, motivation and basic requirements are also needed to set up a project.

The factors formal methodology and the last one (capital budgeting and post-implementation audit) are measures to ensure future project success. The methodology should lead to a standardization and the post-implementation audit will serve as a guide for future projects (Atsu et al., 2009).

Factors of Atsu et al. (2009)	Corresponding success factor
Availability of funds	-
Executive management support	High-level mandate
Training	Learning during the process
Motivation	-
Proper planning	Planning and coordination of the project
Minimized scope	Initiated at a local level
User involvement	Citizen participation
Firm basic requirements	-
Clear statement of requirements	Clear working practices and organizational processes
Formal methodology	-
Ownership	Clear leadership
Culture	Attitude of the organization
Political power	Strategy and legislation
Risk management	Possible negative impacts
Capital budgeting and post-implementation audit	-

Table 9 Comparison of success factors with Atsu et al. (2009)

Each year, the Standish Group creates a report about project success factors and challenges. This is an established yearly report that is highly appreciated by organizations and is frequently used in research (e.g. Atsu et al. (2009) and Ashja, Hadizadeh Moghadam, and Bidram (2013)). The Standish Group defines a project successful when "[t]he project is completed on-time and on-budget, with all features and functions as initially specified" (The Standish Group, 2013b, p. 4).

For each of the factors they identified they provide an explanation of how to implement that factor to enable a successful project (The Standish Group, 2013a), also the research of Atsu et al. (2009) is based on these factors, which explains the overlap between these two sources. In 2013², their top ten of project success factors was as follows (The Standish Group, 2013b):

- 1. User involvement;
- 2. Executive management support;
- 3. Clear statements of requirements;
- 4. Proper planning;
- 5. Realistic expectations;
- 6. Smaller project milestones;
- 7. Competent staff;
- 8. Ownership;
- 9. Clear vision and objectives;
- 10. Hard-working, focused staff.

 $^{^2}$ Note: The report of 2013 was the most recent available report. The report of 2009 used the same success factors, only then they were ranked differently (The Standish Group, 2009). This indicates that the factors are consistent over time.

Within these success factors the ownership and executive management support are attributed to the executive sponsor of the project. The executive sponsor has, as owner of the project, the "full weight and responsibility for the success or failure of the project squarely on his or her shoulders" (The Standish Group, 2013a, p. 5). As stated above, the community is responsible for an open data project, implying that there is not one owner of the project. Therefore, this does not apply to open data projects. However, it does raise the issue of accountability.

"The BAG (key register of addresses and buildings). Within the government files are linked, municipalities have worked together, it is done within a reasonable period, and it uses a national control point from which everything orchestrated. is Without it we would not have been able to catch a large number of fraudsters."

Interviewees 4 & 8

The role of the community is expressed in 'User involvement'. In the list of prioritized success factors in Table 7, the community is identified as most important factor. In 2013, 'User involvement' was also the most important factor according to the Standish Group (2013b).

The factors can be roughly divided into two categories, the first one is focused on the execution of project skills and those factors are most beneficial for the success rate of a project (The Standish Group, 2013a). These are factors such as 'User involvement' and 'Executive management support'. The other category contribute to success, but provide less benefits (e.g. 'Clear business objectives').

As mentioned above, the research of Ashja et al. (2013) is also based on the findings of the Standish Group. Based on a literature study in which eleven

papers are concluded, they identified sixteen success factors for large information systems. Their factors are spread amongst the phases of the Open Data Process, however, they also do not include factors of the phases 'Creating data' and 'Findin open data'. The factors of Ashja et al. (2013) that cannot be mapped to the success factors for open data projects include change management, effective communication, careful selection of appropriate software package, fit business to software, business mission and vision, and external consultant.

This comparison provides a first indication that there are some differences between open data projects and other type of projects. Moreover, a common issue with open data projects is that they are abandoned. This is specific for open data projects as the community is responsible for keeping the project alive, when the community does not put in the required effort the project becomes inactive (Khondhu, Capiluppi, & Stol, 2013).

8.1.2 FAIL FACTORS

In addition to the success factors, the Standish Group also composed a ranked list of fail factors (i.e. factors that challenge the success rate of a project). The list of 2013 is displayed below (The Standish Group, 2013b).

- 1. Lack of user input;
- Incomplete requirements and specifications;
- 3. Changing requirements and specifications;
- 4. Lack of executive support;
- 5. Technology incompetence;

- 6. Lack of resources;
- 7. Unrealistic expectations;
- 8. Unclear objectives;
- 9. Unrealistic time frames;
- 10. New technology.

Only three of these factors are also mentioned during the interviews as fail factors. Factor 7 'Unrealistic expectations' corresponds to the fail factor 'Differences in expectation', which was mentioned twice during the interviews. The factors 'Unclear objectives' and 'Unrealistic time frames' correspond to 'Mismanagement' as mentioned in the interviews.

8.2 SUMMARY

An open data project differs from other projects by the fact that the open data is key for its existence. As not much is written about characteristics of projects, the comparison is mostly aimed at the success and fail factors. There is some overlap between the factors aimed at projects, as found in literature, and the factors for open data projects, found in literature and in the interviews.

As can be seen in Appendix C (p. 108) the success factors of other fields cover some of the success factors of this research, spread over the table. However, the first two success factors of the prioritized list are not covered at all, while they are most mentioned for open data projects.

Considering the phases, the first phase 'Creating data' is only covered by two papers, however, this could be explained as there are only two success factors in this phase. The third phase ('Finding open data') is not covered by any of the papers, this provides a first indication that this category is specific to open data projects. In addition, a lot of factors as mentioned in the paper of Belassi and Tukel (1996) cannot be mapped to the five phases and can be found in 'Other'. The other phases are well covered in all four papers.

It is important to realize that some factors might have been overlooked by the interviewees. In addition, the interviewees were specifically asked for success factors of open data projects, which might have led them to not mention more general factors. Moreover, only four papers are included in this comparison, to grasp the differences. Therefore, it is only a first indication into the differences.

9. EVALUATION SUCCESS FACTORS

In this chapter the success factors will be evaluated. The first step in this evaluation is the creation of a workable list of success factors for the evaluation. The second step of the evaluation is to select projects from the GOOD DATA database. The third step is the actual evaluation, after which a conclusion will be provided.

9.1 CREATING A WORKABLE LIST

In order to create a workable list to perform the evaluation, the prioritized list of success factors (see Table 7 on page 4850) is divided into factors that are externally perceptible and factors that are only internally perceptible. For this research it is only possible to look at the external perceptible factors, as it is not possible to join a project team that is working on an open data project to see whether the internal factors are considered. The same division is made for the fail factors, however, it turned out that those are all not suitable for the evaluation. The externally visible factors are listed below.

- 1. Ecosystem
- 2. Reuse
- 3. Collaboration
- 4. Strategy and legislation
- 5. Continuity, up to date data
- 6. Publicity
- 7. Availability of data
- 8. Availability of metadata

- 9. Market acceptation
- 10. Value creation
- 11. Citizen participation
- 12. Combining data
- 13. Initiated at local level
- 14. Innovative project
- 15. Location of the project

The prioritization of these factors is, again, based on the number of mentions in literature and the interviews. Basically, the only difference with the complete list is that the non-externally perceptible factors are taken out of it.

At first, more factors were selected to be evaluated, however, those factors could not be identified. Factors such as timeliness, accuracy and completeness are completely focused on the data. For the selected projects, this data was not available, therefore, this factor could not be determined. For other factors it would be necessary to interact with users of the application, however, for this thesis it was not possible to do so.

9.2 SELECTING PROJECTS FROM THE DATABASE

In this section, the selection of the projects from the database will be explained.

The database contains 541 projects and 529 unique titles (in the version of February 3rd, 2015). At the moment the number of projects in the database is not presented directly, therefore, a query was needed. The following query is used to retrieve the amount of unique projects in the database:

SELECT COUNT(1) FROM(SELECT COUNT(1), `title` FROM `node` GROUP BY `title`) IN_QUIRY

This query groups duplicates based on the title, therefore the returning number equals the number of unique projects in the database.

Table 11 provides an overview of all fields in the GOOD DATA database. Per field the corresponding table in the database, a description of the field, whether the field is useful for this research, the input, and the number of records is indicated. The number of records sometimes exceeds the number of unique titles, this is because for some fields a project can have multiple values, for example, a project can be available in various languages. In such a case the project occurs more often, resulting in a higher number of records. Moreover, the projects
depicted in Table 10 have multiple appearances in the database, explaining the dissimilarity between the number of projects and number of unique titles.

Project	Number of occurrences
Buienradar	3
FindToilet	2
Funda	2
Health Choices	2
MAGIC	2
Öffentliche WC Wien	2
PetrolPrices	2
See UK	2
Tax Calculator	2
UK Economic Data Dashboard	2
Znasichdani	2

As stated before, Table 11 contains an indication per field whether it can be useful to get that information out of the database. Those fields, which are also listed below, are chosen because they can serve as a first indication for the success factors. When the information from those fields would not be available, it indicates that the people who looked up the projects and filled the database were not able to find the information. It would be very hard to perform an evaluation with the projects for which that information is not easy accessible. Therefore, these factors serve as a first selection criteria. Below are the fields from the database that are used to select suitable projects to be used in the evaluation of the success factors:

- English Title
- Product Type
- Business Type
- Intended effect
- Geographic Level of Application
- Location of use (meso/macro)
- Location of use (micro/local)
- Language
- Year of Publication
- Status

- Pricing model (for user)
- Themes (fixed)
- Topic covered by Data (COFOG)
- More than one data source?
- Source(s) of Data
- Data collecting organization
- Format of data
- Update Frequency
- Geographical level of dataset
- Funding

These fields are extracted into an Excel-file³ to create a more workable solution. This is needed for the selection of projects for this evaluation, as the database itself is not easily editable.

When a field does not have a value, it returns as a NULL value in the database. So, when the location of use for a project is not provided, it will return NULL. If all NULL values are deselected for the above mentioned fields (i.e. only fields which have a value are included), only one project remains for the evaluation: Unlish.

As can be seen in Table 11, format of data has the least number of records. So, these NULL values are first included for the selection of projects. This results in one project as well: Unlish is again the only project. By also including the NULL values of update frequency 47 projects are selected. Update frequency is chosen next as this is the field with the second least number of records.

To make a selection out of these 47 projects, an online random sampling number generator is used⁴. The result of this generator can be found in Figure 8. Each set consists of five projects, where the number indicates which

³ When interested in this Excel-file, please contact the author.

⁴ <u>www.randomizer.org</u>

project to take. After each set it will be determined whether more projects are needed for the evaluation. When no new information is found, the evaluation will come to an end.

5 Sets of 5 Unique Numbers Per Set Range: From 1 to 47 Set #1 p1=40, p2=26, p3=25, p4=3, p5=33 Set #2 p6=39, p7=47, p8=37, p9=27, p10=16 Set #3 p11=18, p12=7, p13=13, p14=41, p15=16 Set #4 p16=32, p17=34, p18=19, p19=33, p20=40 Set #5 p21=44, p22=38, p23=33, p24=45, p25=32

Figure 8 Result random sampling generator

Available Fields	Table	Description	Useful?	Input	Number of
General					records
English Title	field_data_field_english_title	Title of the project in English.	Yes, to know which project it is.	Value	368
Project Link	field_data_field_project_link	Link to the website of the project.	No, the website will be looked up during the evaluation.	Value	536
Description	field_data_field_description	Description of the project.	No, the content will be looked up during the evaluation.	Value	446
Product Type	field_data_field_product_type	Indication of the kind of product.	Yes, might be a success factor.	ID	528
Owner	field_data_field_owner	Owner of the project.	No	Value	469
Developer	field_data_field_developer	Developer of the project.	No	Value	503
Business Type	field_data_field_business_type	Indicates whether the business is commercial, individual, government, NGO, or non-profit.	Yes, might be a success factor.	Value	319
Intended effect	field_data_field_goal	Indication of the intended effect of the project, such as to inform, social societal impact, transparency.	f the intended effect of the Yes, might be a success factor. as to inform, social societal sparency.		272
Geography and					
Language Geographic	field data field geographic lovel	Indicator the level to which the project	Vos indicatos the application lovel of		E27
Level of		applies.	the project.	ID.	332
Application		- F F			
Location of use (meso/macro)	field_data_field_location_of_use_meso_macro	Indicates the country in which the project is used.	Yes, indicates the exact application.	ID	493
Location of use (micro/local)	field_data_fieldlocation_of_use_micro_loc	Indicates a more specified location of use, such as city or region.	Yes, indicates the exact application.	ID	193
Language	field_data_field_language	Main language of the project.	Yes, language might be of influence.	ID	538
<u>Project</u>					
Year of Publication	field_data_field_year_of_publication	The start year of the project.	Yes, indicates the duration of the project.	Value	317
Status	field_data_field_status	Indicates the current status of the project.	Yes, indicates whether the project is still live.	ID	501

Available Fields	Table	Description	Useful?	Input	Number of records
Pricing model (for user)	field_data_field_business_model	Indicates whether the user has to pay for using the project.	Yes, indicates whether the project is freely available.	ID	541
Themes (fixed)	field_data_field_themes_fixes_	21 themes to categorize the projects, for instance Public Spaces.	Yes, might give an indication for fields where the chance at success is higher.	ID	497
Tags	field_data_field_freetags	Tags to categorize the projects, for instance Geographic, Urban planning, Facilities.	No	ID	483
<u>Data</u>					
Topic(s) covered by Data (tags)	field_data_field_nature_of_data	50 topics that are covered with the data in the projects.	No	ID	439
Topic covered by Data (COFOG)	field_data_field_nature_of_data_cofog_	Categorized tags of data coverage.	Yes, provides a categorized overview of topics covered in the project.	ID	539
More than one data source?	field_data_field_more_than_one_data_source_	Indicates whether the projects combines multiple sources.	Yes, that might indicate the contribution of a specific project: combining sources in a unique way.	0/1	541
Source(s) of Data	field_data_field_source_s_of_data	Sources which are used in the project.	Yes, indicates which sources are used.	ID	472
Data collecting organisation	field_data_field_data_collecting_organisati	Organization that collected the used data.	Yes, type of organization might be an indicator.	ID	420
Format of data	field_data_field_format_of_data	Format of the original data used in the project.	Yes, gives an indication of the data is open.	ID	50
Update Frequency	field_data_field_update_frequency	Frequency of updates.	Yes, indicates whether the project is still live, a frequent update might indicate that the project is used a lot.	ID	103
Geographical level of dataset	field_data_field_geographical_level_of_data	The level of the dataset (country, world).	Yes, indicates the level on which the data is collected.	ID	480
Funding	field_data_field_fundrev	Indicates whether there is a funding for the project.	Yes, indicates type of funding and whether there is funding.	ID	534

Table 11 Overview fields GOOD DATA database

9.3 SELECTED PROJECTS FOR EVALUATION

In the previous section the selection method for the projects for evaluation is discussed. In this section, these projects will be evaluated using the success factors acquired from literature and the interviews. The approach of the evaluation is already described in chapter 2.

Table 12 provides an overview of which success factors were found for each project and if they had an influence on the potential success of the project. The explanation for each success factor for each project can be found in Appendix D (p. 111). The fail factors are not included in the evaluation as they are often the inverse of the success factors, or not externally perceptible.

9.4 CONCLUSION OF EVALUATION

After ten projects, all success factors were found, this can be seen in Table 12. Although it is not possible to make firm statements about the success factors, these findings do serve as a first indication that the success factors are important for organizations to think about. Each success factor was found in practice, within a small sample group of only ten projects. In the remainder of this section, each success factor will be discussed briefly.

9.4.1 ECOSYSTEM

The presence of an ecosystem was found for one project: Jeco Guides. For two projects, the presence of an ecosystem could be beneficial to keep the application up to date (i.e. Dog Toilet Map Salzburg and shinyMig OÖ). For the other projects it is not necessary to have an ecosystem surrounding the data, as they are just (static) representations of the dataset, which can be automatically linked to the application.

9.4.2 REUSE

The factor reuse was found for two projects: Vienna City Bike and Dog Toilet Map Salzburg. For those two projects the data is reused in other projects. For three other projects (i.e. An analytical view on data, Jeco Guides, shinyMig OÖ), it is possible to reuse the data of the projects, but there were no indications found for this. For the other projects, the data of the projects themselves are not available, therefore it is impossible to reuse it.

9.4.3 COLLABORATION

Three projects have collaborated to create their application: Vienna City Bike, Jeco Guides, Dog Toilet Map Salzburg. Vienna City Bikes collaborates with other developers to spread their project over multiple operating systems, whereas Dog Toilet Map Salzburg collaborated with developers of another application to make their application suitable for theirs. Jeco Guides works together with its users to fill the application with information. The other projects are set up by one (sometimes two) individual(s) and do not show any indication that they have worked together with other parties.

9.4.4 STRATEGY & LEGISLATION

For four projects some form of legislation is mentioned: Taxi Wien, Wanderwege Wien, Sozialmärkte Wien, shinyMig OÖ. The first three projects mention a privacy policy on their website. In addition, Taxi Wien refers to 'applicable legislations', however, it is not clear which legislations are meant. shinyMig OÖ is the only project that is available under an open data license, which specifies the reuse possibilities. The other projects did not mention any strategy or legislation that might be influencing their project.

	InstaPLAY	Vienna	Hochwasser	An analytical	Тахі	Wanderwege	Jeco	Sozialmärkte	Dog Toilet	shinyMig OÖ
		City Bike	Krems	view on data	Wien	Wien	Guides	Wien	Map Salzburg	
Ecosystem							\checkmark			
Reuse		\checkmark							\checkmark	
Collaboration		√					\checkmark		\checkmark	
Strategy & Legislation					✓	\checkmark		\checkmark		\checkmark
Continuity		\checkmark	\checkmark		✓			✓	\checkmark	
Publicity		\checkmark					\checkmark			
Availability of data	\checkmark	\checkmark								
Availability of metadata							\checkmark			
Market acceptation		\checkmark					\checkmark			
Value creation	\checkmark				\checkmark		\checkmark	\checkmark	\checkmark	
Citizen participation							\checkmark			
Combining data										✓
Initiated at a local level	✓	\checkmark	\checkmark	\checkmark	✓	\checkmark	\checkmark	\checkmark	√	✓
Innovative project		\checkmark			\checkmark		\checkmark			
Location of the project	✓	\checkmark	\checkmark	\checkmark	\checkmark	√	\checkmark	\checkmark	\checkmark	\checkmark

Table 12 Overview of success factors found per project

9.4.5 CONTINUITY

Five projects have up to date data: Vienna City Bike, Hochwasser Krems, Taxi Wien, Sozialmärkte Wien, Dog Toilet Map Salzburg. Most of these projects are up to date as they are automatically linked to their data source, which is up to date. The projects An analytical view on data and shinyMig OÖ are not up to date. For the other three projects, the continuity cannot be determined.

9.4.6 PUBLICITY

For two projects publicity was found: Vienna City Bike, Jeco Guides. For Vienna City Bike, an overview is presented of all press related to the project. Due to a language barrier, the publicity for An analytical view on data could not be determined. The other projects are only presented at a national data portal, more publicity was not found for those projects.

9.4.7 AVAILABILITY OF DATA

For two projects the data was available: InstaPLAY, Vienna City Bike. The data of Vienna City Bike is only available after the developers are contacted. The data of Jeco Guides and Hochwasser Krems are unavailable, where for the other projects only their source data is available.

9.4.8 AVAILABILITY OF METADATA

For one project the metadata was available: Jeco Guides. Since the data itself was only available for two projects, it can be expected that there is not much metadata available. However, it is remarkable that Jeco Guides provides the metadata but not the data itself. For Vienna City Bike it is unclear whether one also gets the metadata after contacting the developers.

9.4.9 MARKET ACCEPTATION

Market acceptation was found for two projects: Vienna City Bike, Jeco Guides. For the other eight applications it was not possible to determine the market acceptation, as there were no figures available concerning the usage of these applications. Moreover, during the evaluations it became clear that the platform for which the application is built might be influencing the market acceptation. Some of the projects were only available for BlackBerry, however, at the moment, the market share of BlackBerry is very low. Therefore, it will not reach a large public.

9.4.10 VALUE CREATION

The factor value creation was found for five projects: InstaPLAY, Taxi Wien, Jeco Guides, Sozialmärkte Wien, Dog Toilet Map Salzburg. For the other five projects, the value creation cannot be determined without involving actual users of the applications.

9.4.11 CITIZEN PARTICIPATION

Citizens are participating in one project: Jeco Guides, which depends on the input from its users. This factor is highly linked to 'Ecosystem'. As stated above, most of the projects are mere visualizations of the data that forms the basis for their project. Since the data is probably automatically linked to the project, there is no need for citizen participation.

9.4.12 COMBINING DATA

One project combines datasets for its creation: shinyMig OÖ. Most projects are representations of their data source, the exception can be found for Jeco Guides, which is based on the input of its users. shinyMig OÖ is a representation of the data as well, however, it combines the data from two datasets to do so.

9.4.13 INITIATED AT A LOCAL LEVEL

All ten projects are initiated at a local level. However, during the evaluation it became clear that if the projects are not spread to a higher level, it restricts the maximum number of users. Moreover, if the project is expanded it can be assumed that another language is added. By adding another language, the application is usable for more people. During the evaluation, almost every project needed to be translated, which poses a barrier for the use of the application.

9.4.14 INNOVATIVE PROJECT

Three projects are considered to be innovative: Vienna City Bike, Taxi Wien, Jeco Guides. For the other projects, they were mostly a representation of the data. Even though shinyMig OÖ combines two datasets, which indicates that it is somewhat innovative, the project is just a representation of the data as well.

9.4.15 LOCATION OF THE PROJECT

For all ten projects, the location of the project was found. As with the factor 'Initiated at a local level', the location can restrict the maximum number of users that might be using the application. Jeco Guides was the only project for which the location of the project itself does not limit the number of users, as people from all over the world can join the project.

9.5 SUMMARY

In this chapter, the database of GOOD DATA is used to select ten projects to evaluate the success factors found in literature and the interviews. As not all factors are externally visible, only fifteen of them were included in the evaluation. The selection of the projects is based on the amount of data available in the database, so the

projects of which the most fields were available were selected. After this first selection a random generator was used to select the ten projects which are eventually used in the evaluation. The results of the evaluation can be found in Table 12.

The evaluation shows that there are some dependencies between the success factors. For instance, all projects are initiated at a local level and are also limited by their location in the maximum number of users they can achieve. Other dependencies are indicated in section 7.2, where the success factors are mapped onto the Open Data Process. "Bestwelsnel.nl, one of the employees of a company that works with our data started crafting. He created a website which depicted the speeds at the roads across the Netherlands. So, you could see the top speed per road. Well, the government was not very pleased with that. So, maybe that is the downside of open data."

Interviewee 14

10.CONCLUSION & DISCUSSION

In this concluding chapter, the conclusion and discussion are presented. In the conclusion, the research questions are answered, whereas the discussion puts a critical view on this thesis and provides insights for future research.

10.1 CONCLUSION

In this conclusion, the main research question will be answered. To be able to do so, the answers of the sub research questions will be provided first.

10.1.1 DEFINING OPEN DATA

The first sub question is 'What is open data?' and is answered by conducting a literature study. As there was not yet a consensus about the definition of open data, a definition is created in this thesis. This definition originates from several sources, including both scientific papers as well as documentation from institutions such as the Knowledge Foundation. The definition is as follows:

Open data is data that is publicly available, easy accessible and has no restrictions for the users.

The foremost characteristic of this definition is that open data is not a type of data, it is about the way the data is disclosed. The data is released in such a way that the public can access it, it is also created for a public purpose. This entails that the data should be easy to find, offered in a format that can be processed with an open-source software tool, moreover, it should be released within open standards, and accompanied by an open license.

During the interviews, the definition is also discussed. All interviewees agreed with the core of the definition.

10.1.2 THE MOMENT HAS COME TO CONSIDER OPEN DATA

To answer the second sub question, which is 'Why should open data be of interest for organizations?', the findings from literature are combined with those from the interviews. The reason of why organizations should look at open data at this moment is considered. Promised benefits of open data, motivations of other organizations, as well as the barriers to move towards open data are discussed. Furthermore, the economic value of open data is considered.

Currently, open data is gaining momentum, and, as interviewee 15 indicates, it is easier to join the movement than to start it. Therefore, this is the moment for organizations to start getting involved with open data, this can either be opening their own data or reusing open data from others. Moreover, the focus is shifting from opening as much data as possible towards realizing the benefits of open data projects. This implicates that it is more beneficial to join the movement now, as more attention will be paid to its benefits.

The benefits are promised to be huge. Despite the cautionary note of interviewee 7, both literature and the interviewees are confident that disclosing data will result in an increase in innovation. Furthermore, the relation between an organization and its customers will intensify, which might result in better products and services. The largest benefit is that of transparency, as the data is open and available for everyone, people will be able to see what is done within an organization. For example, interviewees 1 and 2 would like the data of city councils to become open, to be able to see what is happening within those councils. They indicate that this would be beneficial for democracy. An overview of all (promised) benefits is provided in Table 1 on page 13.

On the other hand, there are still some barriers for organizations to open their data. Interviewee 15 has collected 45 barriers he has heard from organizations. As indicated by Huijboom and Broek (2011) and Vosough (2013), barriers often arise from within the organization. This implies that there is still some work to be done before organizations are convinced they should open their data. However, as indicated in the interviews each organization deals differently with those barriers. Furthermore, in order to remove some of the barriers, it is important that more organizations are disclosing their data, so that they can set an example for other organizations. Moreover, the potential economic value of open data is unknown. The added value of open data entails more than just the economic value that can be achieved, however, for organizations it is unclear what they can gain from open data.

10.1.3 SUCCESSFUL PROJECTS

The first part of the third sub question is 'What defines a successful project?' and is answered using the literature study of the comparison in chapter 7. The Standish Group defines a project successful when "[t]he project is completed on-time and on-budget, with all features and functions as initially specified" (2013b, p. 4). As this requires knowledge of the internal processes of a project, it was not possible to state whether the open data projects used in the evaluation are successful.

Success is also defined by the success factors, which are covered in the second part of this third sub question: 'Which success factors can be distinguished?'. This has been answered using the literature study and the interviews. Combining the success factors found in both activities resulted in a list of 42 success factors and 20 fail factors, which can be found in chapter 7 where a prioritized list of factors is created. For the evaluation of the factors, only those factors that are externally perceptible are included, resulting in a list of fifteen success factors, which is presented below. All factors were found in existing projects. However, even though the list presented here is ranked, the top factor 'Ecosystem' was only found for one project. The factors 'Initiated at a local level' and 'Location of the project' were found for all ten projects. However, this does not provide an indication of importance, as these two factors might be easier to find for the projects that are currently in the GOOD DATA database. The goal of the evaluation was to find indications of the success factors found in literature and the interviews in existing projects, therefore, the prioritizations remain unchanged.

Success factor	Number of mentions [*]	Found in # projects
Ecosystem	10	1
Reuse	7	2
Collaboration	6	3
Strategy and legislation	5	4
Continuity, up to date data	4	5
Publicity	4	2
Availability of data	3	2
Availability of metadata	3	1
Market acceptation	3	2
Value creation	3	5
Citizen participation	2	1
Combining data	2	1
Initiated at local level	2	10
Innovative project	1	3
Location of the project	1	10

Table 13 Evaluated success factors

^{*}The number of mentions is a combination of the mentions found in the literature study and in the interviews, by an equal amount of mentions, the factors are placed in alphabetical order.

10.1.4 CHARACTERISTICS OF OPEN DATA PROJECTS

Sub research question 4 is answered in chapter 8, and is solely based on literature found for that chapter. Unfortunately, not much is written about characteristics of projects in general, therefore, it is difficult to make a comparison of open data projects and other projects. The most important distinction is that open data projects rely on open data, without the open data, the project would not exist.

Regarding the success and fail factors, some overlap of the general factors can be found with the factors found in this research. Some of the factors found in this research are specific for open data projects. These factors include the presence of an ecosystem, even though the general factors do indicate a need for motivated users, the ecosystem has a more extended role as it needs to support the project; availability of data, as the open data project cannot exist without the data; and reuse of the data, as this is an important aspect of open data.

10.1.5 MAIN CONCLUSION

The conclusions to the sub questions lead to the main conclusion, which answers the main research question. The main research question within this thesis is:

'What are success factors of open data projects?'

From sub questions 1 and 4 can be learned that open data projects are dependent on the open data, which is defined as 'data that is publicly available, easy accessible and has no restrictions for the users'.

Even though it is clearly stated that this is the moment for organizations to get involved with open data, the answer to sub question 2 is not convincing. There are still some barriers to overcome and it is not clear what can be gained from open data, the list of success factors might be helpful for organizations to have an indication of what they should focus on.

Currently, the list of success factors contains 42 factors, they can be found in Table 7 on page 50. When the success factors as well as the fail factors are compared to success and fail factors for generic projects, some overlap can be found. Unfortunately, not much has been written about the subject, therefore, it is not possible to make a sound comparison. Some of the success factors are specific for open data projects, however, as there were only a few sources used in the comparison, also the overlapping factors are taken into account for the evaluation. It was not possible to validate all success factors, as not all factors are externally perceptible. Therefore, a workable list of fifteen factors is created for the evaluation. However, the answer to the main research question is the complete list of success factors, with the notion that only fifteen of them are used in the evaluation and were all found in existing open data projects. These fifteen factors are indicated in the list below by displaying them in bold and italic. A prioritization has been made for the success factors based on the number of mentions in literature and in the interviews.

The 42 success factors for open data projects are mapped onto the Open Data Process. The ranking of the success factors is taken into account by the order in which they are presented. The success factors as mapped onto the Open Data Process can be found in Figure 9 on the next page.



10.2 DISCUSSION

Within this research an overview of success factors for open data projects is provided. By doing so, the fact that open data is a rising phenomenon is underlined.

10.2.1 CONTRIBUTION OF THIS THESIS

First of all, the contribution of this thesis is the clear definition of open data, as this was not available before. The scientific contribution of this thesis is that this is the first research into success factors of open data projects, as stated in the introduction (p. 1), such an overview of success factors was lacking. This research can in its turn be used by organizations to see what barriers and motivations exist. Even though, it is still unclear what organizations can gain from open data, this research in combination with the GOOD DATA database shows that open data is gaining momentum and there are a lot of projects already based on open data. Furthermore, the potentials of open data are discussed and the success factors can be used as a guidance for organizations which are opening their data. The success factors are evaluated by looking into existing open data projects. By looking into just ten projects, all fifteen factors were found. This provides organizations with a first indication of what they have to think of when getting involved with open data.

10.2.2 LIMITATIONS

LITERATURE

As there was no literature available about success factors and open data projects, a lot of the information, as presented in chapter 4, is induced from the literature. Unfortunately, a lot of the literature does not originate from peer-reviewed journals, which influences the quality of the used literature. Many papers come from conferences or public organizations. This indicates that open data is still in its beginning state. The papers that

can be found in journals are mainly about open data policies, however, this is a limited number of papers. However, the number of papers used in this thesis and their year of publication indicate that open data is getting increasingly popular.

INTERVIEWS

By choosing to use expert interviews to validate and complement the findings from literature, the findings become more grounded. However, the use of such interviews also implies some limitations. First of all, the ability of the researcher to conduct an interview is of influence on (the quality of) the results. As the researcher did not have much experience in conducting interviews, this threat is reduced by the creation of an interview protocol, which was checked and improved by one of the supervisors of this thesis.

Another pitfall in using expert interviews is that it might turn out that the chosen expert is not the right expert (Flick, 2009). In the beginning of the interviews, this did occur as it turned out that some of the interviewees did not work with open data and only knew a little about it. However, this did result in other insights. To overcome this in the remaining interviews, the (first) interviewees were asked who would be interesting to interview for this research. The persons that were mentioned were looked up by the researcher to get a first indication of their expertise.

The last limitation of the expert interviews is that of time pressure (Flick, 2009). Within the interviews, only one interviewee mentioned that he only had 50 minutes for the interview. The other interviewees had calculated more time for the interview, resulting in no time pressure during the interviews.

GOOD DATA DATABASE

The database of the GOOD DATA project is an initiative to provide an overview of open data projects. However, at the moment, the database does not succeed in its purpose. First of all, not all projects in the database are open data projects. An example is provided by multiple interviewees, as they correctly pointed out that Buienradar is not an open data project, as it is not created with open data and without the open data, the project would still exist. For Buienradar, the researcher knows that this is not an open data project as it is a familiar project, however, for other projects, it is not possible to tell by just looking at the title of the project and the information that is provided in the database. Moreover, the goal of the database is to collect open data projects that originate within Europe. However, some (i.e. 46) of the projects in the database are from the United States.

Another remark towards the database is the fact that some projects have multiple occurrences in the database. As can be seen in Table 10, eleven projects have multiple occurrences. It is unclear why these multiple occurrences are present in the database. It might be due to the fact that two people are responsible for the projects in the database. This has also influenced the trustworthiness of the database, as they have made some interpretations for the database. For instance, based on their estimations the update frequency is determined. As they have indicated themselves this is prone to interpretation. They even stated that for some projects they just filled in a value, as they were not sure about what to fill in precisely. For instance, when asked about when a project is frequently updated, they did not know what to answer. In addition, for everything that they were not able to find, either a NULL value appears in the database or the field is empty, which makes it hard to make a selection to exclude fields with no value. At the moment, only one project does not have a NULL value, this indicates that there are too many empty fields in the database.

The data in the database is also not up to date. This was already indicated in the beginning of the project, as this is one of the main struggles for the database: 'how do you keep the database up to date'? Moreover, the two persons who have filled the database with projects indicate that they were not able to select all options necessary. For instance, they were unable to select the language Turkish, so for projects that are available in

Turkish, the information is not complete. This will be mainly covered when the database will become open and people will be actively contributing to the database. The database will be supported by those people, who (ideally) will keep it up to date. However, regarding this thesis project it means that the data on which the projects for evaluation are selected might be incorrect.

On the positive side, the interviewees indicated that they are willing to use the database as long as it is nicely structured. Within the GOOD DATA database they have used the COFOG classification scheme. COFOG stands for 'Classification of the Functions of Government' and is used as a standard for "classifying the purposes of government activities" (Vries et al., 2011, p. 8). By using this classification, the database adheres to classification standards used in the public domain, making the database easier to use. Within the standard classifications of COFOG they defined lower level tags to extend the classification and increase the searchability of the database.

10.2.3 FUTURE RESEARCH

As there was no research conducted into success factors of open data projects, this research is just the beginning. It leads to indications of what is important, but also to the notion that open data is indeed gaining momentum and the success factors resulting from this research are a first step into more thorough research into open data. The gained knowledge can be used in further research.

For future research it would be advisable to conduct the interviews with two persons. Even though the interviews are recorded and therefore, all information is taken into account during the analysis, by doing the interviews together, both interviewers can come up with questions, which might result in more findings. Furthermore, it might be interested to involve more interviewees, as they would provide a more complete view on the present state of open data. The interviewees for this research already showed some diversity, as some are at the driver-side of open data, where others are more or less forced to join the movement. Still, it would be interesting to involve even more parties, such as individuals working with open data, employees of organizations who are anti open data, and someone involved in the creation of European or American legislation.

Regarding the success factors, they need to be validated in real life practices. In this research, only fifteen of them are used in the evaluation. However, this is not enough to be able to make statements about their importance. It would be interesting when multiple projects would be used to test the complete list. This includes both projects that are already up and running as well as projects that are in their startup phase. In order to be able to do so it is necessary to get in touch with the initiators of the project, for the new projects it would be nice to be able to participate in the project. In this way, it is also possible to indicate which of the success factors is doable for organizations and what the practical implications are of those factors.

REFERENCES

- Ashja, M., Hadizadeh Moghadam, A., & Bidram, H. (2013). Comparative study of large information systems' CSFs during their life cycle. *Information Systems Frontiers*, *17*(3), 619–628. doi:10.1007/s10796-013-9445-6
- Atsu, M., Andoh-Baidoo, F. K., Osatuyi, B., & Amoako-Gyampah, K. (2009). Examining the Success Factors of ICT Projects in Developing Nations : A Case Study of AB Networks. In *AMCIS 2009 Proceedings* (pp. 1–10). San Francisco, CA.
- Attila, M., Michel, A., & Blegind, J. T. (2013). Reframing open big data. In *ECIS 2013 Proceedings of the 21st European Conference on Information Systems* (p. 12). Copenhagen Business School, Frederiksberg, Denmark. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-84905842153&partnerID=40&md5=3fb7ebbb7a943ff24183e3187ac525bf
- Auer, S., Bizer, C., Kobilarov, G., & Lehmann, J. (2007). *Dbpedia: A nucleus for a web of open data*. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-76298-0_52
- Bachmann, A., & Bernstein, A. (2009). Software process data quality and characteristics: a historical view on open and closed source projects. In *Proceedings of the joint international and annual ERCIM workshops on Principles of software evolution (IWPSE) and software evolution (Evol) workshops IWPSE-Evol '09* (p. 119). Amsterdam. Retrieved from http://dl.acm.org/citation.cfm?id=1595808.1595830
- Bates. (2012). The Politics of Open Government Data: a Neo-gramscian Analysis of the United Kingdom's Open Government Data Inititive.
- Bates, J. (2014). The strategic importance of information policy for the contemporary neoliberal state: The case of Open Government Data in the United Kingdom. *Government Information Quarterly*, *31*(3), 388–395. doi:10.1016/j.giq.2014.02.009
- Bauer, F., & Kaltenböck, M. (2011). Linked Open Data: The Essentials. *Edition Mono/monochrom, Vienna*. Retrieved from http://bls.buu.ac.th/~f55361/08Aug22/LOD/EBFFFd01.pdf
- Belassi, W., & Tukel, O. I. (1996). A new framework for determining critical success. *International Journal of Project Management*, *14*(3), 141–151.
- Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data-the story so far. *International Journal on Semantic Web and Information Systems*. Retrieved from http://eprints.soton.ac.uk/271285/
- Böhm, C., Schmidt, M., Freitag, M., Heise, A., Lehmann, C., Mascher, A., ... Haase, P. (2012). GovWILD: Integrating Open Government Data for Transparency Christoph. *Proceedings of the 21st International Conference Companion on World Wide Web - WWW '12 Companion, 321–324.* doi:10.1145/2187980.2188039
- Bonina, C. M. (2013). New business models and the value of open data: definitions, challenges and opportunities. London.
- Borges, M. R. S., De Faria Cordeiro, K., Campos, M. L. M., & Marino, T. (2011). Linked open data and the design of information infrastructure for emergency management systems. In 8th International Conference on Information Systems for Crisis Response and Management: From Early-Warning Systems to Preparedness and Training, ISCRAM 2011. Graduate Program in Informatics, Universidade Federal Do Rio de Janeiro, Brazil. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-84905662212&partnerID=40&md5=45396d5ab21d3a0f624655379e8a7e83

- Capgemini Consulting. (2013). The Open Data Economy: Unlocking Economic Value by Opening Government and Public Data.
- Chernoff, M. (2010). What "open data" means and what it doesn't. Retrieved from http://opensource.com/government/10/12/what-"open-data"-means---and-what-it-doesn't
- Chui, M., Farrell, D., & Kuiken, S. van. (2013). Generating Economic Value through Open Data. In B. Goldstein & L. Dyson (Eds.), *Beyond Transparency: Open Data and the Future of Civic Innovation* (pp. 163–173). San Francisco, CA: Code for America Press.
- Conradie, P., & Choenni, S. (2012). Exploring process barriers to release public sector information in local government. In *Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance ICEGOV '12* (p. 5). New York, New York, USA: ACM Press. doi:10.1145/2463728.2463731
- Conradie, P., & Choenni, S. (2014). On the barriers for local government releasing open data. *Government Information Quarterly*, *31*, S10–S17. doi:10.1016/j.giq.2014.01.003
- Cooke-Davies, T. (2002). The "real" success factors on projects. *International Journal of Project Management*, 20, 185–190. doi:10.1016/S0263-7863(01)00067-9
- Cucciniello, M., Nasi, G., & Valotti, G. (2011). Assessing transparency in government: Rhetoric, reality and desire. In *Proceedings of the Annual Hawaii International Conference on System Sciences* (pp. 2451–2461). doi:10.1109/HICSS.2012.123
- Davenport, T., & Prusak, L. (1998). Working Knowledge: How Organizations Manage What They Know. *Ubiquity*, 1–15.
- Davies, T. (2011). Open Data : Infrastructures and ecosystems.
- Davies, T. (2014). Open Data Policies and Practice: An International Comparison. *Available at SSRN 2492520*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2492520
- Davies, T., & Edwards, D. (2012). Emerging Implications of Open and Linked Data for Knowledge Sharing in Development. *IDS Bulletin*, 43(5), 117–127. doi:10.1111/j.1759-5436.2012.00372.x
- Dekkers, M., Polman, F., te Velde, R., & de Vries, M. (2006). *MEPSIR Measuring European Public Sector Information Resources, Part 1*. Retrieved from http://epractice.eu/files/media/media2575.pdf
- Ding, L., Peristeras, V., & Hausenblas, M. (2012). Linked Open Government Data. *Intelligent Systems, IEEE*, (January 2010), 11–15. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6237454
- European Commision. (2011). Open data: An engine for innovation, growth and transparent governance. Brussels.
- European Commission. (2000). Europe's public sector information.
- European Commission. (2010). Riding the wave: How Europe can gain form the rising tide of scientific data. Final report of the high Levlel Expert Group on Scientific Data.
- European Parliament, & European Council. (2013). Directive 2013|37|EU. *Official Journal of the European Union*, 175(8), 1–8.
- Evans, A., & Campos, A. (2013). Open Government Initiatives: Challenges of Citizen Participation. *Journal of Policy Analysis and Management*, 32(1), 172–203. doi:10.1002/pam

- Ferro, E., & Osella, M. (2012). Business Models for PSI Re-Use: A Multidimensional Framework. In Workshop "Using Open Data: Policy Modeling, Citizen Participation" (pp. 1–5). Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Business+Models+for+PSI+Re-Use+:+A+Multidimensional+Framework#0
- Fioretti, M. (2010). Open Data, Open Society. Project about openness of public data in EU. Retrieved from http://www.dime-eu.org/files/active/0/ODOS_report_1.pdf
- Flick, U. (2009). An Introduction to Qualitative Research (4th ed.). London: SAGE Publications Ltd.
- Gurstein, M. (2011). Open Data: Empowering the empowered or effective data use for everyone? *First Monday*, *16*(2), 1–7. Retrieved from http://journals.uic.edu/ojs/index.php/fm/article/view/3316/2764
- Hall, W., Shadbolt, N., Tiropanis, T., O'Hara, K., & Davies, T. (2012). Open data and charities. Oxford.
- Harrell, C. (2013). The Beginning of a Beautiful Friendship: Data and Design in Innovative Citizen Experiences. In
 B. Goldstein & L. Dyson (Eds.), *Beyond Transparency: Open Data and the Future of Civic Innovation* (pp. 151–163). San Francisco, CA: Code for America Press.

Heimstädt, M. (2014). The British Open Data Ecosystem. University of St. Andrews.

- Hogge, B. (2010). Open data study. *Transparency & Accountability Initiative*. Retrieved from http://www.transparency-initiative.org/wp-content/uploads/2011/05/open_data_study_final1.pdf
- Huijboom, N., & Broek, T. Van Den. (2011). Open data: an international comparison of strategies. European Journal of ePractice, (April), 1–13. Retrieved from http://www.epractice.eu/files/European Journal epractice Volume 12_1.pdf
- lemma, R. (2012). Open Government Data: a focus on key economic and organizational drivers.
- Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, *29*(4), 258–268. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-84867251336&partnerID=40&md5=b0f5814bd1a88c3c15f3b5c1141b5180
- Janssen, M., & Zuiderwijk, A. (2012). Open data and transformational government (pp. 1–7). Brunel University, United Kingdom.
- Jetzek, T., Avital, M., & Bjørn-Andersen, N. (2013). Generating value from open government data. In International Conference on Information Systems (ICIS 2013): Reshaping Society Through Information Systems Design (Vol. 2, pp. 1737–1756). Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-84897679692&partnerID=40&md5=5cac1d6137ac8136bfa6657f99cb0c5a
- Kalampokis, E., Hausenblas, M., & Tarabanis, K. (2011). Combining social and government open data for participatory decision-making. *Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 6847 LNCS,* 36–47. doi:10.1007/978-3-642-23333-3_4
- Kaschesky, M., & Selmi, L. (2014). 7R Data Value Framework for Open Data in Practice: Fusepool. *Future Internet*, 6(3), 556–584. doi:10.3390/fi6030556
- Kassen, M. (2013). A promising phenomenon of open data: A case study of the Chicago open data project. *Government Information Quarterly*, 30(4), 508–513. doi:10.1016/j.giq.2013.05.012

Khondhu, J., Capiluppi, A., & Stol, K. (2013). Is It All Lost? A Study of Inactive Open Source Projects. In *Proceedings of the 9th International Conference on Open Source Systems* (pp. 1–19).

Kulk, S., Loenen, B. van, & Ploeger, H. D. (2012). Open data and beyond.

Lapsi Network. (2010). What is PSI? Retrieved November 25, 2014, from http://www.lapsi-project.eu/what#3

Lee, D., Cyganiak, R., & Decker, S. (2014). Open Data Ireland: Best Practice Handbook.

Leeuw, M. De, & Lemmer, T. (2011). Hergebruik en Open Data. The Hague, The Netherlands.

- Linked Data Community. (2014). Linked Data Connect Distributed Data across the Web. Retrieved November 19, 2014, from http://linkeddata.org/home
- Manyika, J., Chui, M., Groves, P., Farrell, D., Van Kuiken, S., & Doshi, E. A. (2013). Open Data: Unlocking Innovation and Performance with Liquid Information. McKinsey. doi:http://www.mckinsey.com/~/media/McKinsey/dotcom/Insights/Business%20Technology/Open%20d ata%20Unlocking%20innovation%20and%20performance%20with%20liquid%20information/MGI_Open_ data_Full_report_Oct_2013.ashx
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*, *339*, 13. Retrieved from http://www.bmj.com/content/339/bmj.b2535.abstract
- Möller, K., Hausenblas, M., Cyganiak, R., & Handschuh, S. (2010). Learning from Linked Open Data Usage: Patterns & Metrics (pp. 1–8). Raleigh, NC, USA.
- Murray-Rust, P. (2013). Open Data in Science. *Serials Review*, 34(1), 52–64. doi:10.1080/00987913.2008.10765152
- Neuroni, A. C., Riedl, R., & Brugger, J. (2013). Swiss executive authorities on Open Government Data Policy making beyond transparency and participation. In *Proceedings of the Annual Hawaii International Conference on System Sciences* (pp. 1911–1920). doi:10.1109/HICSS.2013.19
- Nilsen, K. (2010). Economic theory as it applies to public sector information. *Annual Review of Information Science and Technology*, 44, 419–489. doi:10.1002/aris.2010.1440440117
- Obama, B. (2013). Executive Order -- Making Open and Machine Readable the New Default for Government Information. Retrieved May 11, 2015, from https://www.whitehouse.gov/the-pressoffice/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government-
- Oliver, M., Palacin, M., Domingo, A., & Valls, V. (2012). Sensor information fueling Open Data. *Proceedings* -*International Computer Software and Applications Conference*, 116–121. doi:10.1109/COMPSACW.2012.31
- Open Data Research Network. (2013). *Open Data Barometer: 2013 Global Report*. Retrieved from http://www.opendataresearch.org/project/2013/odb
- Open Knowledge Foundation. (2014). Open Definition. Retrieved December 3, 2014, from http://opendefinition.org/od/
- Peled, A. (2011). When transparency and collaboration collide: The USA Open Data program. *Journal of the American Society for Information Science and Technology*, *62*(11), 2085–2094. doi:10.1002/asi.21622

- Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, 45(4), 211. doi:10.1145/505248.506010
- Pollock, R. (2006). The Value of the Public Domain, (July). Retrieved from http://www.dspace.cam.ac.uk/handle/1810/242072
- Pollock, R. (2009). The Economics of Public Sector Information. Search. Retrieved from http://www.dspace.cam.ac.uk/handle/1810/229487
- Rijksoverheid. (2014). Wat is open data? Retrieved December 3, 2014, from https://data.overheid.nl/wat-isopen-data-concept
- Roy, J. (2014). Open Data and Open Governance in Canada: A Critical Examination of New Opportunities and Old Tensions. *Future Internet*, *6*(3), 414–432. doi:10.3390/fi6030414
- Shadbolt, N., O'Hara, K., Berners-Lee, T., Gibbins, N., Glaser, H., Hall, W., & Shraefel, M. C. (2012). Linked open government data: Lessons from data. gov. uk. *IEEE Intelligent ..., 12,* 16–24. Retrieved from http://eprints.soton.ac.uk/340564/
- Shenhar, A. J. (2001). One Size Does Not Fit All Projects: Exploring Classical Contingency Domains. *Management Science*, 47(3), 394–414. doi:10.1287/mnsc.47.3.394.9772
- Socrata. (2014). Why does my organization need open data? *Open Data Field Guide*. Retrieved December 23, 2014, from http://www.socrata.com/open-data-field-guide-chapter/why-does-my-organization-need-open-data/
- Streeter, L. A., Kraut, R. E., & Caby, L. (1996). How Open Data Networks Influence Business Performance and Market Structure. *Communications of the ACM*, *39*(7), 62–73.
- STT Netherlands. (2002). *Dealing with the data flood*. (J. Meij, Ed.) (First.). The Hague, The Netherlands: VNU Publishers.

The European Thematic Network on Legal Aspects of Public Sector Information. (2010). What is PSI.

The Standish Group. (2009). CHAOS Summary 2009. Chaos.

The Standish Group. (2013a). CHAOS 2013: Think Big, Act Small. The Standish Group International. Retrieved from http://www.standishgroup.com

The Standish Group. (2013b). The Standish Group Report - Chaos, 16.

- Tongco, M. D. C. (2007). Purposive sampling as a tool for informant selection. *Ethnobotany Research and Applications*, *5*, 147–158.
- Ubaldi, B. (2013). Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives. *OECD Working Papers on Public Governance*, *NO.22*(22), 61. doi:10.1787/5k46bj4f03s7-en
- Verhulst, S., Noveck, B. S., Caplan, R., Brown, K., & Paz, C. (2014). *The Open Data Era in Health and Social Care*. Retrieved from http://images.thegovlab.org/wordpress/wp-content/uploads/2014/06/nhs-full-report.pdf

Vosough, P. (2013). Implementing an open data system and showing its benefits.

Vickery, G. (2011). Review of recent studies on PSI re-use and related market developments. Paris.

- Vries, M. De, Kapff, L., Achiaga, M. N., Wauters, P., Osimo, D., Foley, P., ... Whitehouse, D. (2011). European Commission Information Society and Media Directorate-General, 90.
- W3C. (2013). Linked Data. Retrieved from http://www.w3.org/standards/semanticweb/data
- W3C. (2014). Linking Open Data. Retrieved from http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData
- Wang, R. W., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal* of Management Information Systems, 12(4), 5. doi:10.2307/40398176
- Weiss, P. (2002). Borders in Cyberspace : Conflicting Public Sector Information Policies and their Economic Impacts. *Federal Register*, 6428, 1–92. Retrieved from http://www.weather.gov/sp/Borders_report.pdf
- Wohlin, C., Runeson, P., Höst, M., Ohlsson, M., Regnell, B., & Wesslén, A. (2012). Case Studies. In *Experimentation in Software Engineering* (pp. 68–69). Springer Heidelberg. doi:10.1007/978-3-642-29044-2
- Wright, G., Prakash, P., Abraham, S., & Shah, N. (2010). Open government data study: India.
- Zuiderwijk, A., & Janssen, M. (2014). Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly*, *31*(1), 17–29. doi:10.1016/j.giq.2013.04.003
- Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., & Alibaks, R. S. (2012). Socio-technical impediments of open data. *Electronic Journal of E-Government*, 10(2), 156–172. Retrieved from http://www.ejeg.com/issue/download.html?idArticle=255

APPENDICES

In the Appendices the interview questions and summaries of the interviews are presented.

APPENDIX A: INTERVIEW QUESTIONS

For the purpose of this thesis, the questions are translated into English. For the original questions, please contact the author.

INTRODUCTION

- I. Introductory round
- II. In what way are open data and open data projects embroiled in your daily activities?

GENERAL

- 1. Why are you interested in open data?
- 2. What do you think of when you think of open data and open data projects?
- 3. What do you consider as open data?
- 4. What is your definition of open data?

DEFINITION:

Open data is data that is publicly available, easy accessible and has no restrictions for the users.

//DEPENDENT ON INTERVIEWEE

5.

CONTEXT OF OPEN DATA

- 6. What kind of project, based on open data, would you like to be initiated?
- 7. How do you stay up to date regarding open data?
- 8. Do you feel a need for more overview, as the GOOD DATAdatabase, in which open data projects are indexed?

SPECIFIC QUESTIONS

- 9. What is success for you?
- 10. What are examples of success stories of open data projects?
- 11. When does a project fail?
- 12. Do you have an example of an open data project that failed?
- 13. In what way would you measure the success of open data projects?
- 14. What factors are involved by the success of an open data project?
- 15. What factors are involved by the failure of an open data project?

Do you have any tips regarding my research? Or people that are interesting for me to interview?

APPENDIX B: INTERVIEW SUMMARIES

In this Appendix, the summaries of the interviews are presented.

B.1 SUMMARY INTERVIEWEE 1

- IT consultant
- Project manager on a secondment basis
- Starting a new, specifically aimed at data, company
- Based on data service provision, focused mainly on journalism, but also SMEs do not possess enough knowledge to provide in the data needs of customers. Becoming intermediary party to enable businesses to answer within a week to the data needs of the customer.
- Help people to find what they need in the enormous amount of available data. Because that is tricky, you can have 400 files, but it is hard to see what is precisely in there.
- It takes special skills to be able to disclose the data that you are looking for. This is mostly acquired by experience, think about the data that you got, what do you know about that data, and read the meta description of the provider carefully. Subsequently, you work with the data to discover its quality.
- Open data mainly as a hobbyist.
- By opening data the job of an data journalist becomes easier, as more data and sources are available. It will be easier for such a journalist to look critically at the statements made by politicians.
- Interested in all data. Until three or four years ago it was problematic to get some data from the government, you needed a WOB-request or you happen to know someone who could help you. However by opening the data, and when this happens in a structured way, existing questions can be answered quicker and easier.
- The publication of data happens at such a high pace that it is hard to determine which data is interesting.
- First project that comes to mind is the open data of RDW and KNMI.
- By opening the data it is possible to really dive into the data and increase the depth of the analysis. Moreover, it is possible to make a comparison between neighborhoods or even within a neighborhood.
- It is a challenge to map the enormous amount of data and provide the correct context for that data.
 - <u>Example</u>: In crime statistics, if car thefts increase 1% in one year, is that a bad sign? When it increases 1% in value, but the fleet rises 2% in that year, than the chance of car theft decreases. It illustrates that you need even more data to give it the correct context.
- It is interesting to find patterns in the dataset that were not part of the reason to collect the data. For instance, by collecting the data from pharmacies you can deduce patterns as that people buy more lipstick on Thursdays. Doing so, you disclose a commercial significance. Another example was that you could see that the standard package contains 25 pills, but the prescription is almost always for 21 pills. This means a lot of work for the pharmacists as they have to cut the right amount of pills. Therefore, it could save money and time to change the standard packages to 21 pills.
- <u>Definition open data</u>: everything that is put in the public domain in a structured way, without restrictions regarding government or businesses.
- <u>Reaction definition:</u> Easy accessible is a nice addition. I was considering that element, personally I refer to it as usable. You have to present your data in the most accessible, usable way.
- <u>Successful open data project</u>: In logistics they use on board systems and planning software to calculate the most efficient route for trucks. Nowadays, the planning software is based on road data (the amount of traffic present on a specific road) and weather forecasts (as the probability of rain influences the duration of a trip). By using both those open data sources, the software is able to adapt

the routes dynamically. This is of strategic importance for logistical service providers as they will be able to work more efficiently.

- When designing new processes within organizations or new products, it is important to see how can you use open data or any form of data in the creating process to be more clever or quicker than your competitor.
- It is for large organizations easier to adapt open data, as they are already used to making use of large amounts of data.
- <u>Value of open data</u>: being able to offer the data in a usable way is the profit of open data, as that will be the business model of my new organization. However, the value of open data cannot be seen as just in monetary terms, the democratic value is also there.
- Open data from the government increases the transparency of the democratic process. It provides the citizens, who will elect the people, with a more equivalent information position. Making it possible for those citizens to decide more informed.
- Rather slow and better usable than fast, too much, and useless.
- Governments do a lot of things/measurements out of legal obligations, those data is also being opened. However, the findability is pretty bad, as it is not possible to find those measurements by searching for the law that obligated them to do the measurements in the first place. So, you need a lot of extra knowledge to know which search terms you have to use.
- This can be a barrier for a lot of people, as it takes a lot of time. Which is a shame as a lot of data is left unexamined.
- <u>Which project would you start:</u> complete voting behavior of all city councils in open data; to see what their voting behavior is, but also to provide citizens more insight in how they deal with local issues. So that the citizen can see whether their party voted for or against a particular motion. They can see what the influence of their vote is, and where it actually ends up.
- This is mainly interesting, because interviewee 1 thinks there is a democratic deficit at the level of municipalities. This is caused by the decreasing interest of journalists.
- Another project would be the mapping of permits, such as planning permissions. This is hard as each municipality has its own way of handling permits.
- <u>How do you stay up to date with open data news:</u> Twitter, open data portal of the government (check 2-3 times a week whether there are new files), professional journals (Automatiseringsgids), websites (WebWereld), once a while visit a conference.
- <u>Need for a database that provides an overview of (active) open data projects:</u> preferably a news site that highlights the most interesting stories. A database might lead to an information overload. The database does serve as a means to find out whether your idea for a new project is already being done by somebody else.
- The most interesting stories are the ones with a high impact on the society, that prove to be useful although that was not the expectation. So, the realized potential is unexpectedly large.
 - <u>Example</u>: In England, the purchasing data of hospitals are published. Some retired purchasers created their own database and started to provide unsolicited recommendations about among other things the toilet paper. This resulted in enormous savings, as they found a supplier which offers the same toilet paper for a lower price. This is an unforeseen effect of opening the data.
- <u>What is project success for you</u>: when the deliverable (i.e. product or service) is useful for the customer. It has to contribute to the function that the customer serves.
 - \circ $\;$ $\;$ For a company this means that they have to be able to make a profit.
 - For a journalist this means it has to lead to a publication.
 - \circ \quad For government it might entail the adjustment of policy.

- <u>Why do you consider it a success</u>: it is both beneficial for the environment and for the organization. The success becomes bigger as multiple factors are combined, for example when it is a success from a societal perspective as well as from an economic perspective.
- <u>When do you consider it a failure:</u> if the effort put into the project is disproportionate to the outcome of the project. When there is no result at all, the project is failed, but it can also happen that halfway down you realize that the yield is too small and that it is better to terminate the project. When it is not a commercial project, you can only determine whether it is a failure by comparing it to other (similar) projects.
- <u>Measurement of success by open data projects:</u> factors that you can score on. So, is it beneficial for the democracy on a scale from 0 to 10, et cetera. It serves as an indication whether something is good or not.
- <u>Success factors:</u>
 - Usefulness and cooperation of the supplying party.
 - Continuity; you need the guarantee that the data stays up to data, consistent and relevant over time.
- Fail factors:
 - Wrong team members.
 - Not possessing the right knowledge; you have to have knowledge of the domain you are working in.
 - Erroneous communication.
 - Differences in expectations.
 - Unrealistic expectations at the beginning.
 - Changes in legislation and regulations.
 - Knowledge drain: what if the person with all the knowledge leaves?
- <u>Example of a failed project</u>: watstemtmijnraad.nl, failed because they did not have any money left when they had indexed only 16 municipalities. So, they had other expectations, they expected more enthusiastic responses.

B.2 SUMMARY INTERVIEWEE 2

- Project manager at Open State Foundation
 - Goal is to promote democratic transparency and participation with digital means. Ensuring that the government becomes more transparent, releasing data as open data, organize projects to make sure that the data is actually used. Improving the interaction between government and citizens.
- Former employee of ProDemos (creator of Stemwijzer)
- We organize masterclasses for institutions to promote the idea of open data and make their data online available as open data.
- For another project, I was the lobbyist, to actually go to the municipalities and explain why it is important to open that data. In divers ways I try to create support for transparency.
- We try to make all data at municipality websites more structured, machine readable, so people can find it in a structured way, but also that developers can use an api to create a product on that data.
- Decision making should be more controlled by citizens, a decision will be better when the people which are subject of the decision know what it is about, the decision moments are up to them.
- Well informed citizens lead to a better government and a better society.
- Governments do a lot of things with data and information, using our tax money. So we have a right to
 see what they do with it. There are instances known that two ministries asked one research bureau to
 do a research for them, while they did not know that the other ministry also asked them to do so. So
 the bureau gets to do one research for which it gets paid twice. This would not happen if all gates

were open, which would in this case save tax money. And if the results are shared with the rest of society, there might be some economic yield.

- <u>Example</u>: Buienradar results in a lot of money. That is only possible thanks to the Royal Netherlands Meteorological Institute (KNMI). At first they thought, opening the data will cost us a lot on yearly basis, but now Buienradar and all the other apps built on that data result in so much money, that it multiplies the tax money it takes for KNMI and the resulting tax money (because of those new businesses) with a factor 100.
- I am convinced that opening data will result in economic yield for our society, not just democratic yield, but also economic yield.
- A project such as Buienradar is interesting for a lot of people, but I am more interested in the right you have as a citizen to see into the democratic decision process.
- The current disinterest in politics can be resolved partly by opening data, as it will be more present in the media, however, I cannot imagine that it will increase the turnout for the elections. Whilst an app as Buienradar is used by almost everyone. This also makes it a really good example to use when explaining open data to people.
- <u>Definition open data</u>: Open data is information that is financed, and gathered, completely or partly with public means. Subsequently, this is released in such a way that it has no restrictions on the reuse, so no obligations to pay, no login screen. Preferably in an open standard as defined by Berners-Lee five star model. Also in such a way that it is machine readable, ensuring that developers can reach it easily, automatic readout. So the general definition with the public means and those three criteria.
- <u>Example</u>: With the municipal elections people often vote for the front runner, however, if it would be possible to see from each member what his or her actions were the past 4 years, you could see which member matches your preferences best.
- We try to stimulate the actual reuse of data by organizing hackatons, app challenges and so forth. We do this also to show public organizations that their data is really used and that it is applied in innovative ways.
- When data is opened, people still need a little boost to actually use the data. That is what we try to do by hosting those challenges.
- It is important that the data is well structured, otherwise it acts as a barrier for people to start using the data, as it takes them too much effort to be able to use it.
- <u>Thoughts over current policy</u>: A lot of talking and only a few actions. Put your money where your mouth is. In general, the Ministry of the Interior and Kingdom Relations (BZK) created a fine policy regarding open data. They even created a group within the ministry which only concern is open data, although they mostly focus on networking. However, at the end of the day it is very hard to get funding for a project. It takes a lot of time to acquire the money needed for a project.

The policy concerning open data may become slightly more severe, by forcing public organizations to supply a data register. It should become more proactive instead of reactive. Data should be opened as a standard, in that way when an information request comes along, the data is already open. In addition, it forces the municipalities to actively think about open data.

- It is important to train the municipalities in their open data use, even the little ones. At the moment
 there is a lot of incomprehension and fear of opening data at the municipalities. Therefore, it would be
 very helpful if the municipalities would get a training from BZK on how to open their data and deal
 with information requests. This will be necessary to enable a culture change and get over the fear of
 opening. However, it is not up to the municipalities to train reusers, the municipalities only need to be
 open for feedback.
- The Open Barometer is really unreliable as each government and organization has its own agenda. When governments have not opened their data, the activists may benefit from a low score as it is a signal to the government to do something about it. Whilst governments that have opened a lot of data

want everyone to say that they are doing awesome, even though they might have just opened the data and nothing more.

- On European level, the focus is too much on linked open data, we are not there yet. Let us first focus on opening the data, linking the data is a next step.
- Opening the data can result in less venation as the government is more transparent by opening its data.
- <u>Which project would you start</u>: Open Council information. For me, it is important to make the actions of the city council publicly available, not just their voting behavior, but also their planning. This would benefit the local democracy.
- <u>How do you stay up to date with open data news:</u> people within the organization are often up to date, conferences, Twitter, magazines, reports, I like to read about it but I need to have time left to actually do something.
- It is not a danger that too much data is opened, as everyone has its own niche, so they know what they have to search for. As long as it is categorized properly, it will not be a problem. There is always a little risk that it will be too much, however, entrepreneurs often know what data set they need, so it will probably work out just fine. As long as the data is presented in a structured way, easy to find, and on a centralized location.
- At the moment a lot of links provided via data.overheid.nl are broken. So that should be replaced by a central register where you have categories that facilitate the search process.
- Developers need to be able to trust that the data they use is up to date.
- <u>Need for a database that provides an overview of (active) open data projects</u>: Such a thing already exists at open-overheid.nl, where the top cases are highlighted. But such a portal would be useful to see what is already being done, making sure things are not done twice, moreover, it can be used to contact developers to start a collaboration. It could also function as a central point, from which all other information can be found.
- It is a success when three things are accomplished.
 - 1. Create support for transparency using open data;
 - 2. Open data that was not open before;
 - 3. Stimulate reuse.
- <u>Example of a failed project:</u> ArtsHolland was a project to present touristic information about a museum, but also the route to that museum. Millions are used for that project from the Ministry of Economic Affairs. It cost a lot of money, while there were almost no results. Another example failed because it was thwarted by copyright law.
- The focus should not be on the business models, but on the opening of data itself.
- Fail factors
 - Momentum.
 - Putting words into action, sometimes a lot of time is wasted by discussing what can be done, writing an advice, and then no one looks at the advice and the project dies.
 - Copyright issues.
 - Mismanagement.
 - Fear towards opening data.
 - A lot of IT projects within government fail, governments should think more like private owned organizations, to reduce the number of failures.
- The economic value of open data is realized by the fact that people disclose data, which can be used by other people, which leads to new products and services. These products and services are in their turn the basis for other products and services.

B.3 SUMMARY INTERVIEWEE 3

- Information advisor, making sure that all employees of the municipality can access the information needed for their job.
- Project leader of the project open data within the municipality. In the long-term, open data should be a way of working instead of being a project.
- The municipality is opening data for selected areas. They created a plan of action in which they state that they focus at first on the data which are most often requested via WOB-requests. Other criteria were whether it is easy to realize (so, what can be opened?), what is available, what is topic of discussions right now. Aim to open all data that can be opened according to the law.
- <u>What is the motivation to open the data:</u> That is a combination of motivations. For me, the main reason is transparency. Transparency can enhance the trust of citizens in government, as the citizen can observe and check what the government is doing. This can also ensure that the government pays more attention to the way it is acting. This might decrease the number of instances in which the citizen does not understand a specific action of the government, as the government's actions are more closely watched. Others name the possibilities to economic added value. I talked mainly with app developers who want to work with our data.
- <u>Economic added value</u>: Organizations create value by connecting value to data, that data itself is free, because it is already paid for. It can be combined, present in a different way, unlocking its value, making it available at places in which it is not available right now.
 - Zie hiervoor ook <u>http://www.lelystad.nl/Docs/Actieplan%20open%20gemeente.pdf</u>
 "Efficiëntere informatie-uitwisseling met collega's en met ketenpartners. Veel tijd gaat nu verloren omdat informatie slecht vindbaar is, de kwaliteit of actualiteit niet duidelijk is dus moet worden nagevraagd, of omdat het gegevensformaat niet past bij de toepassing van de gebruiker. Hiermee zou al gauw 1% op de totale uren van de organisatie bespaard kunnen worden, mogelijk nog veel meer.

Afname van het aantal WOB-verzoeken. Als door proactieve publicatie het aantal WOBverzoeken met een kwart zou afnemen, van 120 naar 90 per jaar, kan naar schatting 300 uur per jaar gespaard worden. (1 WOB-verzoek kost tussen 1 en 50 uur - bron: FB) Besparingen door verbeterde gegevenskwaliteit of door constructieve participatie door burgers. Deze zijn op zich reëel haalbaar maar inschattingen zouden speculatief zijn."

- <u>Definition open data</u>: Taking away as much barriers as possible for reuse, that can be done in various ways. Such as making sure that you do not have to pay for it, ensuring that it is easy to find, present it in a structure that is usable (for example the 5 star model of Berners-Lee).
- <u>Role of municipalities:</u> They have to do their best to present their data as open as possible. Citizens already paid for the data, so they have to open it and should not make a fuss about it. If municipalities have their data stored in a structured way in which everything is classified, it does not take much time to open the data. However, this is not the reality, which makes it hard to go further with open data. However, this does not mean that we should focus on structuring our data before we focus on opening the data, as this would only slow down the opening process. We should use the opening of data as a means to get more structured data.
- <u>Thoughts over current open data policy</u>: It only states that if a citizen requests certain data, that you are obliged to provide that data, but it does not state that you have to actively disclose the data. However, currently a lot of municipalities and other governmental organizations are actively working on opening data. So it is not necessary to include this in a policy. It might be easier to use agreements and if that is not working, then try to facilitate it from the central government and as a last resource change the policies. It also works if you see others doing and succeeding in it, it works as an extra motivation.

- <u>Example</u>: If every municipality would register the trees that appear in the municipality and there is no standard way in doing so, then some will use Latin names, other Dutch, which makes it very hard to combine the data.
- <u>Which project would you start</u>: I would create something regarding sustainability and the environment. Like a website to compare different municipalities on their energy usage. Or something to see whether a policy has reached the intended effect. For instance, you can create new bike lanes, but does that really increase the number of bikers? It would be nice to be able to show such a thing. Or something else like the spending of municipalities, where does the money go, is that to organizations surrounding or close by the municipality or is the money going even further?
- How do you stay up to date with open data news: Twitter, Google, sometimes visit a conference.
- <u>Need for a database that provides an overview of (active) open data projects:</u> I would like to see who is using our data. However, I would definitely use it, also to contact reusers to see whether they are satisfied with the data and are interested in other datasets.
- Some projects have a really insipid business model, in which you have to pay when you want more information. For example the first 10 are free and then you have to pay for each extra data point you want access to. It is really insipid, why don't they just open up their complete database.
- Success factors:
 - Increase of trust in government, which can be measured using a survey.
 - Reuse of the data, that the data is used in several applications.
 - Supports citizen participation.
- It is important that a discussion arises, that discussion between government and citizens improve by making it about facts instead of feelings.
- Fail factors:
 - When nobody looks at the data, meaning that it is pointless to publish the data.
- We release datasets based on WOB-requests, but I also release datasets based on what I think is an interesting dataset.
- <u>Example successful project</u>: Basic registration large-scale topography (BGT), a lot of municipalities are in the middle of that process, releasing datasets. Everything is released in the same way and the same format, and is collected in a national receptacle, from which it is published as a web service. It is useful for citizens, but also for internal use within municipalities as it simplifies our architecture, since all data is available at that receptacle.
 - Open Streetmap is also a nice project, resulting in very detailed locations, however the BGT is more precise.
 - The earthquake project in Groningen is also a nice example. They used data from the KNMI, persuaded the KNMI to release their data in an open format, and even found a structural error in the data of the KNMI.

B.4 SUMMARY INTERVIEWEE 4

- I encountered the problem that no one had an overview of timetables.
- I wanted to access the national road file (NWB) of Rijkswaterstaat, however it was not available for reuse. After 7 years of trials and waiting I was the first individual to gain access to that file. In those 7 years I started to look for projects that were collecting that data, to be able to access that data quicker, that is when I encountered OpenStreetMap. I started with collecting data myself by walking around with a GPS and informed by organizations whether bus stops locations were available. So, I ended up at public transportation.
- OpenStreetMap is a map that is acquired by people just as Wikipedia does as encyclopedia.
- When I wanted to access the data of 9292, I had to pay for it. They were not interested in working together to provide a better service to the customer allow others to reuse timetables for their own

projects. I went to the Ministry of Infrastructure and the Environment and they showed me that if I set up travel information system I could demand transit information from all transport organizations. So, openOV was born.

- OpenKvK is another project in which the government possesses the information, but does not want to release it. So I made a copy of the data to open it.
- <u>Example</u>: When you have a long road and another road is connected to that road in the middle, a new point arises. In all systems this was a new point what would stay there even though the road would be removed. My model describes a situation where the type of road changes, but no extra data points are added.
- I see a need and act upon it to get that data. I focus on disclosing the data, then I leave it to others to commercialise or extend the project.
- By creating openOV, existing organizations also start to innovate, because the market is innovating. Until then the market could not innovate as all parties were at the same level.
- <u>Definition open data</u>: I use the strict definition of opendefinition.org. That means that open data does not have to be free, but it should not create barriers. However, if it is possible to make it free, it should be free. Personally I try to get a creative commons zero weaver for my projects. That entails that one is not limited to a certain use. The attribution clause, using a reference to the data source, is the only limitation that is allowable.
- <u>Reaction definition</u>: In that definition there is nothing about the guarantee towards the user. You can consider open data as is, so you get what is offered. However, you could also say that if you get governmental data you should be able to trust the correctness of that data. This means that you should get a guarantee that the data is correct, from there on it is up to you what you want to do with the data.
- Both openOV and openKvK have had an impact. That is important to me, that the establishment realizes that there is a force that is there to stay. They either do it themselves or they take into account that there is another party that is willing to do so.
- The most important impact is that it should have value for the end users. So that they get something new that was not available beforehand.
- Personally, I like to combine datasets to create new insights.
- By doing so organizations realize that not all data needs to be bought, a lot of the things can be done by themselves. That is another form of knowledge sharing.
- For openOV much more parties were interested to join than we expected. In fact, we are allowed to calculate €1000,- per participant, which would return in €92.000,- for the current 92 participants (May 20 2015). So in potential it has a lot of economic value.
- To disclose the true economic value of open data you need to go to micro level. By doing measurements on a micro level, which will become significant later on in the process by combining them in a project on a higher level. In the future the most rewarding projects are the ones for which no measurement is needed, but in which the data can be used as a given fact.
- <u>Example:</u> Measuring the temperature of all houses, which leads to knowing all ground temperatures. Providing you with a lot more measuring stations, which can be used for the weather predictions.
- <u>Which project would you start:</u> Energy is very interesting, like smart devices. At the moment this does not exist, because they cannot measure when there is a shortage of capacity. According to the energy companies people are not interested in such a thing as they think no one will look at that to determine whether to turn on their washing machine or wait for another hour.
- <u>How do you stay up to date with open data news</u>: This is quite an issue with open data. I make a lot of use of my network, I get a lot of proposals that people want me to read, LinkedIn, and I collect a lot of that for fun for myself to send to others.
- <u>Need for a database that provides an overview of (active) open data projects:</u> Maybe as a sort of pr activity. It is interesting if you manage to create an overview of which data is used by whom and in

which way. It needs to work in such a way that when you search you will actually find what you are looking for. However, it would be very nice to see who is using openOV for instance. We do not know about all projects based on it, that would be nice to know. I would definitely use it and supplement the data about my projects.

- <u>Definition of success</u>: Reuse. That your data is being reused in other projects, that would be my biggest criteria.
- <u>Example of success project</u>: BAG. Within the government files are linked, municipalities have worked together, it is done within a reasonable period, and it uses a national control point from which everything is orchestrated.
- <u>Success factors:</u>
 - Impact of reuse.
 - Ecosystem that arises around the project, this can consist out of feedback, but also out of money.
- Economic value is something that originates from a successful project and the success factors mentioned before.
- <u>Definition of failure:</u> When it is not possible to link it with current data. Or when data is not ready to use, it requires another party to edit the data. Although this other party indicates that there is some economic value, it also indicate the failure of the data source. When the released data does not correspond to the needs. When data is just published, there will be no ecosystem, which was one of the success factors. Most importantly, when data is sold as open data, but is not really open.
- <u>How to measure success</u>: The biggest issue arises when data is used in a combination. So you should take a look at a dataset and award points for certain characteristics, such as number of views. However, often that information is not publicly available. The startup costs or costs it takes to create such a project versus the impact it has on users. You can use additional information such as, what did it take, why is it possible to create another project on top of this one.

B.5 SUMMARY INTERVIEWEE 5

- Researcher for my own research agency, mainly focused on socio-economic research and if it is possible I try to apply a geographical information system (GIS).
- I don't necessarily make a difference between data that is truly open and data that is easy accessible. However, it would be nice if all data was open.
- <u>Definition of open data:</u> Accessible via internet, for everyone, free.
- In the Netherlands there are a lot of data sources, so there is always some information available that you can use for a research. However, in countries such as Sierra Leone the information that is available is limited. A source as OpenStreetMap is very useful in such situations, provided that they have an internet connection.
- <u>Example:</u> In Sierra Leone we help the local aid organizations to map investments of western organizations in roads by using a GIS. We learn them how to deal with GPS et cetera. The project is based on OpenStreetMap. Organizations as the Red Cross and Medecins Sans Frontieres are also using OpenStreetMap to coordinate their help. They go even further by establishing a project called Missing Maps in which suggestions can be made to complement OpenStreetMap. Although OpenStreetMap is based on people driving around with GPS equipment, the Missing Maps project uses satellite pictures to map an area. In such areas, project like OpenStreetMap are great, as the standard information is not detailed and is often erroneous.
- <u>Reaction definition</u>: I agree with that, you could add to it that, especially when it concerns geographical information, it should meet certain standards, to make it more usable and better structured.

- It can be a barrier as data sources are not well structured, or not in an easy to use format, as it will take extra effort to be able to use those data.
- If more data becomes available, it will benefit the democracy. If all municipalities increase their transparency, the government becomes more auditable, which will benefit the democracy.
- <u>Added value of open data</u>: In the context of developing countries the added value is quite clear. In more developed countries it can function as a replenishment.
- I don't think that the average citizen is going to do something with open data, it is more meant for other governments or commercial organizations.
- It will take some time for all data to become open as there are commercial parties which are now bringing the data together in between, as they have to make money from that.
- You should establish some kind of protocol to ensure that the data that is open is correct. Although that is not a necessity as other users will step in and correct the data if needed. In the end you are just satisfied with the fact that the data is available, that it might contain some errors is not a major issue, by taking a critical look at the data you will spot faulty data.
- <u>What project would you start:</u> I would like to have the employment information available for everyone on organizational level. Then you can make an indication of the concentration of a specific sector in a particular area. By combining this information with a GIS, you can create nice overviews.
- A lot of stakeholders are involved with open data, all with their own interests.
- <u>How do you stay up to date with open data news:</u> I just search for datasets and if they are open it is a nice bonus. When you need specific information, you look at the current status, but it is just what you happen to see and find.
- <u>Need for a database that provides an overview of (active) open data projects</u>: It would be useful to keep up to date with the active projects. It will allow you to find other sources that might be useful for you in a particular research. Now you use Google for that, but you get lost in all the search results and just check a few of them. I think it should work with references, so on the website of OpenStreetMap they refer you to that website and vice versa.
- As long as you know what you are looking for, you won't get drowned in the enormous amount of data.
- Success factors:
 - The customer needs to be satisfied.
 - The reusers are satisfied, or better served.
- <u>Example success project</u>: OpenStreetMap, because it is worldwide, and everywhere it is the same. It might not be complete for all places, but the infrastructure is the same. Moreover, the data is correct.
- Fail factors:
 - Erroneous data, it is important that data is checked before it is opened to the world.
 - Incomplete data, for instance in Sierra Leone villages have different names; one village can have five different names due to language differences. If this is not complete, it is a challenge for aid organizations to coordinate their actions, especially in crisis situations.
- Projects don't fail because of the data, it is because of the actions performed with or on that data.

B.6 SUMMARY INTERVIEWEE 6

- Entrepreneur, involved in a lot of projects.
- <u>Example</u>: ZoekendGOED, this is a project in which vacancy rates are collected. We are now in the process of finding out what is possible with the data that we can find.
- <u>Example</u>: Crop-R.com, in this project, the founder maps crop fields to facilitate the exchange of data between farmers. They all have a common interest, which is to minimize the number of diseases in their crops. Even though they don't share their profit, they all benefit from a good harvest. By mapping those data you get more precise predictions, but also a better overview of historical data.

- You need people who act as brokers to connect the different datasets of various people.
- When you have enough data, you are able to pinpoint it to one person, so there is an enormous privacy issue at stake here. For instance, based on the pace at which you type on a keyboard is unique. So if you know the pace, you can figure out who is typing.
- <u>Definition open data</u>: Data that is acquired by the government and opened towards the society. However, you can also acquire the data yourself and open it for other people to work with it. Not all organizations are willing to open their data
- Data is a means to achieve your business model, not a business model itself. However, it is an important component.
- <u>What project would you start:</u> It would be interesting to take a look in some health data, to see how certain diseases progress. In Groningen you have the lifelines thing, which means that people are born in Schuppekutteveen and they die in Schuppekutteveen. That means that they can look into generations of people that lived there. It would be very interesting to have access to that data.
- <u>Need for a database that provides an overview of (active) open data projects</u>: Yes that would be very useful. If you know your way with data, it will be very helpful. However, I do see an issue, as a lot of innovative organizations are unwilling to share their ideas with others. And now you learn about projects because you happen to be present at some occasion, it would be a shame if you would miss out on things because you are not there.
- It can be a danger that some parties will acquire too much knowledge by which they will become a danger towards others.
- <u>How to measure success</u>: You should define open data as a KPI in your project, so you are able to tell what the added value of open data within your project is.
- Success factors:
 - Market acceptation.
 - Everything depends on your customer. A business model is a flexible given fact, and it should stay flexible, it should be adapted for each customer, each situation.
- Fail factors:
 - Too focused, by being too focused you lose track of the right path. You cannot think of all the things you will encounter up on front.
- But I don't think you fail completely, I believe in fail for war. From every failure you learn something.
 - Example success project:
 - o OpenKvK
 - o Crops-R
 - Openspending, because governmental organizations are pressed to release their data.

B.7 SUMMARY INTERVIEWEE 7

- Entrepreneur.
- Because of inside information what was possible in the United States with weather data, I got interested in weather data.
- Started Buienradar in 2006, if the merger in the previous year had happened, Buienradar would not exist. After selling it in 2011 to a commercial party. 2014 was the year that it was once again allowed to start working with weather data again.
- The main difference between them and their competition is that they look at the end product from a customer perspective. Trying to perfect the data to make sure the customer has the best data available. To achieve this, multiple radars are combined. By using also the radars from the United Kingdom, it is possible to give a more accurate prediction for a longer period. Compared to some competitors, who use prediction data, their models are based on actual data. However, this is not open data.

- On the one hand, it is an advantage of open data, that it is available and ready to use. On the other hand, the disadvantage is that it might lead to mistakes in interpretation.
- I started from the notion that everyone wants to have an actual overview of the weather, so I started looking where can I find that data and how reliable is it.
- Luck has also been a factor in the success of Buienradar.
- Try to communicate with the intended end user, by listening closely to his needs, the product can be optimized.
- <u>Definition open data</u>: Freely accessible data without restrictions. That you are not forced to pay for the data itself, you could pay for the delivering, or infrastructure, of the data, but not for the data itself.
- <u>Success of Buienradar</u>: The momentum was right. We had the courage to make a big investment, as there was no open data yet. Nowadays part of the data is openly available, not all of it. We were convinced of the product we developed.
- Open data does not lead to more innovative projects, the only project based on open data that I think is really good, is OpenOV. All the other projects are more of the same thing, nothing is really innovative. I think that if someone is creative, he will look for the data and if he really wants to go for it, it does not matter whether it is open or not.
- The European legislation that data should be open is a joke, because when you look at weather data, a lot of it should have been opened by now, but instead I see all the prices rise and for a great part only delayed data is "open".
- If an organization opens its data, it will get a lot more competition, so why would they open it?
- When you look at the United Kingdom, they say they are very mature in opening the data, but the data that is opened, is not the data that you want to use, as it is not up to date for real time services.
- It is dangerous that app users always assume that the data behind the app is correct and contains the right information.
- For some data, the human interpretation is very important. So being completely machine readable is not always a good thing.
- <u>What project would you start:</u> It would be nice to plot the crowds
- Open data that is truly missing is data from telecom providers, for instance Vodafone and TomTom use the open data of NDW, but they do not open their data, and that is a shame.
- The added value can be found in combining data in unique possibilities and the interpretation of that combination, not in the data itself.
- Barriers can be found in organizations that are unwilling to open their data and in current privacy discussions.
- <u>Need for a database that provides an overview of (active) open data projects:</u> Yes that would be nice, but I don't think it will happen. I think that all stories will be a global description, I don't think companies are going to tell you their detailed plan.
- I think the metadata is correct, but that is exactly why it is all so technical. The explanation of how it should be interpreted is missing. Therefore, you cannot be sure that the data that you get to see as an end user is based on the correct interpretation of the source data.
- So one of the dangers is that opening the data empowers the media to garble the interpretation, and because the data is openly available, no one will correct them.
- <u>Success factors:</u>
 - Knowledge of the data. Although it is not a necessity.
 - Current data, a lot of successful projects contain a real-time component.
 - Combining data. That allows you to filter the data, although the danger of real-time data is that you don't have the time to filter the measurement errors out of it.
- Fail factors:
 - Reliability of delivery of the data. If your data source has a lot of down moments, your product loses quality.

• Apply nuance to the data.

B.8 SUMMARY INTERVIEWEE 8

- It is very nice what can be done with data. You can discover links between a person and an organization. In one case they asked if we could find a link between Person X and a motor club. And indeed we found a link. Although it stays a chasing game as they will always find a new way to commit fraud.
- The possibilities of data are unknown. But I guess they are enormous. We are moving towards a data based internet, which unlocks an innumerable amount of possibilities.
- We are just now discovering the value of combining datasets. It helped us to reveal an enormous scheme that was going on in the Netherlands in which people pretend to have a right to subsidy, which they did not. This costs us a lot of money, and by combining datasets we were able to reveal this scheme.
- I think data should be open by definition, unless there are ground reasons not to. We should grow towards a new culture in which giving is the new acquiring and sharing is the new having. In which you do not ask yourself what do I gain from this, because you know someone will return the favor.
- <u>Definition open data</u>: Availability. Being able to get to the data, although there are gradations within availability. These gradations are set by the licenses, of which I am a fan of creative commons. For me it is open if I can reach it, so if this is within the government, it is not open towards everyone, but for me it is. This is also were the five star classification of Berners Lee comes in. For me it is all about the first star, availability.
- <u>Reaction definition:</u> You certainly have a point, but for me it is open if I can use it.
- There is always a context problem, the meaning for one system is different from the meaning of another system. Although the systems might think they understand each other, they don't.
- At the moment, you are already very happy if you have access to data. Whether it is polluted, or your interpretation is erroneous is of minor importance. However, interpretation is extremely important in situations where a decision is based on one data entry, by combining a lot of databases, it is important that all databases understand each other's meaning, so an address could be different in all systems. The context determines the interpretation.
- On the internet, we know that if you tweet something, it is just a stupid tweet of Anniek. But when data is used for serious business, we still think that everything has one reality.
- Privacy is an issue that we cannot control within the revolution of data (which is: data can be used for everything). Legislations concerning privacy are not dealing with the combination of data. So I am allowed to see your income and I am also allowed to know the worth of someone's house, but it does not say whether you are allowed to know the combination of those two.
- Even though it is allowed by law, possible by technology, that does not mean you also have to do it. And that last part is very important when dealing with data. At Belastingdienst we all share a lot of norms and values. However, we do not yet share a carefulness with the use of data: "I can use it, I am allowed to use it but still, will I use it?" This is a question we need to ask ourselves everytime we use data.
- The rationale behind the system of key registers is very nice, linking all data using single source, however that is not what is being done at the moment. Wat niet gedaan wordt is multirealiteit: wat waar is voor jou, is niet automatisch waar voor mij. Op het internet vinden we multirealiteit normaal, maar niet voor data. De subjectiviteit (of context-gevoeligheid) van data moet onderdeel worden van het Stelsel zonder dat het Stelsel wordt geblokkeerd.
- <u>What project would you start</u>: Poverty alleviation. Firstly, to be able to see that someone is heading towards poverty and try to stop that process. Secondly, if we are handling it, that we are handling it

together. We don't know which data we will be using for this purpose, we should just start hacking and see what that brings us.

- I started to link our data with data of other governmental organizations, because I think there is a lot of added value to be achieved.
- <u>Example success project</u>: First of all, the BAG (key register of addresses and buildings), without it we would not have been able to catch those fraudsters. Another good example is the one of the firemen in Amsterdam, they use the energy labels of houses to fight fire. Each label has other implications on the type of fire they might encounter.
- It shows that you don't know all applications that can come from your data, so just open your data and see what happens. Even though you will not know what applications, they will be there.
- <u>Success factors</u>: Availability of data. Some also discuss the format of the data, but you can fix that yourself, the availability is what matters.
- <u>Fail factors</u>: There is only one and that concerns the database law, as that prohibits you from using the data and creating something new.

B.9 SUMMARY INTERVIEWEE 9

- Part of management of ministry of Infrastructure and the Environment (including spatial planning), involved in coordination center open data.
- Contact within the ministry to make sure we comply to the European guidelines.
- <u>Definition open data</u>: All structured data collected by public sector bodies (such as ministries) that is provided to the public for reuse without any costs, according to the CC zero license, with only in a single case a reference. It is open, freely available, machine readable, possible to acquire at once, and it is up to the market to make something out of it i.e. to generate business with it.
- In response to governmental organizations who are having cold feet, we just say that if the data is good enough for you to perform your public task, it is good enough for reuse by commercial enterprises. The few lawsuits that were filed against governmental organizations regarding to liability are lost by the organizations that started them.
- <u>Reaction definition:</u> Also for reuse, you could add that. Indeed with maximum provision costs.
- <u>Example</u>: The lung fund has patients who are troubled by air pollution. So it would be useful for those patients to know whether it is safe going outside. If it is possible for citizens to measure the air quality, the number of measure stations will increases enormously and patients can decide whether they can go outside or not. By providing information to these patients, they are better prepared, and are likely to visit the doctor less often. Thus, it leads to a decrease in costs for society as a whole.
- By opening and using open data, the whole city becomes smarter, based on the fact that citizens get more informed, hence smarter.
- <u>Thoughts current policy</u>: The European guidelines are perhaps too weak in certain aspects, as it states that provision costs may be charged. They are also not clear about the references. Both the costs and the references are not necessary according to me. However, with this amount of EU-member states it is likely that the EU-guideline on open data were a compromise and leaves some room for interpretation. The policy as it is now provides a helping hand for municipalities, so a stricter policy is not necessary for the first years.
- It is important that organizations change their perspective, it is not a bad thing to receive feedback on your data, as it allows you to improve your data. You could even arrange a helpdesk that handles the feedback. By spreading examples like the RDW, you create more understanding and acceptance of open data.
- <u>What project would you start:</u> Visualization of roads and buildings for firemen. When they set out for a fire, they will be able to see what the situation is on the spot, with surrounding buildings, access roads etc.

- <u>Example success project</u>: It is not a good sign that I really have to think about this question. Well, I was interested in the theft rate in my neighborhood, so I asked some people to take a look at that. And it turned out in a very nice overview of the number of thefts in my neighborhood. That is very nice, because I could not get that information from the police, but for a data project they were willing to provide the data. The application was made in less than a day.
- <u>Success factors</u>: People and their networks. It all starts with enthusiasts, by taking the initiative, bringing organizations together, a community originates.
- <u>Fail factors:</u> The approach of skepticism, being afraid that people will sue you, the cold feet attitude. But also, not thinking about your target group, if your target group consists of people over 60, how likely is it that they will be able to use a smartphone or a tablet. It is important that we don't forget about the elderly in society, because we cannot exclude them by using techniques that are not understood by them.
- <u>Example failed project</u>: A laborious project was a project that was initiated to allow elderly people to live longer in their own house instead of moving to elderly homes. The issue has so many aspects and involves a delicate group in difficult times that solutions to make it work were hard to find. I am sure it will need a multi discipline approach to tackle such an issue.
- Projects don't fail because of the data, they fail because of external factors.

B.10 SUMMARY INTERVIEWEE 10

- Open data coordinator at municipality.
- Data driven steering, being able to make decisions based on data with short feedback loops, so you don't have to wait four years before a change can be made based on policy theory. For instance, they based their policy on the assumption that the number of lampposts is correlated to the number of intrusions. However, this turned out to be a wrong assumption. By applying data driven steering it is possible to act upon it.
- It is not only about data, it is also about organizational change, changing the mindset of people, so that they understand what data can do for them.
- Reusers appreciate it better when there is one large pool from which they can pick their datasets, than when they have to go to several websites to gather the datasets.
- At first our goal is to become smarter en more efficient, that will result in a cheaper way of working.
- Before you can know what gold you potentially have and where it is hidden, you need to map your data landscape.
- It is questionable whether quality should be a parameter for determining whether a dataset should be opened.
- <u>Definition open data</u>: Data that is appropriate to share with the public with the goal to give meaning to the city.
- <u>Example success project</u>: We released a dataset about the locations of voting booths, someone created an app based on that data, in which you could see where the nearest voting booth was. And it was actually used during the elections.
- <u>Reaction definition:</u> I think the most important aspect is that it is useful and value adding for the city.
- Open data provide the building blocks to create something new and innovative. The mindset needs to be changed, the data does not belong to one person or one organization, it belongs to all of us. So, we should realize that it should be opened in order to really become building blocks.
- I feel like it is beneficial that I am the open data person within the organization, people know that if they have something with open data, they need to come to me. Then, I will help them with their issue or question, but it is very useful that we have one central person for open data issues.
- We open our datasets from our own perspective, we look at datasets and determine whether it can be opened. We also look within other organizations which have data about Utrecht to see whether those could be opened.
- Opening data contributes to trust in the government, but also transparency of the government. Especially since we don't want it to be an internal party, we want the city to be involved in open data. So that will increase participation.
- By creating a community we can show that it is beneficial to open your data. We are collaborating with other municipalities that are in the process of opening their data, not just to learn from each other, but also to show other municipalities that we are working on it.
- <u>Example:</u> I want to create the same effect as when you are at primary school. You see some kids playing football and you think, hey I also want to play, because it seems fun.
- <u>Thoughts about current policy</u>: The collaborations that are in place are fine, however, we could all put in some extra effort to find the togetherness. The policy is fine, but we should apply a more hands on execution.
- Currently there are a lot of broken links at the national open data portal.
- <u>What project would you start:</u> I would create an app in which all buildings in Utrecht are visualized on a map with additional information like visiting hours. By doing so it is easier for people to decide whether they want to do something, which reinforces consumer behavior. I want to get rid of the feeling of never mind (Dutch: laat maar).
- <u>Need for a database that provides an overview of (active) open data projects</u>: Yes, especially since we want the city to do something with our data, it would be very nice to be able to see what is being made. But I don't think that the database should be created and facilitated by a municipality, it is up to the community to do so.
- <u>Success factors:</u>
 - Knowing what you have, mapping your data landscape.
 - Knowledge sharing with other municipalities, ministries, and other parties.
 - Hearing new insights and doing something with those insights.
 - One big pool in which all data is collected, as it will take too much time for reusers to go to all separate portals.
 - When it has added value for the city.
 - \circ $\,$ Learning during the process. And being brave enough to act upon the things that you have learnt.
 - Involving the city.
- Fail factors:
 - \circ $\$ If the business would have resistance against opening the data.
 - When we would lose the communication with, collaboration with, and networks within the city.
 - When there is no reuse, although that does not mean it failed from our perspective, but I guess it failed as a whole.
 - When an impasse occurs and we are no longer able to work together with other parties.
- <u>How to measure success</u>: One thing is to make a distinction between data that you have identified and data that you have not identified. You can even make another distinction between different parts of the organization. You could measure the number of interventions that are based on open data and the number of interventions that would have took place anyway. The number of opened datasets and the number of successful apps that emerge from it. For us it is mainly about making the data available and the inventory of data. Maybe the number of partners is an indication.

B.11 SUMMARY INTERVIEWEE 11

- Master Business & IT
- Working at the one hand with data, at the other hand making data more insightful and understandable. At the one hand I work as a web developer, at the other hand as an information analyst.
- The amount of data is increasing, unfortunately I see that it is not used that much, as it is often not insightful. And that is a shame, I try to fix that.
- Open data is a nice transition of the last years. Making more data available, increases the chances of data being used. It also contributes to the focus of the data releaser. At this point we don't know which data is out there, so we cannot start linking the data.
- <u>Definition open data</u>: The provision of data, that exist within an organization, for a larger audience. And this is always general data, so no critical business data. No data that might disrupt the market or brings privacy at stake.
- <u>Reaction definition</u>: At data.overheid.nl you also have a list with characteristics to determine how open a dataset is. So a pdf file gets a lower score than an xml file. Another important principle is that as a data discloser you should not make assumptions about the level of interpretation of the data user. That is very hard for organizations as they often like to think that only they know what to do with the data, resulting in already processed data that is being opened instead of raw data. Sometimes this is for the best, because it increases the possibility that someone is going to use the data. I would like to see a combination, so organizations that already process parts of their data, but also the raw datasets, so users can choose what to use.
- By opening data more people will take a look at the data, leading to individuals that see possibilities to link your data with another dataset.
- The use of standards will increase the usability of open data. For instance, if you have a datasets which involves municipalities, it would be nice if you use the same abbreviations or codes as for example the CBS does.
- <u>Motivation for zorgkostenopdekaart.nl:</u> I saw the data and just wanted to do something with it. On the other hand, I knew that I was going to start my own company and this could be a nice project of reference.
- I do consider it as a pretty successful project. I got a lot of responses and I even got a new assignment out of it. This is one of the added values of open data.
- It is not an issue of open data, but for the whole internet culture that some things are misinterpreted.
- Within organizations it takes a while for people to be convinced of open data.
- Just opening datasets is not the way to go. The datasets really have to be useful for someone.
- <u>What project would you start:</u> You have to make an appointment at the city hall, but still you have to wait for your turn. It would be nice if that data would be open. So you could see the average waiting times, and do something with that data. This will also be a boost for the processes of the municipality itself.
- <u>What project would you start:</u> Something regarding mobility. To see how bikers are moving through the city. And maybe optimize the traffic lights, so that people are willing to wait for the green light and will not ignore all traffic lights anymore.
- <u>Need for a database that provides an overview of (active) open data projects:</u> It would be nice to have such an overview. So it is possible to see what is being done with particular datasets. As a reuser it is nice to give the feedback that you are actually using the data, that is an extra motivation for the data discloser to keep working on their open data. At this point it is nice to take a look at the database to see who is working with open data. It is also a stimulant to start working together on a project.
- <u>How do you stay up to date:</u> Twitter, I like to see what new projects are developed.

- <u>Example success project</u>: The project in Groningen, regarding the earthquakes. That will be a success if their findings can be used in negotiations.
- Sometimes you are just practicing your skills with open data, you don't have a specific goal, then it is hard to say whether it is a success.
- It is not a success if the dataset is not feasible. When the dataset contains a lot of pollution and there is no support to get it out of the dataset.
- If organizations open their data they should be able to answer question from reusers, and not just leave the user in the dark.
- <u>Success factors:</u> The quality of the data, how much effort does it take to get it usable for yourself. It is much easier when it is machine readable, I often don't even open pdf files. Towards the future, the usage of standards will become more important. In general, you have three aspects: 1) timeliness, 2) accuracy, and 3) completeness. For each of these aspects it is nice if the dataset scores high.
- <u>Example:</u> At CBS they open data about borders of municipalities once a year, however in march 2016, the data of 2015 is released. Since January 2015, ten municipalities have merged. One of my customers wants to plot data based on the current borders of the municipality. However, since this data is not yet available, I had to adapt the data of last year to the current situation myself.
- <u>Fail factors:</u> Insufficient knowledge about the dataset, not knowing what you are talking about. With the amount of data that is available it is tempting to start working with data about which you don't have any knowledge. You need to have a goal you want to reach. Lack of documentation of information accompanying the dataset.
- <u>Example success project</u>: Plannerstack, I am not sure whether they are successful but it is an interesting project. And OpenStreetMap is an example of a large project, but they don't use data from one organization.

B.12 SUMMARY INTERVIEWEE 12

- Technical program manager
- Bachelor thesis on open data
- Interested in open data from a political view of point. To make people more involved, which turned out to be a need for a political party. The combination between technology and transparency is interesting.
- <u>Definition open data</u>: All information that is collected by organizations and publicly released, in which it is made machine readable and contains no restrictions for reuse.
- <u>Reaction definition</u>: Maybe add that it is about data that from which it is not possible to deduce who it is about. It is about non personal information.
- The research performed by Deloitte about the added value of open data of the KNMI is very interesting. In their conclusions they do not only mention a lot of money but also improvement on employment.
- The value of the data is most prominent for organizations or persons that want to do something with that data. Other value can be found in transparency.
- <u>Example:</u> Regarding to recent happenings, the maps of Nepal have been redrawn. To do so, we opened the access to satellites.
- One of the issues with open data is that it is not always visible. For instance the train delays are included in Google Maps, that is based on open data, however, not much people realize that. This is also the beauty of open data.
- <u>Success factors:</u>
 - Legal status of information. When there is no license available it is hard for organizations to assess whether they are allowed to use the information from governments. Therefore it is best to use a license that is as open as possible.

For instance, OpenStreetMap does not allow the restrictions on data of having to use references to the original source. If they would accept that, their product becomes unusable as the complete right side will consist out of references to original sources.

There is a difference between Europe and the United States, as the legislation in Europe allows governments to put restrictions on the reuse of their data. In the United States, all data, if not deducible to persons, is open by definition.

- The way in which data is published. It should be machine readable, the quality is determined by the degree of machine readability. For instance if you want to publish data about voting behavior, each municipality uses a different format; ranging from PDF, Excel, CSV, XML or not a file at all. This makes it very labor intensive to work with that data.
- Topicality, is the data still up to date, or is it outdated? This also includes whether there is enough documentation available accompanying the data.
- Examples of success projects:
 - Change.org; is empowering citizens, and a sustainable organization.
 - Socrata; is solving the problem that public organizations face in making their data machine readable.
 - Buienradar; for the weather sector this initiated a new market with a lot of economic potential.
- Fail factors:
 - o Inverse of the success factors
 - See also my bachelor thesis
 - The most valuable data is not opened yet, because data owners often do not see the added value for themselves to open the data. For example the cadastre, they make money because of the data they collect and store. However, the Catalan cadastre opened its data and although it cost some money, now it generates money. Still, other cadastres are not opening their data as they don't see this value.
 - Opening the data shows what goes wrong. Open data sets compared to data sets for which you have to pay sometimes show differences in the number of records, it is unsure which is closer to the truth. This is another motivation to actually open the data, although it might scare of organizations to do so.

B.13 SUMMARY INTERVIEWEE 13

- Started at KNMI as a forecaster in aviation, maritime meteorology and general and guidance shift leader. Followed by lecturer in before mentioned meteorological topics. Worked a couple of years in management. Currently, policy officer, focused on real time data on national and international level.
- Often Buienradar is mentioned as an example of an Open Data project, however, they started their business far before the Open Data concept has been introduced.
- Within the Netherlands, we have about 30 meteorological observation stations. Observations come available in (near) real time and are distributed as such. The data sets are not always complete, we have to accept the loss of data if one of the stations does not provide data in the timeframe of distribution. Beside the (near) real time data, KNMI can provide climatological data sets for the longer studies. These data sets are as complete as possible, making use of on site storage possibilities or any manual addition (as far as possible).
- KNMI runs 2 weather radar sites, several (near) real time data products come available
- KNMI operates a large computer system to run several operational numerical atmosphere models as well as long run climatological outputs.
- We gather a lot of data: meteorological data, seismic data, data from satellites.

- An online website Catalogue displays the data products that are available for commercial reuse. One needs to register in order to get access to the data. An agreement is set up, which contains the description of the service level as well as the fee that is to be paid. Since 2009 we only charge for the service and no longer for the data itself (so called license fee). To be able to provide that service you need to know who your customer is, so you can communicate with your customer in case of malfunction. Service delivery will go together with Open Data distribution.
- At the 18th July of this year EU member states are obligated to start with Open Data, according to the PSI Directive by the European Commission. This entails that member states need to adapt their national policies. In The Netherlands this means a change of WOB (Wet Openbaar Bestuur) and the introduction of a new law WOO (Wet Open Overheid) on 1 July 2015. Some other EU countries are already ready (Finland, Sweden, Iceland). Others go slow and might not be ready in time (France, some eastern Europe countries).
- Following lenM (Ministry where KNMI belongs to), we are obliged to follow on Open Data Policy since 1 January 2015. However, we are not opening all our data yet, not because we don't want to, but because the Dutch meteorological enterprises forbid us to do so (as long as legislation has not changed). The current market might be changing much with the introduction of Open Data, opening up data access at much lower financial thresholds for extra competitors.
- Open Data means availability 'as is'. This implies that if the distribution server goes down after normal working hours, it is much likely that users will have to wait till the next working day for connection. Distribution is for free, no one is paying for service. We expect that organizations that rely on the availability of the data will be keep using the service agreement distribution.
- <u>Definition open data</u>: It's about free availability of raw data that can be processed through a machinemachine interaction, without human intervention. We define raw data as digital information which form the basis for a subsequent adding value process. That might imply that we already perform some actions on the data, but we don't create an end product. We only collect data that is part of our assignment of the Ministry. Within our Ministry instruction it is decided that you are not required to ask for owner reference (IPR – Intellectual Property Rights), but it certainly has value to do so (making clear who is the originator of the data). Although it is not part of the Ministry instruction, it is recommended to distribute the raw data under an open data license. And of course you have to make sure you provide enough metadata. Which is not an easy task in relation to the more or less very specific meteorological data (as well as its data formats).
- For the future we work towards standard data formats, also as part of INSPIRE regulations. However, also new standards are not very easy to use for general users (some specific knowledge about the meteorological process will be very helpful).
- We have a lot of data and it is too much to release all of it at once. A lot of climatological data products are already under Open Data available. Near real time products will follow soon, following the preselecting of all data products listed in the KNMI Catalogue. Numerical output from the computer models is far more as the preselected data products in the Catalogue, but makes is very accessible for the users. You can think in this respect of the situation where you go to the butcher and ask for an ounce of meat. The butcher then comes back with a complete cow and tells you to cut a piece off yourself. The same situation applies to the use of numerical output. In most cases the user is only interested in a small subset, with a limited number of parameters and a restricted domain. So to say, only a stamp of the full available collection.
- Regarding Open Data distribution the Ministry has stated that we need to open a kind of helpdesk, which does not involve a phone number per se, but at least an email address.
- Opening up more data products will go faster if we would have more budget to spent. .
- The revenues that we have missed by terminating the license fees (retrieval of money for the collection of the data) are not refunded by the Ministry, which means that we had to cut that loss within our own organization budget. Partly by reduction of staff member. After terminating the

license fees (2009) we didn't see the number of customers very much increase. It didn't lead to that many new revenues from service fees. That is partly due to the fact that the meteorological market is pretty satisfied. Introducing the Open Data Policy will lead to a drop of revenue from services in the next couple of years, but in the long term it might get back to the revenue from before Open Data was introduced. Remark: the revenue form services should balance the cost. Open Data Policy does not allow additional income as a revenue.

- It is very much a political statement to move towards Open Data with the expectation that it will bring a high economic result. However, the principle of having governmental data for free is a valuable thought that counts. The data that is involved, has already been paid for by taxes.
- The most important barrier was back in 2009 when we dropped our license costs and we had to compensate for it through our organization budget cuts. So, barriers now are mostly about the costs for introducing the full concept of Open Data. They are related to infrastructure and staff availability. Since general budget has been shrinking over about 10 years as a political statement (government is to big), also KNMI (as Ministerial agency) has to deal with that.
- <u>Definition of success</u>: It is a success when the data is used more frequently. So, when the number of customers increases.
- <u>Success factors:</u> More publicity would be beneficial, organizations are often surprised about the data that we have. But still you encounter the issue of domain knowledge; knowing what can be done with the data. So not only publicity saying we have that data, but also how to use it, although I don't have an idea on how that should be done.
- <u>When does a project fail:</u> When the data is not used at all. That would mean that we set up a platform but it is not used by anyone. I don't expect that to happen, but we can influence this by promoting our data and distribution platform (KNMI Data Centrum KDC).

B.14 SUMMARY INTERVIEWEE 14

- Every minute the traffic on all important roads in the Netherlands is counted. This data comes from Rijkswaterstaat, provinces, and metropolitan cities as Amsterdam, Rotterdam, The Hague, and Utrecht.
- At the moment there are two options. We have a closed data service with a license agreement, information out of our production process is sent to organizations as ANWB, TomTom, and the traffic information service. They receive data that was created one minute ago. The agreement is mostly based on what they can expect by incidents. They can contact us 24/7 and rely up on us to repair it. That information also goes to Rijkswaterstaat and road management so they can use it for their own purposes. On the other side we have an open data service, which can be accessed by everyone. It is exactly the same dataset, with the difference that you don't have the 24/7 service.
- A lot is expected from open data, I wonder whether this will become reality. I see that there are indeed new ideas, but the individuals that come up with those ideas are not capable of putting it into action, so it is sold to large organizations. So we see the same organizations working with our data, while the ideas might come from new parties.
- <u>Definition open data</u>: All data that is made available by the government. And we go somewhat further by making it available as soon as possible.
- <u>Reaction definition</u>: Easy accessible is very difficult. I mean it is machine readable, so if you take that as easy accessible, okay than it is easy accessible. But we don't process the data, XML comes out and we don't put an API on it.
- We just open the data that we have. Although the ministry states that if the data is not useful as it is, we should make it more usable, but I think that this is something for the future. And if there is some money available to support that, we will be able to do so. Otherwise we won't do it probably, I am not sure whether the society would benefit from it, maybe only the individuals that want to use the data.

- We have opened our data stepwise. We first opened our raw measurements. Moreover, we have a product which already contains some processed data, that contains traffic jams and construction zones. Since January we opened historical data.
- We see that a lot of data is being downloaded, but we don't know who is doing that and what products are created from the data.
- We didn't experience any issues with opening our data, we were already providing these data to existing customers, we just created another environment for the open data, to protect our production environment.
- Small organizations benefit most from open data. They can access our data more easily, and it also saves us time and effort, because we don't have to create another contract.
- I think all our customers will go for open data in the end, because it will get a societal relevance, so we will eventually build in the 24/7 service for open data.
- Our current agreements are based on the fact that we expect a service in return. If all customers go to the open data service, we will lose these services, however, some of them have a societal importance, so maybe the service will continue to exist. At least we will try to keep communicating with those parties.
- Maybe you don't need strict agreements with open data. If your product is usable, customers will start using it.
- <u>What project would you start:</u> Environmental analysis. For each city or area you map the number of trucks, then you take the average emission of a truck and a passenger car to calculate the total emissions in a particular period of time.
- <u>Need for a database that provides an overview of (active) open data projects:</u> I think that is very usable and helpful information for us. I think our communication department would look at it.
- <u>Definition of success</u>: When data is used in such a way that it has a societal relevance, that it is a necessity. When the portal is down, you get a lot of reactions.
- <u>Success factors</u>: Wide usage within society. It is assumed that it is always there. It is a matter of course. For an individual it is nice if his idea is sold to a large organization.
- <u>When would a project fail:</u> If our server is down for a weekend and no one notices it. Is it worth the effort if no one even notices that the server is down for a weekend. Or when you notice a lot of activity on the portal but there are no products made based on our data.
- <u>Example:</u> Bestwelsnel.nl, one of the employees, of a company that works with our data, started crafting. He created a website which depicted the speeds at the roads across the Netherlands. So, you could see the top speed per road. Well, the government was not very pleased with that. So, maybe that is the downside of open data.
- <u>Example success project</u>: The employees of NS were looking at an app of a competitor to see which train to get. That is a real success, especially since NS bought the app themselves later on.
- What would be nice is when you would be able to show what open data is used within large organizations. So, where do they mix open data and internal data for internal use. I wonder whether organizations use that.

B.15 SUMMARY INTERVIEWEE 15

- Leer & Expertisepunt Open Overheid, its goal is to help public professionals with open government. It contains four sections: open approach, open responsibility, open data and open contact. For me the focus is on open data. I am working for quite some time now to get open data working at the Dutch government. I was also involved in the first version of the national open data portal as a project manager.
- For a long time I have been involved in human computer interaction.

- In 2009 open data came to my attention and the Ministry of the Interior and Kingdom Relations asked me to be actively involved in the start of open data in the Netherlands.
- <u>Definition open data</u>: Open government information that is reusable without any restrictions. It is build up out of two parts. The first part open government information refers to the WOB (Dutch: Wet Openbaarheid van Bestuur). The second part reusable without any restrictions refers to a law that is being implemented at the moment, and that is a law on reuse.
- <u>Reaction definition:</u> I think we mention the same aspects, in your definition they are more elaborate. I created the definition based on the two laws, so format will be in the reuse law. I used open government information as a synonym for open data to give more focus to the definition. It is very hard to capture in the definition that it is about bytes and not about policy information.
- At the moment every governmental organization is realizing that they have to do something with open data. So it is no longer about whether they should open their data, but it is about how and when. It is often an subject of discussion that is allocated to one person and not to a part of the organization.
- I got a list of 42 objections. Some people put those objections easier aside than others. Within the boundaries of the current legislations a lot is possible, but it is very difficult to steer it. It is also hard to depict the actual utility of open data. Municipalities that were one of the first to open their data don't see a lot of use of their data, but that is partly due to the datasets that they have opened.
 - o Zie lijst.
- But the conversation is started, organizations are aware of open data and are trying to get their data ready to open.
- We try to give a good example by starting at the departments of ministries. It is easier for lower governments to join the movement than to start it themselves. In addition, data from ministries have a larger impact (i.e. national) than data of one particular municipality.
- There are three lines of benefits. The first one is transparency within democracy. This one is the most important one for me. The second one is reuse, both societal and economic reuse. This one is also important, because we often forget that if we would close all datasets again, a lot of applications will stop working. At some points we forget that open government information is used. The first order effect is the app developer that makes some money with his app, the second order effect could be someone's happiness and is therefore much bigger than the first order, but also immeasurable. The third one is efficiency of the government. By opening the data you prevent that two organizations will perform the same research as they can now use each other's data.
- The billions that are promised to arise from open data might be found in costs that are saved, for example when there would be no weather apps it might be that more people will be having a cold because they are outside in the rain all the time. By using weather apps this number is reduced and will in its turn save out on medical costs.
- We tell all organizations to disclose their data as is. When we are working towards open formats they will have to use those formats also for internal use, so when they disclose their data as is, it is already in open format.
- It is hard to measure the added value of open data, mostly because we try to measure a new phenomenon with traditional methods.
- <u>What project would you start:</u> I would not start thinking from the data. Projects should start from a societal issue. It would be nice if a municipality would arrange some space where developers can work on such issues. Without the municipality influencing the result of those sessions. So just the first order effect by spending some money to arrange such a space, the second order effects are up to the developers to accomplish.
- <u>Need for a database that provides an overview of (active) open data projects:</u> I would definitely use it to search for new examples.
- <u>Success factors:</u> When it really helps people, so when the second order effect is achieved. Being able to show the effects, it is not possible to show everything, but at least show some parts of the achieved

effects. That when you are organizing an open data event nowadays that you have a lot of registrations, as that indicates that the topic is hot.

- <u>Fail factors:</u> It is a shame that hackatons often don't lead to viable products. So they really create smart and intelligent applications, however there is no practical use for it, so it disappears again. Therefore it is important to think taking societal issues as a starting point.
- <u>Example success project</u>: Weather apps, it is remarkable that the government has nothing to do with those apps, they only provide the data. However, it is unfortunate that we still rely on these examples. The IV3 files that are opened by CBS are also an interesting project. Although they need greater depth to make it really interesting for citizens of a particular town or city.

APPENDIX C: MAPPING ON SUCCESS FACTORS

In this Appendix the success factors for non open data projects are mapped onto those for open data projects.

C.1 MAPPING OF THE SUCCESS FACTORS TO THE PRIORITIZED LIST

The success factors are first mapped to the prioritized list of success factors.

Success factors	Captured (Belassi & 1996)	by Tukel,	Captured by (Atsu et al., 2009)	Captured by (The Standish Group, 2013a,	Capturedby(Ashjaetal.,2013)
Ecosystem				2013b)	
Reuse					
Collaboration					✓
Consideration of notential	✓				
value					
Strategy and legislation	✓		✓		
Continuity, up to date data					
Publicity					
Satisfied stakeholders	✓				
Availability of data	✓				
Availability of metadata					
Clear working practices and			✓	✓	✓
organizational processes					
Market acceptation	\checkmark				
Presence of an infrastructure					
Value creation	\checkmark				
Active presence of the public					
sector in the market					
Appropriate release channels	\checkmark				
Citizen participation			√	✓	
Clear leadership	\checkmark		\checkmark	\checkmark	✓
Combining data					
Effects					
High-level mandate	\checkmark		\checkmark	\checkmark	✓
Initiated at local level			\checkmark		
Access to data					
Attitude of the organization			\checkmark		✓
Being able to sell your idea	\checkmark				
Cooperation of the publisher					\checkmark
Different perspective on					
(existing) datasets					
Increase of trust in					
government					
Innovative project	✓			,	
Knowledge of data			,	✓	✓
Learning during the process			✓		✓
Legal status of information					
Location of the project					,
Mapping your data landscape					✓
One central datacenter			,	,	
Planning and coordination of	\checkmark		✓	\checkmark	\checkmark

Success factors	Captured by (Belassi & Tukel, 1996)	Captured by (Atsu et al., 2009)	Captured by (The Standish Group, 2013a, 2013b)	Captured by (Ashja et al., 2013)
the project				
Possible negative impacts		\checkmark	\checkmark	
Proper security of the data				
Support for transparency				
Timeliness, accuracy and				
completeness				
Usefulness				
Embedded in standard practice				
Number of 🗸	12	10	7	10

Table 14 Mapping to the prioritized list

C.2 MAPPING OF THE SUCCESS FACTORS TO THE PHASES

To capture the comparison in one table, each paper will be represented by a number.

Source	Number
(Belassi & Tukel, 1996)	1
(Atsu et al., 2009)	2
(The Standish Group, 2013a, 2013b)	3
(Ashja et al., 2013)	4

Table 15 Paper and number matching

Phase	Success Factor	Covered by
Creating data	Continuity, up to date data	-
Creating data	Citizen participation	2, 3
	Consideration of potential value	1
	Availability of data	1
	Availability of metadata	-
	Presence of an infrastructure	-
	Appropriate release channels	1
	High-level mandate	1, 2, 3, 4
Opening data	Attitude of the organization	2, 4
	Legal status of information	-
	Location of the project	-
	Mapping your data landscape	4
	Possible negative impacts	2, 3
	Proper security of the data	-
	Timeliness, accuracy and completeness	-
	Publicity	-
Finding open data	Access to data	-
	One central datacenter	-
	Reuse	-
	Clear working practices and organizational processes	2, 3, 4
	Market acceptation	1
Using open data	Clear leadership	1, 2, 3, 4
	Combining data	-
	Effects of opening datasets	-
	Different perspective on (existing) datasets	-

Phase	Success Factor	Covered by
	Embedding in standard practices	-
	Knowledge of data	3, 4
	Usefulness	-
	Ecosystem	-
	Collaboration	4
	Satisfied stakeholders	1
	Value creation	1
Discussing open data	Active presence of the public sector in the market	-
	Cooperation of the publisher	4
	Increase of trust in government	-
	Learning during the process	2, 4
	Support for transparency	-
	Strategy and legislation	1, 2
	Initiated at a local level	2
Other	Being able to sell your idea	1
	Innovative project	1
	Planning and coordination of the project	1, 2, 3, 4

Table 16 Mapping to the phases

APPENDIX D: EVALUATIONS

In this Appendix the expanded evaluations of the projects are presented.

D.1 INSTAPLAY

The next project for evaluation is InstaPLAY⁵. Table 17 provides an overview of all metadata that is available in the GOOD DATA database.

DESCRIPTION

InstaPLAY is a website where the playgrounds in Vienna are mapped. The project makes use of Google Maps, which enables planning a route towards the playground. It is also possible to find out which playground is the closest to one's current position. Moreover, it is also possible to view a list of all playgrounds in Vienna.

As stated in Table 17, the goal of the project is to inform. This is quite obvious from taking a look at the website, as it is not possible to interact with it. It seems like the website is not finished yet. This is remarkable as the project already started in 2014.

Feature	As in GOOD DATA
Product Type	Website
Business Type	Non-profit
Intended Effect	To inform
Geographical Level of Application	City
Location of Use (meso/macro)	Austria
Location of Use (micro)	Vienna
Language	German
Year of Publication	2014
Status	Live
Business Model	Free
Themes (fixed)	Public spaces
Topic Covered by Data	Recreational and sporting services
More than one Source	No
Sources of Data	Government generated
Collecting Organization	Stadt Wien
Formats	-
Update Frequency	-
Geographical Level of Dataset	City
Funding	Unknown

Table 17 InstaPLAY as in GOOD DATA

SUCCESS FACTORS

In this section the project of InstaPLAY will be described using the list of success factors, as created in section 9.1. This section will be an expanded version of the second column of Table 12.

⁵ <u>http://instaplay.struggl.com/</u>

ECOSYSTEM

As the goal is to inform people about the playgrounds in Vienna, an active ecosystem is not necessary. When new playgrounds are constructed, it does not take much time to add it to the map.

There are no indications found of an ecosystem.

REUSE

The data of InstaPLAY are not published, therefore it is not possible to reuse the data straightaway. However, as the list does provide an overview of the available playgrounds, the data can be taken from here to be reused.

There are no indications that the data of InstaPLAY is being reused.

COLLABORATION

The project is set up and executed by one individual. The data used for the project also originates from one source, and entails exactly one dataset.

There are no indications found for collaboration with other parties.

STRATEGY & LEGISLATION

At the website itself, there are no references to legislations or strategies. At the websites where the project is mentioned (see 'Publicity'), there are mentions of how open data legislations apply, but this is not done by the developer of the project.

There are no references to legislations.

CONTINUITY, UP TO DATE DATA

As the researcher is not familiar with the playgrounds in Vienna, it is not possible to determine whether the data is up to date. However, the source data was first uploaded in 2011 and the latest adaption was made in June 2014⁶. It might be that there are no changes in the playgrounds in Vienna since then, however this cannot be stated for sure.

It is unknown whether the data is up to date as the latest update of the dataset was already a year ago.

PUBLICITY

The project is included on the governmental open data website of Austria. Moreover, it is included in an overview of open data projects on a website from the United Kingdom⁷. However, no publicity acts can be found.

There are no indications of publicity for InstaPLAY.

⁶ https://www.data.gv.at/katalog/dataset/b5e8ad35-dbbe-483b-bd99-2700beba64bf

⁷ http://opendata.oii.ox.ac.uk/apps/instaplay

AVAILABILITY OF DATA

As stated at 'Reuse', the data itself is not available in a raw format. However, the data can be retrieved from the website. So the data is available, just not in an easy to use format.

The data is available.

AVAILABILITY OF METADATA

There is no metadata available. However, as the application is not very difficult, metadata might be redundant.

There is no metadata available.

MARKET ACCEPTATION

There is no information available about the usage of the website. Therefore, it cannot be stated whether it is accepted by the market.

There are no indications for market acceptation.

VALUE CREATION

The project does create value as it makes it easier to find playgrounds, albeit within Vienna. For people who are not familiar with Vienna, the website can be helpful to find a playground nearby. However, the amount of value created cannot be determined without involving actual users of the application.

InstaPLAY does create (at least some) value.

CITIZEN PARTICIPATION

As stated before, the goal of the project is to inform people. In addition, the project seems to be set up and executed by one individual.

There is no citizen participation within this project.

COMBINING DATA

The data for the application comes from one source. Although this information is combined with Google Maps, it is not an example of a project where data is combined. This is mostly due to the fact that the locations of the playgrounds are mapped. So there is some added value created, just not by combining data.

There are no datasets combined.

INITIATED AT A LOCAL LEVEL

The project is located in Vienna in Austria. It is also initiated here, which makes the initiation of the project local, as it is on city level.

The project is initiated at a local level.

INNOVATIVE PROJECT

Another project exists that does exactly the same as InstaPLAY. The only difference is that Spielplätze Wien⁸ is a mobile application, available exclusively for BlackBerry. Moreover, it does not take much effort to indicate playgrounds on Google Maps and present that on a website.

It is not an innovative project.

LOCATION OF THE PROJECT

The location of the project has an enormous influence on the potential users of the application. Only people in Vienna who are looking for a playground nearby will be using it.

The location of the project puts a limit on the number of users.

D.2 VIENNA CITY BIKE

The next project for evaluation is Vienna City Bike⁹. Table 18 provides an overview of all metadata that is available in the GOOD DATA database.

DESCRIPTION

Vienna City Bike is a bike rental service. There are multiple rental stations across Vienna where the bikes can be picked up and returned. Registration for the rental bikes is possible via the Internet as well as at a rental station.

As there are multiple mobile applications based on the data from Vienna City Bike, the website (<u>www.citybikewien.at</u>) is used. Fortunately the website contains all information necessary.

On the website they provide a detailed explanation about every aspect of the system. The project is still active, as the number of rental stations is still expanding. Per station they indicate the availability of bikes and whether it is out of use.

Feature	As in GOOD DATA
Product Type	Mobile application
Business Type	Commercial
Intended Effect	To inform
Geographical Level of Application	City
Location of Use (meso/macro)	Austria
Location of Use (micro)	Vienna
Language	German
Year of Publication	2014
Status	Live
Business Model	Free
Themes (fixed)	Tourism
Topic Covered by Data	Transport
More than one Source	No
Sources of Data	Organization generated
Collecting Organization	Gewista GmbH

⁸ <u>https://www.data.gv.at/anwendungen/spielplaetze-wien/</u>

⁹ <u>www.citybikewien.at</u>

Formats	-
Update Frequency	-
Geographical Level of Dataset	City
Funding	Unknown

Table 18 Vienna City Bike as in GOOD DATA

SUCCESS FACTORS

In this section the project of Vienna City Bike will be described using the list of success factors, as created in section 9.1. This section will be an expanded version of the third column of Table 12.

ECOSYSTEM

To keep the project running, the team does not make use of input of others. Therefore, there are no indications of an ecosystem. It seems that all updates are done by the team members themselves.

There are no indications for an ecosystem.

REUSE

The data of the project is reused by other developers. As the developers of Vienna City Bike do not create their own mobile applications, they encourage other developers to do so.

The data of Vienna City Bike is reused.

COLLABORATION

The website mentions several partners, however it is not clear whether they cooperate within the project. The team behind Vienna City Bike does however work together with application developers to develop the mobile applications, as they do not possess that knowledge themselves.

Vienna City Bike is collaborating with other parties.

STRATEGY & LEGISLATION

The website of Vienna City Bike does not mention any legislations.

CONTINUITY, UP TO DATE DATA

As the researcher does not have any knowledge about bike rental stations in Vienna, it cannot be stated with certainty. However it can be assumed that the data is up to date, as it is used as a selling point for the city of Vienna towards tourists. Moreover, as the website displays the current state of each station, it has to be up to date for users to be able to rely on it.

The data seems to be up to date.

PUBLICITY

The project is shown at the website of the municipality of Vienna. It is also incorporated in the map of Vienna, in addition to the locations of train stations, the city bike stations are also visible. The project is also presented

at the tourist information website of Vienna. In addition, the website contains an overview of all press related to the project.

The project is presented at several official websites of Vienna.

AVAILABILITY OF DATA

The raw datasets on which Vienna City Bike is based is not available. However, as there are multiple applications based on the service of this project, the data should be available somewhere so that developers can build a mobile application. However this is not publicly available, and therefore not open data. The data is only available when the developers are contacted via mail.

The data is available, after the developers are contacted via mail.

AVAILABILITY OF METADATA

There is no metadata available.

MARKET ACCEPTATION

Users have to register to be able to use the application, as there are many registered users, the application seems to be accepted by the market. Especially since both tourists as well as locals are using the system, as this indicates that it goes beyond its original goal to let tourists go through the city by bike.

Vienna City Bike is accepted by the market.

VALUE CREATION

Since there are no studies performed into the effect of this bike rental service, it cannot be stated what the effect of the application is. It might be that more people are biking through Vienna, however, this cannot be stated with certainty.

The impact of reuse (i.e. value creation) cannot be determined with certainty.

CITIZEN PARTICIPATION

The project was able to move to other cities because citizens as well as tourists are using the application. However, they do not contribute to the project itself. It is not possible for citizens to contribute to the project, other than using the service.

Citizens do not participate in the project.

COMBINING DATA

The project does not combine datasets. It provides a service based on the locations of rental stations, which comes from one data source.

Vienna City Bike does not combine datasets.

INITIATED AT A LOCAL LEVEL

The project is initiated in Vienna. The reach of potential users is therefore limited to visitors and habitants of Vienna, that want to go through the city by bike.

Vienna City Bike is initiated at a local level.

INNOVATIVE PROJECT

The project is very innovative. On their website it is stated that they were the first to develop such a system. They also list thirteen other projects that have followed their example.

It is an innovative project.

LOCATION OF THE PROJECT

As stated before, the project is initiated in Vienna. This limits the number of users to people who want to ride a bike in Vienna. However, as the project is copied to other cities in Europe, people in other cities can use the same system.

The location of the project influences the potential users of the system.

D.3 HOCHWASSER KREMS

The next project for evaluation is Hochwasser Krems¹⁰. Table 19 provides an overview of all metadata that is available in the GOOD DATA database.

DESCRIPTION

The application provides an alarm for floods of the river Krems in Austria. Hochwasser Krems shows the current state of the river, the observations of the past 72 hours, and it is possible to set values by which the user want to receive a notification.

As can be seen in Table 19, Hochwasser Krems is available as a mobile application, more specifically it is exclusively available for Windows Phone. The main goal of the application is to inform people about the current state of the river Krems. The application covers the area around the river Krems, which covers a complete municipality.

Feature	As in GOOD DATA
Product Type	Mobile application
Business Type	Commercial
Intended Effect	To inform
Geographical Level of Application	City
Location of Use (meso/macro)	Austria
Location of Use (micro)	Krems
Language	German
Year of Publication	2014
Status	Live
Business Model	Free
Themes (fixed)	Geology & Geography

¹⁰ <u>https://www.data.gv.at/anwendungen/hochwasser-krems/</u>

Topic Covered by Data	Unknown
More than one Source	No
Sources of Data	Government generated
Collecting Organization	Gemeinde Kremsmünster
Formats	-
Update Frequency	-
Geographical Level of Dataset	City
Funding	Unknown

Table 19 Hochwasser Krems as in GOOD DATA

SUCCESS FACTORS

In this section the project of Hochwasser Krems will be described using the list of success factors, as created in section 9.1. This section will be an expanded version of the fourth column of Table 12.

ECOSYSTEM

The intentions of the applications are to inform and, when necessary, warn the user about the water levels of the river Krems. The values are collected by the municipality and are, probably, automatically loaded into Hochwasser Krems. As this leaves no room for users to interfere, there is no ecosystem.

There exists no ecosystem for Hochwasser Krems.

REUSE

Hochwasser Krems uses an open data set that is provided by the municipality. There are several other projects that are using this dataset as well. However there are no indications found that Hochwasser Krems is being reused itself.

There are no indications that the data of Hochwasser Krems is being reused.

COLLABORATION

Hochwasser Krems is a project of one individual. It seems that he has worked alone on the project without any partners.

There are no indications found for collaboration with other parties.

STRATEGY & LEGISLATION

As there was no documentation found accompanying the project, there were no strategies and legislations mentioned for Hochwasser Krems.

No strategy and legislations were mentioned.

CONTINUITY, UP TO DATE DATA

As the goal of the project is to provide a real-time overview of the water levels of the river Krems, the data should be up to date. Otherwise, the warnings of the application would be pointless.

The application uses real-time data.

PUBLICITY

Hochwasser Krems is presented at the national open data website of Austria and in the Windows Phone Store. More publicity was not found by the researcher.

There are no indications for publicity for Hochwasser Krems.

AVAILABILITY OF DATA

There is a link towards the dataset on which the application is based provided on the website. However, the access to the dataset is restricted, as the researcher is not allowed to access it.

The data is not available.

AVAILABILITY OF METADATA

The metadata is not available.

MARKET ACCEPTATION

The number of users and other indications of the usage of the application cannot be retrieved, therefore it is difficult to determine the market acceptation.

The market acceptation cannot be determined.

VALUE CREATION

The researcher is not familiar with the water levels of the Krems. Therefore, it cannot be stated with certainty whether the application creates value. When the level never exceeds the critical value, the application would not serve a need. It is probable that the application does create some value, however, this cannot be stated with certainty.

The impact of reuse (i.e. value creation) cannot be stated with certainty.

CITIZEN PARTICIPATION

It is not possible to contribute to the project. As a user it is only possible to see which values are displayed in the application.

There is no option to participate within the application.

COMBINING DATA

Only one dataset is used for the creation of this project.

Therefore, there are no datasets combined.

INITIATED AT A LOCAL LEVEL

The location of the project is still at a local level. It was initiated in the municipality Kremsmünster and this is still its location. The only way to expand to another location is by adding more rivers to the application. However, there are no indications for this to happen.

Hochwasser Krems is initiated at a local level.

INNOVATIVE PROJECT

After a quick search, more projects were found that provide a similar service. For instance, Pegelstand Kremsfluss¹¹ also provides the opportunity to receive warnings by certain levels of the river. Moreover, Hochwasser Krems is only available for Windows Phone whereas Pegelstand Kremsfluss a web application is. However, it cannot be determined which project was first.

Hochwasser Krems does not seem to be an innovative project.

LOCATION OF THE PROJECT

The project is located in Austria, more precisely in the area around the river Krems. This puts a restriction on the number of users of the application, as only those people who are interested in the water levels of the Krems will use the application.

The location of the project puts a limitation on the number of users.

D.4 AN ANALYTICAL VIEW ON DATA

The next project for evaluation is An analytical view on data¹². Table 20 provides an overview of all metadata that is available in the GOOD DATA database.

DESCRIPTION

This project provides an analytical view on data related to the city of Vantaa. It uses the earnings and expenses of the years 2010 and 2011. The data is presented using Tableau Software. By going over the numbers more information is provided, for instance, it is indicated whether it concerns earnings or expenses. The numbers are represented in six different ways, where each representation covers a different topic. Examples of these topics are the difference between budget and the actual costs, and the sources of the earnings.

As stated in Table 20 the intended effect of the project is to inform people. There is little functionality available to interact with the data, as it is possible to change the parameters. Even though the project was launched in

¹¹ <u>https://www.data.gv.at/anwendungen/pegelstand-kremsfluss/</u>

¹² <u>http://public.tableausoftware.com/profile/sharad4086#!/vizhome/Vantaankaupungintulotjamenot2010-</u> 2011/Tuloslakselmaraportti

2013, only the data of 2010 and 2011 is included. However, the title of the project already indicates that it is only about the data of 2010 and 2011. Moreover, the last update was two years ago (i.e. in July 2013).

Feature	As in GOOD DATA
Product Type	Data analytics
Business Type	Non-profit
Intended Effect	To inform
Geographical Level of Application	City
Location of Use (meso/macro)	Finland
Location of Use (micro)	Vantaa
Language	Finnish
Year of Publication	2013
Status	Live
Business Model	Free
Themes (fixed)	Finance
Topic Covered by Data	Economic affairs
More than one Source	No
Sources of Data	Government generated
Collecting Organization	City of Vantaa
Formats	XLS
Update Frequency	-
Geographical Level of Dataset	City
Funding	Self funded

Table 20 An analytical view on data as in GOOD DATA

SUCCESS FACTORS

In this section the project of An analytical view on data will be described using the list of success factors, as created in section 9.1. This section will be an expanded version of the fifth column of Table 12.

ECOSYSTEM

The purpose of this project is to inform people about the economic figures of Vantaa of the years 2010 and 2011. It is not needed to keep the project alive, as it is based on historical data. Therefore, there is no need for an ecosystem.

There is no indication for an ecosystem.

REUSE

The content of the project can be downloaded from the website. However, there are no records of this available to the researcher. Therefore, it is not possible to state whether the data of this project is reused.

There are no indications that the data is reused.

COLLABORATION

The project seems to be initiated and executed by one individual. As there are no other parties mentioned, it can be assumed that there has been no collaboration with other parties to exchange knowledge.

There are no indications for collaborations with other parties.

STRATEGY & LEGISLATION

There is no documentation provided alongside the project. Therefore, strategies and legislations are not mentioned.

Strategy and legislation are not mentioned.

CONTINUITY, UP TO DATE DATA

As stated before, the project is based on historical data. Therefore, the data is not new. For the periods 2010 and 2011 the data is up to date, however, the project is not up to date.

The project is not up to date.

PUBLICITY

As the project is in Finnish it is difficult for the researcher to find out what the publicity is for this project. It seems that the dataset is presented at various websites, as also the data of 2012 is shown. On the website of Avoindata.fi the dataset is presented. Moreover, they ask to comment existing projects based on the dataset to the post. Unfortunately, there are no comments.

For the project itself there seems to be no publicity.

AVAILABILITY OF DATA

The data of the project is available, as it is possible to download it. Moreover, it is already possible to interact with the data on the website.

The data is available.

AVAILABILITY OF METADATA

Even though the data itself is available, the metadata is not provided. On the website there is a mention of metadata, however, this is just a repetition of the diverse visualizations.

The metadata is not available.

MARKET ACCEPTATION

There is no information available about the usage of the project. In addition, the number of users is unknown, therefore it is difficult to determine the market acceptation.

The market acceptation cannot be determined.

VALUE CREATION

The project does provide some value as it contributes to the transparency of the city of Vantaa. The project enables users to look into the financials of the city. However, what the actual impact of the reuse is, cannot be stated with certainty.

The created value cannot be determined.

CITIZEN PARTICIPATION

The goal of the project is to show people what the earnings and expenses are of the city of Vantaa in 2010 and 2011. It is not possible for users to participate in the project. It is possible to play around with the visualization, however, this will only be visible for the user who does so.

There is no citizen participation within this project.

COMBINING DATA

The project is based on one dataset.

Therefore there are no dataset combined.

INITIATED AT A LOCAL LEVEL

The project is initiated at city level. Since the dataset covers the financial actions of this city, the project will probably stay at this level.

The project is initiated at a local level.

INNOVATIVE PROJECT

It seems that this project is one of the first projects to publicly display financial information of a city. However, the project only displays the data, the data itself was made available by the city. As the only extra effort is put into the visualization, the project is not considered innovative.

It is not an innovative project.

LOCATION OF THE PROJECT

The location of the project restricts the maximum number of users. Only people who are interested in the financials of the city of Vantaa will be using this project. Even that is not sure, as they could benefit more from the dataset itself, or at least by adding the data of other years to the two years that are included in this project.

The location puts a limit on the number of users.

D.5 TAXI WIEN

The next project for evaluation is Taxi Wien¹³. Table 21 provides an overview of all metadata that is available in the GOOD DATA database.

DESCRIPTION

Taxi Wien provides the opportunity to look for the nearest taxi stand in Vienna. Taxi Wien is available as a mobile application for BlackBerry. The application is based on open data provided by the city of Vienna.

Taxi Wien is still available for BlackBerry users, indicating that the project is still live. It covers the city of Vienna and is main purpose is to inform people about the taxi stands and guide them towards those stands. One's

¹³ <u>https://appworld.blackberry.com/webstore/content/21746447/?lang=de&countrycode=AT</u>

current position is determined using GPS on his or her mobile phone. The taxi stands are visualized by a blue line that is put on to a digital map.

Feature	As in GOOD DATA
Product Type	Mobile application
Business Type	Commercial
Intended Effect	To inform
Geographical Level of Application	City
Location of Use (meso/macro)	Austria
Location of Use (micro)	Vienna
Language	German
Year of Publication	2014
Status	Live
Business Model	Free
Themes (fixed)	Tourism
	Transportation & Traffic
Topic Covered by Data	Unknown
More than one Source	No
Sources of Data	Government generated
Collecting Organization	Stadt Wien
Formats	-
Update Frequency	-
Geographical Level of Dataset	City
Funding	Unknown

Table 21 Taxi Wien as in GOOD DATA

SUCCESS FACTORS

In this section the project of Taxi Wien will be described using the list of success factors, as created in section 9.1. This section will be an expanded version of the sixth column of Table 12.

ECOSYSTEM

The application is based on (static) data provided by the city of Vienna. It is therefore not based on interaction with the users.

There are no indications for an ecosystem surrounding Taxi Wien.

REUSE

There are no records available about the use of the application. Even though the application reuses open data from the city of Vienna, there are no indications that the data of Taxi Wien is being reused.

There are no indications that the data is being reused.

COLLABORATION

The application is created by two men from Vienna. There are no indications that they worked together with other parties to create Taxi Wien.

There are no indications found for collaboration with other parties.

STRATEGY & LEGISLATION

They do mention a privacy policy on the website where the application can be found. In that privacy policy it is indicated that personal information will not be disclosed to third parties. They make a reference to applicable laws and legislations, but do not specify which.

Legislation is mentioned in the privacy policy which is available at the website.

CONTINUITY, UP TO DATE DATA

It can be assumed that the data gathered from the city of Vienna is automatically put into Taxi Wien. Therefore, the data should be up to date, as the city should have this data up to date for themselves. As can be seen on the website of data.wien.gv.at¹⁴, the dataset on which Taxi Wien is based is recently updated.

The data on which Taxi Wien is based is up to date, it can be assumed that the data is automatically put into the project, therefore, Taxi Wien is up to date.

PUBLICITY

The project is presented on the general open data website of Vienna. More publicity was not found.

No publicity acts can be found for Taxi Wien.

AVAILABILITY OF DATA

The data on which Taxi Wien is based in available via the open data portal of Vienna. However, the data of the application itself is not available for reuse.

The data of Taxi Wien is not available.

AVAILABILITY OF METADATA

There is no metadata provided.

MARKET ACCEPTATION

As there are no records about the usage of the application, it cannot be determined whether the application is accepted by the market.

The market acceptation cannot be determined.

¹⁴ <u>https://www.data.gv.at/suche/?search-</u>

<u>term=taxi+wien&top10Filter_Stadt+Wien=on&publisherFilter_Stadt+Wien=on&connection=and&search-data-only=search-data-only=showresults</u> Retrieved on 28-07-2015

VALUE CREATION

The application provides some value as it helps people find a taxi. However, at the moment this is only possible for Vienna. However, the amount of value created cannot be determined without involving actual users of the application.

Taxi Wien creates (at least some) value.

CITIZEN PARTICIPATION

As stated by 'Ecosystem', there is no purpose for users to contribute to the project. The data is (presumably) automatically loaded into the application.

There is no citizen participation.

COMBINING DATA

Only one dataset is used for the creation of the application. This data is subsequently put onto a digital map.

There are no datasets combined for the creation of Taxi Wien.

INITIATED AT A LOCAL LEVEL

The project is initiated in Vienna, on city level. This is still the focus of the application.

The project is initiated at a local level.

INNOVATIVE PROJECT

To the knowledge of the researcher, this is one of the first applications that provides this service. Since it is not just a visualization of the taxi stands, but it is also possible to plan a route towards the nearest taxi stand, it can be considered as an innovative project.

Taxi Wien can be considered an innovative project.

LOCATION OF THE PROJECT

The project is located in Vienna, which restricts the number of users. Only people looking for a taxi stand will be using the application.

The location of the project limits the maximum number of users.

D.6 WANDERWEGE WIEN

The next project for evaluation is Wanderwege Wien¹⁵. Table 22 provides an overview of all metadata that is available in the GOOD DATA database.

¹⁵ <u>https://appworld.blackberry.com/webstore/content/21746530/?lang=de&countrycode=AT</u>

DESCRIPTION

Wanderwege Wien is a mobile application for BlackBerry. The application is based on data that is generated by the city of Vienna. It was published in 2014 and is still freely available. The application provides an overview of the hiking trails in Vienna. It uses the GPS location of the BlackBerry to guide the user towards or over the trails. A basic map is used and the hiking trails are drawn upon that map.

Feature	As in GOOD DATA
Product Type	Mobile application
Business Type	Commercial
Intended Effect	To inform
Geographical Level of Application	City
Location of Use (meso/macro)	Austria
Location of Use (micro)	Vienna
Language	German
Year of Publication	2014
Status	Live
Business Model	Free
Themes (fixed)	Leisure
Topic Covered by Data	Recreational and sporting services
More than one Source	No
Sources of Data	Government generated
Collecting Organization	Stadt Wien
Formats	-
Update Frequency	-
Geographical Level of Dataset	City
Funding	Unknown

Table 22 Wanderwege Wien as in GOOD DATA

SUCCESS FACTORS

In this section the project Wanderwege Wien will be described using the list of success factors, as created in section 9.1. This section will be an expanded version of the seventh column of Table 12.

ECOSYSTEM

The application is based on (static) data provided by the city of Vienna. It is therefore not based on interaction with the users.

There are no indications for an ecosystem surrounding Wanderwege Wien.

REUSE

There are no records available about the use of the application. Even though the application reuses open data from the city of Vienna, there are no indications that the data of Wanderwege Wien is being reused.

There are no indications that the data is being reused.

COLLABORATION

The application is created by two men from Vienna. There are no indications that they worked together with other parties to create Wanderwege Wien.

There are no indications found for collaboration with other parties.

STRATEGY & LEGISLATION

They do mention a privacy policy on the website where the application can be found. In that privacy policy it is indicated that personal information will not be disclosed to third parties. They make a reference to applicable laws and legislations, but do not specify which.

Legislation is mentioned in the privacy policy which is available at the website.

CONTINUITY, UP TO DATE DATA

As the data is generated by the city of Vienna, it can be assumed that it is up to date. However, as the hiking trails surrounding Vienna are not familiar to the researcher it is not possible to determine whether the application is up to date.

It cannot be determined whether the application is up to date.

PUBLICITY

The project is presented on the general open data website of Vienna. More publicity was not found.

No publicity acts can be found for Wanderwege Wien.

AVAILABILITY OF DATA

The data on which Wanderwege Wien is based in available via the open data portal of Vienna. However, the data of the application itself is not available for reuse.

The data of Wanderwege Wien is not available.

AVAILABILITY OF METADATA

There is no metadata provided.

MARKET ACCEPTATION

As there are no records about the usage of the application, it cannot be determined whether the application is accepted by the market.

The market acceptation cannot be determined.

VALUE CREATION

The application provides some value as it helps people find hiking trails. However, at the moment this is only possible for Vienna. The amount of value created cannot be determined without involving actual users of the application.

The impact of reuse (i.e. value creation) cannot be determined.

CITIZEN PARTICIPATION

As stated by 'Ecosystem', there is no purpose for users to contribute to the project. The data is (presumably) automatically loaded into the application.

There is no citizen participation.

COMBINING DATA

Only one dataset is used for the creation of the application. This data is subsequently put onto a digital map.

There are no datasets combined for the creation of Wanderwege Wien.

INITIATED AT A LOCAL LEVEL

The project is initiated in Vienna, on city level. This is still the focus of the application.

The project is initiated at a local level.

INNOVATIVE PROJECT

To the knowledge of the researcher, this is one of the first applications that provides this service. Since it is just a visualization of the hiking trails in the area of Vienna, it is not innovative project.

Wanderwege Wien is not an innovative project.

LOCATION OF THE PROJECT

The project is located in Vienna, which restricts the number of users. Only people going for a hike in the surroundings of Vienna might be using the application.

The location of the project limits the maximum number of users.

D.7 JECO GUIDES

The next project for evaluation is Jeco Guides¹⁶. Table 23 provides an overview of all metadata that is available in the GOOD DATA database.

¹⁶ <u>http://www.jeco.biz/en/</u>

DESCRIPTION

Jeco Guides is a mobile application that consists of a library of interactive guides. Within each guide, it is possible to use the map, see a list of all areas, tracks, and points of interests that are covered within the guide, but it is also possible to use augmented reality. By framing the surroundings with a mobile phone, information about the surroundings will appear on the mobile phone.

The application is build on the guides that are added by its users. Therefore, the developers do not call Jeco Guides an application, but they refer to it as an IT platform. They encourage people to become local authors. Local authors add places and description from their own neighborhood. The local authors can choose for themselves whether they want their guide(s) to be free. Either way they are obligated to pay a small fee every year to be allowed to publish their guide. Users of the application only have to pay for the guides that are not provided for free by the local authors.

Feature	As in GOOD DATA
Product Type	Mobile application
Business Type	Non government organization (NGO)
Intended Effect	To inform
Geographical Level of Application	Region
Location of Use (meso/macro)	Italy
Location of Use (micro)	Northern Italy
Language	Italian
Year of Publication	2012
Status	Live
Business Model	Free
Themes (fixed)	Geology & Geography
	Tourism
Topic Covered by Data	Tourism (CS)
More than one Source	No
Sources of Data	Citizen generated
Collecting Organization	N/A
Formats	-
Update Frequency	-
Geographical Level of Dataset	Region
Funding	Unknown

Table 23 Jeco Guides as in GOOD DATA

SUCCESS FACTORS

In this section the project Jeco Guides will be described using the list of success factors, as created in section 9.1. This section will be an expanded version of the eighth column of Table 12.

ECOSYSTEM

The application is build upon guides that are provided by users. Without these contributions, the application provides just a map, which would not provide any added value above already established maps for mobile phones. There are quite some guides added within the application. Most of them are within Italy, but also outside Italy some guides are added.

Jeco Guides is dependent on its ecosystem.

REUSE

The guides that are put into Jeco Guides are reused by the users of the application. However there are no indications that other applications are using the data of Jeco Guides.

There are no indications for reuse.

COLLABORATION

The creation of the application itself does not depend on collaboration. The content of the application is highly dependent on collaboration, as the core of the application is to provide a platform for local guides.

Collaboration with local authors is necessary to create its added value.

STRATEGY & LEGISLATION

There are no strategies or legislations mentioned.

CONTINUITY, UP TO DATE DATA

It is not clear whether the guides in the application are checked to find out if they are up to date. It seems to be that local authors are able to create and upload a guide. It is not mentioned whether it is possible to update someone else's guide. It is possible to send feedback to the author of the guide, however, this does not obligate the author to adapt the guide.

The continuity of the application cannot be determined with certainty.

PUBLICITY

Jeco Guides is presented in multiple application stores, there is also a Facebook and Twitter account, which lead to more publicity. Moreover, there are several online articles available in which Jeco Guides is discussed and praised. More importantly, the application is not only offered in Italian, but also in English, which increases the maximum number of users.

Some publicity is found for Jeco Guides.

AVAILABILITY OF DATA

It is possible to download all guides that are in the application separately. However, it is not possible to download them all in once. It is also not possible to get the raw data of the application.

The data is not available.

AVAILABILITY OF METADATA

The guides themselves can be considered to be the metadata, as they described what is meant with the data and how to interpret it.

The metadata is available.

MARKET ACCEPTATION

According to the Google Play Store¹⁷ the application the number of downloads is between 1.000 and 5.000. Unfortunately, these numbers are not available for the iOS Appstore. Nonetheless, these numbers indicate that the application is frequently downloaded. Moreover, there are quite some guides available in the application itself.

It seems that Jeco Guides is accepted by the market.

VALUE CREATION

The added value of the application over established map applications is that it incorporates guides created by local authors. By doing so, the user gets a look into the local culture of the place which he or she is visiting. Moreover, if guides are available for a particular area, the application also presents the highlights in that area.

The application does provide added value.

CITIZEN PARTICIPATION

As stated by 'Ecosystem' and 'Collaboration', the application is build upon the input of its users. For all areas, citizens are able to add a guide which presents their city at its best.

Jeco Guides is dependent on citizen participation.

COMBINING DATA

There are no datasets combined. Even though input from several users is presented on the map, the dataset consists of their input.

Jeco Guides is solely based on user input, therefore no datasets are combined.

INITIATED AT A LOCAL LEVEL

The project is initiated in the north of Italy. This can be seen in the application itself, as most guides are available in this region. However, at the moment, the guides are more spread over Italy.

The project is initiated at a local level.

INNOVATIVE PROJECT

The goal of the application is to bring local guides together in one application. To the knowledge of the researcher, this is the first application that provides this service. Jeco Guides is not just a mapping of areas, but it adds an augmented reality layer on top of it.

Jeco Guides can be considered as an innovative project.

¹⁷ Retrieved on 29-07-2015, from <u>https://play.google.com/store/apps/details?id=biz.jeco.jecoguides</u>

LOCATION OF THE PROJECT

The initial location of the project (i.e. north of Italy) does not limit the maximum number of users. The application is available in both Italian as in English, therefore, everyone who speaks English can use the application. The guides are written by local authors, they can decide in which language they want to publish the guide.

The location does not limit the maximum number of users.

D.8 SOZIALMÄRKTE WIEN

The next project for evaluation is Sozialmärkte Wien¹⁸. Table 24 provides an overview of all metadata that is available in the GOOD DATA database.

DESCRIPTION

Sozialmärkte Wien is a mobile application, available exclusively for BlackBerry. The application helps people to find social markets nearby. Social markets are an Austrian phenomenon, where only people who have a low income are allowed to do their shopping.

The application is based on data generated by the city of Vienna. The goal of the application is to inform people about the nearest social market. The markets are also indicated by a number per neighborhood of Vienna.

Feature	As in GOOD DATA
Product Type	Mobile application
Business Type	Commercial
Intended Effect	To inform
Geographical Level of Application	City
Location of Use (meso/macro)	Austria
Location of Use (micro)	Vienna
Language	German
Year of Publication	2014
Status	Live
Business Model	Free
Themes (fixed)	Business
	Leisure
Topic Covered by Data	Community development
More than one Source	No
Sources of Data	Government generated
Collecting Organization	Stadt Wien
Formats	-
Update Frequency	-
Geographical Level of Dataset	City
Funding	Unknown

Table 24 Sozialmärkte Wien as in GOOD DATA

¹⁸ https://appworld.blackberry.com/webstore/content/21717455/?lang=de&countrycode=AT

SUCCESS FACTORS

In this section the project Sozialmärkte Wien will be described using the list of success factors, as created in section 9.1. This section will be an expanded version of the ninth column of Table 12.

ECOSYSTEM

The application provides a list with social markets in Vienna. As the list is retrieved from a dataset of the city of Vienna, there is no need for users to keep it up to date.

There is no need for an ecosystem.

REUSE

The project reuses the open data as acquired by the city of Vienna. However, there are no indications that the data of the application itself is being reused.

There are no indications for reuse of the data of Sozialmärkte Wien.

COLLABORATION

It seems that the application is made by the developer alone.

There are no indications for collaboration with other parties.

STRATEGY & LEGISLATION

They do mention a privacy policy on the website where the application can be found. In that privacy policy it is indicated that personal information will not be disclosed to third parties. They make a reference to applicable laws and legislations, but do not specify which.

Legislation is mentioned in the privacy policy which is available at the website.

CONTINUITY, UP TO DATE DATA

It can be assumed that the data gathered from the city of Vienna is automatically put into Sozialmärkte Wien. Therefore, the data should be up to date, as the city should have this data up to date for themselves.

The data on which the project is based is up to date, it can be assumed that the data is automatically put into the project, therefore, it is up to date.

PUBLICITY

The project is presented on the general open data website of Vienna. More publicity was not found for the application.

No publicity acts can be found for Sozialmärkte Wien.
AVAILABILITY OF DATA

The data on which Sozialmärkte Wien is based in available via the open data portal of Vienna. However, the data of the application itself is not available for reuse.

The data of Sozialmärkte Wien is not available.

AVAILABILITY OF METADATA

There is no metadata provided.

MARKET ACCEPTATION

As there are no records about the usage of the application, it cannot be determined whether the application is accepted by the market.

The market acceptation cannot be determined.

VALUE CREATION

The application provides some value as it helps people find a social market. However, at the moment this is only possible for Vienna. Moreover, it only creates value for people who are new to the phenomenon. People who are already going to social markets are probably already aware of the locations. However, the amount of value created cannot be determined without involving actual users of the application.

Sozialmärkte Wien creates (at least some) value.

CITIZEN PARTICIPATION

As stated by 'Ecosystem', there is no purpose for users to contribute to the project. The data is (presumably) automatically loaded into the application.

There is no citizen participation.

COMBINING DATA

Only one dataset is used for the creation of the application. This data is subsequently put onto a digital map.

There are no datasets combined for the creation of Sozialmärkte Wien.

INITIATED AT A LOCAL LEVEL

The project is initiated in Vienna, on city level. This is still the focus of the application.

The project is initiated at a local level.

INNOVATIVE PROJECT

To the knowledge of the researcher, this is one of the first applications that provides this service. However, it is only a list with the locations of social markets in Vienna.

Sozialmärkte Wien cannot be considered an innovative project.

LOCATION OF THE PROJECT

The project is located in Vienna, which restricts the number of users. Only people looking for a social market, that do not know where to find one, might use the application.

The location of the project limits the maximum number of users.

D.9 DOG TOILET MAP SALZBURG

The next project for evaluation is Dog Toilet Map Salzburg¹⁹. Table 25 provides an overview of all metadata that is available in the GOOD DATA database.

DESCRIPTION

Dog Toilet Map Salzburg is an application for the mobile application Wikitude. It shows the locations of dog toilets as an augmented reality layer on top of the application Wikitude. Within Salzburg there are approximately 80 dog toilets, which are all under supervision of park managers.

The application is available for Android, whereas Wikitude is also available for iOS. It is based on data acquired by the city of Salzburg and is meant to inform dog owners about the nearest dog toilet. It is possible to show the dog toilets on a map as well as in a list.

Feature	As in GOOD DATA
Product Type	Mobile application
Business Type	Non profit
Intended Effect	To inform
Geographical Level of Application	City
Location of Use (meso/macro)	Austria
Location of Use (micro)	Salzburg
Language	German
Year of Publication	2014
Status	Live
Business Model	Free
Themes (fixed)	Environment
Topic Covered by Data	Unknown
More than one Source	No
Sources of Data	Government generated
Collecting Organization	Magistrat der Stadt Salzburg MA 702 Gartenamt
Formats	-
Update Frequency	-
Geographical Level of Dataset	City
Funding	Unknown

Table 25 Dog Toilet Map Salzburg as in GOOD DATA

¹⁹ <u>https://www.data.gv.at/anwendungen/dog-toilet-map-salzburg/</u>

SUCCESS FACTORS

In this section the project of Dog Toilet Map Salzburg will be described using the list of success factors, as created in section 9.1. This section will be an expanded version of the tenth column of Table 12.

ECOSYSTEM

At the moment there are no indications for a need for an ecosystem. However, as mentioned by 'Continuity', the dataset will be removed. If there will not be a replacing dataset, an ecosystem might be necessary to keep the application up and running.

There is no indication for an ecosystem.

REUSE

The data of Dog Toilet Map Salzburg is used in the application Wikitude, as this project is build in as a layer within Wikitude.

The data is reused.

COLLABORATION

The application is built by one individual. However, to make it applicable for Wikitude, there must have been some kind of collaboration. The developers of Wikitude probably shared their knowledge about augmented reality.

Some collaboration has been done to create the application.

STRATEGY & LEGISLATION

There are no strategies and legislations mentioned.

CONTINUITY, UP TO DATE DATA

At the moment, it can be assumed that the data is up to date, as it is probably directly linked to the data source. However, at the website of the data source²⁰ it is stated that the dataset will be removed in October of this year. By then, the application will no longer be up to date.

The application is based on up to date data.

PUBLICITY

The application is presented at the national open data portal of Austria. Furthermore, the application is mentioned at Twitter once. However, no more publicity was found.

There is (almost) no publicity found for Dog Toilet Map Salzburg.

²⁰ https://www.data.gv.at/katalog/dataset/9d324575-558d-4105-a36e-47d6ba90c6e1

AVAILABILITY OF DATA

The dataset on which Dog Toilet Map Salzburg is based is available. The data of the application itself is not available. It is also not clear how it is integrated exactly in Wikitude.

The data of the application is not available.

AVAILABILITY OF METADATA

There is no metadata provided.

MARKET ACCEPTATION

As there are no numbers available about the usage of the application, it is difficult to determine the market acceptation. However, the application is integrated into another application, which indicates that it is at least accepted by that other party.

The market acceptation cannot be determined.

VALUE CREATION

The application creates at least some value as it is integrated in another application. Moreover, it shows where people can walk their dog. For people who do not know where to walk their dog in Salzburg this application serves as a guide.

The application does provide (at least some) value.

CITIZEN PARTICIPATION

As the data is probably automatically linked to the dataset, there is no need for citizen participation at the moment. However, as stated by 'Ecosystem', this need may arise when the dataset is deleted.

There are no indications found for citizen participation.

COMBINING DATA

The project is based on one dataset, therefore there are no dataset combined.

INITIATED AT A LOCAL LEVEL

The project is initiated in Salzburg, which is still the level of application.

The project is initiated at a local level.

INNOVATIVE PROJECT

The application provides a visualization of the dog toilet locations in Salzburg. As it are just the visualizations of the locations, the application is not considered to be innovative.

The application is not innovative.

LOCATION OF THE PROJECT

The project is located in Salzburg. This limits the number of users, as only people who want to walk their dog and want to know where to find dog toilets within Salzburg might be using the application.

The location puts a limit on the maximum number of users.

D.10 SHINYMIG OÖ

The next project for evaluation is shinyMig $O\ddot{O}^{21}$. Table 26 provides an overview of all metadata that is available in the GOOD DATA database.

DESCRIPTION

shinyMig OÖ is a web application which visualizes migration data for Upper Austria. The data is both visualized in a map, where each color depicts a particular number as well as in a table. It is possible to choose the entire period of 2002-2012 or choose one year to see per area what the migration balance is.

The project is based on two datasets from the national open data portal of Austria. It combines the data of migrations and the areas within Upper Austria. It provides the user an overview of the migration within each area of Upper Austria.

Feature	As in GOOD DATA
Product Type	Other
Business Type	Non profit
Intended Effect	To inform
Geographical Level of Application	Country
Location of Use (meso/macro)	Austria
Location of Use (micro)	Austria
Language	German
Year of Publication	2012
Status	Concept
Business Model	Unknown
Themes (fixed)	Demography
Topic Covered by Data	Recreation, culture and religion
More than one Source	Yes
Sources of Data	Government generated
Collecting Organization	Land Ober Österreich Abteilung Geoinformation und Liegenschaft Land Ober Österreich Abteilung Statistik
Formats	-
Update Frequency	-
Geographical Level of Dataset	Country
Funding	Self funded

Table 26 shinyMig OÖ as in GOOD DATA

²¹ <u>https://donk23.shinyapps.io/shinyMig_UpperAT/</u>

SUCCESS FACTORS

In this section the project of shinyMig OÖ will be described using the list of success factors, as created in section 9.1. This section will be an expanded version of the eleventh column of Table 12.

ECOSYSTEM

Users are not able to contribute to this project. However, if this would be possible, the project might have been up to date, incorporating the data of the last years.

There are no indications of an ecosystem, however, an ecosystem would be beneficial.

REUSE

There are no indications that the data of shinyMig OÖ is being reused.

COLLABORATION

There are no indications found that the developer of shyinyMig OÖ collaborated with other parties.

STRATEGY & LEGISLATION

The application is available under Apache License 2.0. This license takes care of the distribution of the data. In short, the license allows people to reuse the data as long as they give credit to the original source.

The project is accompanied by a license.

CONTINUITY, UP TO DATE DATA

Only the data from the time period 2002-2012 is incorporated in the application. The data of 2013 and 2014 should be available as well at this point in time. Therefore, the application is not up to date.

The project is not up to date.

PUBLICITY

On the national open data portal of Austria, the project is presented. More publicity was not found.

There is no publicity found for shyinyMig OÖ.

AVAILABILITY OF DATA

The raw datasets on which the project is based are available, however, the data of the application itself is not available. Since the application combines two datasets, it would be interesting to see how those are combined exactly.

The data is not available.

AVAILABILITY OF METADATA

There is no metadata available.

MARKET ACCEPTATION

As there are no usage data available, the market acceptation cannot be determined.

VALUE CREATION

The application combines two datasets, by doing so some value is created. However, without involving actual users of the application it is not possible to determine the value creation.

The value creation cannot be determined.

CITIZEN PARTICIPATION

As stated by 'Ecosystem', there is no way for users to contribute to this project in the way it is set up currently.

There are no indications for citizen participation.

COMBINING DATA

Yes. There are two datasets combined for the creation of this project. As stated in the description, a dataset containing the migration figures and a dataset containing the borders of the areas are combined to present this visualization.

Two datasets are combined.

INITIATED AT A LOCAL LEVEL

The project is initiated in Upper Austria, this is still the location of use.

The project is initiated at a local level.

INNOVATIVE PROJECT

The project combines two datasets, which makes it a little bit innovative. However, as the application is just a visualization of that data, the project is not considered to be innovative.

The project is not innovative.

LOCATION OF THE PROJECT

The project is located in Upper Austria. Only people who are interested in the migration figures of the areas within Upper Austria might be using it.

The location limits the maximum number of users.