Intermediation as a dissolver of barriers

How intermediaries can help overcome barriers in open data use

GIMA Master Thesis

Student: Davey den Haan
E-mail: d.denhaan@student.utwente.nl
Student number: s6030963

Supervisor: Lukasz Grus
Responsible professor: Bastiaan van Loenen
Place of research: Wageningen University

Intermediation as a dissolver of barriers

How intermediaries can help overcome barriers in open data use

This master thesis is submitted in partial fulfilment of the requirements for the Msc. program; Geographical Information Management and Applications.

By

Davey den Haan

Student number: s6030963

Date: 1st of March 2018

Supervisor: Lukasz Grus

Responsible professor: Bastiaan van Loenen

Preface

After a near six-month research process, I proudly present the results in this thesis report. It was in September last year when I decided to do my research in the field of open geodata. My research process began with a focus on the dissemination of open data to users. Gradually, this focus shifted towards the different layers in the supply of open data. Eventually, I found the concept of open data intermediaries when I was browsing literature on open data. Once I explored this concept further, my interest began to grow. I realised however that the amount of research on open data intermediaries was rather limited. Nonetheless, I saw the opportunity for a unique research.

I would like to thank my supervisor and responsible professor, Lukasz Grus and Bastiaan van Loenen, for the wonderful guidance during the process. Without their support and feedback this research would not have been possible. I would also like to thank all the people that agreed to participate in the interviews. I would like to thank them for their openness and enthusiasm towards me and the research. Without their participation this research would never have been completed in the first place.

I sincerely hope that my research can be an inspiration for further research on open data intermediaries, as I do believe that they can be a very important factor in unlocking the potential of open data. Furthermore, I hope that my research can be a valuable addition to the body of knowledge of open data and open data intermediaries and to the ongoing research on open data.

Please enjoy the read,

Davey den Haan

Zeist, 28th of February, 2018

Summary

The focus of this research is to add to the body of knowledge of open data by researching the role of intermediaries in the use of open data. Specifically, the goal is to analyse how intermediaries can remove barriers found in the (re) use of open data. Open data intermediaries are believed to play a crucial role in unlocking the full potential of open data, by removing these barriers. However, no research has sought to explore the role of intermediaries on barriers in open data use.

Open data intermediaries are actors positioned at some point in a data supply chain that incorporate an open dataset, are positioned between two agents in the supply chain and facilitates the use of open data that may otherwise not have been the case. In the contemporary theory, six types of intermediaries can be identified: aggregators, communicators, demanders, developers, producers and validators. Aggregators collect data from various sources, to combine with other data to create new datasets or to extract information for clients. Communicators transform open data to other formats, channels and contexts that are matched to the skill and capabilities of their clients. They also provide their clients with information that they derive from (open) data. Demanders can act as communication lines between providers and end users. In this role they can link demand with supply. Developers develop software and applications that can help users use and access data. Producers produce data which they combine with open data. This is done to create new datasets, that suit the needs of their clients better. The last type, validators, assess open data on quality and usefulness for end users. They can improve the quality to enhance the potential of open data for end users.

Several areas have been identified where users can encounter barriers when using open data. These areas are: availability and access, findability, usability, understandability, reliability and quality, linking and combining of data and lastly communication with the providers of open data.

Interviews were held to explore the relationship between intermediary types and the areas where they can remove barriers. Five interviews were held in total with organisations that have an intermediating function.

It was found that aggregators affect the findability of the open data and the problems with linking and combining it with other data. Communicators improve the usability and understandability as well as linking and combining data. Demanders only improve the communication between providers and users. Developers, improve the accessibility and availability of data and the findability. Their software also improves the usability and the understandability of the data. Furthermore, their software also helps users to link and combine data more easily. Producers improve the quality and reliability of open data by combining their self-created data with open data. The last type, validators only had a noticeable link with the quality and reliability of data. Users often find it difficult to determine whether the quality is sufficient, and validators can do this for them. The measures undertaken to improve the quality when the quality is deemed to be insufficient, is outside the role of the validator. Quality could for example be improved by doing additional measurements, in this case the intermediary shifts from a validator to a producer. To conclude, the findings also indicate that there is a type of intermediary that is not described in the literature. This is the 'educator' type of intermediary. Educators provide training and education to end users to improve their knowledge and skills. This can enhance their capabilities to use open data.

Table of Contents

1. Introduction

1.1 Background

The phenomenon of open data has seen an exponential rise in attention over the last years. Often, open data is said to have nearly limitless potential; from enhancing economies, creating transparency and even solving societal problems (Welle Donker & van Loenen, 2016a). Because of these potential benefits, open data has increased rapidly in importance for many organisations, both public and private. Such organisations are even finding themselves under increasing pressure to publish their data openly (Janssen et al., 2012). However, the realisation of the full potential of open data has proven to be difficult in practice and has yet to materialise (Janssen et al., 2012; Welle Donker & van Loenen, 2016a; Zuiderwijk et al., 2012). Research that has been done to find the causes of open data failing to reach its full potential (Janssen et al., 2012) found that open data faces a variety of problems, both on the demand and on the supply side. For example, problems on the supply side include; data quality, the costs of opening-up closed data or lack of capability to sustain open data initiatives (Janssen et al., 2012). On the other hand, problems on the demand side were also found. These include for example, accessibility, data literacy, lack of capabilities to use open data and the inability to create new value with open data (Janssen et al., 2012; Zuiderwijk et al., 2012).

1.2 Problem description

Researchers have advocated the importance of intermediaries in overcoming the barriers experienced by users (González-Zapata & Heeks, 2015; Magalhaes et al., 2013; van Schalkwyk et al., 2015, 2016). Van Schalkwyk et al. (2015) noted that the role of intermediaries in the open data ecosystem is insufficiently explored. Nonetheless, they indicate that intermediaries could play a crucial role in unlocking the full potential of open data. One way in which they can unlock this potential is by removing the barriers faced by users such as insufficient knowledge or skills of users to use open data (Zuiderwijk et al., 2012). Intermediaries share many similarities with so called 'keystone species' of ecosystems (van Schalkwyk et al., 2015). Keystone species are actors of an ecosystem that act as mediators in ecosystems by providing additional services, tools, platforms or technologies that can benefit other actors of the ecosystem (Iansiti & Levien, 2004). However, there is limited knowledge and little empirical research on the role of intermediaries in overcoming the barriers found in open data use. Most of these research on open data intermediaries focus on their business models and describing

different variations in intermediaries (González-Zapata & Heeks, 2015; Janssen & Zuiderwijk, 2014; van Schalkwyk et al., 2015, 2016). The problem is that there is no research analysing how intermediaries can remove barriers found in open data use.

In addition, the role of intermediaries has been described in the ecosystem of developing countries (van Schalkwyk et al., 2015), which have less matured open data ecosystems as can be seen in more developed countries, or none at all. The impact and importance of open data intermediaries has not been explored much in the context of a country that has a relatively developed open data ecosystem (van Schalkwyk et al., 2015). The innovation of this research is therefore aimed at analysing intermediaries in a country that scores relatively well in open data compared to other countries, especially the geo-information sector (Open Data Barometer, 2016). Furthermore, previous research on open data intermediaries has not focused specifically on open geo-data. Rather, research has focused on open data general (Dumpawar, 2015; Janssen & Zuiderwijk, 2014; van Schalkwyk et al., 2015, 2016).

1.3 Goal, objectives and research questions

The focus of this research is to add to the body of knowledge of open data intermediaries by researching the role of intermediaries in the use of open data. The contemporary theory on intermediaries limits itself to classifications of intermediaries, how they can have an effect on the use of open data is not known. The overall goal of the research is to evaluate the roles of intermediaries in the open data ecosystem and to analyse how they can remove barriers found in the (re) use of open data.

To achieve this goal the following research questions will be answered:

1. What are open data intermediaries?
2. What types of open data intermediaries can be identified?
3. What barriers can be found in the use of open data?

The first three research questions are preliminary research questions. These will be answered as a preparation for the case study, the empirical phase of the research. The result of these research questions is a theoretical framework which will be applied in the empirical phase of the research. In this phase the following research question are answered:

4. What types of intermediaries can be identified in the AHN?
5. How do intermediaries help overcome barriers found in the use of the AHN?

1.4 Research case

The concept of open data ecosystems is relatively new and while the ecosystem approach has been used in other fields, the approach with regard to open data is still in its infancy (Heimstädt et al., 2014). In addition, the boundaries of an open data ecosystem can range from a global ecosystem, encompassing all possible open datasets and actors to an internal open data ecosystem within an organisation. Limiting the complexity of the open data concept is important to establish a scope which will ensure that the research is controllable and executable. To establish a scope, a specific case is selected.

The selected case is the AHN, *Actueel Hoogtebestand Nederland.* The AHN is a 3D digital height map covering the whole of the Netherlands, containing the height in relation to the NAP or the Amsterdam Ordnance Datum (Rijkwaterstaat, 2017a). In the past the AHN used to be closed data. Since March 6th 2014, the AHN has been made openly available (AHN, 2015), allowing anybody to create products and services with the data (AHN, 2015; A. K. Bregt et al., 2016; Kadaster, 2017; Rijkwaterstaat, 2017b).

Research has been performed to assess the effects of 'opening-up' the AHN (A. K. Bregt et al., 2016). The research focused on three types of effects: internal effects, external effects and relation effects. Internal effects relate to effect for the organisations that provide the data, external effects are effects on the society and relation effects are effects on the relations between the providers and the users.

For the AHN, it was found that the majority (53%) of users are business organisations, the second largest group were government organisations (22%), education and research (17%) and private users (8%) (A. K. Bregt et al., 2016). Interestingly the research also concluded that users receive support in using the AHN from 3rd parties (A. K. Bregt et al., 2016). In addition, it was found that the role of the providers of the AHN has changed since it became open data. It was for example seen that, ironically, the providers were no longer actually delivering the data to users, but that data was acquired from other sources, such as PDOK (A. K. Bregt et al., 2016).

The results of the research on the effects of opening-up the AHN indicate that the role of the providers has changed. This could also indicate that 3rd parties, like intermediaries, have taken on some of the roles and responsibilities which were previously held by providers. For example, the delivery of data and the support in using the data after it is supplied to the user.

1.5 Research scope

The scope will ensure that the research is do-able as well as focused on achieving its goals. The research will focus only on open data intermediaries found in the ecosystems of the AHN. This implies that the research will focus exclusively on the geo information sector of open data. Nevertheless, the results of this research can potentially be relevant for other sectors of open data as well and outside its original scope. Lastly, measuring quantitative effects is also not included in scope of the research, the aim is to expose relationships between intermediaries and barriers found in the literature on open data.

1.6 Scientific and societal relevance

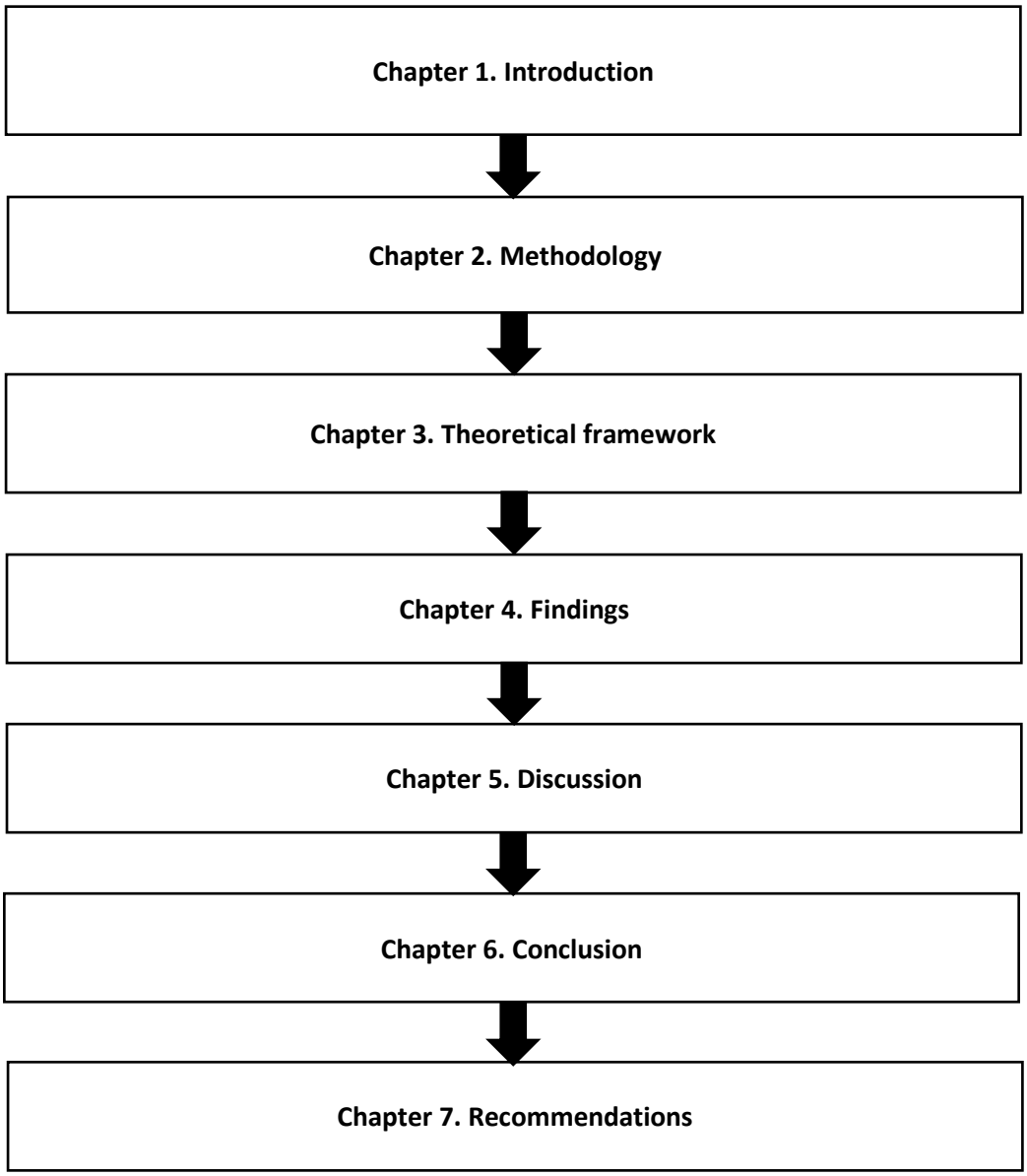
This research can contribute to the body knowledge on open data intermediaries and their role in the open data ecosystem. Currently, no research has been done to find relations between open data intermediaries and barriers of open data use. This research can identify these relationships so that they can be further analysed in the future. This research therefore strives to be a starting point for future research regarding open data intermediaries. Additionally, the research can contribute to the general body of knowledge of open data, as intermediaries could play a crucial role in unlocking its full potential.

From a societal perspective, it is important to analyse whether intermediaries can help overcome the barriers faced by open data. By overcoming these barriers, more people have access to open data and more people will be able to use open data. Open data intermediaries can also help to make open data available and accessible to a broader target group.

1.7 Thesis structure

Figure 1.1 resembles the structure of the thesis. In the second chapter, it will be discussed whether a qualitative or a quantitative approach is chosen and why. Additionally, it will describe the adopted methods of the research. In the third chapter a theoretical framework is created, to serve as a foundation for the research. The fourth chapter will present the results of the analysis. These results are evaluated in chapter five. The second to last chapter will conclude the research by answering the research questions. The last chapter will provide recommendations for future research.

Figure 1.1: Thesis structure



2. Methodology

2.1 Research approach

In general, it is seen that there are options when choosing a approach for scientific research: the quantitative and the qualitative approach.

In quantitative research, one strives to test hypotheses on a large sample size in order to generalize the results to the larger population. If the research would have a quantitative approach, this would imply that there are a large number of organisations or people being studied. The goal would then be to study patterns and trends (Mason, 2002).

In qualitative research, the focus is on studying a smaller number of subjects, but more in depth. Unlike quantitative research, qualitative research focus on textual information rather than numerical information (Mason, 2002). This approach is better suited for studying new subjects. However, unlike quantitative research, results of qualitative research are harder to generalize on the total population of the subject. The research on how intermediaries affect the open data use is a relatively new topic. Subsequently, little research has been performed to analyse intermediaries in an open data ecosystem (van Schalkwyk et al., 2015). Because the research focuses on a relatively new concept, qualitative methods are more appropriate (Mason, 2002).

The overall approach adopted by the research is the case study. Case studies are mostly used in research where *how* and *why* questions are asked, or *what* questions which do not require numerical answers (Yin, 2003). Such questions are often exploratory in nature and require the analysis of connections and operational links between entities, rather than frequencies or values (Yin, 2003). Case studies are also used to study non-historical events or phenomenon, where the researcher cannot influence the behaviour of the actors or the context of the phenomenon to be studied (Yin, 2003).

The case study resembles more of a toolbox; an all-encompassing method which covers the design, data collection and data analysis of the research. The case study is therefore not just a method, but a comprehensive research strategy. Case studies can encompass a variety of different data collection techniques, the choice of selecting collection methods is dependent on the research questions asked. As this research focusses on gaining more knowledge of open data intermediaries, the decision is made to adopt interviews as the main data collecting

method in the case studies. Interviews allow for deeper exploration of subjects, resulting in an in-depth understanding of the subject rather than a understanding of the patterns within the subject (Mason, 2002).

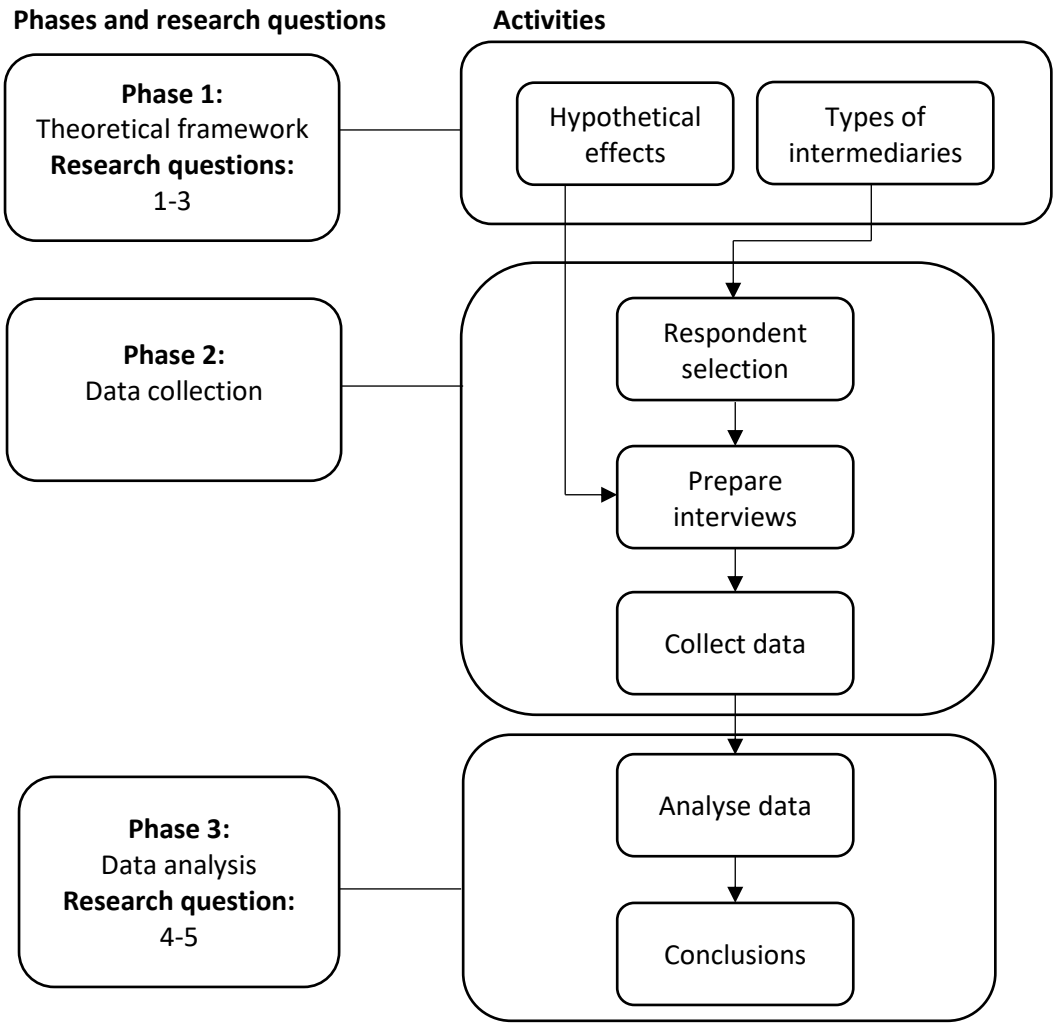
2.2 Research Design

The research was executed in three distinct phases. In the first phase, the theoretical framework was created. This first phase also provided the answers to the first three research questions. The theoretical framework serves as the main input in the second phase. The development of the theoretical framework is further discussed in section 2.3.

During the second phase, the data was collected for the empirical analysis. The data was collected by means of interviews. The concepts of the theoretical framework served as the main input for the respondent selection as well as for the creation of the interview topic list. Details regarding the interview process are discussed in section 2.4.

The last phase of the research was to analyse the data collected in phase two. The results of the analysis provide the answers to research question four and five. The process of the analysis is discussed in section 2.5.

Figure 2.1: Research design



2.3 Methodology for developing the theoretical framework

The theoretical framework serves as the foundation of the research. Its goal is to categorise and describe the concepts that are relevant for the research and the relationships between them. The theoretical framework also provides a reference point for the interpretation of the findings (Rocco & Plakhotnik, 2009)

The theoretical framework was developed by conducting a systematic literature review. The used method the systematic literature review was adopted from Khan et al. (2003). The process of conducting a systematic literature review consist out of five steps, according to Khan et al. (2003). These are:

1. Framing questions for the review
2. Identifying relevant work
3. Assessing the quality of the studies
4. Summarizing the literature
5. Interpret the findings

For the first step, the questions for the literature review where the first three research questions: 'What are open data intermediaries?', 'What types of intermediaries can be identified?' and 'What barriers can be found in the use of open data?'.

A literature search has been done to find the relevant work relating to the research problem and the three research questions. Literature was found using online scientific databases such as; Web of Science, Scopes and google Scholar. In addition, Google Search was used to find literature and documents outside the scientific databases. For the literature search, keywords were used to search the databases. Examples of keywords used are:

- 'Open data intermediaries'
- 'Data intermediaries'
- 'Open data impediments'
- 'Open data intermediation'
- 'Open geodata intermediaries'
- 'Open data broker'
- 'Intermediaries + open data use'
- 'Open data ecosystem'
- 'Open data barriers'
- 'OGD Intermediaries'
- 'Open geodata'
- 'Open data mediator'
- 'Data broker'
- 'Intermediary + effects + open data'

The process of searching for suitable literature was ongoing during the development of the theoretical framework. However, the literature review was performed prior to the second and third phase of the research. Subsequently, no articles published after the 1st of February 2018 were included in the development of the theoretical framework.

The literature search resulted in a collection of articles and books, these were judged on relevance by their titles, abstracts and by scanning the text. The references of the found literature were also consulted and the snow-ball method was then used to find additional sources. In total, seven publications were found that were relevant the first concept of the theoretical framework, the typology of the intermediaries. An overview of the articles is found in table 2.1.

Table 2.1: Articles used in the classification of intermediaries

#	Authors	Title
1	(Deloitte LLP, 2012)	Open growth, Stimulating demand for open data in the UK A briefing note from Deloitte Analytics
2	(Dumpawar, 2015)	Open Government Data Intermediaries: Mediating Data to Drive Changes in the Built Environment
3	(González-Zapata & Heeks, 2015)	Understanding Multiple Roles of Intermediaries in Open Government Data
4	(Janssen & Zuiderwijk, 2014)	Infomediary Business Models for Connecting Open Data Providers and Users
5	(Magalhaes et al., 2013)	Open government data intermediaries
6	(Welle Donker & van Loenen, 2016b)	Sustainable Business Models for Public Sector Open Data Providers
7	(Vancauwenberghe et al., 2017)	Business Models for Geographic Information
8	(Ponte, 2015)	Enabling an Open Data Ecosystem

For the second concept, the barriers of open data use, three publications were used (table 2.2). In their research, Zuiderwijk et al. (2012) made a synthesis of barriers of open data by using three sources: interviews, workshops and a literature review. The literature review included various publications on open data barriers. The result of their research was an extensive list of barriers in different areas. This list has provided the bulk of the barriers used in the theoretical framework.

Table 2.2: Literature used for the barriers in open data

#	Authors	Title
1	(Zuiderwijk et al., 2012)	Socio-Technical Impediments of Open Data
2	(Janssen et al., 2012)	Benefits, Adoption Barriers and Myths of Open Data and Open Government
3	(Welle Donker & van Loenen, 2016a)	How to assess the success of the open data ecosystem?

The publications were summarized, and the interpretation was reported narratively in chapter 3. This chapter will first start with an introduction to open data ecosystems. Next the findings of the systematic literature review are narratively presented for the two concepts: the types of intermediaries and the barriers on open data use. The relationships between the two concepts that were found in the systematic review is schematically represented in the conceptual model in section 3.3.

2.4 Interview methodology

2.4.1 Respondent selection

Interviews were held with organisations that fit one the roles of intermediaries found in the literature. To determine whether an organisation is eligible for selection, the theoretical framework is combined with web research. In the research of A. K. Bregt et al. (2016), several examples were given for AHN applying organisations. Their examples also included organisations that provide services and products for their clients that enable them to work with the AHN. These organisations were the start of identifying intermediaries for the AHN. The web research was performed by using Google Search as well as consulting the member list of the AHN group on LinkedIn. Keywords used for the search are (translated from Dutch):

- 'AHN + services'
- 'AHN + projects'
- 'AHN + products'

Eight organisations were identified using web research (Table 2.3). These eight organisations met the working definition provided in the literature and fit one or more of the different types found in the literature, based on information on the products and services they provide on their websites. These eight organisations were approached for interviews via email. Five

organisations responded positively to the invitations, from the other three no response was received. The five organisations that accepted the invitation were; Esri Nederland, Geodan, Geonext, IntellinQ and Steggink Geo-ICT.

Table 2.3: Overview of intermediary organisations found using web-research

Organisation	**Description**
Esri Nederland	Software developer and supplier of (open) data to their software users.
Geodan	Creator of data and software
Geonext	Specialised in 3D geodata
InfraCAD	Provides software to integrate geoinformation/geodata with AutoCAD.
Mapservices	Provides height information based on the AHN. Also creates applications/software.
Swartvast	Advocacy bureau in the use of (open) geodata.
Steggink Geo-ICT	Freelancer providing services in processing data and advocacy in using (open) geodata.
IntellinQ	Developer of spatial database management software.

2.4.2 Interview process

The interviews were scheduled in period of two weeks. One interview was held on February 7th, two on the 8th, one on the 13th of February and the last one on the 15th of February. Four interviews were held face-to-face, and one interview was conducted by phone. All interviews were held with individuals working at the selected organisations. Respondents were asked permission to record the interview, these requests were accepted. Respondents were also asked permission to use their name or their organisation's name in the research. In the end it was decided not the use the real names of the respondents and quotations, only paraphrasing. The interviews lasted between 30 and 50 minutes.

Before the interviews were held, a list was containing the topics and possible questions. The interviews were semi-structured; some pre-determined questions were asked but the goal was to let the respondents talk freely as much as possible. The overall structure of the interviews was created for two purposes. The first part of the interview was focused at linking the organisation in question to a specific intermediary type. The second part was focused on the areas of effect found in the theoretical framework.

However, while the structure was predetermined, the order in which the questions were asked was not pre-determined. All interviews would start with a description of the organisation and their business. Often, in the descriptions of the organisations links were found with barriers. Therefore, follow-up questions were asked regarding these areas of effect. For the list of topics and question, please refer to appendix 1. It should be noted however that the formulation of the possible questions on the topic list will be different from the formulation during the interviews. The respondents would often give long and detailed answers to questions, which created many opportunities to ask follow-up questions. Often, in the answer to a question regarding an area of effect, links could be noticed with other areas of effect. The interviews went fluidly in all cases and the flow resembled more of an actual conversation. The interviews were ended when all topics on the list were explored. To conclude the interview, respondents were asked if there were any aspects regarding open data use and their organisations that have not been discussed yet. When there were no further aspects to be discussed, the interviews were completed.

2.5 Data and analysis

Interviews were recorded using mobile phone and stored in mp3 format by the researcher. Interviews were then transcribed to textual formats by means of summaries.

The data was analysed based on the topics of the interview. First, each organisation was linked to one or more intermediary types. Then, each intermediary was linked to one or more data practices. For each of the areas of effects, information was collected on how the intermediaries can have an effect on the barriers for their clients. It is also possible that an intermediary has no effect on an area of effect. Furthermore, for each of the areas of effect it is analysed why certain intermediaries have an effect in that specific area. In addition, comparisons were made between the different types of intermediaries. For example, in one area of effect more than one organisation can have an influence, which also means different

types of intermediaries have an effect on the same area. Lastly links between types of intermediaries and their areas of effect were collected.

2.6 Validity and reliability

Validity and relevance are the two main criteria to assess qualitative research. Validity refers to whether one has measured what one wants measure. Reliability on the other hand refers to whether what you have measured is true.

With respect to validity, interviews were held under the same circumstances if possible, with minimal interference from external factors such as noise. In addition, for each interview the pre-determined topic list and structure was used, to ensure that all the areas of effects were covered, even if the organisation in question did not have an observable effect in some areas.

With respect to reliability, all types of intermediaries found in the theoretical framework were covered by the five interviewed organisations. Also, the organisations that were interviewed varied in size and core businesses. Each of the organisations are completely different form one another in this regard. Therefore, the selection of organisations captured a wide range. The results of the interviews were also shared with the persons that were interviewed before the research was completed, for validation. This was done to ensure that there were no inaccuracies in the results.

3. Theoretical framework

3.1 The ecosystem approach to open data

This chapter will provide the foundation for the research, the theoretical framework. It is important to adopt a certain theoretical approach to analyse how intermediaries affect open data ecosystems, as it is necessary to adopt a framework which allows for the analysis of supply and use of data by actors (Heimstädt et al., 2014). The ecosystem approach to open data allows this and is therefore used as the theoretical viewpoint for the research. Within this approach to open data, several elements are important for the research. The first step in understanding and establishing the theoretical framework of the research is to define the concept of an 'open data ecosystem'.

The ecosystem analogy knows many variants and there is not a uniform definition of an open data ecosystem. Heimstädt et al. (2014), concluded that in the literature digital ecosystems were contextualized as cyclical and sustainable environments, that are demand-driven and oriented around different agents that are mutually interdependent for the creation of value. This definition of an open data ecosystem describes it as an environment where the actors are interdependent for the creation of value with open data. In ecosystems, there is a central resource that flows in between the actors. In the open data ecosystem this resources is open data (Heimstädt et al., 2014). Normally, one would see that the demand regulates the supply. However, in an open data ecosystem, there is a natural monopoly of the supply side (Heimstädt et al., 2014). This is because open data is often made available not according to demand, but according to the suppliers will to do so (Dumpawar, 2015; Heimstädt et al., 2014).

Within an open data ecosystem there are different interdependent actors. Each of those groups of actors have specific roles and purposes to sustain the ecosystem (Heimstädt et al., 2014). From an open data point of view, the actors found in the ecosystem are providers, intermediaries and users (Dumpawar, 2015; Heimstädt et al., 2014). Providers are organisations that create data and who make their data openly available. As their names suggests, users are actors that take the open data and use it for example in their work or to create new products and services or improve existing ones (Dumpawar, 2015). Intermediaries will be further discussed in section 3.2.

According to the ecosystem analogy, the central resource, open data, flows from providers to users (Heimstädt et al., 2014). In its simplest form data flows directly from the provider to the user (Heimstädt et al., 2014). When the central resource flows from supply to users, users create value with the open data by using it (Heimstädt et al., 2014; van Schalkwyk et al., 2016). The creation of value could for example be realised by creating products, new services or enhancing existing services (Janssen & Zuiderwijk, 2014). This flow of data along a supply chain is also referred to as a data value chain or information value chain (Heimstädt et al., 2014; van Schalkwyk et al., 2016). However, the flow of data and the use of data is often limited by barriers (Janssen et al., 2012). These barriers indicate a mismatch in what is supplied and what users need. This implies that there are user needs which are not addressed. This is logically the result of the openness of open data, since open data is available to everyone, the group of (potential) users is difficult to define. Therefore it is impossible for providers to address the needs of all (potential) users (Welle Donker & van Loenen, 2016b). This has provided room for intermediaries in the open data market (Dumpawar, 2015; Janssen & Zuiderwijk, 2014).

3.2 Open data Intermediaries

3.2.1 Biased open data

Data is often made openly available with the idea that 'it is available for everyone' (Dumpawar, 2015). This paradigm is especially salient in open government data. However, using open data requires a certain degree of skill and knowledge. The skills and knowledge required to use open data is only possessed by a select group of people (Dumpawar, 2015). The paradigm of making open data available with the idea to make data available to everyone is therefore an illusion (Dumpawar, 2015). In reality, there is only a small group of people, or organisations for that matter, able to use the open data because they possess the required skills and tools to analyse, download and process the open data (Dumpawar, 2015; Zuiderwijk et al., 2012). Therefore, the majority of (potential) users can only work with open data very minimally.

While open data may be driven by the philosophy of access for everyone, in reality there is a digital divide, that Dumpawar (2015) referred to as the 'data divide'. Often, open data initiatives are merely delivery tools, designed to disseminate the open data. However, as Davies (2011) argues, that it is more important to provide technical, social and political resources, in addition to open data, which can be used to support a range of activities around open datasets. This has sparkled the emergence of actors that have a mediation function, to take the responsibility of mobilizing these activities around open datasets (Dumpawar, 2015). These actors are intermediaries.

3.2.2 Intermediaries outside open data

Part of the goal of the research is to address the limited knowledge on intermediaries in open data. There has been research performed about open data intermediaries, but their findings are fragmented (Dumpawar, 2015; González-Zapata & Heeks, 2015; Janssen & Zuiderwijk, 2014; Magalhaes et al., 2013; van Schalkwyk et al., 2016). The result of this fragmentation is that there is not a uniform definition or framework to analyse open data intermediaries. To gain a better understanding of open data intermediaries, it might be beneficial to first look at intermediaries outside open data.

Intermediaries are not new in other fields where an ecosystem approach has been applied. In 1997, Hagel III & Rayport predicted the rise of intermediaries on the information market, as they believed people would take ownership of information about their consumer behaviour. They predicted that companies will buy that information from consumers in the future. They

argued that because consumers do not have time, patience or ability to work out the best deals with information buyers on their own, intermediaries will emerge to negotiate on their behalf with data buyers. This prediction arose from the idea that people would become owners of their own personal data and that many organisations would be prepared to buy this data from the people. Intermediaries in this context act as brokers between data users and providers.

The term infomediary is more common to describe organisations that fulfil an intermediating role between the data providers and users. For example, Sawicki & Craig (1996) describe intermediaries as information providers, giving community groups access to information based on data they cannot work with. In this definition, it is seen that intermediaries operate between a data provider and a user. Their roles can differ, not only do they act as brokers of data (Hagel III & Rayport, 1997), but also as service providers, extracting information from data and delivering it to users, who are not capable to extract this information themselves (Sawicki & Craig, 1996).

3.2.3 Defining 'the open data intermediary'

This section will answer research question 1: *What are open data intermediaries?* To answer this question, the open data intermediary needs to be separated from the intermediary in general. What distinguishes an open data intermediary from an intermediary in general is the engagement with open data. Engaging with open data in this context means that the intermediary actor accesses, delivers or processes the data in some way (van Schalkwyk et al., 2015). Van Schalkwyk et al. (2015), stated that an intermediary does not have to supply or access open data to facilitate the use of open data. It is only implied that open data is present somewhere in the supply chain (van Schalkwyk et al., 2015). However, the products and services created by the intermediaries do not have to be open as well (van Schalkwyk et al., 2015). On the contrary, intermediaries often provide these services and deliver these products for fees to their clients, as seen in their business models (Janssen & Zuiderwijk, 2014). These actors that have taken on the role of intermediaries are actors that operate between the providers and the users.

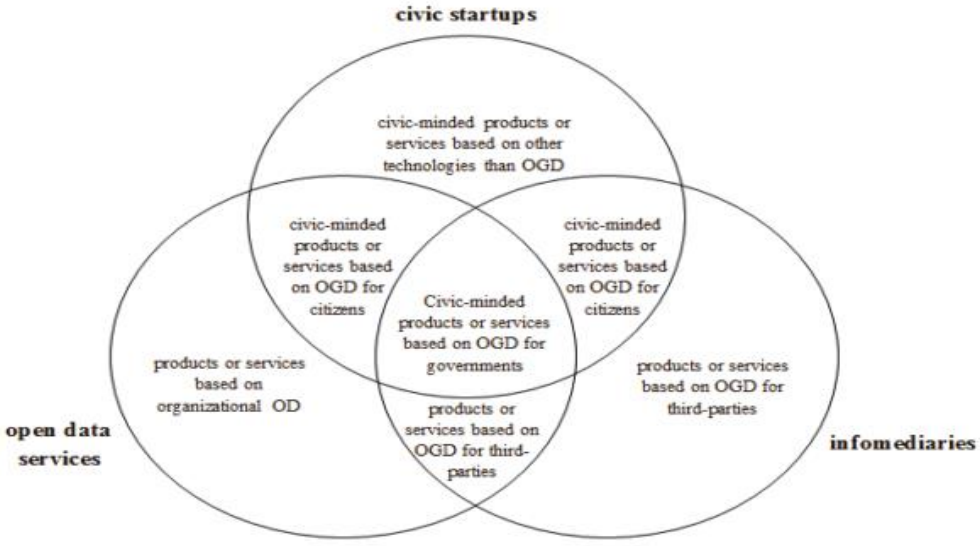
Based on this, van Schalkwyk et al. (2015) formulated a working definition of open data intermediaries: **'An open data intermediary is an actor positioned at some point in a data supply chain that incorporates an open dataset, is positioned between two agents in the supply chain and facilitates the use of open data that may otherwise not have been the case'**. This definition encompasses the general purpose of an intermediary and separates the open data intermediary from general intermediary. Although, it is too ambiguous to be made operational in the case study. Therefore, the next section will dissect the open data intermediary in different subtypes.

3.2.4 Classifying open data intermediaries

This section will provide an answer to research question 2: *What types of open data intermediaries can be identified?* Several researches have been conducted to identify roles of actors in the open data ecosystem and to classify them. Some of these focused specifically on intermediaries (González-Zapata & Heeks, 2015; Janssen & Zuiderwijk, 2014; Magalhaes et al., 2013), while others used a classification of open data actors in general (Deloitte LLP, 2012; Vancauwenberghe et al., 2017; Welle Donker & van Loenen, 2016b). Nonetheless, some actors they identify are consistent with the working definition of open data intermediaries as provided by van Schalkwyk et al. (2015).

Magalhaes et al. (2013) created a framework in which they organize intermediaries into three categories: civic startups, open data services and infomediaries (Figure 3.1).

Figure 3.1: Terminology framework for intermediaries



Source: (Magalhaes et al., 2013)

The main point of this classification is that there is overlap between the three categories (figure 3.1), the researchers then analyse these overlaps. The overlap in this case means that the three types of intermediaries have similar products and services. This makes linking the intermediaries to a specific type more difficult. In addition, it makes identifying the relationship between the intermediary and the barriers of open data use difficult.

Some researchers (Ponte, 2015; Welle Donker & van Loenen, 2016b) used the classification originally created by Deloitte LLP (2012). This classification is aimed at creating a typology of actors in the open data ecosystem. Hence, not all actors identified by Deloitte (2012) fit the working definition as provided in section 3.2.3. Nonetheless, there are three types of actors identified by Deloitte (2012) that fit the working definition of an open data intermediary:

- Aggregators
- Developers
- Enablers

The first type, ***aggregators***, are actors that collect and aggregate open data from multiple sources. Here, suppliers can apply to make their data available and users can download data from the aggregator. In addition, they often provide additional functionalities such as viewers and query tools. Aggregators can also be

The second type, ***developers***, built, design and develop applications and tools that help end-users use open data. Typically, these applications combine multiple open datasets. These applications help to improve the usability and accessibility of the open data for users.

The third type, ***enablers***, are sometimes referred to as intermediaries in other work of research (Ponte, 2015; Welle Donker & van Loenen, 2016b). These actors often fulfil the stereotypical role of intermediaries: facilitating the supply and use of open data. They do typically not use the open data themselves but *enable* (end)users to use open data. It is important to note that enablers do not only service the demand-side, as suppliers can also benefit from the enabler's role of facilitating open data.

However, when considering other sources, problems arise. Vancauwenberghe et al. (2017) divide the group of enablers further into three sub-categories:

- Supply facilitators

- Access facilitators

- Service facilitators.

Supply facilitators facilitate the dissemination of open data by providing technologies or services to provides (Vancauwenberghe et al., 2017). They can also help providers in managing their data and in the publication of open data. The opposite of this are the **access facilitators.** They facilitate the access of open data by users by providing services and products. **The service facilitators** create products and services for specific target groups to offer tailor-made solutions to their needs and requirements in open data use. What is paradoxical, is that Vancauwenberghe et al. (2017) identify aggregators as a sub-type of access facilitator and developers as a sub-type of service facilitators. Both Deloitte (2012) and Vancauwenberghe et al. (2017) refer to aggregators, developers and enablers. However, while Deloitte (2012) regards them as separate entities, Vancauwenberghe et al. (2017) see developers and aggregators as a sub-type of enablers. Another example of the disparity in research is that (Ponte, 2015) classifies developers and aggregators as intermediaries but identify enablers as an individual entity, existing separately from intermediaries. In addition, other researchers add new types in addition to the aggregators, developers and enablers. For example, González-Zapata & Heeks (2015) identified different types of intermediaries. They too identified developers, but also add new types:

- Demanders
- Producers
- Communicators
- Validators

Demanders are actors that stimulate the release of open data by providers, by analysing demand from the users and requesting specific datasets to be released. In addition they can also lobby for new open data policies to be developed in by providers (González-Zapata & Heeks, 2015).

Producers seem paradoxical, but according to González-Zapata & Heeks (2015), these are intermediaries that collect additional data in the field and combine it with open data to offer users new open data.

Communicators are intermediaries that present the open data in formats, channels and contexts that are matched to the skills and capabilities of their clients. Communicators take open data and turn them into information which is meaningful and consumable for the user. In this way they act as infomediaries and facilitate the use of open data in the form of information.

Validators are actors that assess open data on their quality but also on their usefulness. These actors validate the data to help improve the accuracy, quality and potential of the open data.

The problem with the framework of Magalhaes et al. (2013) was that their classifications of intermediaries showed overlap between the three types of intermediaries in the products and services they offer. Due to this overlap there is a risk of different types of intermediaries sharing functions and roles.

This is also seen in the classification provided by González-Zapata & Heeks (2015). The intermediaries they classified as communicators, show similarities in roles with the enablers as identified by Deloitte (2012) and with the access facilitators of Vancauwenberghe et al. (2017). The main issue here is the group categorised as enablers by Deloitte (2012). When considering the definition of enablers: actors that facilitate the use and supply of data, aggregators and developers also meet this definition, as Vancauwenberghe et al. (2017) have indicated. Essentially, the function of enabler is inherently linked to the nature of intermediaries; intermediaries enable open data. Hence, some authors use the term 'intermediary' when referring to enablers (Vancauwenberghe et al., 2017; Welle Donker & van Loenen, 2016b). However, enablers are not necessarily positioned between the data suppliers and users (Vancauwenberghe et al., 2017).

The class of enablers is subdivided as is done by Vancauwenberghe et al. (2017). In addition, the classfication will be expanded with the types of open intermedies proposed by González-Zapata & Heeks (2015). This result the classification provided in table 3.1.

Table 3.1: Types open data intermediaries

Type	**Role**
Aggregator	Aggregating open data from multiple sources, for supply to users or to extract information from multiple data sources.
Communicator	Present open data in formats, channels and contexts that are matched to the skills and capabilities of their clients.
Demander	Stimulate the release or alteration of open data by providers, by aggregating demand and requesting specific datasets to be released.
Developer	Built, design and develop applications and technologies that help end-users use and access open data.
Producer	Collect additional data in the field and combines it with open data to offer users new open data, that suits their needs better
Validator	Assesses open data on quality and usefulness, to improve accuracy, usefulness and quality of the data. This improving its potential.

3.2.5 The functions of intermediaries

Now that a typology is created to classify the types of intermediaries in open data, it can be looked at the functions these intermediaries have in the open data ecosystem. These functions are represented in the practices of open data intermediaries as proposed by Dumpawar (2015). Dumpawar (2015) lists the following data practices that are supported by intermediaries:

- **Practices of Aggregation:** Intermediaries can collect, combine, link and filter data. Often from multiple sources.

- **Practices of Rectification:** Intermediaries can address the inaccuracies, completeness and obsolescence of open data by validation, updating and correcting datasets.

- **Practices of Interpretation:** Intermediaries can analyse open data to reveal biases, assumptions and to extract information.

- **Practices of Representation:** Intermediaries can visualise and contextualise open data, to match the skills of the users.

- **Practices of Dissemination:** Intermediaries can release open data in a reusable way or develop technologies to help users access and use open data.

- **Practices of Augmentation:** Intermediaries can address gaps and limitations in open data by augmenting and annotating existing open data, or by creating new data and combining it with existing data.

- **Practices of Connection:** Intermediaries can support interactions between users and suppliers.

Van Schalkwyk et al. (2015), found that there are multiple layers of intermediation in an ecosystem and supply chain. These layers are related to the capital that intermediaries possess. Capital in this context is what intermediaries specialize in. Van Schalkwyk et al. (2015), list three types of capital: technical, social and cultural/symbolic capital. Technical capital refers to the facilitation of using and working with the data, for example transforming data into other formats or creating tools and applications. Social capital refers to the ability to connect with users and potential users. Cultural/symbolic capital refers to actors that, for example, pressure governments to make their data openly available or. When users have a deficit of these capitals, intermediaries use their capital to complement the deficits of the users (van Schalkwyk et al., 2015). Most importantly, they found that within supply chains more than one intermediary may be present, fulfilling different roles and utilizing different capitals (van Schalkwyk et al., 2015). This can also mean that more than one type of intermediary is present in a single supply chain or that one actor can fulfil multiple roles if it possesses more than one type of capital. Therefore, multiple types of intermediaries might be present between a provider and a user.

3.2.6 Barriers found in open data use

This section will answer research question 3: *What barriers are found in the use of open data?* Welle Donker & van Loenen (2016a) have developed a holistic assessment framework to asses open data, based on existing assessment frameworks. They focused on three elements of success of open data: supply indicators, governance indicators and user indicators. The following areas were selected from the assessment framework provided by Welle Donker & van Loenen (2016a):

- Recognisability: the open data needs to be identifiable by users, for example by search engines.
- Findability: The dataset must not only be identifiable by the user it also has to be found by them, for them to download it.
- Reliability: The data has to be of adequate quality, also metadata has to be present and the data should be published in a reliable way.
- Clarity: Not all users have sufficient expertise to use the data, as it is not clear to them how to use the data. Or they do not have the knowledge or skills to understand the data.
- Manageability: users should be able to use data with the resources available and for the purpose/goal they have intended. The data has to be manageable by the user for the goal the user wants to use the data for.
- Self-organising capacity: This is the capacity to match supply with demand.
- Communication: Communication between users and providers for support with the access and use open datasets.

Zuiderwijk et al. (2012) have performed research to analyse possible barriers making the use of open data by users difficult. They identified barriers in the following categories:

- Availability and access
- Findability
- Usability
- Understandability
- Quality
- Linking and combining data
- Interaction with providers

Both the categories of Welle Donker & van Loenen (2016a) and Zuiderwijk et al. (2012) show similarities. Both list findability and reliability and quality are related. Manageability is similar to usability and clarity shows similarity with the understandability. Lastly, the recognisability is also included in the availability and access and the interaction. Zuiderwijk et al. (2012) also included the ability to link and combine data with other open data. Based on these researches, barriers can occur in the following areas:

- Availability and accessibility of data
- Findability of data
- Usability of data
- Understandability of data
- Reliability and quality of data
- Linking and combining of data
- Communication and interaction with the provider

Zuiderwijk et al. (2012) also provide for each category, a list of barriers found in the literature and which were validated by them with the use of interviews and workshops. An overview of the barriers and their correspodning ara of effect is given in table 3.2.

Table 3.2: Barriers found in the area of effects

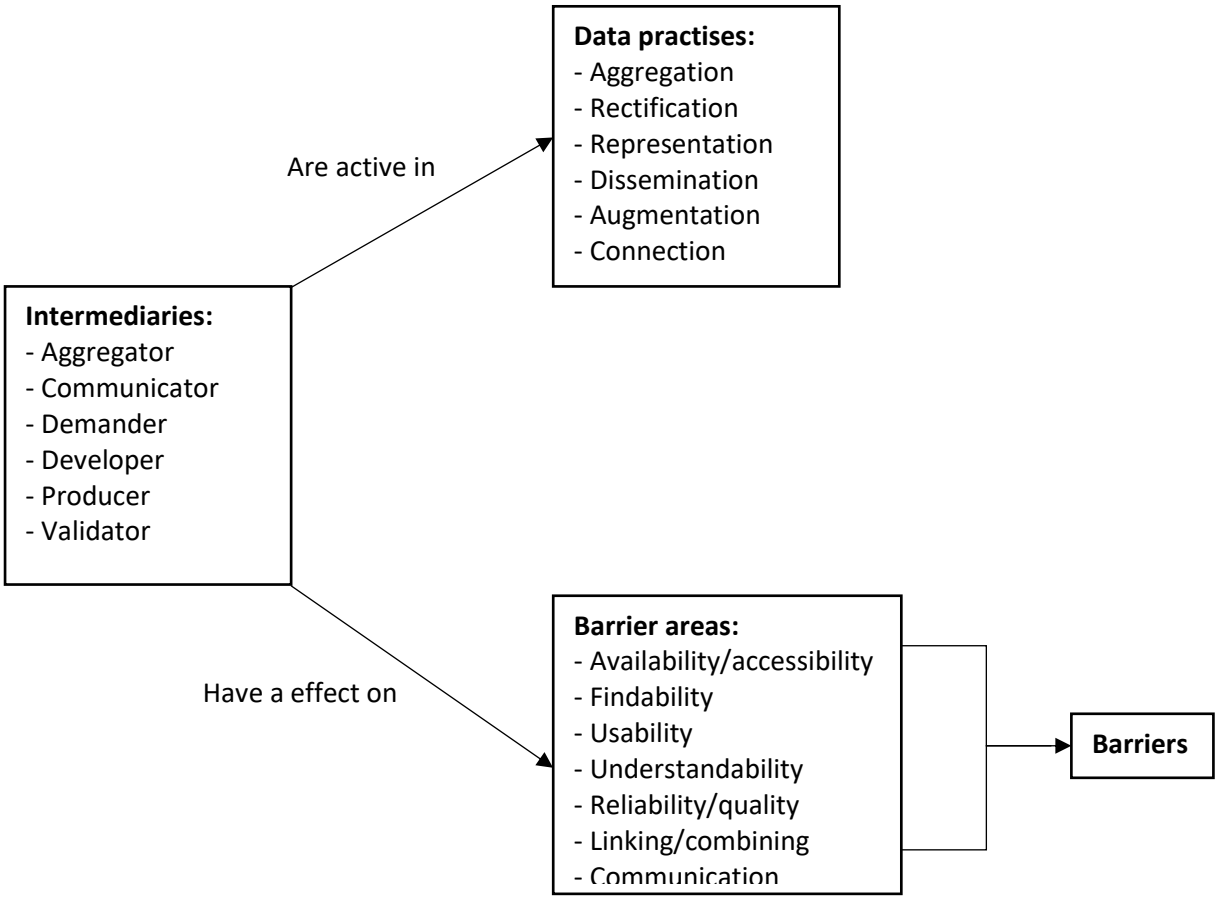
Area	**Barriers**
Availability and accessability	- Data is not provided continously - Data is temporariy not available - Duplication - Too much data to process (overload) - Lack of regular updates - Data temporarily not avialable at the source
Findability	- Data cannot be found - lack of search possibilities available - Lack of index - Metadata is not found - Data cannot be found
Usability	- Lack of knowledge - Lack of supprt/advice in using the data - Lack of services provided by providers - Data transformations are needed to make data usable - Heavy workload to use the data - Too much softrware - No standard software to process/analyse data - Lack of time

	- Potential users do not see the icentive to use the data
Understandability	- Lack of knowledge - Inssuficient explanation of the meaning of the data - Data is not visualized - Lack of skills and capablility to use the data - Lack of knowledge to interpret the data - No information on the context of data - No support in data use
Reliability and quality	- Data is not reliable - No knowledge whether data is reliable - Difficult to derermine the quality - Limited quality - Lack of accuracy
Linking and combining data	- Difficult to link data or not possible - Lack of tools - No knowledge to combine/link data with other data - Providers focus on single datatsets - Lack of unique identifiers
Communication and interaction with the provider	- No discussion possible between users an providers - Not platforms available for participation

3.3 Theoretical framework summary

In the theoretical framework, two important concepts have been described; the types of intermediaries and the barriers. Together they form the theoretical framework for the research (figure 3.2). However, the relationships between these concepts in the framework are hypothetical. Central in the framework are the intermediary types. It is assumed that intermediaries are active in certain data practices. Additionally, it is assumed that intermediaries have a certain effect in areas where barriers can occur. However, which areas and what barriers specifically is not known. These relationships are the main assumptions of the research. The empirical research will determine whether the assumed relationships are confirmed.

Figure 3.2: Schematic representation of the theoretical framework

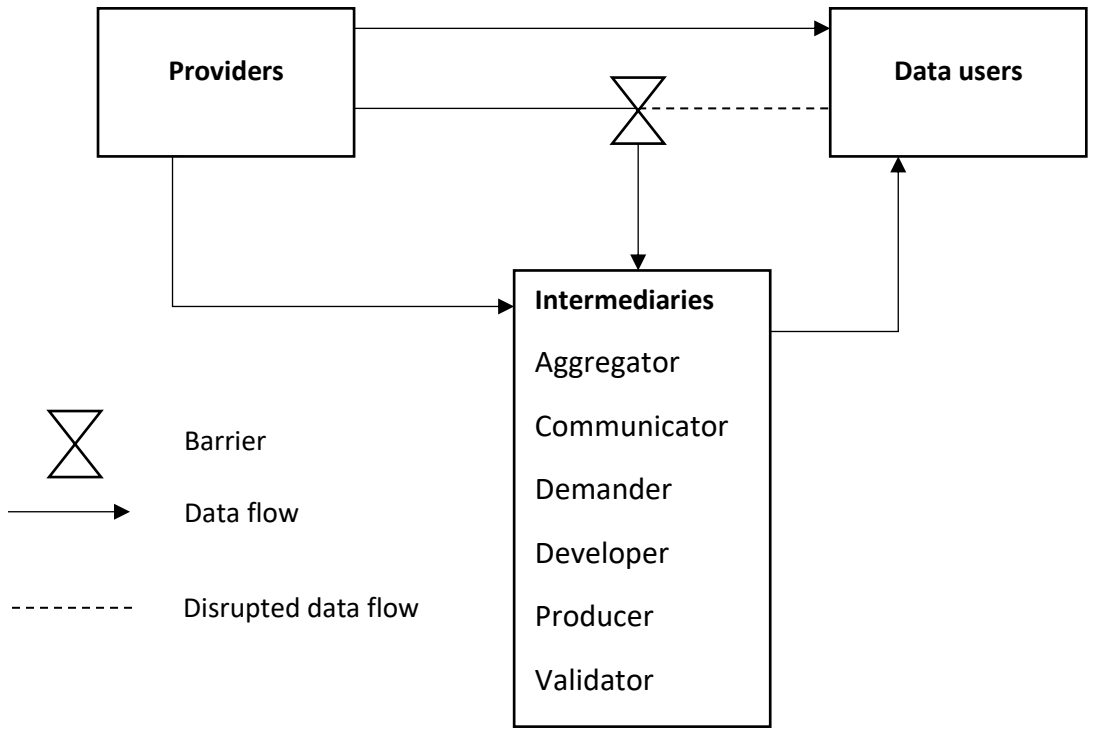


3.4 Conceptual model

In the literature study, several important concepts and relations have emerged. The conceptual diagram resembles an abstraction of the studied phenomenon (figure 3.3). It also schematically resembles the findings of the literature study and their context. For this research, open data is viewed from an ecosystem perspective. Within this ecosystem, data flows, or is expected to flow, from a provider to a user. The open data is then used for various purposes, such as creating new products, services or enhancing existing ones. However, the flow and use of data is often hindered by barriers. Intermediaries have emerged to fill in the gaps, facilitating the flow and use of open data to/by users.

This can potentially indicate that intermediaries might be of crucial importance for the success of open data. However, how they are able to facilitate the flow and use of data, and thus have effects on the open data ecosystem, is related to the type of intermediary and what function/role they fulfil in the ecosystem. Each type has its own function and therefore can have an effect on different factors influencing the flow and use of open data to and by users. Where the user is not able to use the data in the normal situation, due to barriers, intermediaries enable them to do so. However, it is not known if the types of intermediaries found in the literature actually exist in the cases. Not only that, it is also not know which type of intermediary has an effect on which barrier regarding open data use.

Figure 3.3: Conceptual model



4. Findings

4.1 Intermediary types and data practices

The first step in analysing the interview results is to determine what type of intermediaries the organisations are (table 4.1). This answers the fourth research question: 'What types of intermediaries can be identified in the case?'. In addition, the intermediaries will also be linked to the data practices found in the literature (table 4.2).

Four of the five interviewed intermediaries are **aggregators**: Geonext, Geodan, Esri and Steggink Geo-ICT. The aggregators engage in the data practices of aggregation, dissemination and augmentation. Aggregation means that they collect, combine, link and filter data from multiple sources. In addition, they are involved with dissemination by helping users access open data from different sources at a central location. Furthermore, by combining datasets together, they are augmenting open data. The only non-aggregator was IntellinQ. IntellinQ was also the only intermediary that did not use open data directly. However, the software that they develop, GeolinQ, allows users to collect, combine and link spatial data from multiple sources. The GeolinQ software is built to enable organisations to act as an aggregator, but IntellinQ itself is not fulfilling this role.

The same four intermediaries that are aggregators also have the role of **communicator**. As communicators they are engaged with the data practices of interpretation and representation. With respect to interpretation, they analyse open data to extract information for clients. They are also in the practice of representation as they can visualise and contextualise open data to match the skill of the user.

Two of the five intermediaries are **demanders**; Esri and Geodan. As demanders, they are engaged in the practice of connection; both interact or communicate with providers of open data. Esri for example, reports errors and problems that are reported by their software clients to providers. In addition, Esri also organises events where providers and users come together. In this way they are also in the practice of connection. Geodan works together with providers in projects, where they also pass on issues and problems they find in using the data.

Four of the intermediaries are **developers**; Esri, IntellinQ, Geodan and Steggink Geo-ICT. As developers they are active in the practices of dissemination and representation. They develop software that allows users to access and use geodata including open data. However, the software they provide differs in functionality. For example, Geodan develops software that is predominantly focused on visualizing data. Often, the software accompanies the data products that they create and sell to their clients. IntellinQ develops spatial database management software, named GeolinQ. This software allows users to access open data from various sources, create a standardized way of handling this data for their organisation and distribute data. This enables their clients to also be a provider for their data.

Of all the intermediaries, Geonext is the only one that is a **producer**. As a producer, Geonext is active in the practice of augmentation. Geonext often produces their own additional data by making extra measurements. By producing their own data and combine it with open data, Geonext is augmenting open data. Therefore, they are in the practice of augmentation.

Four of the intermediaries are **validators;** Geonext, Esri, Steggink Geo-ICT and Geodan. As validators they are active in the practice of rectification. They asses the quality of open data for their clients. This is done to determine whether the quality is sufficient to be used for the intended purpose of the client. However, the methods of coping with insufficient quality differs substantially between these three.

Table 4.1: Interviewed organisations and intermediary types

Organisation	Type
Geonext	Aggregator Communicator Producer Validator
Esri Nederland	Aggregator Communicator Demander Developer Validator
IntellinQ	Developer
Steggink Geo-ICT	Aggregator Communicator Developer Validator
Geodan	Aggregator Communicator Demander Developer

Table 4.2: Intermediary types and their data practice

Type	Data practise
Aggregator	Aggregation
Communicator	Interpretation Representation
Demander	Connection
Developer	Dissemination Representation
Producer	Augmentation
Validator	Rectification

The majority of the intermediaries cannot be categorised in a single intermediary type. Only one can be considered purely a developer. This majority of intermediaries is specialised in a variety of different areas. Often, these companies have various departments or divisions aimed at specific subsections of the market. However, some focus their business on one market in particular. IntellinQ for example can be classified as a developer, as their core business is focused on their software platform GeolinQ. Their software is a spatial database management system that allows users to collect data from various sources, including open sources, and store and manage these. In this example, IntellinQ is a developer. Esri on the other hand is also a developer, but fits other roles as well, given the services they provide to their software users on top of their software.

In addition, some intermediaries showed characteristics which were not part of any type identified in the theoretical framework. Esri for example also acts as a sort of 'educator'. They provide training and education to clients. This improves the client's skill and knowledge, improving their capability to work with geodata.

4.2 Areas of effect

4.2.1 Availability and accessibility

Availability and accessibility was only found in two interviews, both intermediaries were **developers**. Esri makes the AHN available for their software users via their own servers. If the official platform of the providers, PDOK, is offline, then the AHN is still available to Esri users for access. Furthermore, Esri updates datasets when providers release new versions, to ensure that their clients have the most up-to-date data. On the other hand, the GeolinQ software of IntellinQ gives users easier access to various data sources. In addition, it provides a flexible and configurable data model, allowing more efficient access to various data sources. Their software also warns their users when updates for the datasets are available at the source.

4.2.2 Findability

Findability, like availability and accessibility was only found in the interviews with the two **developers**. Their software allows users to search for data. In the case of Esri, they apply smart searching methods using keywords that allow their users to search for data quicker. Often, people do not know the specific name of a dataset like the AHN but have a general idea about what data they are looking for. By implementing smart keywords, datasets are easier to find by (potential) users.

4.2.3 Usability

The major limitation of the usability of open data is the lack of knowledge on how to use the data. In addition, working with open data requires substantial amounts of workloads, something which is at this moment not feasible for a lot of potential users that do not have divisions specialised in working with geodata. The respondent from Geonext stated that their clients are not able to work with geodata as reliable and efficient as them 'at this moment'.

Because of the lack of knowledge many potential users of open data do not see the added value or incentives to use it. Intermediaries on the other hand do possess the knowledge that is required. Because of this they are able to work with open data faster, more reliable and

more efficiently than many of their clients. **All** of the interviewed intermediaries enhance the usability of the data to a certain extent. However, when looking at **communicators**, the limitations regarding the usability are mostly removed completely, because they do the work with the data for the client. Communicators deliver data to clients in formats or mediums that are usable by the client. For example, Geodan derives information on building heights using the AHN in combination with other data. The information is then distributed to clients in excel format. This format is easier to use by clients that do not have specific knowledge of geo information. Similarly, Geodan maps the coverage of the mobile internet network for internet providers using the AHN. Because the information is presented in maps, the usability for their clients is increased.

Another problem regarding usability is that there is a gap between raw data on one side, and information on the other side. To retrieve the information that is needed, many different software packages exist. However, users that do not have knowledge on geo information get lost in the maze of software packages. Intermediaries can remove this problem because they have knowledge and experience on what software to use and when. **Communicators** remove the need for this knowledge because they retrieve the information out of the data for their clients.

Developers improve usability by creating software that allows users a standardised way of working with data. This is seen with IntellinQ as well. They provide flexible datamanagement functionality for organisations using their software to implement a standard way of handling and using open data and proprietary data. This functionality includes client portals, webservices and data export functions to export data to other formats. Esri on the other hand makes open datasets available for their clients in formats that are usable by their software packages. This improves the usability of open data for their clients.

4.2.4 Understandability

The barriers found regarding understandability are also caused by the knowledge gap on the side of the users. Users often do not know what the data means or cannot interpret the data. In addition, lack of skills and capabilities to work with the data are also limitations of the understandability. **Communicators** can extract the information out of the data for their clients so in these cases the client does not need to have to knowledge, skills or capabilities to do this themselves. Geodan for example indicated that a substantial part of their client base has no

expertise in geo-information. Geodan offers them the information that they need without them having to retrieve it themselves. In addition, Geodan rents out their expertise. Geodan employers are often detached to work at clients.

4.2.5 Reliability and quality

Quality and reliability of data has importance for each intermediary. However, the way for they cope with quality and reliability shortcomings differs substantially. The first difference is noticeable between the two **developers**. IntellinQ is not actively involved in checking or enhancing the quality of the data that their software users use. However, their software offers users functionality to automatically safeguard the quality of their data by providing warnings when datasets are updated and by providing a standardised way of handling and using data within their organisations. IntellinQ also enables organisations to get into control of the (open) datasources they use by configuration of the GeolinQ software.

In contrast, the other developer, Esri, is actively involved in the quality of their client's data. However, this activity is part of the extra services they provide in support of their software. Here, Esri also acts as a validator. The open data that they provide to their software users are all checked using quality control measures. When an open dataset like the AHN is updated at the source, but the update has caused issues regarding the quality of the data, Esri can decide not to update the AHN and AHN products they make available to their software users because the impacts for their clients will be too high. Clients are also informed regarding issues on data quality.

Geonext checks whether the quality of the data they use is sufficient and reliable enough to ensure that their clients get reliable results. For Geonext, the timeliness of the AHN is considered a significant problem. The AHN is released in bulks with some years between the releases, the consequence of this is that the measurements for the AHN can be several years old. Due to changes made to the built environment and infrastructure, such as new roads being built or buildings being demolished, the accuracy of the AHN can decrease rapidly. Geonext's clients often need to have more accurate and up-to-date information. As a **producer,** Geonext makes additional measurements in the field in combination with the AHN. The measurements made by them are often a factor two to three better in quality and accuracy than the AHN itself.

Geodan on the other hand answered that although they check the quality and reliability of the data they use, they are not actively enhancing the quality of the data themselves. They are transparent towards their clients regarding their data sources and quality of the data, but they will for example not actively pursue to enhance the quality.

4.2.6 Linking and Combining of data

Linking and combining data was the only area of effect that came forward strongly in **all** the interviews. The AHN is often considered an intermediate good for many applications. Intermediate goods that are used as input for the creation of other goods. Therefore, for the majority of applications the real value of the AHN comes from combining it with other data. This can be data from clients, open data or self-created data. In the case of self-created data, the intermediary acts as a **producer**. Geonext is an example of the producer. Some clients of Geonext require more accurate and up-to-date information than the AHN can provide. In some areas of the Netherlands, the soil is sinking several mm each year. Nonetheless, the creation of a nation-covering AHN takes several years. The measurements for the AHN in some regions of the Netherlands are therefore several years old. Geonext combines their own more accurate data with the AHN.

Another example is that the AHN is often combined with other data to retrieve information or to create new data products. In this case the intermediary acts as a **communicator**. This can be seen with Geodan. One of Geodan's business is to sell data products. These are mostly created by combining open data sets. One example is that they combine the AHN with the BAG to create a dataset on building heights, which they in turn sell to their clients. To get the information the client needs, the building heights, the AHN has to be combined with another dataset.

In addition, **developers** can enable users to link and combine data more efficiently. In the example of IntellinQ, their software allows users to collect, combine and publish data from different sources using a configurable and flexible data model. Esri also creates links between datasets that can be used together or that are used together often. In this way they can help their users to combine the data that they deliver.

Linking and combining of data is also related to **aggregators**, aggregators collect data from multiple sources to resupply to end users or to combine data. In the case of Esri, they collect data from various open sources, check the quality and make it suitable to be used in their

software for their software users. These users can then obtain the data through Esri instead of from the providers directly. In addition, Esri creates new datasets by combining existing datasets. These new datasets are also made available for their software users.

4.2.7 Communication between providers and end-users

The communication between providers and end-users was found in two intermediaries. Both of these were the **demanders**. Esri organises two events each year where providers and also their software users come together; the GIS Tech and the GIS Conferentie. In this way Esri can be seen as a bridge between providers and users. In addition, Esri strives to preserve the quality at the source. As was explained in the section about quality, Esri's reports problems that their clients have with open data to the provider. In addition, Esri shares user statistics with data providers to give them more insights in what happens with their data once it is released. These statistics also help providers to estimate what the effects are of their open data.

Geodan is also involved in projects where they work together with providers of data, also there are meetings regarding the possibilities of the data between providers and other parties for the base registers. However, Geodan stated that these were not yet seen for the AHN.

4.3 Summary

In the previous section, for each of the areas of effect it was explained how the different types of intermediaries can influence that area of effect. In this way, each intermediary type is linked with the area of effect. These links are resembled in the matrix below (table 4.3).

Table 4.3: Intermediary types and the area of effect based on the interviews

	Availability/ accessibility	Findability	Usability	Understandability	Reliability/ Quality	Linking/ combining	Communication
Aggregator		X				X	
Communicator			X	X		X	
Demander							X
Developer	X	X	X			X	
Producer					X	X	
Validator					X		

Aggregators are associated with linking and combining of data, they collect data from various sources including open sources and combine this data. They can then resupply this data to others as is seen in the case of Esri or combine it to create new data products as is seen in the case of Geodan. Additionally, Esri provides smart searching for their software users to search within the datasets that they provide. This also increases the findability of the data.

Communicators transform the data in other formats, channels and media. In addition, they also retrieve information out of datasets for their clients. This type of intermediary is associated with the usability and understandability. For both these areas of effect, knowledge is an important factor. Knowledge on how to use data and to interpret the data is required to retrieve information out of the data. Their clients are not able to do so or not as fast and efficient as the intermediaries. In addition, communicators also address the understandability and usability by presenting the data or information in different formats. Information is for example delivered in the form of excel files which are easier to use and understand by their clients. Maps are also important in this regard. Almost always, to retrieve the information that end users want, the open data is combined with other datasets. Therefore, communicators also remove the barriers related to the linking and combining of data.

Demanders are only linked with barriers regarding communication between providers and users. Often users find it difficult to reach providers and vice versa. Intermediaries that have the role of demander can function as a communication bridge between the users and providers. For example, they will share user statistics or organise events where providers and users meet. They can also pass down issues their clients have with using the data to the providers. Lastly, they are involved in projects where they work together with providers to help improve the supply of open data.

Developers are intermediaries that create applications and software that help end users to work with (open) data. Also, their software helps users to find and access open data. In addition, their software helps end users to combine and link data more easily. Usability is also improved because the software provides a standardised way of using data.

Producers are intermediaries that create their own data with measurements and combine this with open data. This enhances the quality and reliability of the data. Often the data that is provided is not accurate or up-to-date enough to be used for the purpose that end users want to use it for. While the open data provides a solid base, additional measurements are needed to check for inaccuracies. Here the open data is also combined with self-created data by the intermediaries, removing the issues with linking and combining data for their clients.

Lastly, validators have a link to the quality and reliability of data. Before open data is used by intermediaries or resupplied to end users by intermediaries, the quality is assed. How validators cope with data quality differ. However, when measures are taken to improve the data, for example by doing additional measurements or combing it with other data, the intermediary shifts roles from validator to producer and communicator respectively.

When all the findings are put together, an overview can be created of the found results. Table 4.4 provides an overview of the intermediary types, their data practices, areas of effect and which barriers they affect.

Table 4.4: Overview of the findings

Intermediary type	Data practice	Area of effect	Affected barriers
Aggregator	Aggregation	Findability	- Data cannot be found - Metadata is not found - Difficulty in searching/browsing: lack of index - Lack of advanced search possibilities
		Linking/combining	- Lack of tools - lack of know how - Lack of unique identifiers - Focus on single datasets by providers: real value is in combining
Communicator	Interpretation Representation	Usability	- Lack of knowledge - Forced data transformations - Huge workload - Too much software - Lack of time - No incentives
		Understandability	- Lack of visualization - Lack of knowledge - Lack of skill/capability - Lack of support in the use of data - No explanation of the meaning
		Linking/combining	- Lack of tools - lack of know how - Lack of unique identifiers - Focus on single datasets by providers: real value is in combining
Demander	Connection	Communication	- No discussion with provider - No platform for participation available
Developer	Dissemination Representation	Availability/accessibility	- Data is not provided continuously - Temporary not available - Lack of updates - Too much data to process - Duplication of datasets
		Findability	- Data cannot be found - Metadata is not found - Difficulty in searching/browsing: lack of index

			- Lack of advanced search possibilities
		Usability	- Forced data transformations - Huge workload - Too much software - Lack of time
		Linking/combining	- Lack of tools - Lack of unique identifiers - Focus on single datasets by providers: real value is in combining
Producer	Augmentation	Reliability/quality	- Data is not reliable - No knowledge whether the data is reliable - Difficult to determine the quality - Limited quality - Lack of accuracy
		Linking/combining	- Lack of tools - Lack of unique identifiers - Focus on single datasets by providers: real value is in combining
Validator	Rectification	Reliability/quality	- Data is not reliable - No knowledge whether the data is reliable - Difficult to determine the quality - Limited quality - Lack of accuracy
Educator	*Education*	*Usability*	*- Lack of knowledge*
		Understandability	*- Lack of knowledge*

5. Discussion

5.1 Discussion of results

This chapter discusses the implications of the research based on the findings. The aim of the research is to fill a gap in the knowledge of the role of intermediaries on the use of open data. With the use of qualitative research methods relationships between two important concepts have been found: open data intermediaries and barriers for using open data. However, the importance of the found relationships as well as the impact of the effects on the use of the AHN in general were not found in this research.

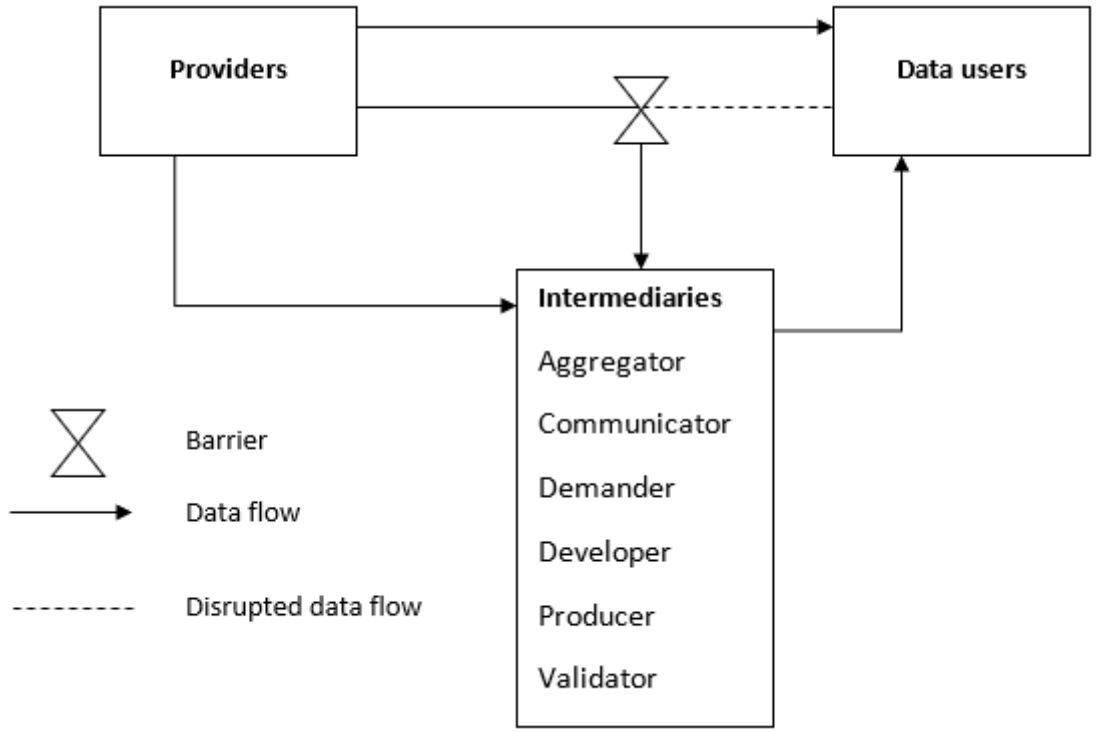
Due to the limited research on open data intermediaries, there is a limited amount of theory on open data intermediaries available. Therefore, it is possible that besides the types of the intermediaries found in the theoretical framework, other types may exist. Evidence for this was also found in the interview results. For example, some intermediaries provide their clients with training and education. But, the role of an 'educator' intermediary was not found in the theory.

The results indicate that intermediaries are not confined to single type. They often perform multiple roles at the same time. This relates to the layers of intermediation explained by van Schalkwyk et al. (2015). They found that multiple intermediaries may operate in a supply chain, because a single intermediary is not capable of fully unlocking open data. The results of this research show that this is true as Esri indicated that some of their clients resupply the open data they get from Esri to other users. However, van Schalkwyk et al. (2015) stated that each intermediary in a supply chain is specialised in a specific area. The results of this research show something different: a single intermediary can be specialised in different areas. It is true that multiple types are present in a supply chain, but this does not mean that there are also more than one intermediaries operating in a supply chain.

Also, when considering the group of developers, a distinction had to be made between what they as an organisation do and what their software does. In the case of IntellinQ, they are purely a developer. However, their software can act as an aggregator: their software can collect data from various sources and combine them. But, it is important to note that IntelllinQ in itself is not acting as an aggregator, they do not collect the data, but users can do this with their software.

In the first phase of the research, a conceptual model was created based on the systematic literature review in the theoretical framework. However, the results of the interviews have given new insights into the concepts of the conceptual model and their relationships to one another.

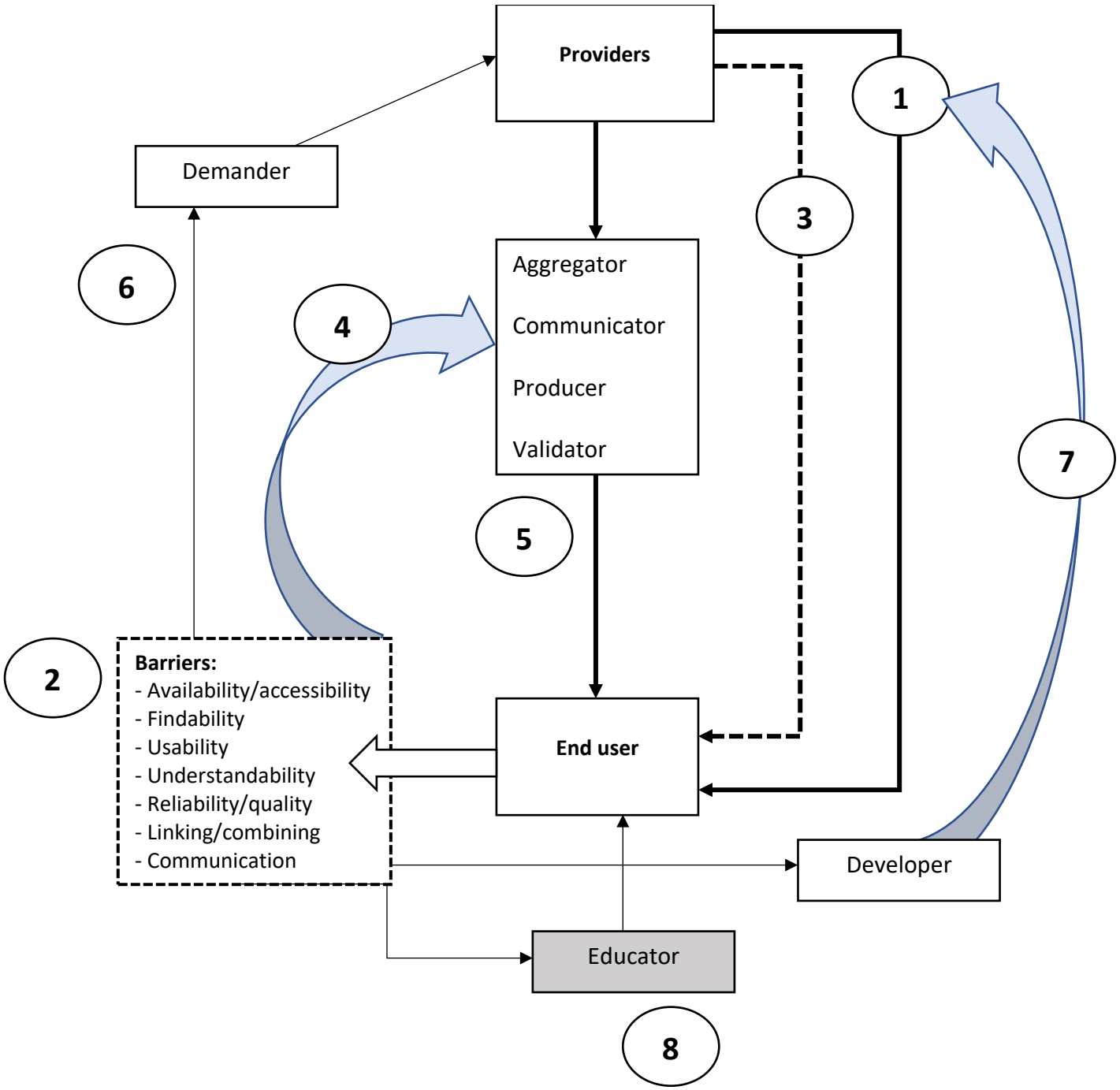
Figure 5.1: Original conceptual model



Because of these new insights, a new conceptual model has been developed. This new model also includes the intermediary type of educator, which was originally not found in the literature review. In the new conceptual model (figure 5.2), it is seen that in a normal situation (1), data flows from a provider to an end user. However, the end users encounter barriers (2) that limit or prevent them from using the data. They can consult intermediaries (4) to help them use the data, or the intermediaries can use the data for them and provide the end user with information (5). The demander has a different position, they do not help the end user in using the data, nor do they provide them with information. Instead, they act as a communication line between the end users and the provider. They can report issues that end users have with the data to the provider. The developer also has a special position. Like the demander, this intermediary also does not work with the data directly in their role as developer. Instead, they create software and applications that can support the end user in accessing, finding or using the data. This allows users to use open data as is the case in the

normal situation (7). The newly found type, the educator, can improve the end user's capabilities in working with the data. This can help reduce the barriers experienced by end users (8).

Figure 5.2: New conceptual model



5.2 Discussion of the methodology

The areas of effect and barriers are derived from literature on barriers of open data use in general. These barriers are therefore not specific for the AHN. Nonetheless, many of the barriers and areas of effects were confirmed in the interviews.

It also turned out that the majority of barriers are caused by a lack of knowledge on the user's side. Because intermediaries are specialists, they possess knowledge on how to use data for specific purposes. Users are therefore not required to have this knowledge, because the intermediary removes the need of knowledge. However, there are also clients that do have knowledge of geodata. Nevertheless, intermediaries can still facilitate the use of open data by making it more efficient and easier to work with for users with knowledge and skill. This shows that there can be differences in the effects of intermediaries based on the skill level of the user. Different skill levels of users were not taken into consideration in the research.

It is also likely that there are more intermediaries than the selected ones in the research. Furthermore, the intermediaries found in this research are specifically selected for the AHN. However, the AHN is rarely used on its own. These intermediaries will most likely use multiple open datasets in addition to the AHN, making their potential reach much wider than the AHN.

For the research, five interviews were held in total. While there are no fixed rules for the number of interviews to be held in qualitative research, it is generally considered best to do as many interviews as possible. More interviews improve the validity and reliability of the research. Given the time constraints of the research, it was not possible to schedule more interviews. Future research should therefore strive to interview more organisations. However, even though five interviews were held, these provided sufficient data to answer the research questions. It is therefore arguable whether more interviews will produce more results.

The interviews were also semi-structured. In the future, more structured interviews might be necessary to improve the comparability of the interviews. The results indicate differences between intermediaries within a single type. Future research could therefore focus on a single type of intermediary and use more structured interviews to compare organisations found within that type.

The research focused only on qualitative analysis, this means that values or the impact of the effects are not researched. Measuring effects quantitatively is difficult and some effects are

not quantifiable. It is decided that at this moment, exploratory research is more appropriate (Mason, 2002). Accurate measurements of the quantifiable effects that intermediaries have on open data is assumed more useful when the body of knowledge has matured further.

The case study methodology also faces some limitations and criticism. Foremost case studies are often criticised because it is assumed they produce results which cannot be generalised on the studied phenomenon in general (Yin, 2003). Consequently, case studies are often discarded as preliminary methods, to serve hypothesis building rather than theory building (Yin, 2003). However, although the context in which this research is performed is fixed, intermediaries may have effects that are common for intermediaries in other open data ecosystems.

6. Conclusion

The overall objective of the research is to evaluate the roles of intermediaries in the open data ecosystem and to analyse how they can remove barriers found in the (re) use of open data. To achieve this goal, five research questions will be answered using the results of the research.

The first research question is: 'Wat are open data intermediaries?'. From the literature study it was established that an open data intermediary is an actor positioned at some point in a data supply chain that incorporates an open dataset, is positioned between two agents in the supply chain and facilitates the use of open data that may otherwise not have been the case.

The second research question aims to further explore the definition of open data intermediaries to find classifications that could be used in the empirical phase of the research; 'What types of open data intermediaries can be identified?'. The result of this research question is a classification of different types of intermediaries. These types of intermediaries each have a specific role within the open data ecosystem. In total, 6 types were found in the literature study. These are: aggregators, communicators, demanders, developers, producers and validators. Aggregators collect data from various sources, to combine with other data to supply to the end users. Communicators transform open data to other formats, channels and context that are matched to the skill and capabilities of end users. They also provide their clients with information that they derive from (open) data. This is mostly done by combining open data with other data. This can be self-created data, client's data or open data. Demanders can act as communication bridges between providers and end users. In this role they can link demand with supply. Developers develop software and applications that that help users use and access data. Producers produce their own data which is in turn combined with open data. This is done to create new data, that better suits the needs of their clients. Validators assess the open data on quality and usefulness for the end users. They can improve the quality to enhance the potential of open data for end users.

The third research question; 'What barriers are found in the use of open data?', aims to establish possible area of effects where intermediaries can have an influence. This was the second input for the empirical phase of the research. In the literature, barriers are found in several different areas that all complicate the use of open data by end users.

These areas are: availability and access, findability, usability, understandability, reliability and quality, linking and combining of data and lastly communication with the providers of open data.

These results of the first three research questions are combined to create the theoretical framework for the empirical research. With the use of interviews, the last two research questions were answered.

The first research question answered with the interviews is: 'What types of intermediaries can be identified for the AHN?'. The results indicate that all of the types are present for the AHN. However, several intermediaries are several types at the same time. In addition, even if two organisations are the same type, the way in which they fill in that role can differ substantially. Also, the type of demander was found the least, possibly because providers cannot maintain communications with every organisation involved with their open data. The type of aggregator was found the most, this is because these organisations are not using only the AHN. Almost always, they work with several datasets. The majority of intermediaries are also communicators; the open data is almost never resupplied unaltered. At the least, the data is transformed into other formats that can be used by specific software packages. Three of the intermediaries also developed software and applications for their clients. Almost all of the intermediaries were validators as well since they check the quality and usefulness of the open data. Besides the types found in the literature, a new type was identified as well; the educator. Educators are intermediaries that provide training and education to improve the capabilities of the users to work with open data.

The fifth research question; 'How do intermediaries help overcome barriers found in the use of open data?', is aimed at establishing links between the types of the intermediaries found in research question two and the areas of effect in research question three. It was found that aggregators affect the findability of the open data and the problems with linking and combining it with other data. Communicators improve the usability and understandability and also link and combine data. Demanders only improve the communication between providers and users. Developers, improve the accessibility and availability of data. Their software also improves the usability and the understandability of the data. In addition, they can help users to link and combine data more easily with their software. Producers improve the quality and reliability of open data by doing additional measurements. Producers also combine data for

their clients, the problems regarding the combining of data are therefore removed for the users. Validators only had a noticeable link with the quality and reliability of data. Users often find it difficult to determine whether the quality is sufficient, and validators do this for them. The measures undertaken to improve the quality when the quality is deemed to be insufficient, is outside the role of the validator. Quality could for example be improved by doing additional measurements, in this case the intermediary shifts from a validator to a producer. Lastly, educators can improve the usability and understandability by enhancing the knowledge of users. This enhances their capability to understand and use the data.

7. Recommendations for future research

The aim of the research is to understand the role of intermediaries in the open data ecosystem. Although this research provided new insights in the role of intermediaries in an open data ecosystem, it should be considered a first small step towards a full understanding of open data intermediaries. This research therefore aspires to provide a base for future research. In addition, due to the qualitative nature of the research, more information was found than was required to answer the research questions. This secondary information could provide valuable starting points for future research. These will be discussed in this section.

First, it was found that some of the intermediaries also provide training and education to clients in different areas and act as a 'educator'. This role has however not been described in the literature on open data intermediaries. Future research can therefore be focused on further describing this type of intermediary.

Secondly, despite links found in the interviews, further research can be dedicated to analysing in depth how certain types of intermediaries affect specific areas. Besides looking more in depth to the links found in this research, it is also possible to conduct more quantitative research to determine quantifiable effects; how many users of open data are able to work with open data because of intermediaries? How many potential users do intermediaries reach? How much of the open data use can be contributed to intermediaries?

Thirdly, in some of the interviews it was found that there is still much to improve in the area of communication. Because intermediaries have a better view of what certain groups of users want, providers can learn a lot from them to improve their open data overall. Future research

can be dedicated to analysing how the communication between providers and intermediaries can be improved and whether it is beneficial for open data.

Also, it was found in the interviews that many clients of the intermediaries do not have much interest for the data. The clients want specific information that is beneficial for their business. How and with what data is used is not a priority for them. In addition, open data like the AHN are often seen as intermediate goods; the real value comes from combining it with other data. Sometimes the desire was expressed that providers start to improve the capability of linking and combining it with other data.

Some intermediaries also felt that providers should not interfere too much with the business of intermediaries. For example, when an intermediary sells open data in different formats packages and a provider starts to do the same; then the intermediary is at risk of losing its business. This shows that there is an interesting tension field between how providers can improve open data while at the same time not doing too much to hinder the market. Future research can be dedicated to investigating the boundaries; how much services should open data providers provide? or should providers be limited to just releasing open data in open formats?

To conclude, future research could be aimed at analysing whether the found effects are also experienced by the clients of the intermediaries. This research limit itself to the viewpoint of intermediaries. However, by involving their client's viewpoint, the found effect can be validated. This could also lead to the discovery of more effects.

References

AHN. (2015). *Handleiding 'AHN downloaden van PDOK'*. Retrieved from http://www.ahn.nl/binaries/content/assets/ahn-nl/downloads/handleiding-ahn-downloaden-via-pdok.pdf

Bregt, A., Grus, L., & Eertink, D. (2014). *Wat zijn de effecten van een open basisregistratie topografie na twee jaar?* Wageningen.

Bregt, A. K., Grus, L., Beuningen, T. Van, & Meijeren, H. Van. (2016). *Wat zijn de effecten van een open Actueel Hoogtebestand Nederland?* Wageningen.

Davies, T. (2011). *Open Data : Infrastructures and ecosystems*. *University of Southhampton*.

Deloitte LLP. (2012). Open growth Stimulating demand for open data in the UK A briefing note from Deloitte Analytics. *Deloitte*, 1–9.

Dumpawar, S. (2015). *Open Government Data Intermediaries: Mediating Data to Drive Changes in the Built Environment*. Massachusetts Iinstitute of Technology.

González-Zapata, F., & Heeks, R. (2015). Understanding Multiple Roles of Intermediaries in Open Government Data. *13th International Conference on Social Implications of Computers in Developing Countries*, 1–10.

Hagel III, J., & Rayport, J. (1997). The coming batle for customer information. *The McKinsley Quarterly*, *3*, 64–77. Retrieved from http://go.galegroup.com/ps/i.do?id=GALE%7CA20341539&sid=googleScholar&v=2.1&it=r&linkaccess=fulltext&issn=00475394&p=AONE&sw=w&authCount=1&u=utrecht&selfRedirect=true

Heimstädt, M., Saunderson, F., & Heath, T. (2014). Conceptualizing Open Data ecosystems: A timeline analysis of Open Data development in the UK. *In Proceedings of the International Conference for E-Democracy and Open Government (CeDEM2014)*, (2014/12). Retrieved from http://hdl.handle.net/10419/96627

Iansiti, M., & Levien, R. (2004). Strategy as Ecology. *Harvard Business Review*, *82*(3), 68–81. https://doi.org/10.1108/eb025570

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. https://doi.org/10.1080/10580530.2012.716740

Janssen, M., & Zuiderwijk, A. (2014). Infomediary Business Models for Connecting Open Data Providers and Users. *Social Science Computer Review*, *32*(5), 694–711. https://doi.org/10.1177/0894439314525902

Kadaster. (2017). BRT. Retrieved 11 November 2017, from https://www.kadaster.nl/brt

Khan, K. S., Kunz, R., Kleijnen, J., & Antes, G. (2003). Five steps to conducting a systematic review. *Journal of the Royal Society of Medicine*, *96*(3), 118–121. https://doi.org/10.1258/jrsm.96.3.118

Magalhaes, G., Roseira, C., & Strover, S. (2013). Open government data intermediaries. In *ICEGOV 2013 Proceedings - 7th International Conference on Theory and Practice of*

Electronic Governance (pp. 330–333). https://doi.org/10.1145/2591888.2591947

Mason, J. (2002). *Research Methodology* (Vol. 2).

Open Data Barometer. (2016). The Open Data Barometer. Retrieved 3 November 2017, from http://opendatabarometer.org/?_year=2016&indicator=ODB

Ponte, D. (2015). Enabling an Open Data Ecosystem. In *ECIS 2015 - Research-in-Progress Papers -23rd European Conference on Information Systems* (pp. 1–7). Retrieved from http://aisel.aisnet.org/ecis2015_rip/55

Rijkwaterstaat. (2017a). Actueel Hoogtebestand Nederland. Retrieved 11 November 2017, from https://www.rijkswaterstaat.nl/zakelijk/open-data/actueel-hoogtebestand-nederland/index.aspx.

Rijkwaterstaat. (2017b). Open data Rijkswaterstaat. Retrieved 11 November 2017, from https://www.rijkswaterstaat.nl/zakelijk/open-data/actueel-hoogtebestand-nederland/index.aspx.

Rocco, S. T., & Plakhotnik, S. M. (2009). Literature reviews, conceptual frameworks, and theoretical frameworks: Terms, functions, and distinctions. *Human Resource Development Review*, *8*(1), 120–130. https://doi.org/10.1177/1534484309332617

Sawicki, D. S., & Craig, W. J. (1996). The democratization of data: Bridging the gap for community groups. *Journal of the American Planning Association*, *62*(4), 512–523. https://doi.org/10.1080/01944369608975715

van Schalkwyk, F., Chattapadhyay, S., Caňares, M., & Andrason, A. (2015). *Open Data intermediaries in Developing Countries*. Retrieved from http://dx.doi.org/10.6084/m9.figshare.1449222

van Schalkwyk, F., Willmers, M., & McNaughton, M. (2016). Viscous Open Data: The Roles of Intermediaries in an Open Data Ecosystem. *Information Technology for Development*, *22*(1), 68–83. https://doi.org/10.1080/02681102.2015.1081868

Vancauwenberghe, G., Donker, F. W., & Loenen, B. Van. (2017). Business Models for Geographic Information. In J. B. Kruse, J. Crompvoets, & F. Pearlman (Eds.), *GEOValue: The Socioeconomic Value of Geospatial Information.* (pp. 85–115). CRC PRess.

Welle Donker, F., & van Loenen, B. (2016a). How to assess the success of the open data ecosystem? *International Journal of Digital Earth*, *10*(3), 284–306. https://doi.org/10.1080/17538947.2016.1224938

Welle Donker, F., & van Loenen, B. (2016b). Sustainable Business Models for Public Sector Open Data Providers. *JeDEM-eJournal of eDemocracy and Open Government*, *8*(1), 28–61.

Yin, R. (2003). *Case Study Research: Design and Methods* (5th ed., Vol. 5). Thousand Oaks: Sage.

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. https://doi.org/10.1641/b570402?ref=search-gateway:885882d1830675b0f27af0760faeaef8

Appendices

Appendix 1 – Interview topic list

Topics	Possible questions
Type/Role	1. Describe you organisation for me? 2. Wat products and services do you provide? a. With respect to the AHN? 3. To what extent do you collect data yourself? 4. Do you gather data form different sources? 5. Do you develop software for users to make easier use of open data? 6. Do you have contact with providers?
Effects in general	7. What are the foremost effects and benefits for your clients that you can offer them?
Availability/accessibility	8. To what extent is accessibility and issue for your clients? 9. How do your services and products help the accessibility of data?
Findability	10. To what extent is findability and issue for your clients? 11. To what extent are your clients known with open geo data like the AHN?
Manageability/usability	12. To what extent are your clients capable of using open data on their own? 13. Do you also reach clients outside the Geo domain?
Quality/reliability	14. How do you cope with the quality of the data? 15. How do you enhance the quality?
Linking and Combining data	16. Is linking and combining data important? 17. How do you combine and link the AHN with other data?
Communication with provider	18. To what extent is there communication with the providers? 19. How can this be improved for the future?
Concluding questions for closure	20. Besides the topics discussed, are there other aspects of open data that you feel are important but have not been discussed? 21. How important would you consider the position of the intermediary?