# THE EFFECTS OF OPEN DATA COMMUNITIES ON OPEN DATA BENEFITS AND BARRIERS

**MSc Complex Systems Engineering and Management (CoSEM)**

**Tom Schuurmans**

**September 2022**

This page intentionally left blank

# The Effects of Open Data Communities on Open Data Benefits and Barriers

---

Master thesis submitted to Delft University of Technology

in partial fulfilment of the requirements for the degree of

**MASTER OF SCIENCE**

in **Complex Systems Engineering & Management**

Faculty of Technology, Policy and Management

by

Tom Schuurmans

Student number: 4349288

To be defended in public on September 16th 2022

**Graduation Committee**

Chairperson : Prof.dr. M.E. Warnier, Section Systems Engineering and Simulation
First Supervisor : Dr. A.M.G. Zuiderwijk- van Eijk, Section ICT

# Colophon

**Graduation thesis**

| | |
|---|---|
| University | Delft University of Technology |
| Faculty | Technology, Policy & Management (TPM) |
| Master program | Complex Systems Engineering and Management (CoSEM) |

**Author**


| | |
|---|---|
| Name | Tom Schuurmans |
| Student number | 4349288 |
| E-Mail | t.f.schuurmans@student.tudelft.nl |


**Graduation Committee**

| | |
|---|---|
| Prof.dr. M.E. (Martijn) Warnier | Chairperson, Second Supervisor |
| Dr. A.M.G. (Anneke) Zuiderwijk- van Eijk | First Supervisor |

# Preface

Before you lies the dissertation “The Effects of Open Data Communities on Open Data Benefits and Barriers”, the basis of which is a series of qualitative interviews with the managers and users of the Dutch open data community and two open data researchers. It has been written to fulfil the graduation requirements of the Complex Systems Engineering & Management Master at the Technical University of Delft.

My research questions were formulated together with my first supervisor, Dr. Anneke Zuiderwijk- van Eijk. Executing the research turned out to be much more demanding than I thought beforehand, but eventually I was able to answer the questions that we identified. I would like to thank both of my supervisors, Prof.dr. Martijn Warnier and Dr. Anneke Zuiderwijk- van Eijk. They have guided me excellently throughout the process of writing this thesis and helped me to head in the right direction. I also want to express my gratitude to all of the interviewees, without their help I would not have been able to execute my analysis.

While writing my thesis I came into contact with completely different, but all very motivated, individuals. They all believed in the potential of open data initiatives. Despite the fact that some of them still had clear reservations about the current contribution of the open data community, they had no doubt about its potential. This made me realise that the potential success of open data communities, as well as the success of open data initiatives in general, does not only depend on the policies that are written. I have come to realise that the success of these kinds of innovations, which have a long run-up, is also attributable to the determination of individual civil servants and professionals. Their efforts ultimately makes for great changes.

I hope you enjoy your reading.

Tom Schuurmans
Delft, September 4, 2022

## Summary

Governmental organisations worldwide are working on their open data strategies and are openly sharing their data with the public. Over the last two decades, both society and government have digitised drastically. Current research focusses on many aspects of open data, including the benefits and barriers of open data. Recently also open data communities have been set up by several national governments, on their existing open data portals. The effects of setting up these data communities on open data benefits and barriers and whether an effect exists at all is not yet clear, because there has been little scientific research into data communities. This thesis therefore focuses on the effects of setting up open data communities and is therefore filling a part of the current gap in scientific literature. The main research question is formulated as:

*"What are the potential effects of open data communities on open data benefits and barriers?"*

In order to answer this question, existing literature on open data was used. Benefits and barriers of open data were retrieved. After that, the benefits and barriers that could be potentially affected by introducing an open data community were identified and selected for this research. To be able to better understand the potential effects of the open data community, institutional theory was introduced. This theory allows to gain more insight into the effects, by looking at three levels of instruments: (1) formal, (2) informal and (3) enforcing. The instruments within these levels can contribute to the value creation of the open data community.

A single-unit case study was conducted to gather qualitive information of the effects of data communities. The case study focussed on the Dutch open data community and consisted of two parts, first a brief document analysis and then in-depth interviews. The document analysis focussed on existing policy documents, to gather more information of the Dutch open data community.

Interviewees were then systematically questioned about the effects of the Dutch open data community on the open data benefits and barriers identified from literature. Interviewees were also asked to reflect upon institutional instruments on all three levels. Two open data researchers were interviewed to validate the results.

The document analysis used five policy documents and eight in-depth interviews took place. The results of the analysis show how the Dutch open data community contributes to enhancing open data benefits, including (indirectly) creating more informed citizens, increasing the access to capacity and resources outside of the data publishing organisation and a higher problem-solving capacity.

Furthermore, the interview participants agree to a large extent that communities contribute to intragovernmental collaboration and the use of collective intelligence to solve public problems.

According to the interviewees, the community also (potentially) mitigates open data barriers such as the lack of interest in using open data (by governmental organisations). The interviewees stated that the community managers made sure every question that was posted got a sufficient answer within a reasonable amount of time and therefore the barrier stating that the data provider ignores requests and suggestions of data users could also be mitigated, as well as difficulties in the interaction with the data provider. Both researchers and most community users and managers argued that the community could also contribute to mitigating low engagement of public managers with open data and increasing the knowledge and skills of employees to use the open data. Lastly, according to the interview participants, the community can decrease difficulty in discovering/locating data and not being able to combine and connect datasets.

The interviewees were also questioned about how institutional instruments could increase the value of the open data community. Although the participants concluded that the contribution of formal instruments (such as rules) is limited, they indicated informal rules (such as norms) and enforcing instruments (such as rewards) can contribute to the value that is created by an open data community.

This research indicates that setting up and maintaining/ supporting data communities can lead to very positive effects, by both enhancing open data benefits and mitigating barriers. Policy makers can use the conclusions to determine better-grounded objectives when setting up new open data communities. The community-specific challenges identified in the analysis can contribute to the process of designing an open data community, because policy makers can compare scope and design choices with the empirical experiences.

This thesis also provides an exploratory scientific contribution. There is little research within the field of data communities and this research provides insight into how open data communities can contribute to enhancing open data benefits and the mitigation of open data barriers.

Although the results are promising, certain limitations are applicable to the research. First of all, the thesis only studied one open data community and only two community managers, five community users and two researchers were interviewed. Furthermore, the full list of open data benefits and barriers was reduced based upon an assessment of the author of the thesis and validated by an open data researcher. Last of all, the interviews focused on qualitative data and therefore the results cannot

be generalised. However, combining these results with other case studies can improve generalisability.

Future research is needed, to validate the conclusions of this thesis. Concrete suggestions for future research are (1) to select a larger set of different case studies and (2) to increase the number of interviewees per case study. Furthermore, future research should (3) execute a quantitative review of the effects of open data communities on open data benefits and barriers, (4) specifically look at the negative effects of open data communities, and (5) perform a more comprehensive validation of the open data benefits and barriers that are potentially affected by open data communities.

# Table of contents

# List of figures and tables

# List of Abbreviations

The table underneath (Table 1) describes the meaning of various abbreviations and acronyms used throughout the thesis. The page on which each abbreviation is defined or first used is also given.

*Table 1: List of abbreviations and their meaning*

| ***Abbreviation*** | **Meaning** |
|---|---|
| *API* | Application Programming Interface |
| *CKAN* | Comprehensive Knowledge Archive Network |
| *CoSEM* | Master's degree programme in Complex Systems Engineering and Management |
| *DCAT* | Data Catalog Vocabulary |
| *DMP* | Data Management Plan |
| *DOI* | Digital Object Identifier |
| *GDRP* | General Data Protection Regulation |
| *HREC* | Human Research Ethics Community |
| *I#* | Interviewee # |
| *ICT* | Information and Communication Technology |
| *KOOP* | Dutch Knowledge and Operation Centre for Official Publications (part of Ministry of the Interior and Kingdom Relations) |
| *NODA* | National Open Data Agenda |
| *OECD* | Organisation for Economic Co-operation and Development |
| *OGD* | Open Government Data |
| *OGDI* | The Open Government Data Initiative |
| *OGP* | Open Government Partnership |
| *OGDP* | Open Government Data Portal |
| *PSI* | Public Service Information |
| *SSI* | Semi-structured Interview |
| *URL* | Uniform Resource Locator |

# 1 Introduction

## 1.1 Problem indication

Governmental organisations worldwide are openly sharing their data with the public. Over the last two decades, both society and government have digitised drastically. This transformation of governments is also called E-government. This transformation includes the process in which procedures, documents, and services are digitised to improve governance, by using technology (Unisys, n.d.). E-government can also be broader defined as the combination of using Information and Communication Technology (ICT) in government services and organisational changes plus new skills in order to improve public services (European Commission, 2005).

The information that those (central and local) governments share (but also create, collect and/or use themselves) is called Public Sector Information (PSI). The OECD, the Organisation for Economic Co-operation and Development, defines Public Sector Information (PSI) as "information, including information products and services, generated, created, collected, processed, preserved, maintained, disseminated, or funded by or for a government or public institution" (OECD, 2008, p. 4). Open Government Data (OGD) can be seen as a subset of Public Sector Information (PSI), according to Ubaldi (2013). OGD contains two main elements, which are government data and open data. Those elements are generally defined as follows: government data is "any data and information produced or commissioned by public bodies" (Ubaldi, 2013, p. 6) and open data can be defined as "data that can be freely used, re-used and distributed by anyone, only subject to (at the most) the requirement that users attribute the data and that they make their work available to be shared as well" (Ubaldi, 2013, p. 6).

Governmental organisations are sharing their data with the public to attain various objectives, such as increasing transparency, data usability, participation, and economic growth (Alexopoulos et al., 2013; Attard et al., 2016; Kassen, 2013; Zeleti et al., 2016). Many governments openly share their data through national open data platforms. In the Netherlands for instance, the national Open Government Data Portal (OGDP) can be found on data.overheid.nl. Other examples of countries with open data platforms are the United Kingdom (data.gov.uk), Spain (datos.gob.es) and the United States (data.gov).

Governments are looking for ways to increase the benefits and value generated through their open data initiatives. However, open government data initiatives often do not succeed in achieving their goals (Zuiderwijk-van Eijk & Reuver, 2021). One initiative to increase the value of open government data initiatives is the creation of so-called data communities.

In 2020 the Dutch national open government data platform started with the data communities pilot, following best practices from other countries such as France and Spain (Ministry of the Interior and Kingdom Relations, 2020). The communities connect data owners, providers, users and researchers. In total, five data communities have already been created on the topics of mobility, social security, education, energy, and the migration chain (KOOP, n.d.).

According to the information on the Dutch OGDP (KOOP, n.d.), at least one leader and a number of experts take part in each community. The initiator of a community is also its manager, regulating the day-to-day affairs and, when necessary, draws up rules about the community. The experts provide the content on specific topics within the community. Based on other countries and early insights from the first data community on data.overheid.nl, the Ministry of Interior and Kingdom Relations (2020) concludes that the domain centred approach for communities increases insight into the impact of the open data and also strengthens transparency and participation.

However, the effects of data communities have barely been investigated. Furthermore, participation of actors with Open Government Data is not even a part of many of the different OGD benchmarks (Zuiderwijk, 2021). Morelli (2017, p. 129) implies a positive effect on usage of open data, by stating that the existence of a community is "essential for that resource to be used as a common". Related research, for instance into the effects of citizen engagement on open data is present (Purwanto et al., 2020). These effects can be positive, but there may also be negative effects. Moreover, there may be potential to enhance the positive effects of open data communities. Thus it becomes clear that further research is needed in order to understand all the effects of open government data usage.

## 1.2 Defining open data communities

Cooper and Springer (2019, p. 8) define a data community as follows: "a fluid and informal network of researchers who share and use a certain type of data". The definition of Cooper and Springer (2019) focusses on researchers, while the definition used in this thesis will be broader. Since this research focusses on communities that can include other data users than just researchers, the definition of Cooper and Springer (2019) needs to be altered by specifying what a community is. Merriam-Webster (n.d.-b), a dictionary, defines a community as "a body of persons of common and especially professional interests scattered through a larger society".

Data communities can be focussed on open data, as well as focussed on closed data. For instance, Forbes (2020) describes the case for data communities in the commercial sector, focussing on business opportunities for using data communities around closed data. Government data can be both open and closed, the main difference is that closed data cannot be generally accessed and is also

not circulated (Birchall, 2016). According to Birchall (2016, p. 2), government data can be closed for various reasons: “it is withheld from general access and circulation for reasons concerned with diplomacy, stability, power play, or security”.

Open Government Data combines the concepts of linked data, open data, big data and government data (Charalabidis et al., 2018). Since this research focusses on data communities around Open Government Data, it will only examine open data communities.

Combining the definitions of both Cooper and Springer (2019) and Merriam-Webster (n.d.-b), and rephrasing data to open data, results in the following definition of an open data community: ‘a fluid and informal network of persons with a common (professional) interest, that share and use a specific sort of open data’. This will be the definition used in this research.

## 1.3 Objective, problem definition and research questions

The problem statement of this research concerns the lack of knowledge about the effects of data communities on national Open Government Data Portals. This lack of knowledge makes it difficult to understand the potential contribution of the open data communities to national open data value creation, as well as other (dis)advantages that should be taken into account when adding the communities to existing data portals. Policy makers and developers of new and existing data portals want to know whether or not it is worth investing in setting up and maintaining/supporting data communities. After all, setting up an open data community does cost time and money.

This problem statement culminates in the following main research question:

*“What are the potential effects of open data communities on open data benefits and barriers?”*

The answer to this question provides insight into the specific effects of open data communities and also includes recommendations for designing and implementing open data communities in order to enhance the effects. The recommendations are systematically derived from a case study. To answer the main research question, multiple sub questions were formulated. All sub questions are briefly explained in the remainder of this paragraph.

*Sub question 1: What are the benefits and barriers of open data that can potentially be influenced by open data communities?*

The first sub question aims to elaborate on the concept of open data communities in general. After that, a list of open data benefits and barriers that are possibly affected by an open data community can be compiled. To compile this list, existing research, theories and frameworks regarding

benchmarks of open data and the various open data benefits and open data barriers are discussed. The theoretical framework constructed in the next chapter will answer this question.

*Sub question 2: Which open data benefits and barriers are influenced by current national Open Government Data communities?*

The second sub question aims to study the effects that open data communities have on the open data benefits and barriers that were identified earlier by existing research. It is important to gain more insight into the (possible) effects that are empirically visible on open data communities, thus a case study will be used to answer this question.

*Sub question 3: Which institutional instruments contribute to the value creation of current national Open Government Data communities?*

Institutional theory can be used to better understand how the behaviour of open data communities' users affects the contribution of the open data communities on open data benefits and barriers. According to Scott (2005, p. 2) institutional theory "considers the processes by which structures, including schemas, rules, norms, and routines, become established as authoritative guidelines for social behaviour". The case study will identify whether institutional instruments are currently visible on the open data community.

*Sub question 4: Which challenges and recommendations can be derived from the studied Open Government Data Communities?*

The last sub question focusses on the best practices that can be derived from the case study. By cross-referencing the benefits and barriers of open data found by previous research with the insights from actual community users, a broader perspective can be offered. The results can be used as input for a future evaluation or comparison method for open data communities.

## 1.4 Scientific and societal relevance

A lot of research has been done on open data, for instance regarding open data benefits (Janssen et al., 2012), as well as regarding open data barriers (Barry & Bannister, 2014). Open data initiatives can for instance contribute to increasing transparency and economic growth (Alexopoulos et al., 2013; Zeleti et al., 2016). However, there is a knowledge gap when it comes to the value and functioning of open data communities.

As already explained in paragraph 1.2, the current definition by Cooper and Springer (2019) does not suffice for data communities that can be found in the practice of sharing open data. Cooper

and Springer (2019) identify three features of successful data communities but, since their definition is not suffice for open data communities, these features do not focus on open data specifically. It is also unclear how open data communities should ideally be set up to lead to as many positive effects as possible and it is unclear what these positive effects are at all.

Not only is scientific literature lacking, but it is also unclear for policy makers what can be achieved with an open data community. Open data has enormous potential, from improved efficiency of public administration, economic growth for businesses to improving social welfare (Data Europe EU, n.d.-b; The World Bank, n.d.). Policy makers and developers of existing and new Open Government Data Portals want to know whether it is worth investing in setting up and maintaining/supporting data communities. Open data has benefits for both citizens and governments (Data Europe EU, n.d.-b) and the question for policy makers is whether this applies for open data communities as well. Open Data Initiatives, such as an open data community, "require substantial commitment regarding cost, personnel, and skills" (Berrone et al., 2016, p. 59) and therefore the added value of the open data communities should also be proven.

## 1.5 Thesis outline

In the next chapter the theoretical framework of this thesis is introduced, including open data barriers and benefits found in literature. This chapter, chapter 2, also describes institutional theory. Chapter 3 elaborates upon the chosen research methods and selection methods for the case-study informants and documents. This chapter also discusses the validity of the research. The fourth chapter entails the execution of the document analysis. Chapter 5 focusses on the analysis of the in-depth interviews, which are used to add information that is missing in the policy documents. Chapter 6 reflects upon the results of the analysis first and subsequently reflects upon the used literature, theories and methods. Chapter 7 delineates the conclusions of the analysis and provides the societal and scientific contribution of the research, as well as providing suggestions for future research. The case study protocol and interview protocols can be found in the appendix.

# 2 Theoretical framework

This chapter frames the research by describing the relevant concepts, theories and models in relation to the problem statement and research questions that were formulated in the previous chapter. First, the literature search strategy is presented and the relevant definitions are stated, for instance the definition of Open Government Data. The characteristics of successful data communities are introduced next and the implementation of the communities on OGDPs is described. Moreover, the absence of open data communities in open data benchmarks is discussed.

The second part of the theoretical framework focusses on open data benefits and barriers and institutional theory. Open data literature is used to identify those benefits and barriers. Finally, institutional theory is introduced and its relevance elaborated.

## 2.1 Literature search strategy

In order to identify relevant literature, a search strategy is formulated. The literature review consists of academic resources that were found using Google Scholar, SCOPUS and Web of Science. It also contains few policy documents, found by using Googles normal search engine, as well as open-overheid.nl and data.overheid.nl. The following search queries were used (some of them translated to Dutch as well):

- Open government data (portals)
- Open (government) data communities
- Open (government) data engagement
- Open government AND communities
- Government Data communities
- Effects OR barriers OR benefits OR … open data communities

The brackets indicate that the queries were both used with and without the term in-between and everywhere the word communities is used, the singular form (community) has also been entered.

The literature review process consisted out of three phases, as shown in table 2. After searching with numerous variations based on the different queries above (phase 1), 33 papers and reports were identified as relevant for this proposal. Then, the snowballing method was applied (phase 2). Snowballing refers to the process of using the citations of a paper to identify new relevant literature.

After searching and snowballing in total 45 relevant papers and reports were found. The papers were scanned and categorised (phase 3). This resulted in the selection of 23 papers and reports, based on their relevance and link to the research subject. Besides scientific literature and

reports, also five other sources are part of the literature review (for instance definitions from a dictionary). The amount of papers per selection phase can be found in table 2.

*Table 2: Literature reviewing process*

| ***Phase*** | **Phase name** | **Total number of papers and reports** | **Total number of other sources** |
|---|---|---|---|
| *1* | Initial selection, scanning | 33 | 8 |
| *2* | Snowballing | 45 (+12) | 9 (+1) |
| *3* | Categorizing and relevance selection | 23 (-22) | 5 (-4) |

The goal of the categorisation and selection was to provide a clear overview of relevant literature, starting from a very broad perspective and ending at research that is closely related to the challenge mentioned in the introduction. The following paragraphs will summarise the relevant research, define the shortcomings in literature and provide a synthesis of what new research should incorporate.

## 2.2 The landscape and implementation of (open) data communities

Cooper and Springer (2019) describe the characteristics of three studied successful data communities, which are all within the scientific field. According to them the features can be divided into three different categories. The first characteristic is bottom-up development. The communities all started a long time ago, as small-scale collaborations and to facilitate communication between researchers and grew naturally as colleagues empirically witnessed the benefits their follow researchers derived from sharing their data (Cooper & Springer, 2019). They also found out that the communities expanded narrowly around research interests, mirroring the involved researchers' professional networks.

The second characteristic is the absence of technical barriers to sharing, or this should be mitigated (Cooper & Springer, 2019). There should be no technical barriers to uploading, transferring, and reusing data. This also includes the file format of the datasets. The third and last characteristic involves community norms. Cooper and Springer (2019) indicate the importance of noticing the way in which data sharing is rewarded or encouraged within those scientific data communities.

Although the communities studied by Cooper and Springer (2019) are small-scale, closed scientific research communities, their features are still relevant for studying open data communities that

operate on a national level. The next paragraph will elaborate on Open Government Data Portals, which correspond with the second described feature.

To publish their datasets, governments use Open Government Data Portals (OGDPs). Many governments, such as the Netherlands, Italy, Spain, the United Kingdom, and the United States have those portals in place. Open data communities are implemented on Open Government Data Portals, which are web-based interfaces. The OGDPs are designed specifically to find reusable information in an accessible way (European Commission, n.d.). Like library catalogues, the portals include metadata of each dataset. Metadata is data that "provides information about other data" (Merriam-Webster, n.d.-c), by describing the context, content, and structure of the dataset published.

The portals also feature extensive search functionalities, helping users to find applicable datasets. To lower the barrier of (re)using open data even more, Application Programming Interfaces (APIs) are frequently available on the portals (European Commission, n.d.). Through APIs users have direct and automated access to data which can be used in software applications.

## 2.3 Open data communities in open data benchmarks

Globally and specifically for Europe, the efforts of national governments in the field of open data are measured regularly by benchmarks. A benchmarks is defined as "a point of reference from which measurements may be made" (Merriam-Webster, n.d.-a). The Open Data Barometer is one of the many initiatives that benchmark governmental open data efforts. In their last report, of the 115 countries studied, open data in some form or another is available in 97% of all countries (Brandusescu et al., 2016). However, less than three quarters of the datasets are up to date and less than one quarter of the datasets can be accessed and downloaded without a license. Furthermore, the usability of these portals can also vary greatly. The Open Data Barometer is just one of the many initiatives that try to benchmark the open data efforts from national governments worldwide, or specifically in Europe. Other well-known benchmarks are:

- The Global Open Data Index
- The EU Open Data Maturity Report
- The OECD's OURData index

According to Zuiderwijk et al. (2021), the metrics used to benchmark Open Government Data (OGD) initiatives differ considerably. Furthermore, most OGD benchmarks do not include the participation of actors in OGD use (Zuiderwijk et al., 2021). The way in which the involvement of open data users is stimulated or interaction is promoted, for example through the use of open data communities, is absent in most benchmarks. However, in an evaluation framework for open data portals from Chu and

Tseng (2016) the use of data communities can be found. Although this is not a benchmark, the developed framework assesses open data platforms as well. Within the framework, the presence of an open data community is ranked as the highest implementation of the indicator 'Discussion' (one out of five indicators in the category other functionalities). The scores for this indicator are as follows (Chu & Tseng, 2016 pp. 8):

- 0 None of any mechanism for users to discuss
- 5 Provide remark or feedback mechanisms
- 10 Provide forums
- 15 Provide data communities

Attard et al. (2015) have compiled a list of aspects in open government initiative evaluations. They considered the following aspects: Data, functionality, features, stakeholder participation, initiative maturity, stakeholder feedback. Only two of the twenty-five evaluations included the participation of stakeholders. One of the two evaluations, the paper of Sayogo et al. (2014) specifically mentions data communities.

Sayogo et al. (2014) introduce a new set of factors to describe the stage of open government data portal development of which one is Engagement Capability. Together with the factor Data manipulation Capability it reviews the progress of open data portals. The OGD portals are divided into three categories: No Features, Participative and Collaborative. One of the conclusions of Sayogo et al. (2014, p. 1904) is that "The types of user engagement provided in the OGD portal might correlate to the level of user's engagement".

## 2.4 Open data benefits and the effects of data communities

The benefits of data communities are hardly described in scientific literature. According to the OECD OURData Index (2019) stakeholder engagement is required for data quality and completeness. According to Bruce and Shelley (2010), stakeholder engagement is "the interaction between an organisation and those individuals and groups that are impacted by, or influence, the organisation". When interpreting the data-uploader as the organisation named in this definition, the interaction of this party with other individuals and groups on a community could be a means of stakeholder engagement. This suggests that setting up data communities impacts certain benefits of open data and that data communities can contribute to the successfulness of OGDPs. This raises the question which benefits of open data are present in literature. The World Bank (n.d.) lists a few of the benefits of open data including transparency, public service improvement, innovation and economic value and efficiency.

Numerous other sources describe the benefits of open data. The Centre for Strategic Economic Studies at Victoria University (2011) examined both costs and benefits of making Public Sector Information (PSI) data freely available. The publication quantitatively approached the benefits of open data and introduced an equation to calculate benefits. Janssen et al. (2018) identify both the benefits and barriers of opening government data. Benefits were (categorised as) either Political and Social, Economic or Operational and technical.

Kawashita et al. (2022) used most of these papers, amongst others, to identify benefits, barriers, drivers and enablers of open data. The table beneath (Table 3) presents the category of the benefit, the name of the benefit, and the papers in which the benefit is mentioned (all according to Kawashita et al. (2022)), and whether the benefit is possibly affected by open data communities.

Since there is a lack of literature on open data communities, the assessment of whether a benefit is affected by open data communities is based upon the judgement of the author of this research. The definition of open data communities was used. In order to increase validity, the assessment is verified by an open data expert. The assessment resulted in one out of the following three options: the benefit is either presumably not related to open data communities, possibly related to open data communities or presumably related to open data communities. The choices are briefly explained.

*Table 3: Open data benefits and categories present in literature, according to Kawashita et al. (2022)*

| ***Category*** | **Benefit** **(according to Kawashita et al., 2022, pp. 2537 – 2538)** | **Sources** **(according to Kawashita et al., 2022, pp. 2537 – 2538)** | **Affected by data communities** |
|---|---|---|---|
| *Political and social* | Increased transparency | (Albano et al., 2017; Zuiderwijk et al., 2019; Jamieson et al., 2019; Janssen et al., 2012; Albano & da Silva Craveiro, 2015; Albano & Reinhard, 2014; Albano & Reinhard, 2015; da Silva Craveiro & Albano, 2015) | **Presumably not**, the benefit appears related to the publication of data |
| | Increased social control | (Albano et al., 2017; Zuiderwijk et al., 2019; Jamieson et al., 2019; Janssen et al., 2012; Albano & da Silva Craveiro, 2015; Albano & Reinhard, 2014; Albano & Reinhard, 2015; da Silva Craveiro & Albano, 2015) | **Possibly**, the community allows feedback mechanisms (including criticism) on open data |

| | | | |
|---|---|---|---|
| | Increased civic participation and public engagement | (Albano et al., 2017; Albano & Reinhard, 2014; Jamieson et al., 2019; Janssen et al., 2012; Zuiderwijk et al., 2019; Albano & da Silva Craveiro, 2015; da Silva Craveiro & Albano, 2015) | **Presumably**, the community allows for participation and engagement between publisher and user |
| | More informed citizens | (Jamieson et al., 2019) | **Possibly**, the community allows to inform community members |
| | Increased accountability | (Zuiderwijk et al., 2019; Albano et al., 2017; Jamieson et al., 2019; Janssen et al., 2012) | **Possibly**, the community allows taking ownership of the produced data |
| | Gained new knowledge and insights into the public sector | (Janssen et al., 2012; Zuiderwijk et al., 2019; Albano & da Silva Craveiro, 2015; da Silva Craveiro & Albano, 2015) | **Presumably not**, the benefit appears related to the publication of data itself |
| *Economic and financial* | Increased administrative efficiency | (Zuiderwijk et al., 2019; Jamieson et al., 2019) | **Presumably not**, the benefit appears related to the publication of data itself |
| | Reduced operating costs | (Zuiderwijk et al., 2019; Chorley, 2017) | **Presumably not**, the benefit appears related to the publication of data itself |
| | Gained access to external capacity and resources for solving problems | (Zuiderwijk et al., 2019) | **Presumably**, data communities can help in getting in touch with external capacity |
| | Offer of improved processes, products, and services | (Janssen et al., 2012; Zuiderwijk et al., 2019; da Silva Craveiro & Albano, 2015) | **Presumably not**, the benefit appears related to the publication of data itself |
| *Operational and technical* | Development of improved public policies | (Janssen et al., 2012; Zuiderwijk et al., 2019; da Silva Craveiro & Albano, 2015; Jamieson et al., 2019; Parkes et al., 2018; Albano & da Silva Craveiro, 2015) | **Presumably not**, the benefit appears related to the publication of data itself |
| | Increased intra-governmental collaboration | (Albano & Reinhard, 2014; Janssen et al., 2012) | **Presumably**, data communities can include different governmental parties |
| | New processes, products, and services developed | (Janssen et al., 2012; Albano & da Silva Craveiro, 2015; da Silva Craveiro & Albano, | **Presumably not**, the benefit appears related to the publication of data itself |

| | | |
|---|---|---|
| | 2015; Albano & Reinhard, 2014) | |
| Improved processes | (Janssen et al., 2012; Zuiderwijk et al., 2019; Albano & da Silva Craveiro, 2015; Albano & Reinhard, 2015; Albano & Reinhard, 2014; Melin, 2016) | **Presumably not**, the benefit appears related to the publication of data itself |
| Improved data management | (Janssen et al., 2012; Zuiderwijk et al., 2019; da Silva Craveiro & Albano, 2015; Albano & da Silva Craveiro, 2015) | **Presumably not**, the benefit appears related to the publication of data itself |
| Improved access to public services | (Parkes et al., 2018; Chorley, 2017) | **Presumably not**, the benefit appears related to the publication of data itself |
| Increased efficiency in making changes in service delivery | (Janssen et al., 2012; Jamieson et al., 2019; Parkes et al., 2018; Albano et al., 2017) | **Presumably not**, the benefit appears related to the publication of data itself |
| Decision-making process more informed | (Janssen et al., 2012; Zuiderwijk et al., 2019; da Silva Craveiro & Albano, 2015) | **Presumably not**, the benefit appears related to the publication of data itself |
| Innovation support processes deployed | (Janssen et al., 2012; Zuiderwijk et al., 2019; Albano & Reinhard, 2014; Jamieson et al., 2019) | **Presumably not**, the benefit appears related to the publication of data itself |
| Increased problem-solving capacity | (Janssen et al., 2012; Chorley, 2017) | **Presumably**, communities can allow for problems to be addressed and solved (specifically related to the data) |
| Use of collective intelligence to solve public problems | (Janssen et al., 2012; Zuiderwijk et al., 2019; da Silva Craveiro & Albano, 2015; Albano & Reinhard, 2014) | **Possibly**, collective intelligence appears mainly used on the open data itself, but communities may increase usage of it |

Based upon the table above, the benefits that were either assessed as presumably or possibly were extracted. The table underneath (Table 4) shows the benefits possibly or presumably related to open data communities.

*Table 4: Open data benefits presumably or possibly related to data communities*

| *Category* | Benefit (according to Kawashita et al., 2022, pp. 2537 – 2538) | Benefit # | Related to data communities |
|---|---|---|---|
| *Political and social* | Increased social control | **1** | **Possibly**, the community allows feedback mechanisms (including criticism) on open data |
| | Increased civic participation and public engagement | **2** | **Presumably**, the community allows for participation and engagement between publisher and user |
| | More informed citizens | **3** | **Possibly**, the community allows to inform community members |
| | Increased accountability | **4** | **Possibly**, the community allows taking ownership of the produced data |
| *Economic and financial* | Gained access to external capacity and resources for solving problems | **5** | **Presumably**, data communities can help in getting in touch with external capacity |
| *Operational and technical* | Increased intra-governmental collaboration | **6** | **Presumably**, data communities can include different governmental parties |
| | Increased problem-solving capacity | **7** | **Presumably**, communities can allow for problems to be addressed and solved (specifically related to the data) |
| | Use of collective intelligence to solve public problems | **8** | **Possibly**, collective intelligence appears mainly used on the open data itself, but communities may increase usage of it |

## 2.5 Open data barriers and the effects of data communities

Objectives of Open Government Data Initiatives (OGDI) are not always achieved, but it is unclear to which extent those objectives are actually reached (Zuiderwijk et al., 2019). The work of Zuiderwijk et al. (2018, p. 667) also adds a distinction on different scale levels, stating that "the objectives of state- and national-level OGDIs are more often achieved compared to those of local- and regional-level OGDIs".

Various authors describe barriers for open data, which may explain why objectives are not met. Also, the effects of data-communities possibly resolve some of those barriers. Therefore, this paragraph will discuss literature found on this topic.

Van Veenstra and Van Den Broek, (2013) argue that the drivers they investigated, as related to opening up data, are the same for the whole process of opening data. The barriers change however. At the start of this process, organisational factors gained much attention, while later in the process the attention shifted to factors related to the re-use of the open data. Van Veenstra and Van Den

Broek (2013, p. 3) divided the barriers were into four groups: “information technology, organisational and managerial, legal and regulatory, and institutional and environmental”.

Smith and Sandberg (2018) also researched open data barriers, resulting in the identification of three OGD user archetypes, namely: (1) employees, (2) entrepreneurs and (3) hobbyists. Furthermore, Smith and Sandberg (2018) argue that the significance of the barriers differs across lifecycles and depends on the several properties of the OGD users. These properties include the motivation, objective, pre-conditions and approach of the users (Smith and Sandberg, 2018). Using a case study approach, focussing on the importance and function of training users, Gascó-Hernández et al. (2018) derived three conclusions regarding promoting OGD usage. The last conclusion focusses on the properties of open data users, similar to the just mentioned statement of Smith and Sandberg (2018). Gascó-Hernández et al. (2018, p. 1) indicate that “embedding the training interventions in the specific contexts and considering the unique characteristics, interests, and expectations of different types of users is critical to success”.

Janssen et al. (2018) identify barriers of opening government data. The barriers can be categorised into six areas. Barriers can be related to: (1) institutional aspects, (2) task complexity, or (3) use and participation. The barriers can also be related to (4) legislation, (5) information quality, or (6) technical aspects. There is much more relevant literature regarding barriers of open data, including the work of Barry and Bannister (2014), Martin (2013) and Zuiderwijk et al. (2012), all using different viewpoints, respectively top view, risk analysis and social-technical.

Kawashita et al. (2022) also used most of the papers above to identify barriers in open data. The table beneath (Table 5) presents the category of the barrier, the name of the barrier and the papers in which the barrier is mentioned (all according to Kawashita et al. (2022)), and whether the barrier is possibly affected by open data communities.

Similar to the assessment of the open data benefits, there is a lack of literature on open data communities relating to open data barriers. The assessment of whether a barrier is affected by open data communities is therefore based upon the judgement of the author of this research. The definition of open data communities was used. In order to increase validity, the assessment is as verified by an open data expert. The assessment resulted in one out of the following three options: the barrier is either presumably not related to open data communities, possibly related to open data communities or presumably related to open data communities. The choices are briefly explained. In order to increase validity, the assessment of the barriers related to the communities was verified by an open data expert as well.

*Table 5: Open data barriers and categories present in literature, according to Kawashita et al. (2022)*

| ***Category*** | **Barrier** **(according to Kawashita et al., 2022, pp. 2538 – 2539)** | **Sources** **(according to Kawashita et al., 2022, pp. 2538 – 2539)** | **Affected by data communities** |
|---|---|---|---|
| *Policy and legal* | Open data is not a political priority | (Cahlikova & Mabillard, 2020) | **Presumably not**, the barrier appears related to politics |
| | Strategy and/or leadership do not support open data use | (Ra & Lam, 2019; Kučera, 2017; Chorley, 2017) | **Presumably not**, the barrier appears related to leadership |
| | Open data policy is inadequate or lacking | (Kučera, 2017; Huang et al., 2017; Ra & Lam, 2019; Shao & Saxena, 2019) | **Presumably not**, the barrier appears related to policy |
| | Open data use laws are inadequate or lacking | (Albano & da Silva Craveiro, 2015; Albano & Reinhard, 2015; Rhoa et al., 2019; Shao & Saxena, 2019) | **Presumably not**, the barrier appears related to policy |
| | Open data use standards and guidelines are inadequate or lacking | (Shao & Saxena, 2019; Smith & Sandberg, 2018) | **Presumably not**, the barrier appears related to policy |
| | Use license does not exist, or unclear, or restrict use | (Janssen et al., 2012; Smith & Sandberg, 2018; Kučera, 2017; da Silva Craveiro & Albano, 2015; Mockus, 2016; Rhoa et al., 2019) | **Presumably not**, the barrier appears related only to open data itself |
| *Cultural* | Lack of organisational culture favorable to open data | (Janssen et al., 2012; Albano & da Silva Craveiro, 2015; Kučera, 2017; Cahlikova & Mabillard, 2020; Shepherd et al., 2019; Kim & Eom, 2019; Wieczorkowski, 2019; Smith & Sandberg, 2018) | **Presumably not**, the barrier appears related to organisational culture |
| | Management and public managers do not know what open data is | (Chorley, 2017; Shao & Saxena, 2019) | **Presumably not**, the barrier appears related to a basic knowledge level |
| *Economic and financial* | There is no budget, no financial resources to use open data | (Janssen et al., 2012; Kučera, 2017; Rhoa et al., 2019; Shao & Saxena, 2019) | **Presumably not**, the barrier appears related to finances |
| | Cost to hire people able to use open data is very high | (Kučera, 2017; Cahlikova & Mabillard, 2020; da Silva Craveiro & Albano, 2015; Chorley, 2017; | **Presumably not**, the barrier appears related to finances |

| | | | |
|---|---|---|---|
| | | Shepherd et al., 2019; Wieczorkowski, 2019) | |
| *Organisational and institutional* | Upper management does not support open data use | (Chorley, 2017; Rhoa et al., 2019) | **Presumably not**, the barrier appears related to organiational support of open data |
| | The organisational structure does not support open data use | (Kučera, 2017; Albano & Reinhard, 2014; Cahlikova & Mabillard, 2020; Huang et al., 2017; Shao & Saxena, 2019) | **Presumably not**, the barrier appears related to organisational support of open data |
| | The organisation is not interested in using open data | (Janssen et al., 2012; Kučera, 2017; Albano & Reinhard, 2014; Cho & Lee, 2020; Huang et al., 2017; Crusoe et al., 2019; Ra & Lam, 2019) | **Possibly**, open data communities may increase interest |
| | The organisation does not have the organisational capabilities, routines, and processes to use open data | (Janssen et al., 2012; Kučera, 2017; da Silva Craveiro & Albano, 2015; Rhoa et al., 2019) | **Possibly**, data communities may increase capabilities (of the public) |
| | Low engagement of public managers with the use of open data | (Albano & Reinhard, 2014; Cahlikova & Mabillard, 2020; Rhoa et al., 2019) | **Possibly**, managers may also be engaged through communities |
| | Lack of non-financial resources (time, equipment, etc.) to use open data | (Janssen et al., 2012; Kučera, 2017; da Silva Craveiro & Albano, 2015) | **Presumably not**, the barrier appears related to recources that cannot be provided by a community |
| | There is no definition of competencies for the use of open data | (Shao & Saxena, 2019) | **Presumably not**, the barrier appears related only to open data itself |
| | There are no human resources with the knowledge, skills and/or capabilities to use open data | (Janssen et al., 2012; Kučera, 2017; Chorley, 2017; Huang et al., 2017; Luna-Reyes, 2018; Ra & Lam, 2019; Rhoa et al., 2019; Shao & Saxena, 2019; Smith & Sandberg, 2018) | **Possibly**, communities may connect (more) skilled people to the open data |
| | Data provider ignores my requests and suggestions | (Janssen et al., 2012; Kučera, 2017; Zuiderwijk et al., 2012a; da Silva Craveiro & Albano, 2015; Saxena, | **Presumably**, the barrier appears related to engagement and responsiveness |

| | | | |
|---|---|---|---|
| | | 2018b; Luna-Reyes, 2018; Ra & Lam, 2019) | |
| *Operational and technical* | There is no centralized portal or open data infrastructure | (Albano & Reinhard, 2014; Luna-Reyes, 2018) | **Presumably not**, the barrier appears related to infrastructure |
| | Difficulty in discovering/locating data | (Janssen et al., 2012; Crusoe & Melin, 2018; Luna-Reyes, 2018) | **Possibly**, communities can provide guidance for the data portal |
| | Data availability issues | (Janssen et al., 2012; Crusoe et al., 2019; Wieczorkowski, 2019; Kučera, 2017; Saxena, 2017; Rhoa et al., 2019) | **Presumably not**, the barrier appears related to the infrastructure/open data itself |
| | Data accessibility issues | (Janssen et al., 2012; Kučera, 2017; Wieczorkowski, 2019; Eberhardt & Silveira, 2018; Gebre & Morales, 2020; Luna-Reyes, 2018; Ra & Lam, 2019; Smith & Sandberg, 2018) | **Presumably not**, the barrier appears related to the infrastructure/open data itself |
| | Poor data quality (incomplete, inaccurate, obsolete, and/or inconsistent) | (Janssen et al., 2012; Crusoe & Melin, 2018; Albano & Reinhard, 2014; Albano & Reinhard, 2015; Shepherd et al., 2019; Cho & Lee, 2020; Eberhardt & Silveira, 2018; Crusoe et al., 2019; Wieczorkowski, 2019; Saxena, 2017; Huang et al., 2017; Gebre & Morales, 2020; Saxena, 2018b; Luna-Reyes, 2018; Ra & Lam, 2019; Rhoa et al., 2019; Shao & Saxena, 2019; Smith & Sandberg, 2018) | **Presumably not**, the barrier appears related to the infrastructure/open data itself |
| | Low data usability (fragmented data, no user-friendly format, not machine readable) | (da Silva Craveiro & Albano, 2015; Chorley, 2017; Huang et al., 2017; Janssen et al., 2012; Wieczorkowski, 2019; Huang et al., 2017; Crusoe & Melin, | **Presumably not**, the barrier appears related to the infrastructure/open data itself |

| | | |
|---|---|---|
| | 2018; Luna-Reyes, 2018; Shao & Saxena, 2019) | |
| Metadata insufficient and/or lacking | (Janssen et al., 2012; Crusoe & Melin, 2018; Kučera, 2017; Albano & da Silva Craveiro, 2015; Shepherd et al., 2019; Crusoe et al., 2019; Saxena, 2017; Rhoa et al., 2019; Shao & Saxena, 2019; Smith & Sandberg, 2018) | **Presumably not**, the barrier appears related to the infrastructure/open data itself |
| Data are not compatible and cannot be compared | (Crusoe & Melin, 2018; Janssen et al., 2012; da Silva Craveiro & Albano, 2015; Crusoe et al., 2019; Rhoa et al., 2019) | **Presumably not**, the barrier appears related to the infrastructure/open data itself |
| Data cannot be combined and connected | (Crusoe et al., 2019) | **Possibly**, data communities can give guidance in connecting datasets |
| Data are not sustained or maintained | (Huang et al., 2017; Smith & Sandberg, 2018) | **Presumably not**, the barrier appears related only to open data itself |
| Data sources are unreliable | (Janssen et al., 2012) | **Presumably not**, the barrier appears related to reliability of the datasets |
| Results obtained from different sources differ | (Janssen et al., 2012) | **Presumably not**, the barrier appears related to platform characteristics/ differences between data portals |
| Uploading data is not supported | (da Silva Craveiro & Albano, 2015) | **Presumably not**, the barrier appears related to platform capabilities/ data portal functionality |
| No processing power | (Wieczorkowski, 2019; da Silva Craveiro & Albano, 2015) | **Presumably not**, the barrier appears related to technical barriers |
| Lack of proper tools to manipulate data | (Shao & Saxena, 2019; Smith & Sandberg, 2018) | **Presumably not**, the barrier appears related only to (availability of) tools |
| Difficulties to interact with the data provider | (Gebre & Morales, 2020; Huang et al., 2017) | **Presumably**, related to engagement/interaction (can be facilitated by data communities) |

Based upon the table above, the barriers that were either assessed as presumably or possibly were extracted. The table underneath (Table 6) shows the barriers possibly or presumably related to open data communities.

*Table 6: Open data barriers presumably or possibly related to data communities*

| ***Category*** | **Barrier** **(according to Kawashita et al., 2022, pp. 2538 – 2539)** | **Barrier #** | **Related to data communities** |
|---|---|---|---|
| *Organisational and institutional* | The organisation is not interested in using open data | **1** | **Possibly**, open data communities may increase interest |
| | The organisation does not have the organisational capabilities, routines, and processes to use open data | **2** | **Possibly**, data communities may increase capabilities (of the public) |
| | Low engagement of public managers with the use of open data | **3** | **Possibly**, managers may also be engaged through communities |
| | There are no human resources with the knowledge, skills and/or capabilities to use open data | **4** | **Possibly**, communities may connect (more) skilled people to the open data |
| | Data provider ignores my requests and suggestions | **5** | **Presumably**, the barrier is related to engagement and responsiveness |
| *Operational and technical* | Difficulty in discovering/locating data | **6** | **Possibly**, communities can provide guidance for the data portal |
| | Data cannot be combined and connected | **7** | **Possibly**, data communities can give guidance in connecting datasets |
| | Difficulties to interact with the data provider | **8** | **Presumably**, related to engagement/interaction (can be facilitated by data communities) |

## 2.6 Institutional theory and its influence on communities

By using benefits and barriers, the effects of data communities on open data value creation can be measured. However, to gain understanding into how these communities function, a different kind of analysis is needed. Cooper and Springer (2019) found that the community norms were one out of three features of successful scientific communities.

This paragraph will introduce institutional theory, a theory which can be used to better understand which properties of data communities enhance or diminish the effects of the open data community on the benefits and barriers of open data. According to Scott (2005, p. 2), institutional theory "attends to the deeper and more resilient aspects of social structure. It considers the processes by which structures, including schemas, rules, norms, and routines, become established as authoritative guidelines for social behaviour." He adds that "It inquiries into how these elements are

created, diffused, adopted, and adapted over space and time; and how they fall into decline and disuse" (Scott, 2005, p. 2).

Fogarty (1996) describes the added value of institutional theory by indicating that there is a distinction between what an organisation achieves and what should be potentially achieved. The latter can be uncovered by institutional theory, by looking at what is suggested to the environment by the structures of the organisation.

Kim et al. (2009) uses three dimensions (or forces) of institutionalisation, which are coherent yet also clearly distinguishable. The first dimension introduced is the regulatory or coercive mechanism. This force is based upon influences from both politics and legislation. "The regulatory factors are affected by politics and legislations and influenced by firms' decisions to adopt a specific organisational practice" (Kim et al., 2009, p. 43).

The second force described by Kim et al. (2019) is the mimetic mechanism. This mechanism entails how principles of different systems' can be copied (Scott, 2013). The mechanism becomes visible if uncertainty reigns, making organisations more likely to copy behaviour, or as Kim et al. (2019, p. 43) states "model themselves on other organisations or refer to culturally presumed meanings and ideologies".

Normative mechanisms form the last dimension of institutionalisation. This force is focussing on the norms that are prevailing and perceived in the organisations domain. These institutions consist of several elements (Scott, 2013) and those mechanisms structure choices, stress how something in an organisation should be done.

These three dimensions were applied in a single case study, part of the work of Kim et al. (2009). The case study entailed e-government and institutional theory was used because it could "identify challenges surrounding the implementation of e-government systems" (Kim et al., 2009, p. 44). The theory "elucidates how an innovation or new system developed in an organization is diffused, adopted, or copied by others" (Kim et al., 2009, p. 44). These three dimensions could therefore be used to review open data communities.

# 3 Research approach

This chapter discusses the research methods applied in this study. First, the reasons for choosing an explorative, qualitative research setup will be substantiated. After that, the methods used to collect data are explained. Then, the reason for conducting a single-unit case study is explained, followed by how both the different documents and interviewees were selected for the case study. The last part of this chapter will discuss the different types of validity regarding this research, as well as the reliability.

## 3.1 An explorative, qualitative case study

This research focusses on the contribution of open data communities to value creation through open data. The added value of open data communities has barely been described in the scientific literature.

According to Yin (2018), a case study can either be (1) exploratory, (2) descriptive, or (3) explanatory. Yin (2018) identifies three relevant properties when studying a situation and selection a research method, including the form of the research question, the control over behavioural events and the focus on contemporary events. Case studies are a suitable research method when the form of the research question is how or why, there is no need to control the behavioural events and the focus lays on contemporary events (Yin, 2018). This explains the exploratory nature of the research.

Furthermore, there is little to no scientific literature available on open data communities beforehand, which also matches with exploratory research (Brown, 2006). On the other hand, the advantages and disadvantages of open data have been described. Because the open data community is currently implemented by multiple governments, it is expected that the community potentially affects some of the advantages and disadvantages of open data. For that reason, the literature on the benefits and barriers of open data was used as a starting point for the data collection and analysis.

A selection of possibly relevant factors was made in advance (the assessment can be found in Chapter 2). In-depth interviews are used to test which benefits and barriers actually apply to the open data community. Because this research uses concepts found in literature in a practical situation, the research can be categorised as deductive (Azungah, 2018).

Because of the exploratory nature of the research, conducting qualitative research is more logical than quantitative research: qualitative research can be typed as exploratory and quantitative research as conclusive (Ahmad et al., 2019). Therefore, a qualitative research approach is selected.

Besides the exploratory nature, the amount of open data communities that is currently active is limited. According to Sekaran and Bougie (2016), a sample size ranging from 30 (up to 500), depending on various factors, is sufficient for studies within the field of social science. Since the

amount of active national open data communities is considerably lower, we can conclude that there are not enough suitable examples to obtain statistically reliable information.

Qualitative research focuses on collecting and analysing linguistic material such as existing documents or interviews (Yin, 2015). The research contains both primary and secondary data sources, since the document analysis is combined with in-depth interviews. Especially in the search for the added value of open data communities, it is important to analyse in what way managers and users of the open data community see that this community is or can be of added value. A qualitative research approach is therefore most appropriate. Since the interviews will discuss the Dutch open data community, the research will be descriptive (and not experimental).

## 3.2 Methods of data collection

This research used two research methods. The first method used is a document analysis. Information missing in the documents was then added by conducting in-depth interviews. The interviews were focused on three groups: (1) the open data community managers, (2) the open data community users, and (3) open data researchers. By using different independent data sources, which is also called method triangulation, the impact of potential biases that may occur is decreased (Bowen, 2009). The two next paragraphs discuss the document analysis and the in-depth interviews in more detail.

### 3.2.1 Document analysis

Document analysis is a systematic procedure that can be used for reviewing or evaluating documents (Bowen, 2009). The five documents that are selected for this analysis offer insight into the origins of the data community, expectations for its output or added value, and policy choices. This is an effective way of mapping out how processes have gone and at which moments certain decisions were made by the actors involved (Bleijenbergh, 2015).

This research specifically sought policy documents that are relevant to the chosen case study. Therefore, the documents should contain information regarding the open data strategy of the selected case study and the relationship between this strategy and the implementation of an open data community. The documents were retrieved in two different ways. First, publicly available government sources were consulted. The interview participants were also asked to provide documents relevant for this research. This can prevent 'biased selectivity', which is described in an organisational context by Bowen (2009, p 32) as "the available (selected) documents are likely to be aligned with corporate policies and procedures and with the agenda of the organization's principals".

### 3.2.2 In-depth interviews

To complement the document analysis, eight in-depth interviews are conducted with nine individuals who are either a managers or users of the data community or are researchers in a field related to open data engagement. In order to be able to make statements about the added value of this community, it is important that respondents are frequently active on the open data community.

The activity of those users may originate from different perspectives: either from the user perspective, the community management perspective or the research perspective. It is expected that most respondents will fall into the user group. These individuals are expected to be able to substantiate the added value they experience when using the open data community.

According to Yin (2018, p. 161), “Case study interviews will resemble guided conversations rather than structured queries”. The type of interviews used in case studies are also called unstructured interviews (Weiss, 1994). In this research however, semi-structured interviews (SSIs) will be conducted. This way of interviewing allows to ask participants about a particular phenomenon they have experienced (McIntosh & Morse, 2015). In this case, the phenomenon is the open data community and its effects open data benefits and barriers.

At the same time, the semi structured format allows to go in depth regarding how the interviewee perceives the effects. Although the main questions and interview topics are predetermined, the interviewee can still determine to a reasonable extent what to say and also how much to say (Drever, 1995).

The interviews took place online via Teams. This made it possible to record and automatically transcribe the interviews using Teams' built-in transcription tool. As the transcription tool of teams is not perfect each interview transcript was tested against the recording and altered when necessary to match the interviewees exact meaning of words. The transcriptions were used for further analysis.

The recordings were temporarily stored on the environment of Delft University of Technology. In order to guarantee the privacy of the respondents, the video recordings were removed after transcribing the interviews. The anonymised transcriptions are saved on the TU Delft environment and can also be requested (see appendix 3), in order to keep the research replicable.

## 3.3 Sample and selection

According to Yin (2018), a multi-unit case study is in many cases more favourable than a single case study. Therefore, the initial goal was to select two or three cases. However, the amount of active open

data communities is very limited. Many of the communities are either not active anymore, taken offline in the recent history or only open to civil servants.

Although preferring a multiple case study, conducting it appeared impossible. The criteria for the case selection, which can be found in the next section, resulted in just one appropriate case.. Therefore, the decision was made to switch to a single-unit case study.

### 3.3.1 Case selection

Different techniques can be used for selecting cases (Patton, 1990). For this research intensity sampling will be used. The sampling strategy can be defined as "The case is information rich but not an extreme case" (Patton, 1990, pp. 182-183). This sampling strategy is comparable to extreme case sampling because it uses the same logic: a lot of interesting information is available on the extremities of the distribution range (Patton, 1990). However, when choosing for intensity sampling the focus lies less on the extremes. There are three criteria formulated for selecting the cases:

1. Case study factor criterion: The first criterion focusses on the set data points of the case studies. The open data communities differ, but trends may be found. The criterion aims on identifying a uniform set of data points that are applicable for the different open data communities. Identifying these data points ensures that when executing the analysis, results can be compared with other case studies (Yin, 2018). The nation to which the open data community belongs to, its openness and roles are all relevant for this criterion.
2. Case study quality factor criterion: The second criterium focusses on the details of the individual open data communities. By defining the details of each case study, the information can be used to draw similarities and differences between cases. This criteria seeks to describe the different details of each community. Quality plans, policy documents, evaluations and use of audits are all examples of the details that are important to this criterion.
3. Case study performance information criterion: Lastly, it is important to quantify the success of the open data community of each country. As far as possible, the established performance criteria for the community's decisiveness and openness are used.

After an extensive search only three cases were found that could comply with the first and last tier of the criteria for the project selection, being (1) the United Kingdom (Energy themed) open data community, (2) the United States open data community, and (3) the Dutch open data community. All cases feature an open data community directly connected to the Open Government Data Portal (OGDP). Implementation of both the portal and community differs however. The table below (Table

7) compares the three data communities. Information that could not be found resulted in an empty cell.

*Table 7: National open data communities compared*

| ***Country*** | **NL** | **UK** | **US** |
|---|---|---|---|
| *URL to OGDP* | https://data.overheid.nl/ | https://data.gov.uk/ | https://data.gov/ |
| *Platform Standard* | DCAT | | |
| *Platform API* | CKAN | CKAN | CKAN<br>WordPress |
| *Open data license* | CC-0 | CC BY | U.S. Government Works (Public Domain).<br>Outside of the US: CC-0 |
| *URL to community* | https://data.overheid.nl/datacommunities | https://opendatacommunities.org/home | https://digital.gov/communities/ |
| *Location of community* | Centralised | Decentralised | Centralised |
| *URL to forum* | https://datacommunities.nl/ | https://support.environment.data.gov.uk/hc/en-gb/community/topics | https://groups.google.com/g/us-open-government |
| *Community system* | Huddle Forum | | LISTSERV Email list (Government only) & Google groups (public) |
| *Openness of community* | Open | Open | (Mostly) Restricted government |

The United Kingdom Open (Government) Data Portal contained decentralised communities. There was no central place where information regarding communities could be found and contact information was also not found.

Contact information was found for the US open data community. Multiple informants, including all community managers of the US Data Community were e-mailed to ask for more information and an interview. The open data community managers in the United States referred to their government policy regarding interviews, which unfortunately states that they are not allowed to participate in an interview. The US Community was furthermore limited in its openness, being available for Civil Servants with an United States government e-mail address only. Therefore, the second tier could not be complied with. As a result the open data community of the Netherlands was selected as the only suitable case. The last paragraph of this chapter (reliability and validity) will discuss the study limitations, due to this decision.

### 3.3.2 Document selection

In the process of selecting the cases numerous documents were found, twelve in total. Not all documents turned out to be relevant, or added new information, to the document analysis. The (five) documents that were selected for the document analysis are included in the table below (Table 8). For each document, besides the name and type of the document, also the source, publishing organisation and document language are stated.

*Table 8: Selected documents and their characteristics*

| ***Document Type*** | **Name** | **Author/Publishing Organisation** | **Document language** |
|---|---|---|---|
| *Letters to parliament* | Letter to parliament about Nationale Open Data Agenda | Ministry of the Interior and Kingdom relations | Dutch |
| *Policy documents* | Open Data Action Plan 2020 - 2022 | Ministry of the Interior and Kingdom relations | English |
| | Pilot open data communities concept (NL0048) | Open Government Partnership | English |
| *Community documents* | Starting document Energy community | Interviewee 1 | Dutch |
| | Community rules | KOOP | English (translated) |

### 3.3.3 Informant selection

After the case study was selected according to the paragraph above, relevant managers and users of the selected open data community were approached for an interview. Some of their names were also found in documents identified in the previous paragraph. To be able to answer the research question, the goal was to speak with approximately ten participants. There were three categories of potential interviewees:

1. Open data community managers: This category of informants includes those who are specifically managing the Dutch open data community;
2. Open data community users: This category of informants includes those who are members of communities within the selected case. It includes both civil servants and other parties and can be both from the side of the data-owner as the data-user;
3. Open data researchers: This category of informants includes those who are connected to an university as a researcher, specifically within the field of Open- or E-Government

In total, 32 relevant informants were identified. Most participants (23 in total) were contacted via e-mail or LinkedIn directly, based on their contact information present on the community. The other

nine participants were contacted via other informants, which for instance gave names during the interviews. The table beneath (Table 9) lists the informants that agreed on participating in an interview.

*Table 9: Interview participants and their characteristics*

| ***Community Role*** | **Anonymised name** | **Interview** | **Job position** | **Interview language** |
|---|---|---|---|---|
| *Community managers* | Interviewee 1 | Interview 1 | Open data community manager at KOOP | Dutch |
| | Interviewee 2 | Interview 1 | Open data community manager at KOOP | Dutch |
| *Community Users* | Interviewee 3 | Interview 2 | Community member at VIVET | Dutch |
| | Interviewee 4 | Interview 3 | Community member at VIVET | Dutch |
| | Interviewee 5 | Interview 4 | GEO-data specialist | Dutch |
| | Interviewee 6 | Interview 5 | Programme advisor data at Rijkswaterstaat | Dutch |
| | Interviewee 7 | Interview 6 | Data engineer at Rijkswaterstaat | Dutch |
| *Open data (community) researchers* | Interviewee 8 | Interview 7 | Open data PhD candidate | English |
| | Interviewee 9 | Interview 8 | Open data PhD candidate | English |

## 3.4 Method of analysis

The analysis used three methods to label the interview transcripts: open coding, axial coding and selective coding (Khandkar, 2009; Williams & Moser, 2019). Qualitative data analysis software is used to make the coding process more accurate and also more efficient. The transcripts of the interviews and the existing documents were imported into the qualitative data analysis program Atlas.ti (version 22.1.5). The interview protocol structured the interviews: each participant was first asked a few questions which were based on their community role. In this part of the interview, for each participant type a different set of questions was used.

All participants were the same set of questions regarding benefits, barriers and institutional instruments. The concluding questions of the interview were also the same for all community roles. The interview protocols can be found in appendix 2.

The coding process included labelling all interview answers into different categories. The table below (Table 10) shows the codenames and descriptions used in Atlas.ti.

*Table 10: Atlas.ti Code tree*

| ***Dimensions*** | **Codename** | **Code description** |
|---|---|---|
| *Open data Benefits* | Benefit 1 | Increased social control |
| | Benefit 2 | Increased civic participation and public engagement |
| | Benefit 3 | More informed citizens |
| | Benefit 4 | Increased accountability |
| | Benefit 5 | Gained access to external capacity and resources for solving problems |
| | Benefit 6 | Increased intra-governmental collaboration |
| | Benefit 7 | Increased problem-solving capacity |
| | Benefit 8 | Use of collective intelligence to solve public problems |
| *Open data Barriers* | Barrier 1 | The organisation is not interested in using open data |
| | Barrier 2 | The organisation does not have the organisational capabilities, routines and processes to use open data |
| | Barrier 3 | Low engagement of public managers with the use of open data |
| | Barrier 4 | There are no human resources with the knowledge, skills and/or capabilities to use open data |
| | Barrier 5 | Data provider ignores my requests and suggestions |
| | Barrier 6 | Difficulty in discovering/locating data |
| | Barrier 7 | Data cannot be combined and connected |
| | Barrier 8 | Difficulties to interact with the data provider |
| *Institutional Instruments* | Formal Instruments | Formal instruments, such as laws and regulations |
| | Informal Instruments | Informal instruments, such as norms and values |
| | Enforcing Instruments | Enforcing instruments, such as punishments and rewards |
| | Managerial Challenges | Challenges related to (for instance) motivating others or coaching, measures and managing performance, accountability, development and communication |
| | Reactions of the Public | Reactions from for instance direct colleagues, community users, open data enthusiasts. |
| | Change in reactions over time | Reactions from for instance direct colleagues, community users, open data enthusiasts. |
| *Other labels* | Other type of community value creation | Open answers by interviewees |
| | Community specific issue/ challenge | Open answers by interviewees |
| | Referring to other documentation | Open answers by interviewees |

**Deriving community-specific challenges and recommendations**

For answering the last sub question of the research (Which challenges and recommendations can be derived from the studied Open Government Data Communities?), just labelling the answers according to the table above (Table 10) was not enough. After executing the document analysis and in-depth interviews, best practices were derived systematically. This is a matrix based analytic method (Ritchie et al., 2013). The analysis strategy is built upon the Explanation Building method by Yin (2018), who identifies five techniques for case study analysis. The Explanation Building technique is specifically relevant because it aims "not to conclude a study but to develop ideas for further study" (Yin, 2018, p. 228). This is in line with the exploratory setup of this thesis.

## 3.5 Construct validity, internal and external validity and reliability

Explicit attention was paid to the quality of this research. Four important factors are: (1) construct validity, (2) internal validity, (3) external validity, and (4) reliability (Yin, 2018). The sections below discuss the four factors and the measures have been taken to ensure or increase compliance.

### 3.5.1 Construct validity

The first out of four factors important for the quality of the research is construct validity, and according to Yin (2018) this is especially challenging when conducting a case study. Construct validity regards to whether the measure of a construct sufficiently measures the concept and furthermore has been proven as a necessary component of the research process (O'Leary-Kelly & Vokurka, 1998).

According to Yin (2018), there are two steps to cover for construct validity and specific case study tactics that can be applied. The first step is to define specific concepts and the second step to identify operational measures. In this research, this was ensured by first defining open data community contributions by using open data literature and then operationalising by selecting open data benefits and barriers out of existing literature.

The operationalisation step consisted of reducing the long lists of open data benefits and barriers and due to the lack of literature this was done based upon the assessment of the author of this research. The assessment was validated by an open data expert, but still limits the validity of the construct since possibly not all benefits and barriers are included in the interview questions. Therefore, the interview participants are also asked if they can think of other open data benefits or barriers that are possibly affected by an open data community. Furthermore, multiple sources of evidence have been used (multiple interviewees and documents) to meet one of the tactics described.

### 3.5.2 Internal validity

According to Yin (2018), internal validity is specifically a concern for case studies with an explanatory character. Steckler and McLeroy (2008, p. 9) state that internal validity focusses on "whether or not observed covariation should be interpreted as a causal relationship". Reis and Judd (2000) address this question by explaining the third-variable problem. An unobserved or unmeasured factor may explain the variations in outcome, instead of the input that is tested in the research (Reis & Judd, 2000). Thus, in order to increase internal validity it is important to work systematically and third variables must be held constant.

According to Yin (2018), tactics to increase internal validity include explanation building and addressing rival explanations. Both of these tactics are included in the analysis. Another measure used to strengthen internal validity is to include the possibility to ask extra questions to an interview participant (Thiel, 2015). This can be done when the participants interprets a question wrong or the answer is not complete. The interviewer can ask clarifying questions in order to obtain the necessary information.

### 3.5.3 External validity

External validity concerns the generalisability of research. Steckler and McLeroy (2008, p. 9) phrase this as "whether causal relationships can be generalized to different measures, persons, settings, and times". Because this research is built upon a single-unit qualitative case study, it cannot be excluded that the selection of other cases would lead to different results.

According to Yin (2018), a case study tactic to increase external validity in single-unit case studies is to use theory. Institutional instruments will be used to better understand how interaction on an open data communities works. Furthermore, Yin (2018) indicates the external validity of the single-unit case study can be increased by combining the findings of this research with other case-study research.

### 3.5.4 Reliability

Reliability of research relates to the consistency of measures. According to Yin (2018), the objective of reliability is minimizing both the errors and the biases in research. When a researchers follows the described data collection and analysis procedures, under similar circumstances, the same findings and conclusions should arise.

To improve the reliability of this study, a case study protocol is used. This is also one of the tactics Yin (2018) described. Furthermore, multiple methods have been used in the analysis, earlier referred to as method triangulation.

# 4 Document analysis results

This chapter contains the results from the desk research examining information on the case, including information available on the platform and in other public governmental papers (e.g. letters to Parliament). Missing and additional information is then collected via interviews with open data community managers and end-users, both connected to the Dutch open data community, as well as experts within an open data related research field. The interviews results can be found in chapter 5. The results of the document analysis and the interviews are separated on purpose, since the two parts are of a very different nature and the analysed information barely overlapped.

## 4.1 National Open Data Agenda

In 2015 the Second Chamber of the Dutch Parliament was informed by the Ministry of the Interior and Kingdom Relations about the status of the implementation of the Dutch open data policy (Ministry of the Interior and Kingdom Relations, 2015), which is written down in the Cabinet's National Open Data Agenda (NODA). This agenda, which was executed in 2016, consisted of six items. These are divided into three clusters: The first cluster, inventory and disclosure of datasets, focuses on the numbers and prioritisation of datasets. The second cluster, monitoring progress and quality, contains measures to monitor the quality of the metadata and the progress of the disclosure.

The last cluster, support of accessibility, technology and users, aimed to offer support to the data managers. The actions in this cluster also included to channel the wishes and questions of open data users through a platform. In 2015, a start has already been made on organising user meetings focused on the functioning of data.overheid.nl (the Dutch national OGDP) under the working title 'data as a utility'. In 2016, three follow-up meetings were organised in which in particular technical and operational improvement opportunities were discussed with users of open data. In consultation with the data hubs, consultations with various user groups were be bundled and reinforced. The letter to parliament shows that already back in 2015 users were consulted on their experiences with the open data.

## 4.2 Start of the communities

In September of 2019, the open data community was announced on the Dutch Digital Government platform. The website stated “The communities are now actively supported and are, for example, offered groups of datasets. It will soon be possible to start up new communities” (Digitale Overheid, 2019).

Besides the National Open Data Agenda, the Netherlands also publishes the Open Government Action Plan. This Action Plan describes the ambitions of the government and other action holders. According to the website of the Open Government team, it should help government organisations, social organisations and citizens to work together towards a more open government (Ministry of the Interior and Kingdom Relations,2020).

The (Dutch) Open Government Action Plan 2020-2022 (Ministry of the Interior and Kingdom Relations, 2020), which is the fourth revision of the plan, contains seven focus areas that each feature one or several commitments. The focus areas are (1) Open Democracy, (2) Open Information, (3) Open Communication, (4) Open Organisation, (5) Open Procurement, (6) Open Technology, and (7) Open Justice. The open data communities commitment is one out of three commitments that are part of focus area 6 (Open Technology). The commitment consists of setting up the communities, sharing information on how data can be reused and the development of a framework to measure the impact of the data that is made available (Ministry of the Interior and Kingdom Relations, 2020).

This last point seems to be related to the datasets and portal itself, but is listed under the community commitment. The Action Plan states “Indicators are drawn up and these are made visible on the data portal. In this way, more insight is provided into the impact of government data” (Ministry of the Interior and Kingdom Relations, 2020, p. 25).

According to the Open Government Partnership (n.d.-a), the Netherlands is, amongst 76 other countries, a member of the Open Government Partnership (OGP). This is an alliance of countries from around the world aimed at promoting "open government. Participating countries commit to promoting transparency, increasing citizen participation and fighting corruption. They are using new technologies to make governments more open, effective and accountable, and to improving service delivery to citizens (Open Government Partnership, n.d.-a). Countries join the initiative to demonstrate their efforts to making their data free and easy to use, reuse and redistribute in accordance with principles of open data (Attard et al., 2015). The OGP does not benchmark countries.

According to the commitment template (Open Government Partnership, n.d.-b), filled in by the Ministry of Interior and Kingdom Relations, setting up communities will increase insight into the data impact, increase data availability, get supply and demand more attuned, contribute to the shareability of the data. It will “benefit transparency and the possibility of participation” and “stimulate technological innovation” (Ministry of the Interior and Kingdom Relations, 2020, p. 26).

The commitments are submitted to the Open Government Partnership and reviewed by the accountability division of the OGP, called the Independent Reporting Mechanism (IRM). The specific commitments on Open Technology (including open data communities) were also part of that review,

labelled as verifiable and with a modest potential impact. Furthermore the review states that “The commitments in this cluster could improve how the Netherlands uses critical technologies and data” (Ministry of the Interior and Kingdom Relations, 2020, p. 26).

## 4.3 Information on the community

Currently, there are four theme-based open data communities active in the Netherlands at the national level. The energy data is the most active community, with 47 members subscribed to the community homepage and the last post was one week ago (date checked for all community themes is 14th of June 2022). The education data community and mobility data community are of similar size with 17 and 15 active members respectively. The last new post in both communities was one month ago. The last community, regarding social security data has only 3 subscribers and the last post was one year ago.

Besides the theme-pages the community also features a general community board, with general announcements and updates for community-members. All communities have a uniform and simple set of rules, differing slightly by mentioning the data theme in the rules. The translation of the education community can be found in the box below (Figure 1).

*Figure 1: Open data community rules (education data)*

1. We keep it respectful. Everyone here works or is interested in educational data, let's make sure we improve the impact of this together.

2. Ask your questions and start the conversation. This is the place for your questions about education data. And if you are helped yourself, we appreciate it when you help others too.

3. Do you see a great initiative or application with educational data or have you seen something current? Share it with us here!

## 4.4 Conclusions from the document analysis

Based on the document analysis it can be concluded when and why the open data community came into being. The Dutch government has used the open data community as a tool in a wider range of initiatives, as a part of their open data strategy. A number of expectations about the added value of this community have been expressed, but it is unclear whether these were tested during or after the pilot phase.

It is clear that the pilot phase has ended, since it ran until the end of 2021. No explicit decision seems to have been taken regarding the continuation of the pilot, as this has not been found in policy documents. Interviews will shed light on how the pilot worked out in practice and what benefits the community managers and users experienced. The next chapter will focus on the analysis of the interviews.

# 5 Interview results

This section will elaborate on the individual interview results and it will discuss different viewpoints of the participants. The synthesis in the last paragraph of this chapter will then argue how the interview sections and its conclusions relate to each other.

The answers of all of the interviewees regarding both the contribution of the open data community to the open data benefits (5.1) and barriers (5.2) and the role of institutional instruments in creating value with the open data community (5.4) have been labelled and categorised. The table underneath (Table 11) shows the answer given and the corresponding colour.

*Table 11: Legend of interview analysis tables for open data benefits, barriers and institutional instruments*

| Colour/ mark | Interview answers benefits/barriers | Interview answers institutional instruments |
|---|---|---|
| | The open data community currently contributes to this benefit/ barrier (i.e. it is visible on platform) | The institutional instrument currently contributes to the open data community (i.e. it is visible on platform) |
| | The open data community potentially contributes to this benefit/barrier, but it requires a large(r) user group | The institutional instrument potentially contributes to the open data community, but it requires a large(r) user group |
| * | The open data community potentially contributes to this benefit/ barrier, but it requires more or something else than a large(r) user group | The institutional instrument potentially contributes to the open data community, but it requires more or something else than a large(r) user group |
| | The open data community contributes to this benefit/ barrier to a (very) limited extent | The institutional instrument contributes to the open data community to a (very) limited extent |
| | The open data community does not contribute to this benefit/ barrier at all | The institutional instrument does not contribute to the open data community at all |
| | The interviewee did not answer the question | The interviewee did not answer the question |

Because the two open data researchers (interviewee 8 and 9) are not active on the Dutch open data community, their answers are never marked as dark green.

## 5.1 Contribution of the open data community to open data benefits

All interviewees were asked to respond to eight statements regarding possible open data benefits, identified in the theoretical framework. The participants responded to each statement by indicating to which extent the open data community contributes or could contribute to the open data benefit.

### 5.1.1 Increased social control

The first benefit that the interview participants were asked to comment on was the open data benefit of increased social control. All the interview participants answered that there is either a very limited contribution (answered three times) or no contribution at all (answered six times) of the open data community to this benefit.

Participants from both the community management and the user group (I1 and I4) elaborated that the communities are used mainly by professionals and that the participation of other parties is very limited. Furthermore, the purpose of collecting this data was argued as part of research (I1) or for making plans as a government (I3) and not to be used for social control.

Interviewee 5 stated that although the contribution to more social control is less relevant, the community does contribute to simplifying communication about data. Participant 7 clarifies that the data itself can be used to control government efforts (in this case in the field of water management), but that the contribution of the open data community is limited.

The researchers also both think that communities cannot contribution to social control, or that the contribution is very limited. Interviewee 9 does add to that that other factors might as well be relevant. The freedom of press for instance, which is very strong in the Netherlands, may be beneficial to the way communities can possibly contribute to social control.

### 5.1.2 Increased civic participation and public engagement

The second benefit of open data was the increase in civic participation and public engagement. The community managers and one of the community users see a potential contribution for the community to this open data benefit. All other participants, including 5 users and both researchers think the contribution of communities to the open data benefit is very limited.

Interviewee 1 states that once you have everyone on the community platform, you can start with interaction between the different groups, including the data owners, data experts, students and scientists. Interviewee 6 also doesn't see this contribution right now, but can imagine that the contribution will become visible if there are enough civilians asking questions on the platform, and they are answered. A similar answer was provided by interviewee 5, however here the potential contribution is considered to stay limited.

Interviewees 3 and 7 both argue that there is only a very limited contribution. According to interviewee 3, the amount of people that is actually working with the data itself it limited and interviewee 7 sees much more effects on proving information than actual participation. The same applies to one of the researchers (I8), who also thinks a community can be of a bigger contribution to informing rather than facilitate participation.

Interviewee 4 explains that there are occasional discussions within project meetings of the open data team. These discussions are not about social control, but do include how certain things can be explained in the right way.

The second open data expert (I9) adds that via a community is it not possible to reach out to digitally unskilled civilians, therefore making the community not a suitable platform for participation.

### 5.1.3 More informed citizens

The third open data benefit regards more informed citizens. The interviewees are unanimously positive about the potential contribution of an open data community to this benefit of open data. However, most of the answers do include requirements in order to unleash the potential.

Interviewee 1 remarks that only citizens that want to be informed, will actually be informed by the community. Interviewee 2 adds that interested civilians have access to high amounts of data. This participant also states that the information may be too heavy or too complicated for the average citizen, and the information is more beneficial to professionals or people from the municipality.

Interviewee 3 adds the requirement of good data quality, interviewee 4 of civilians actually reading information that is already provided with the datasets (such as metadata).

Both interviewee 3, 7, 8 and 9 conclude that the communities indirectly contribute to more informed citizens. For instance by consultancies or municipalities that are looking for data, which they turn it into information products (I3), carnival, swimming and surf associations that are using the community to gather information (I7). One of the researchers (I9) refers to open data intermediaries, such as journalists, that can transform this complex data into more digestible information to their audience Also the other researcher (I8) mentioned the intermediary role, giving the example of mobility apps.

### 5.1.4 Increased accountability

Participants were asked to which extent the open data community contributed or could contribute to increased accountability. The answers provided are varying to a great extent, however the sentiment in most answers is fairly negative.

Participants 2 and 5 think this contribution is potentially there. Either since its creates an easier way to ask questions about data (I5) or because of the fact that it is possible to ask questions via the community at all (I2). Participant 2 adds that the wording sounds heavy, but true nonetheless, since people are approachable about the data.

Multiple participants (I1, I4, I8) argue the contribution is limited since not all data is being published currently and therefore the government can be accounted on the data they decide to

publish. Participant 4 elaborates that the community can give cause to check the wording of certain aspects or ask why things are missing.

One participant (I3) thinks that the contribution is limited, since the type of published data is not management information. The published data is for monitoring, research and statistics

Some participants also state that there are other mechanisms for accountability, and that using a community to enhance governmental accountability is not (most) suitable (I3, I6, I9).

Last of all, participant 7 thinks that there is a limited contribution, inherent to the domain. However, data that is not published for accountability can in some cases still be used to, for example biological and chemical data that is being shared.

#### 5.1.5 Gained access to external capacity and resources for solving problems

All participants responded positive to the contribution of an open data community for the open data benefit of gained access to external capacity and resources for solving problems. Only one of the platform users sees a potential contribution, all the other participants (researchers excluded) already see the contribution of this benefit.

For instance participant 1, this participant is describing a list of researchers within the energy community that is already large and still growing. By using that list, new data requests can be redirected to the right person. Participant 2 adds that the communities brainpower is enormous, because everyone knows something and you can get in touch with specialists.

The communities users agree with that viewpoint. Interviewee 3 has already spoken to new people, which are part of the so-called 'playing field' and interviewee 4 has colleagues who have asked questions in the community themselves and sees how consultancies are also using the open data. Interviewee 7 sees how a surfing association has made their own data viewer, by using open data. Other interested parties have asked the question how they could get access to the data as well.

Interviewee 5 made the sidenote that the contribution is related to having a forum function on a community, and the contribution would therefore be attributable to that function rather than the community.

#### 5.1.6 Increased intra-governmental collaboration

The community managers and researchers were also very positive on the contribution of an open data community to an increase in intra-governmental collaboration. Most of the users were positive as well, but some others were more reticent.

Two of the participants (I4, I6) indicate that communities do not lead to increased collaboration but that a good collaboration leads to forming communities. The community can be seen

as a means of shaping that collaboration (I4). Interviewee 6 argues that “The intensive collaboration that leads at some point to results on that community. It will not very quickly lead to new collaboration that would not otherwise occur”.

Interviewee 2 explains that the collaboration of different organisations in VIVET and the parties active actually carry the community around energy. Interviewee 3 thinks that the communities should lead to more collaboration, but doesn’t see it happening in practice yet. The participant furthermore states that intra-governmental collaboration is very difficult, especially when concerning different layers within de government or includes decentralised governments.

Interviewee 7 uses the example of their relation with a water authority. This authority uses the API of the participants organisation. However, this collaboration did not originate from data.overheid.nl. Both researchers see potential contributions of the community for this open data benefit.

#### 5.1.7 Increased problem-solving capacity

The benefit increased problem-solving capacity scored about the same as factor 5, gained access to external capacity and resources for solving problems. All participants gave similar answers.

Participant 2 argues that increasing the problem-solving capacity is the whole purpose of the community. To provide it to people that are searching for something or want to spar with someone. Interviewee 4 agrees, also stating that this was the goal of the community. Participant 5 states that the community makes getting in contact more accessible, adding that accessibility is a key word when describing the added value of the community.

Participant 6 is a bit more reluctant, but does also see a lot of potential and uses the metaphor of ‘teaching people to fish instead of giving a fish all the time’. However, to achieve that, there has to be a critical mass first according to this participant.

Exactly that metaphor can be used for an actual example that participant 7 introduced. The interviewee stated that if the community works together, they can get around practical limitations to get to the goal they have. The community that started on GitHub wrote a Python script, to get more data than is available in the data portal. Since there only the last five years of data is stored, someone wo wants to receive data for ten years can’t use the portal. With the script from the community, it can still be done.

Both researchers see the potential contribution of the community. Researcher 9 indicates that there is a difference between coming up with solutions and deciding which solution to implement. The latter is a policy process, something that should not be part of the community. For data or technical problems (rather than societal) communities can also add value.

### 5.1.8 Use of collective intelligence to solve public problems

The last benefit of open data introduced was the use of collective intelligence to solve public problems. Both community managers and researchers see a (potential) contribution of the data community. Users are divided into a group of three users which do see a contribution and two who see a very limited contribution. Interestingly, both of the latter group are other users than the users who were hesitant at benefit 6.

Both community managers see a large contribution of the data community. Interviewee 1 sees that the providers and demanders come together and everyone thinks with each other. As a result, there are always some people who know more about something. Interviewee 2 agrees and thinks that the core task of a community is also to solve problems. Getting that brainpower out of that community, that's what the Community does.

Multiple users also see the contribution. Interviewee 3 does, if coupled with domain expertise. Because 'very often you have data expertise, people who know a lot about quality of data and metadata, but are not necessarily able to interpret the outcomes. For that you need domain expertise, so also content-wise'. Interviewee 4 also sees the contribution, but indicates that there are different speeds. The questions need an answer directly, but the team members within a project need more time to be able to answer

Interviewee 6 sees the potential, but does not see the contribution on the community yet. This participant argues that before you have collective problem-solving capability, you have to have a collective.

Both interviewee 5 and 7 indicate that they think the contribution is limited. Interviewee 5 doesn't expect anyone to post a problem in a community and ask the group as a whole to come to a solution. Interviewee 7 has similar thoughts, arguing that the group has as much knowledge as community members have individually. Members help each other but reinforce each other much less.

Both researchers argue that there is a large potential for this contribution. However, one researcher (I9) adds if an identified problem is about certain marginalised groups which is not active on the open data community, the solution does not necessarily address the root causes of the problem.

### 5.1.9 Conclusions regarding open data benefits

The interviews showed that there are large differences in the extent to which the individual open data benefits are enhanced by the open data community. Three of the eight discussed benefits are considered significantly by the interviewees. These three open data benefits are: more informed citizens, gained access to external capacity and resources for solving problems and increased problem-

solving capacity. The first of those, more informed citizens, was considered not present currently on the community. All interviewees did however see potential for the contribution to become visible and gave additional comments or requirements. Those included that the contribution could indirectly lead to more informed citizens, for instance via intermediaries, that the data should be of good quality and that the information has a too high threshold for the average citizen.

The other two contributions, were currently visible according to some of the interviewees and contained much less comments/requirements. Both researchers also judged that there is potential for these two contributions. These contributions are increased intra-governmental collaboration and the use of collective intelligence. Some of the community users were more sceptical about the potential. Per statement two community users, being different users for both benefits, judged the (potential) contribution would be limited.

For the increased intra-governmental collaboration, two community users stated that good collaboration was needed to lead to use of the community by the organisations, not the other way around. Two other users argued that the contribution of the community to the use of collective intelligence was limited, either because the interviewee did not expect anyone to post a problem in a community to ask the group as a whole to come up with a solution or because the community at large has as much knowledge as community members have individually.

There are three benefits left, where the majority of respondents indicate that the contribution of communities to the benefit is limited or non-existent. The first being the increase in social control, where all community managers, users and researchers indicate a very limited or absent contribution. Although some interviewees were more positive by indicating the community creates both a new and easier way to ask questions about data, one of those participants also said the word accountability was too heavy. Researchers judged the potential contribution would be limited.

Finally, most users and both researchers indicate that the increase in civic participation and public engagement because of the community would be very limited. The community is seen as a way to inform people rather than to facilitate participation, the amount of people active is considered too low and their background not diverse enough. The community managers however did see potential for this contribution.

The table on the next page (Table 12) shows how the interviewees scored the different benefits. It is noteworthy that there is much consistency in the assessment of the interviewees, indicating a high amount of agreement regarding the contribution of the community on enhancing the various open data benefits.

*Table 12: Open data benefits enhanced by an open data community, according to interview participants*

| *Viewpoint* | Interview | Benefit 1 | Benefit 2 | Benefit 3 | Benefit 4 | Benefit 5 | Benefit 6 | Benefit 7 | Benefit 8 |
|---|---|---|---|---|---|---|---|---|---|
| *Community managers* | Interviewee 1 | | | * | | | | | |
| | Interviewee 2 | | | * | * | | | | |
| *Community users* | Interviewee 3 | | | * | | | | | |
| | Interviewee 4 | | | * | | | | | |
| | Interviewee 5 | | | * | | | | | |
| | Interviewee 6 | | * | * | | * | | | |
| | Interviewee 7 | | | * | | | | | |
| *Researchers* | Interviewee 8 | | | * | | | | | |
| | Interviewee 9 | | | * | | | | | * |

## 5.2 Contribution of the open data community to mitigating open data barriers

All interviewees were also asked to respond to eight statements regarding possible open data barriers, identified in the theoretical framework as well. The participants responded to each statement by indicating to which extent the open data community contributes or could contribute to (partially) mitigating the open data barrier.

### 5.2.1 The organisation is not interested in using open data

All interviewees responded positive to the first statement, meaning that the open data community can (potentially and/or partially) mitigate the open data barrier of the organisation not being interested in using open data.

Interviewee 2 argues that users are directing others to the community and people have to realise that there is much more data available on that subject. The community can make those people a little bit more knowledgeable, because it is very accessible. One of the community users (I4) agrees by saying that in practice it really works this way. This user saw how someone was making attractive visualisations with someone else's data, encouraging the data-owner to make their own products even better. In that sense, some mutual competition is not bad.

Three other community users (I3, I5 and I6) argue that contribution will be present when more people will become active on the community. Interviewee 3 often has to explain to people what the data portal (and community) is.

Interviewee 6 adds that the data-uploading organisations do not know who is using their data, they are just publishing it. Once more people start asking questions in the communities, then the realisation that people are actually using the data will follow.

Interviewee 7 is convinced that the answer differs depending on which person you're asking. A lot of the work concerning open data has been made compulsory by the European Union and not everyone is motivated intrinsically.

The researchers both indicate that there is a contribution. One researcher (I8) compares communities to the pre-event phase of hackathons where participants of hackathons engage with governmental organisations. Out of this interaction, the governmental organisations become more interested. After a hackathon the artifacts that the hackathons has produced generate interest as well. The other researcher (I9) indicates that having a community would potentially increase the amount of ideas on how different sorts or types of data can be used.

### 5.2.2 The organisation does not have the organisational capabilities, routines, and processes to use open data

The second statement, concerning organisational capabilities, routines and processes to use open data, was answered much more hesitant by the interview participants. Although the community managers and one of the interviewees answered positive (I1, I2 and I4), all other users think there is a very limited contribution (I7) or even no contribution at all (I3, I5 and I6).

Both community managers think that this contribution is present, interviewee 1 is explaining this by illustrating the bi-directional effect between data.overheid.nl (the data portal) and the open data communities, leading users from the community on the portal or the other way around. The second community manager thinks that once you know that you can find everything you want on the community, people will start to use it as a knowledge database. The community can help to not get lost in the immense amount of data and act as a signpost. Interviewee 4 agrees and indicated that more tailor-made information is being provided after questions on the community. This also has downsides, since this can be a one-time gesture only. However, if the same questions get asked again, the customisations to the data can be standardised, thus increasing the routines and processes to also use open data.

Interviewees 3, 5 and 6 all argue that the capabilities, routines and processes of their organisation are not part of the scope of the communities. Interviewee 3 substantiates this by indicating that there are other networks and communities are not present on a lot of subjects, while interviewee 5 argues that the skills, routines and processes are already required into the phase before actually publishing the data.

Interviewee 7 thinks the contribution is limited, because although they are using the externally made python-scripts of the community and are referring to them, this solution was not made by them and regards external capabilities.

Interviewee 8 did not answer this statement, since the researcher did never work for a governmental organisation. The second researcher (I9) was also hesitant and claimed this depends on the amount of civil servants that are active in a community. People who are not part of a governmental organisation wouldn't know the procedures and resource constraints in the government.

### 5.2.3 Low engagement of public managers with the use of open data

The third open data barrier that was incorporated into a statement was regarding the low engagement of public managers with the use of open data. Most interviewees agree that this is potentially contributing to mitigation (6 times) or the contribution is already visible (2 times). Interviewee 4 does not see an contribution.

One of the community managers (I1) indicates that the community helps in showing the enthusiasm, by showing the type of questions being asked in the community and also specifying if the questions concern a specific subject or portfolio. Interviewee 2 goes one step further and suggests if a manager could become active as a specialist or ambassador on specific topics. This does not require frequent actions, but can help in convincing other managers about the community or let them get in touch.

Interviewees 3 and 6 also see a potential contribution. Both are indicating that in practice the contribution is not visible at all, calling the contribution an 'ideal scenario' (I3) or indication that to achieve this contribution a really big community is necessary (I6).

Two participants already see a contribution of the data community. Interviewee 5 states that "managers do not need to have specific knowledge about a dataset, but they just see a nice example, for instance a Success Story". Interviewee 7 mentions that in certain periods water information is used much more frequently and notices that their services are used all over the world. The community helps in showing where the data is used for.

Both researchers also thinks that an open data community could contribute to mitigating this barrier, however Interviewee 9 adds that this depends on the priority of the individual department. When open data is prioritised managers would be interested in the community.

#### 5.2.4 There are no human resources with the knowledge, skills and/or capabilities to use open data

Most interviewees (six in total) see a potential contribution for the fourth barrier, the absence of human resources with the knowledge, skills and/or capabilities to use the open data. One of the interviewees already sees this contribution in the community (I6), one did not answer this statement (I1) and one interviewee thinks the contribution is limited (I7).

One of the community managers (I2) argues that this contribution will became visible after people start finding their way on the community. Furthermore, other organisations should actually also know about the existence of communities. The manager suggests "you should actually almost start advertising for that. A kind of marketing campaign, I think".

Most community users (I3, I4, I5) also indicate there is a potential contribution. Interviewee 3 mentions how many organisations are just publishing their data demand-driven or because they have to, because of a regulation for instance and thinks that the question whether the user also benefit from it is asked not often enough. Interviewee 4 explains that this is already happening behind the screens, so not on the community itself yet and interviewee 5 adds the requirement of the community questions to be strongly content-related.

Interviewee 6 already sees in practice how the contribution is being achieved. This participant has answered thousands of questions over the last decade, which he or she couldn't be answering without first delving into the subject matter. That created a certain expertise.

The researchers both argue that there could be a significant contribution. One researchers (I8) thinks that especially the learning part is important, because that is why you enter a community and the other researcher (I9) sees that the contribution is only possible if the community can reach out to more people who are interested to develop their digital skills.

#### 5.2.5 Data provider ignores my requests and suggestions

Every interview participant answered positively on the mitigation of this barrier by the data community. In total six participants (including both managers and four users) already see how the community is contributing to mitigating this barrier.

Interviewee 1 mentions how this is currently being rolled out in the Energy Community. Data requests can be submitted via the data portal website, and those request are directly shared on the community forum as well. Interviewee 5 mentions how the community manager contacting the submitters of those requests by mail and does add that communicating via mail does feel more personal than in the community. Interviewee 4, who is active within the energy community, also confirmed the claim of the community manager (I1). However, this interviewee added that there is a difference in pace. The question submitters need quick answers, while the data provider often needs time to fulfil the request.

Both interviewee 6 and 7 also see how the contribution is made by the community. Interviewee 6 is mentioning how the community can be an eye-opener for some data-owners, who simply never see or hear from an end-user. They start to see that their data is used in many different ways and that they therefore need to offer the data in different ways as well.

Interviewee 3 mentions how this factor is not only constrained by whether organisations want to fulfil requests and suggestions, but are often limited by their legal mandate, "Users want a lot and we also want a lot, but we are not allowed because of a ministerial regulation or because of a policy rule. So in addition to being open to, it is also about being allowed to. Actually it is about wanting, being able and being allowed".

The researchers both indicate that this contribution can be present, because when you have a community, it's easier to go to the data provider (I8) and members of the community can then tell the government what sort of data is needed and in which format (I9).

### 5.2.6 Difficulty in discovering/locating data

Most participants are also positive about the sixth barrier a community can mitigate, the difficulty in discovering or location data. Three interviewees currently see this contribution, five others see a potential contribution (of which three have requirements) and one of the data community users (I5) does not see a contribution.

Interviewee 5 argues that discovering or locating the data is part of the data portal itself, not of the community. One of the researchers (I9), who does see a contribution of the community, argues that the community can be used by users to give feedback on the data portal itself and its filters for instance, which would also contribute to maximizing open data use. Furthermore the researcher argues that 'the community can function like a decision three. Based on your questions other can help in navigating through the data portals and thinking of which titles of datasets might be relevant'.

The community managers (I1 and I2) both see how this contributes and interviewee 2 elaborates by stating that you can use to community to help, when you are looking for a particular topic and ask for specific tips. Interviewees 3 also sees this contribution and interviewee 4 and 6 see how it could potentially contribute. According to one of them (I3), this is also why the community was founded, people could discover the datasets and also have a good grasp of what belongs together.

Interviewee 4 personally responds when people are asking questions regarding the location of datasets, but only responds if someone has also taken the trouble to search well for information themselves. Interviewee 7 does not see this on the Dutch national open data community, but does see it happening in other communities, for instance the GitHub community.

### 5.2.7 Data cannot be combined and connected

Interview participants were also positive regarding the mitigation of the open data barrier 'data cannot be combined and connected'. One of the community users (I7) argued that the contribution is limited, but all other interviewees indicate that the open data community (potentially) mitigates this barrier.

Interviewee 7 states that when combining datasets is part of a research trajectory, the community is of limited added value because the datasets are already known. Interviewee 7 argues that the community currently focusses on unlocking the data. Interviewee 8 complemented that by arguing that for datathons, this phase is often not reached, "Especially the operation of combining different datasets is technically intense to do. You won't do that in a hackathon usually, because it takes too much time". The researcher added that there is more time when looking at communities, indicating that there may be more potential.

Interviewee 5 refers to the previous barrier regarding locating data, stating that there is more potential for combining datasets. However, the interviewee still considers it secondary to the portal itself, and explains: “You're going to search on the portal first, and if you've got that sorted out there, then such a community would actually be superfluous.”

Other interview participants, including the community managers and other users, are more positive of the contribution. Interviewee 1 argues that the community can draw relationships at data sets and interviewee 3 answered the question from a wider perspective, by saying that often people can locate a dataset but do not know how to use or interpret it. Furthermore this interviewee notices that question articulation is incredibly important. Sometimes community questions can be answered very strictly, but when understanding the underlying goal better help could be provided.

Interviewee 6 distinguished two levels, first of all by individual questions of community users and secondly by data-owners getting the same questions several times and therefore being able to connect different datasets themselves.

The second researcher (I9) indicates that the contribution is quite high because of domain expertise. If someone wants to use data from a different domain (for instance a digital expert wanting to use agricultural data), the data community can help out. Therefore, all the different expertise’s must be present on the community.

#### 5.2.8 Difficulties to interact with the data provider

According to all interviewees, the open data community could contribute or already contributes to mitigating the open data barrier: difficulty to interact with the data provider. One of the community users (I3) does not currently see the contribution on the data community, but all the other users and the community managers do.

Interviewee 1 highlights that the community makes interaction easier and more accessible. Interviewee 5 also states the increase in accessibility to describe how a community affects interaction, adding that preferably it is possible to allow for direct contact or to tag people: “If you write to the data holder directly then it helps because then it's very accessible”.

Community users (I6, I7) and one of the researchers (I8) think that this is the main reason to actually have an open data community. Interviewee 7 adds that besides interaction it is also a way to collect user requirements, “It is the most important way of finding out what an end-user actually wants and what the biggest stumbling blocks are”.

Interviewee 3 does not currently see this contribution, but would like to see it in the future. This participant also indicates that there might be depending of the type of data product. Application-

oriented products are possibly more suitable for interaction with the data provider then more factual data tables.

### 5.2.9 Conclusions regarding open data barriers

When comparing the answers on the statements regarding open data benefits to those regarding the barriers, it is worth noting that the interviewees are more positive about the removal of barriers around open data by the data community. The interviewees are unanimously positive of the mitigation of three barriers, all but one interviewee responded positive to four others. There is one last barrier on which the interviewees are very divided.

All interviewees indicate there is a potential mitigation of barriers 1, 5 and 8. Multiple community users say that the first of those, the organisation is not interested in using open data, is not visible on the community yet but potentially could become. Most users think the community needs to gain more active users first.

For the other two barriers, the data provider ignores my requests and suggestions and difficulties to interact with the data provider, four out of five community users indicated that the contribution is already visible.

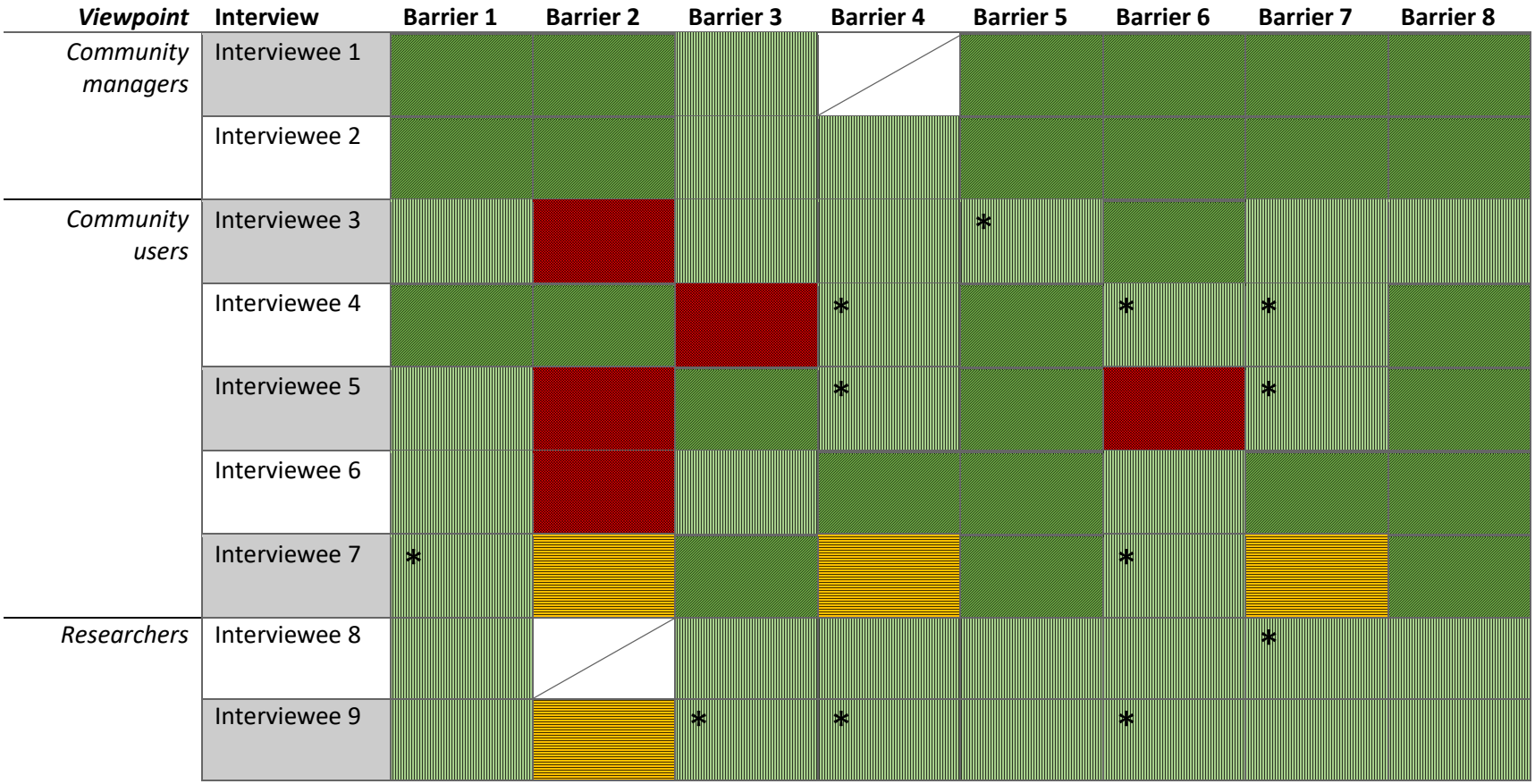
One of the participants, a community user, does not think the community contributes to mitigating the barrier low engagement of public managers with the use of open data (barrier 3), while most other participants see a potential contribution. The barrier difficulty in discovering/locating data is answered similarly (barrier 6), in this case also one of the users didn't see any contribution and explained that this is entirely within the scope of the data portal itself.

For the other two barriers, the lack of human resources with the skills and/or capabilities to use the open data (barrier 4) and the data cannot be combined and connected (barrier 7), the same user answered that there would only be a limited contribution twice. According to this participant, the mitigation of the first barrier is limited because the organisation in which the participant works does not use the community to increase skills or capabilities and the latter barrier is not mitigated because when researchers want to combine datasets for their research, this is already done in the preliminary phase of the research. Both researchers conclude that the community can potentially contribute to the mitigation of all four barriers.

The interviewees are in disagreement whether the second barrier, the organisation does not have the capabilities, routines, and processes to use the open data, can be mitigated by the open data community. Most of the community users think the contribution is either very limited (one interviewee) or that there is no contribution at all (three interviewees). Interviewees explained that other networks are more applicable to achieve the mitigation of this barrier and that the skills, routines and processes are already required in the phase before publishing the data. One of the researchers was not able to answer this question and the other researcher indicated that the contribution would be limited.

The table on the next page (Table 13) shows how the interviewees scored the different barriers. As with the assessment of open data benefits, the interviewees here also largely agree on the contribution of the community.

*Table 13: Open data barriers mitigated by an open data community, according to interview participants*



## 5.3 Other ways open data communities stimulate open data value creation

All interviewees were asked if there were other benefits that the open data community contributed to and/or barriers that the open data community mitigated. Since most of the answers could be formulated as both, it was chosen to formulate them as contributions to open data value creation (thus benefits).

### 5.3.1 Making interaction more accessible and the data provider better approachable

Although closely related to eight barrier of open data (difficulties to interact with the data provider), interviewees seem to make a difference between the actual difficulty and the perceived difficulty to get in contact with the data provider.

Interviewee 7 mentioned how open data users can experience a threshold in the first contact with their service desk. “You notice that when people know how to find us, they describe a problem much more concisely. They already know that they are going to get good help, because there are smart people behind the service desk. So the community removes a threshold. That threshold is actually not there, but it is experienced as such”.

Interviewee 5 gives another example to describe the same phenomena: “If you find the right e-mail address in the metadata, then you can start emailing. But by using the community it feels structured and you have a broader channel than when you send an email. You can also reach out to other users”.

Although both community users above describe the increased accessibility of a community when contacting a data-owner, one of the researchers (I8) argues that a community can possibly also create a new threshold, because communities are very specific places where professionals meet online. “A social kind of paradigm is that entering the community can be a bit intimidating. For example, you can also contribute to Wikipedia but there is a threshold”.

### 5.3.2 Providing new, broad insights

According to one of the community managers (I2), the community can lead to new thoughts and insights, also enhancing the opportunity for new themes to get active on the community. These insights are wider than insights regarding the open data itself, such as insights regarding the locating or combining of datasets, but rather on new use cases for open data or improved formulation of a question.

This possibly also relates so something that Interviewee 9 described. This participant described how the community could help solve problems with the different stakeholders, rather than not including them in even defining the problems they face.

### 5.3.3 Reducing "data anxiety”

In almost all interviews the current audience of the community was described as mostly professionals, including civil servants from different governmental organisations, but also businesses, entrepreneurs and other institutions.

According to interviewee 3 the community could also decrease the so-called ‘data-anxiety’. People who are not data savvy can't find everything easily. That is why the interviewee started to develop a kind of signpost. In order to achieve this contribution (reducing data-anxiety), the interviewee came up with several requirements. More people should find out about the community and join it, and the community platform itself should be more accessible and user-friendly. Interviewee 4 came up with a similar thought, stating that the community can get people into action faster and added that data itself is really not ’sexy’.

### 5.3.4 Reducing individual disclosures of data and offer them grouped/thematically

Several participants mentioned the fact that their primary use of the community was to cluster datasets. This happens for instance in the Energy Community where VIVET, a Dutch Partnership for Improving Information for the Energy Transition, bundles all sorts of data and information relevant for the energy transition.

According to participants that are working for VIVET, it could soon become a one-stop shop for many governments, a central place for discussion of data (I6) and also interview participant 4 firmly believes in thematizing. Interviewee 7, which is not involved with VIVET also sees how a community can help with being able to access the data properly and easily.

### 5.3.5 Increasing the amount of published data

One of the researchers (I9) argued how an open data community can help with identifying ‘high value data sets’. Data users sometimes need other data than the government releases and distributors don't realise that there is demand for this sort of data. The community could function as “a collective voice, demanding a certain dataset or type of data”. The data providers are, without a community, not aware of that demand. Interviewee 5 gave a similar statement, saying “it can be a big stick for people to start publishing data more, by raising user goals”, referring to the community as a whole.

### 5.3.6 Increasing understandability and readability of the metadata and manuals

Both interviewee 4 and interviewee 6 discuss how data users often lack to read the manuals and/or the metadata connected to a dataset. Interviewee 4 argues that, instead of looking in the manual themselves, data users often sit back and just ask a question. Interviewee 6 also sees this but adds

that metadata is hard to read and also hard to find. The open data community could help data users to actually find and read the metadata and manuals and furthermore also provide feedback to the author of the metadata or manuals when they are for instance too complicated.

#### 5.3.7 Providing insight in data users and uses

Two interviewees (I3, I6) stated that organisations do not know who is using their data, they are just publishing it. Once more people start asking questions in the communities, then the realisation that people are actually using the data will follow (I6). According to Interviewee 6, therefore the data owners will have "to offer, package or explain the data in different ways".

Interviewee 3 came to the same conclusion, saying that users maybe want a "slightly different breakdown or a different format", and envisioning a place "where supply and demand come together though. Where the uploader gets to hear from the downloader, I can use this data I can use for this purpose, but it would be even more useful if that, that and that could also be done".

#### 5.3.8 Increasing open data importance awareness

According to interviewee 9, the community could also contribute to open data value creation by raising awareness about the importance of open data. The interviewee stated that "The open data community can reach out to people who are outside this community to then raise awareness about why data is important".

#### 5.3.9 Exposing latent demand for data

Last of all, Interviewee 9 mentioned how the community could show the latent demand for data. Open data users do not know what (type of) data they need or will need in the future. The participant calls this a latent demand for data, indicating that the user itself does not know yet what type of data he or she needs, but does know that there is a problem which may be solved by using open data.

## 5.4 Contribution of Institutional instruments to the open data community

The next paragraphs will discuss the three levels of institutional instruments. Interview participants have been asked to answer statements about to which extent each of these levels contributes to the community creating open data value.

### 5.4.1 Formal instruments

According to almost all of the interviewees formal instruments, such as laws, rules and regulations, have a very limited effect or no effect on the way an open data community can contribute to open data value creation.

Only one interviewee (I6) stated that formal rules had an effect. This interviewee saw the effect on two different levels. First of all, if the amount of users on a forum increases some rules are needed to keep the interaction somewhat decent. Secondly, if questions and answers are given according to a certain form, it makes it easier to find the questions already asked. However the interviewee also said that this may raise a threshold to ask a question, so a balance should be found on that matter.

Both community managers (I1, I2) and two of the platform users (I3 and I5) did not think that formal rules would be of any added value. Interviewe3 3 did however think of regulations on the data itself, that do limit open data value creation by the means of a community. For instance laws regarding privacy or microdata limit what can be shared publicly.

Other community users (I4, I7) and both researchers (I8 and I9) thought that formal rules would have a limited impact. Interviewee 7 gave several arguments why formal rules are not important. First of all, the community users are 'like-minded', since they have the same goal. Often it are professionals as well, or at least people with a reputation to hold. Lastly, the community is fully public.

The hackathon researcher (I8) elaborated that at a hackathon everything happens without rules and mentioned a paper called "There's no rules, it's a hackathon". However, structures in general can help. For instance, Hackathons often have a very clear timeline. The other open data researcher (I9) thinks the contribution is quite limited, because it's hard to formalise rules in an open data community.

### 5.4.2 Informal instruments

When being asked about the influence of informal instruments, participants answered notably more positive. All interviewees are convinced that informal rules on an open data community are contributing to a limited extent (I5 and I7) or to a significant extent (I1, I2, I4, I6, I8 and I9).

One of the platform users that answers questions on the community (I4) sees that there are people who prefer to just be served, without too much effort. This participant does have difficulty with it. However the participant also says that a kind of duty of care exists: if a question is asked, an answer must also be given.

Interviewee 6 argues that because the community currently consists of professionals things go quite naturally, but it the platforms shifts more towards civic participation and informing then moderators should become active. Interviewee 5 agrees on the fact that currently mostly professionals are active on the community, stating that “the organisations on such a community are all fairly old, like municipalities, for example”. The participants elaborates on this because those organisations “already have their own integrity code and rules of conduct”, and those are not formed by a community.

Interviewees 1, 2, 8 and 9 also think that informal rules do contribute. Interviewee 1 gives the example of not starting unnecessary discussions, and that you respect each other. “That's also what often goes wrong on Facebook, but that's not the intention there either”, according to the community manager.

Both researchers also think that informal rules can contribute to the creation of value. One researcher (I8) draws parallels with hackathons, for instance on the procedures for team formation, but also the culture that is present in a hackathon. The second researcher argues that we need to think about open data communities as how we look at other communities. “These develop very organically. Growth is not something that you can you can really force. But of course you have certain nudges that a government uses to nourish this community”.

#### 5.4.3 Enforcing instruments

Interviewees also reacted more positive to whether the use of enforcing instruments on data communities could help in creating open data value (compared to formal instruments). Three of the platform users (I3, I4, I6) think there is a limited contribution, although two of them also acknowledge that other users are possibly more sensitive to these mechanisms than they are themselves.

One of the community managers (I2) elaborated upon the active enforcing instruments, currently levels are being used on the platform. If you post a lot of things then you get a higher level, which is an incentive to contribute. The platform also allows for the use of trophies, but those are currently not activated. The other community manager added that “if someone comes to the forum with bad intentions or says crazy things, they get a warning and if it really isn't acceptable, they are simply blocked or removed”.

Both Interviewee 4 and 6 personally do not do not attach much value to levels. Interviewee 4 thinks that perhaps companies would like a high level behind their name and interviewee 6 is even more convinced that others do want a badge, for example when they answer many questions. This interviewee suggests that community users could be rated: “Or that, for example, they answered 1,000 questions to their satisfaction and received an average rating of 8.9”.

Interview participants 5 and 7 both think that rewards could encourage users. Interviewee 5 thinks focussing too much on sanctions could create a negative atmosphere and would also warn people in advance for sanctions. Interviewee 7 adds that a badge would also be useful to show whether someone is answering as an employee from an organisation or on a personal title. For instance when a certain type of data use is not desirable and someone indicates that on the community, the interviewee comments: “That is viewed very differently when someone posts that from our organisation”.

Interviewee 3, as well as one of the researchers (I9) thinks that intrinsic motivation is very important and that certain rewards on that level are more applicable. Interviewee 3 states “The reward is that you get much more insight. Someone has to be convinced that it has added value”. The researcher suggests that community members might be motivated through mutual interests, for instanced a shared passion for climate change or infrastructure in their own neighbourhood.

The other researcher sees both networking opportunities as financial incentives in hackathons. The same actors that are usually present at a hackathon (are institutional actors, participating companies and other stakeholders) are also active on open data communities.

### 5.4.4 Managerial challenges and reactions from the public

The interview participants were also asked about managerial challenges and reactions from the public. Most did not come up with any managerial challenges but instead mentioned more general challenges of communities. Even the community managers, who stated that operational challenges were more relevant, indicated keeping the community active as one of the main challenges. These challenges are mentioned in the next section (5.5), were community-specific challenges that were mentioned in one of more interviews are summarised.

One of the researchers (I9) however did think about possible managerial challenges. The researcher suggested that possibly bureaucracy within the government or the public sector in general is limiting the amount of impact of the open data community and open data in general. The researcher added that this depends on the commitment of the government and individual departments that are more enthusiastic about open data.

### 5.4.5 Conclusions regarding institutional instruments

When combining the answers of the interviewees, it is clear that institutional instruments have an important role in the functioning of the community. However, not all levels of instruments have an equal level of importance.

Formal instruments, including rules and regulations, are of limited added value for open data communities. Both the community managers and users and researchers agree. Only one of the interviewees argues that these rules do add value, for instance when standardizing question formats. Other interviewees state that due to the audience of the community formal rules are less relevant than on other communities, such as Facebook. The researchers also agree and both indicate that the contribution of formal rules to the community creating value is limited.

Interview participants saw informal and enforcing instruments as more important, although respectively two or three users did remain hesitant and stated that the contribution would be limited. Rules that are enclosed in for example a code of conduct, are often present in the organisations the community users work for. Community users argue that those rules will be kept in mind when using the open data community. Applying those rules to the community can be seen as a norm.

Enforcing rules, such as badges, trophies and levels, are also relevant. Several users think they are of other value and although other interviewees indicate they personally do not care that much about this type of incentives, they also think that other users possibly do. However, the interviewees agree that in the end the motivation to use the community should be intrinsic, and not from for instance financial incentives.

Table 14 shows how the individual interviewees scored both the formal, informal and enforcing instruments.

*Table 14: Institutional instruments relevant for open data community contribution to open data value creation, according to interview participants*

| ***Viewpoint*** | **Interview** | **Formal** | **Informal** | **Enforcing** |
|---|---|---|---|---|
| *Community managers* | Interviewee 1 | | | |
| | Interviewee 2 | | | |
| *Community users* | Interviewee 3 | | | |
| | Interviewee 4 | | | |
| | Interviewee 5 | | | |
| | Interviewee 6 | | | |
| | Interviewee 7 | | | |

| | | | | |
|---|---|---|---|---|
| *Researchers* | Interviewee 8 | | | |
| | Interviewee 9 | | | |

## 5.5 Community-specific challenges and recommendations

This paragraph addresses other challenges, issues or comments regarding the open data community made by the interviewees. It lists considerations that are neither part of the possible open data benefits/barriers influenced by the communities nor the institutional instruments, all elaborated upon in previous paragraphs.

### 5.5.1 Critical mass

All interviewees of both the community management and user group (I1 to I7) did at some point in the interview reflect upon contributions that were currently not present in the open data community but did have a lot of potential. Most of the time clear requirements were mentioned that needed to be fulfilled in order for the contribution to become present and most of the time this requirement was reaching a 'critical mass'.

Interviewee 3 for instance said that many people don't know about the initiative yet and interviewee 5 stated that the platform has to grow, elaborating: "… it will only have added value if there are a lot of users on it, asking a lot of questions".

### 5.5.2 Visibility, advertising and marketing

In order to achieve the so-called 'critical mass', or a significant active user group, some of the participants suggested to start with a marketing campaign or to advertise for the community.

One of the community managers (I2) stated that "The open data community and the open datasets must be made known. When I came to work here, I didn't really know what I could find and where … Other organisations should also know this, you should almost advertise it. A kind of marketing campaign, I think".

Interviewee 3 agrees to the latter idea, saying that advertisement is needed in order to increase the amount of users on the community and make the initiative more widely known. interviewee 5 directly mentioned visibility when asked for recommendations, saying "Yes, especially in terms of visibility. Ensuring that this community is presented in more places and that you are also stimulated more".

### 5.5.3 Danger of becoming an inactive community

Both community managers highlight the danger of the community becoming dormant. One of the managers (I2) finds this very challenging for communities and its users: “To get them active and especially to keep them active. It is often not in people's system, you have to keep prodding them”.

The other community manager adds that they are working on encouraging people to remain active in the community. Besides the community user this also applies to the ambassadors. The managers are giving form to this currently.

### 5.5.4 Definition and interpretation of what a community is

Two of the interview participants explicitly made clear that when they were referring to the community, this was not the online forum. Interviewee 3 elaborated that they felt a need to organise and group datasets, so that the relevant datasets could be easily found and added that “At the same time, it was said that a lot of questions are coming in that are not so much about a data request, but about where you can find data or a specific request for help. This is what we then called the forum”.

Both interviewee 3 and 4 do not think that interaction should be the main goal of the open data community. Interviewee 4 wants the community to achieve that disclosures of datasets stop existing and all data is disclosed based on its theme. Interviewee 3 states that “The focus is very much on the forum section though. That is not our primary goal, which is to make the datasets accessible and to organise them properly”. Other users do interpret the main focus of the community as the forum, which also includes some more statical information that is made available via this forum.

### 5.5.5 Difference between data portal and data community

In addition to the different ideas about the focus of the community itself, interview participants also confuse the open data portal (data.overheid.nl) with the open data community (datacommunities.nl), or see the two services as a totality.

When for instance Interviewee 7 is being asked about the contribution of the community to more informed citizens, the answer entails the different elements in the dataset and service that may be of use to a citizen. Answers from other interviewees also sporadically hint that the interview participants either view the portal and community as a whole or lack to distinguish them.

### 5.5.6 Importance of using KPIs

One of the data community users (I3) indicated that a relevant property of the open data community should be the outcome indicators (clear and measurable targets). The question that goes along with this property is when the community can be called successful.

The interviewee mentioned that in the early stages of the community, Key Performance Indicators (KPI's) were formulated, the most important one being the increase in the use of data and adds "And of course that's also a very high goal, because I don't know how you can link those one to one. But for us, that's the most important thing, an increase in use of data".

### 5.5.7 Platform user-friendliness and usability

When asked about the user-friendliness of the open data community platform software, users respond very differently. Interviewees that see the disclosure of data on the platform as the main goal of the community (I3 and I4), are not satisfied with the platform software. Interviewee 4 states "For having a discussion and asking a question totally fine. For unlocking data that is not directly data it works less well".

Other interviewees are more happy with the platform software. One of the community users (I5) thinks the platform is very user friendly. This user also posted screenshots on the community and did not have any issues: "That was like you would make a 'post' on any platform, like also on Facebook or on LinkedIn. I found it a recognisable interface and therefore easy to use, it explains itself". In addition, the user was pleasantly surprised with the gamification element which is incorporated into the software.

### 5.5.8 Interpretation of government role

Both the community managers and one of the researchers (I9) shared thoughts regarding the role of the government in the open data.

The community managers noticed how community users assumed that because the government initiated the open data community they would also keep it running. Interviewee 1 elaborated "Of course we offer the platform, but sometimes people think about this too easy. For example, that we should also fill the platform; there is still much uncertainty about this".

The open data intermediary researcher (I9) adds a different perspective by indicating that not only the government has valuable data that could be offered on the data portal. Other community users (such as businesses) have lots of valuable data too. The government could act as an intermediary.

### 5.5.9 Target audience

Many interviewees made a comment on the role of citizens on the community. both the community managers, users and researchers did. Interviewee 2 (community manager) thought that "… it is too heavy or too complicated for the average citizen".

Interviewees 3 and 5, both community user, also made remarks about citizens. The first stated that “I don't think that many citizens come to data government and also not that citizens are on the forum” and the latter “I think it is mainly intended for professionals. But suppose citizens should also have a place, I can imagine that”. One of the researchers (I9) stated that, when asked if citizens could get more informed by the open data community, citizens could become more informed indirectly by journalists active on the community.

#### 5.5.10 Communities based on open data themes

Two of the community users have comparable ideas about the future of the open data community. One of them (I4) is in favour of thematizing and I am afraid that data.overheid.nl will fall victim to its own success, because it also wants to cover all fields. It would then become too big. This user believes in thematizing.

Interviewee 5 is a member of several communities in the GEO-domain (the GEO-forum of PDOK and the knowledge base of the municipal GEO-meeting, part of PLEIO) and considers communities as a good source of information. However, this user would rather go to one central place.

Although their statements may sound conflicting at first (one user was sceptical of creating a central place and the other very fond of the idea), it looks like both would like to have one central place to find information regarding their own domain, not necessarily for all different open data domains.

#### 5.5.11 Obliged to or motivated to

Both a community user (I7) and researcher (I9) reflect upon the obligations regarding open data publishing. Interviewee 7 sees that a lot of the open data efforts are attributable to the obligations from the European Union. Although this participant gets very excited when data is actually used for completely different applications, colleagues are less enthusiastic.

The researcher (I9) thinks that when a department within the government doesn’t see the value of open data, but is obliged (by for instance a government regulation) to publish certain datasets, they probably aren't interested in contributing to the open data community. In other words, the researcher suggests that impact will be limited if open data efforts are because of obligations instead of by intrinsic motivation.

#### 5.5.12 Comedy of the Commons

At the end of interview 8 the Tragedy of the commons was discussed, which relates to depletable resources like fisheries: common-pool resources (Ostrom, 2008). Interviewee 9 indicated that data is not depletable. He mentioned another theory, called Comedy of the Commons: Individuals

contributing either their knowledge or their content for the good of the community, rather than extracting resources for their own personal gain (Rose, 1986).

The researcher elaborated upon potential positive effect when the private sector would also contribute more to the open data community, by not only becoming active on the forum but also releasing open data. Both the example and the theory are very relevant when philosophising about the future of the open data platform, its community and the role of the national government.

## 5.6 Conclusions from the interview results

The analysis of the interview has promising results. This section will briefly summarize the findings of the analysis.

Five out of the eight benefits were identified as potentially influenced by data communities. For three of those, all interviewees were positive: more informed citizens, gained access to external capacity and resources for solving problems and increased problem-solving capacity. The other two contributions, were currently visible according to most of the interviewees and contained much less comments/requirements. Those benefits are increased intra-governmental collaboration and the use of collective intelligence.

For the barriers, all interviewees indicate there is a potential mitigation of barriers 1, 5 and 8 (the organisation is not interested in using open data, the data provider ignores my requests and suggestions and difficulties to interact with the data provider). Most interviewees indicated that four of the remaining five barriers are also potentially mitigated by an open data community, being low engagement of public managers with the use of open data (barrier 3), difficulty in discovering/locating data (barrier 6), the lack of human resources with the skills and/or capabilities to use the open data (barrier 4) and the data cannot be combined and connected (barrier 7).

Interviewees also identified other ways data communities stimulate value creation of open data. According to the interview participants, the communities can make interaction more accessible and the data provider better approachable, they can providing new, broad insights, they can reduce "data anxiety" and reduce individual disclosures of data and offer them grouped/thematically. Furthermore, the communities can increase the amount of published data, increase the awareness of open data and increase the understandability and readability of the metadata and manuals. The communities can also provide insight in data users and uses and expose the latent demand for data.

Interview participants were also asked if institutional instruments contribute to the functioning of the community. The interviewees argue that both informal and enforcing instruments can be of

importance, although a few of the interviewees (respectively two or three) were sceptical about this. Formal rules were of less importance, mainly because of the professional background of the data community users and their organisations existing rules and culture.

Last of all, the participants were asked about community-specific challenges and recommendations. Some of the interviewees also mentioned challenges themselves, even before the question was asked. There is a critical mass that should be reached, before the community can contribute to most open data benefits or mitigation of barriers. The community currently lacks visibility, advertising and marketing are needed to get people to know the community and recognise its potential. The community is also in a constant danger of becoming inactive.

The interviewees were using different definitions and interpretations of what a community is. Some referred to the forum as a part of the community, while others see the community as the forum. Not all users were able to distinguish the difference between data portal and data community.

In order for the community to become successful, one of the users stressed the importance of using KPIs. The users also had different interpretations of the government role on the community and on how all-encompassing the community should become, in terms of both the variety of themes and audience. The interviews did not agree upon the platform user-friendliness and usability, some of the users thought the community was very easy to use while other were less positive.

Last of all, one of the experts introduced the theory Comedy of the Commons, where individuals contribute their knowledge for the good of the community, rather than extracting resources for their own personal gain (Rose, 1986).

# 6 Discussion

This chapter will reflect upon the results from the analysis first. After that both the used literature, part of the theoretical framework, and the methodology will be discussed. Both have their limitations and therefore it is important to be careful when interpreting the research results.

## 6.1 Results

The case study analysis primarily focused on the effects of open data communities on open data benefits and barriers. In many of the answers of the interviewees ifs and buts can be found. When interviewees stated that the open data community could contribute to a benefit or to the mitigation of the barrier, they often stressed this would be a potential contribution. The effect was not yet present due to various reasons, such as the active user group being too small or restrictions due to either organisational policy or even laws.

Even when excluding the answers of the researchers (they always spoke about potential contributions since they are not active on the community themselves), about half of all the effects caused by the community are indicated as potential effects. This means that the contribution is not yet visible on the community. It raises the question whether communities in general can actually live up to the full potential that is described by the interviewees.

Interviewees made several remarks regarding the way the community functioned and some of those remarks were contradictory with those of other interviewees. For the community to become successful, some interviewees believed there should be separate communities based upon open data themes while others indicated that the platform should incorporate all open data themes in order to become a central place for open data interaction.

Other contradictories include the different judgements about the platforms' user-friendliness and usability and the interpretation of what a community actually is. Interviewees who stated that the core function of the community was to group and bundle datasets, were not positive on the user-friendliness whilst interviewees who believed the forum functionalities were the core community were more positive on the user-friendliness.

The document analysis described how the community fits in the Open Government Action Plan (n.d.) and is described as a pilot. Some of the contradictories that are mentioned in the paragraph above could be caused by the broad scope of the pilot. Both for this specific case study, but also for other open data communities, some of the community-specific challenges and recommendations can also be used as a list of things to consider when designing an open data community. Examples of focus

areas are (1) to clearly define what the open data community is and does, (2) to clearly demarcate the target audience of the community, (3) to use KPIs to define and measure the successfulness of the community and (4) to define and document the role and responsibilities of the platform administrator.

Even without any ifs and buts, the open data community possibly also has negative effects. Although the analysis of this research focuses on the contribution to open data benefits and mitigation of open data barriers, the negative effects should also be taken into consideration. The analysis made clear that there was an ongoing battle to prevent the community from becoming inactive (paragraph 5.5.3), which could possibly lead to inverted effects. One of the interviewees mentioned that users will face negative experiences getting in touch with the data-owner, when a community is present but questions are not answered. There are undoubtedly more negative effects of inactive open data communities, think about running costs, but there even may be negative effects related to active open data communities. Future research should investigate this.

## 6.2 Literature & theories

In order to study the effects of open data communities on open data benefits and barriers, a long list of both open data benefits and open data barriers were identified in the theoretical framework of this research. The full list of both the open data benefits and of the open data barriers in this work is from Kawashita et al. (2022). A first limitation is that other scientific literature may have identified further benefits and barriers. In addition, not all of the open data benefits and barriers identified by Kawashita et al. (2022) were included in the case study. Since scientific literature within the field of open data communities is scarce, the list of benefits and barriers was compiled on the basis of the personal assessment of the author of this research.

In order to increase the construct validity, the list of possibly relevant benefits was subsequently checked by an open data researcher. Furthermore, interviewees were consistently asked for other benefits of the open data community.

## 6.3 Methods

In this research, a single-unit qualitative case study was used. A limited amount of documents were analysed for the case study, five in total. All the analysed documents were connected to the Dutch case. For the interviews both community managers, users and scientific researchers were interviewed. The total number of interviewees is limited to nine. There are currently also other open data communities active, but they differ significantly. Table 7 compared both the Open Data Portal and the open data community of the Netherlands with the efforts of the United Kingdom and the United

States. The table, that can be found in Chapter 3, shows that there are differences between the different existing Data Portals and communities. Both the Open Data Portal and the community show differences. While the Netherlands and the United States have one location for all data communities, the United Kingdom has communities around themes on different locations. The technical system used to implement the community differs and the openness of the community does as well. This limits generalisation of the results in the previous chapter.

For the various barriers, benefits and institutional instruments, the degree of agreement varies. For most of the factors, the majority of the participants agreed with each other. For a few propositions, there was a great deal of difference between the individual answers. It is difficult to generalise these results, but they may offer clues that can be further investigated in follow-up research (Yin, 2018).

# 7 Conclusions

This chapter first answers the sub research questions. Then, the answers of the sub research questions are combined to answer the main research question. Furthermore, this chapter includes the societal and scientific contribution of the research. Last of all, recommendations for future research are presented.

## 7.1 Answering the research questions

Governmental organisations have various objectives when developing their Open Data Initiatives, such as increasing transparency, data usability, participation, and economic growth (Alexopoulos et al., 2013; Attard et al., 2016; Kassen, 2013; Zeleti et al., 2016). Open data communities could potentially help in achieving those objectives. However, the effects of setting up open data communities are unclear. There is very little knowledge about the benefits of these communities because of a lack of scientific literature. To gain insight into the potential contribution of an open data community to open data value creation, this research aimed to exploratively analyse the added value of an open data community. This section will answer the four sub questions that were introduced in the first chapter and subsequently answer the main research question.

*Sub question 1: What are the benefits and barriers of open data that can potentially be influenced by open data communities?*

Scientific literature on the advantages and disadvantages of implementing an open data community on an Open Government Data Portal is absent. Even a suitable definition for an open data community was not found. Features of successful data communities were found in literature, although not specifically for open data communities. The features of successful open data communities can be divided into three categories (Cooper & Springer, 2019, p. 12): “bottom-up development, absence or mitigation of technical barriers to sharing, and community norms”. Open data communities are currently often implemented on Open Government Data Portals (OGDPs), or are directly accessible via those portals.

Based on existing work, a systematic literature review on open data benefits, barriers, drivers and enablers performed by Kawashita et al. (2022), eight open data benefits and eight open data barriers are identified as potentially affected by open data communities. Due to the lack of literature on open data communities, the assessment was done by the author of this thesis and validated by an open data expert.

The potentially affected open data benefits are: (1) increased social control, (2) increased civic participation and public engagement, (3) more informed citizens, (4) increased accountability, (5) gained access to external capacity and resources for solving problems, (6) increased intra-governmental collaboration, (7) increased problem-solving capacity, and (8) use of collective intelligence to solve public problems.

The potentially affected open data barriers are: (1) the organisation is not interested in using open data, (2) the organisation does not have the organisational capabilities, routines, and processes to use open data, (3) low engagement of public managers with the use of open data, (4) there are no human resources with the knowledge, skills and/or capabilities to use open data, (5) the data provider ignores my requests and suggestions, (6) difficulty in discovering/locating data, (7) data cannot be combined and connected, and (8) difficulties to interact with the data provider.

*Sub question 2: Which open data benefits and barriers are influenced by current national Open Government Data communities?*

The case study answered the second research question. According to the interview participants, open data communities (can) contribute to multiple benefits of open data, including (indirectly) creating more informed citizens (benefit 3), gained access to external capacity and resources (benefit 5) and increased problem-solving capacity (benefit 7). Furthermore, the participants agree to a large extent that communities contribute to intragovernmental collaboration (benefit 6) and the use of collective intelligence to solve public problems (benefit 8).

The open data communities can also (potentially) mitigate open data barriers. All participants agree that the community potentially resolves the governmental organisations lack of interest in using open data (barrier 2), the data provider ignoring requests and suggestions (barrier 5) and it also mitigates difficulties in the interaction with the data provider (barrier 8). Furthermore, both researchers and most community users and managers think that the community can also contribute to mitigating low engagement of public managers with open data (barrier 3), the lack of human resources with the knowledge, skills and capabilities to use the open data (barrier 4), difficulty in discovering/locating data (barrier 6) and not being able to combine and connect datasets (barrier 7).

Besides those contributions, the interviewees also suggested other benefits of the open data communities, which include resolving so-called ‘data anxiety’, making interaction more accessible, reducing individual disclosures of data and increasing the amount of published data.

*Sub question 3: Which institutional instruments contribute to the value creation of current national Open Government Data communities?*

Institutional instruments are important in the data communities, according to the interview participants. Although the participants argue that the contribution of formal instruments (such as rules) is limited, informal rules and enforcing instruments do contribute to the value that is created by an open data community.

The interviewees claim that because most users in the data community are professionals, they already work in accordance with their own organisation's code of conduct. However, interviewees do think that the norms adopted by the open data community help to keep the interaction on the community content-driven and that rewards may incentivise individual user activity.

*Sub question 4: Which challenges and recommendations can be derived from the studied Open Government Data Communities?*

The challenges and recommendations were systematically derived from the interviews. According to the interview participants, community specific challenges include reaching a critical mass of active users, preventing the community from becoming inactive, being able to understand the difference between the data portal and data community, the interpretation of the government role and corresponding target audience and the difference between organisations being obliged to or motived to contribute to the open data platform.

Numerous recommendations were also made by the interview participants, including increasing the visibility of the community by advertising and marketing, using KPIs to measure the success of the data community, increasing the user-friendliness and usability of the community forum and creating more theme-based data communities.

The answers to the sub questions allow the following main research question to be answered.

*Main research question: "What are the potential effects of open data communities on open data benefits and barriers?"*

Open data communities have the potential to enhance open data benefits and also mitigate barriers to open data. The interview participants stated that the communities do this in various ways, for instance by informing citizens via intermediaries, increasing interaction between data users and data owners, and strengthening collaboration between government agencies. Furthermore, the participants indicated that communities can generate more interest in open data, within the organisations of the data owners, also specifically for public managers. Lastly they can increase skills and knowledge, generate increased problem-solving capacity, and help with both locating and

connecting datasets. The case study also showed that there are potentially even other benefits, including mitigating 'data anxiety' and making interaction more accessible.

Informal and enforcing institutional instruments can play an important role in the contribution of the open data communities. Norms help to keep the interaction on the community content-driven and rewards may incentivise individual user activity.

Although the potential effects data communities on open data benefits and barriers seem promising, the question is whether the communities can actually achieve their potential. The analysis described numerous challenges, including reaching a critical user mass, and the reflection indicated that the open data community could also have negative effects.

## 7.2 Societal and scientific contribution of the research

Open data has enormous potential, from improved efficiency and transparency of public administration, economic growth for businesses, to improving social welfare (Data Europe EU, n.d.-b; Alexopoulos et al., 2013; Zeleti et al., 2016; The World Bank, n.d.). This research indicates that setting up and maintaining/ supporting data communities can lead to positive effects, by both enhancing open data benefits and mitigating barriers. Open data communities can increase interaction between the data owner and data user, and lower the threshold to contact the open data owner and thus have enormous potential in the broader open data strategy of national governments.

Policy makers can use the results of the analysis to determine their objectives when setting up new open data communities, since the results indicate the specific open data benefits that are enhanced, as well as the barriers that are mitigated. Furthermore, the community-specific challenges identified in the analysis can already contribute to the process of designing an open data community, because policy makers can use the recommendations and community-specific challenges that have emerged from this study to refine and sharpen their design choices. For instance, choices regarding marketing, the scope of the target audience and also technical choices. In this way, the experiences of current Dutch open data community managers and users collected in this research can contribute to improve both existing and new open data communities.

This thesis can be seen as a first exploratory scientific contribution to the examination of open data communities. A lot of scientific literature is available within the field of open data, for instance regarding open data benefits (Janssen et al., 2012), as well as regarding open data barriers (Barry & Bannister, 2014). However, little is known about the effects of open data communities on the benefits and barriers of open data. Some scientific literature can be found on data communities, but these data

communities often either focus on closed data or are not openly accessible. Cooper and Springer (2019, p. 12) described three characteristics of successful data communities within the scientific field on a high level: “bottom-up development, absence or mitigation of technical barriers to sharing, and community norms”. However, these communities are specifically meant for researchers.

This research provides a first insight into how successful data communities can enhance open data benefits and mitigate of open data barriers. Future research can use the identified benefits and barriers to validate the exploratory results.

**The CoSEM-perspective**

The Master's degree programme in Complex Systems Engineering and Management (CoSEM) aims to explore innovations in complex socio-technical environments (TU Delft, n.d.). Technical innovations also have other (non-technical) important aspects, such as (existing) regulations, subsidies, as well as interests of other stakeholders, a cultural aspect and human behaviour. Dwyer (2011) states socio-technical systems “consist of components, which are social structures, and artifacts, which are technical elements that contribute directly or through other components to a common system goal”. A data community on an Open Government Data Portal is the perfect example of a socio-technical system, since it features both sides of the socio-technical environment:

- The human interaction of the different user groups, including the community managers, community users (which can be professionals, citizens, researchers). The community norms are described by Cooper and Springer (2019) as a feature of successful data communities. The community users also have their own personal set of social norms and rules, mentioned by the interviewees when discussing the formal, informal and enforcing institutional instruments. There norms can be based upon their company culture, or their own values.
- The technical artifact, open data. The open datasets are made available on the OGDP, allowing to easily use the datasets. The system goal can be described as maximizing the benefits of open data, and governments are trying to achieve this by increasing the benefits and value generated through their open data initiatives. Cooper and Springer (2019) mentioned the absence of technical barriers as one of the characteristics of successful data communities.

In order for the open data community to become successful, socio-technical aspects have to be considered and implemented into its technical design. Several aspects were mentioned by the interviewees, including existing regulations and culture. First of all the existing regulations, such as the General Data Protection Regulation (GDRP). One of the interviewees explained how this regulation prevents the uploading organisations (the data owners) from meeting user requests.

The second socio-technical aspect that was mentioned is the workplace culture, for instance the culture of the data-uploading organisations. Community users stated that due to the strong culture at their organisation, the influence of formal rules would be very limited. On the other hand the users were more positive about rewards on the community, to nudge individual community user behaviour.

## 7.3 Recommendations for future research

The previous paragraph listed several limitations of this research. More research is required to validate the results presented. This section provides a brief overview of the next steps that can be taken to further explore the contribution of open data communities to a country's open data strategy, by connecting recommendations for future research to each of the limitations.

- The first limitation is the use of a single case study. The total list of open data community contributions should therefore be validated by a multiple case study, including different types of data communities. Table 7 (part of section 3.3.1) shows that the Open Data Portals and open data communities have differences.
- The second limitation is the limited amount of analysed documents and spoken interviewees. In a future study a larger group of interviewees, from different national open data communities, should participate.
- The current research design was fully qualitative. Although interviewees have already made a difference in the contribution of the open data community to a benefit or barrier, quantitative research (for example with a survey) can reveal to which extent a community can contribute.
- Furthermore, this research did not study the negative effects of open data communities. The research was mainly focussed on how the open data communities could enhance open data benefits and mitigate open data barriers. Although some users indirectly mentioned negative effects, this was not part of the interview protocol. Future research should focus on the potential negative effects of open data communities, as well as how they work and can be avoided.
- Lastly, a shortlist of benefits and barriers of open data was used. Future research should validate the shortlist of open data benefits and barriers that a community could potentially contribute to, by including more open data experts in the pre-interview phase.

# Bibliography


Ahmad, S., Wasim, S., Irfan, S., Gogoi, S., Srivastava, A., & Farheen, Z. (2019). Qualitative v/s. Quantitative research- a summarized review. *Journal of Evidence Based Medicine and Healthcare*, *6*(43), 2828–2832. https://doi.org/10.18410/jebmh/2019/587

Albano, C. S., & da Silva Craveiro, G. (2015). Lessons Learned From The Use Of Budgetary Data Open Format: An Exploratory Study In The Brazilian Ecosystem. *Revista de Gestao e Projetos*, *6*(3), 17-27.

Albano, C. S., de Araujo, M. H., & Reinhard, N. (2017). Motivators and facilitators of relationships in networks: how public managers acknowledge these factors in open government data. *NAVUS-REVISTA DE GESTAO E TECNOLOGIA*, *7*(1), 73-92.

Albano, C. S., & Reinhard, N. (2014). Open government data: Facilitating and motivating factors for coping with potential barriers in the brazilian context. In M. Janssen, H. J. Scholl, M. A. Wimmer, & F. Bannister (Red.), *Electronic Government, 8653*, 181–193. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-44426-9_15

Albano, C. S., & Reinhard, N. (2015). Challenges for governments and society in the Brazilian Ecosystem of Open Government Data/Desafios para governos e sociedade no ecossistema brasileiro de dados governamentais abertos (DGA)/Desafíos para gobiernos y sociedad en el Ecosistema Brasileño de Datos Gubernamentales Abiertos. *Cadernos de Gestao Publica e Cidadania*, *20*(67), 214-235.

Alexopoulos, C., Spiliotopoulou, L., & Charalabidis, Y. (2013). Open data movement in Greece: A case study on open government data sources. *Proceedings of the 17th Panhellenic Conference on Informatics - PCI '13*, 279-286. https://doi.org/10.1145/2491845.2491876

Attard, J., Orlandi, F., Auer, S. (2016). Data Driven Governments: Creating Value Through Open Government Data. *Computer Science*, *9860*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-53416-8_6

Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly*, *32*(4), 399–418. https://doi.org/10.1016/j.giq.2015.07.006

Azungah, T. (2018). Qualitative research: Deductive and inductive approaches to data analysis. *Qualitative Research Journal*, *18*(4), 383–400. https://doi.org/10.1108/QRJ-D-18-00035

Barry, E., & Bannister, F. (2014). Barriers to open data release: A view from the top. *Information Polity*, *19*(1,2), 129–152. https://doi.org/10.3233/IP-140327

Bengtsson, M. (2016). How to plan and perform a qualitative study using content analysis. *NursingPlus Open*, *2*, 8–14. https://doi.org/10.1016/j.npls.2016.01.001

Berrone, P., Ricart, J. E., & Carrasco, C. (2016). The Open Kimono: Toward a General Framework for Open Data Initiatives in Cities. *California Management Review, 59(1)*, 39–70. https://doi.org/10.1177/0008125616683703

Birchall, C. (2016). Shareveillance: Subjectivity between open and closed data. *Big Data & Society*, *3*(2). https://doi.org/10.1177/2053951716663965

Bleijenbergh, I. (2015). *Kwalitatief onderzoek in organisaties* (Tweede druk). Boom Uitgevers.

Bonina, C., & Eaton, B. (2020). Cultivating open government data platform ecosystems through governance: Lessons from Buenos Aires, Mexico City and Montevideo. *Government Information Quarterly*, *37*(3), 101479. https://doi.org/10.1016/j.giq.2020.101479

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, *9*(2), 27–40. https://doi.org/10.3316/QRJ0902027

Brandusescu, A., Iglesias, C., Robinson, K., Alonso, J. M., Fagan, C., Jellema, A., & Mann, D. (2016). *Open Data Barometer: Global Report* (4th ed.). The World Wide Web Foundation. http://opendatabarometer.org/doc/4thEdition/ODB-4thEdition-GlobalReport.pdf

Brown, R. B. (2006). *Doing your dissertation in business and management: The reality of researching and writing*. SAGE

Bruce, P., & Shelley, R. (2010). Assessing Stakeholder Engagement. *Communication Journal of New Zealand*, *11*(2), 30-48.

Cahlikova, T., & Mabillard, V. (2020). Open data and transparency: Opportunities and challenges in the swiss context. *Public Performance & Management Review*, *43*(3), 662–686. https://doi.org/10.1080/15309576.2019.1657914

Chorley, K. M. (2017). The challenges presented to records management by open government data in the public sector in England: A case study. *Records Management Journal*, *27*(2), 149–158. https://doi.org/10.1108/RMJ-09-2016-0034

Chu, P. Y., & Tseng, H. L. (2016). A theoretical framework for evaluating government open data platform. *Proceedings of the International Conference on Electronic Governance and Open Society: Challenges in Eurasia*, 135-142. https://doi.org/10.1145/3014087.3014099

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed). Routledge.

KOOP (n.d.). *Communities.* Retrieved August 24, 2022, from: https://data.overheid.nl/datacommunities

Cooper, D., & Springer, R. (2019). *Data Communities: A new model for supporting STEM data sharing.* Ithaka S&R. https://doi.org/10.18665/sr.311396

Cranefield, J., Robertson, O., & Oliver, G. (2014). *Value in the mash: exploring the benefits, barriers and enablers of open data apps*. https://doi.org/10.26686/wgtn.12910091.v1

Crusoe, J., Simonofski, A., Clarinval, A., & Gebka, E. (2019). The impact of impediments on open government data use: Insights from users. *2019 13th International Conference on Research Challenges in Information Science (RCIS)*, 1–12. https://doi.org/10.1109/RCIS.2019.8877055

Data Europe EU (n.d.-a). The benefits and value of open data. In data.europe.eu. Retrieved August 29, 2022, from https://data.europa.eu/en/datastories/benefits-and-value-open-data

Data Europe EU (n.d.-b). What is open Data?. In data.europe.eu. Retrieved June 11, 2022, from https://data.europa.eu/en/trening/what-open-data

Davies, T. G., & Bawa, Z. A. (2012). The promises and perils of open government data (OGD). *The Journal of Community Informatics*, *8*(2), 1-8. https://doi.org/10.15353/joci.v8i2.3035

Dawes, S. S., Vidiasova, L., & Parkhimovich, O. (2016). Planning and designing open government data programs: An ecosystem approach. *Government Information Quarterly*, *33*(1), 15–27. https://doi.org/10.1016/j.giq.2016.01.003

Denzin, N. K. (1970). *Sociological methods: A sourcebook*. New Brunswick, Transaction Publishers.

Digitale Overheid (2019). *Nieuwe versie van data.overheid.nl gelanceerd.* https://www.digitaleoverheid.nl/nieuws/nieuwe-versie-van-data-overheid-nl-gelanceerd

Drever, E. (1995). *Using semi-structured interviews in small-scale research: A teacher's guide.* The Scottish Council for Research in Education.

Dwyer, C. (2011). Socio-technical Systems Theory and Environmental Sustainability. Proceedings of SIGGreen Workshop. *Sprouts: Working Papers on Information Systems, 11*(3). http://sprouts.aisnet.org/11-3

European Commission (n.d.). *Open data portals*. Retrieved March 11, 2022, from: https://digital-strategy.ec.europa.eu/en/policies/open-data-portals

European Commission (2005). *E-Commission: enabling efficiency and transparency.* Retrieved February 8, 2022, from: https://ec.europa.eu/commission/presscorner/detail/en/IP_05_1474

Forbes (2020). *The case for data communities.* Retrieved February 19, 2022, from: https://www.forbes.com/sites/tableau/2020/11/20/the-case-for-data-communities/?sh=182f9cee1402

Gascó-Hernández, M., Martin, E. G., Reggi, L., Pyo, S., & Luna-Reyes, L. F. (2018). Promoting the use of open government data: Cases of training and engagement. *Government Information Quarterly*, *35*(2), 233–242. https://doi.org/10.1016/j.giq.2018.01.003

Gebre, E. H., & Morales, E. (2020). How "accessible" is open data?: Analysis of context-related information and users' comments in open datasets. *Information and Learning Sciences*, *121*(1/2), 19–36. https://doi.org/10.1108/ILS-08-2019-0086

Gonzalez-Zapata, F., & Heeks, R. (2015). The multiple meanings of open government data: Understanding different stakeholders and their perspectives. *Government Information Quarterly*, *32*(4), 441–452. https://doi.org/10.1016/j.giq.2015.09.001

Jamieson, D., Wilson, R., & Martin, M. (2019). The (im)possibilities of open data?. *Public Money & Management*, *39*(5), 364-368. https://doi.org/10.1080/09540962.2019.1611240

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

Kassen, M. (2013). A promising phenomenon of open data: A case study of the Chicago open data project. *Government Information Quarterly*, *30*(4), 508–513. https://doi.org/10.1016/j.giq.2013.05.012

Kawashita, I., Baptista, A. A., & Soares, D. (2022, January). Open Government Data Use by the Public Sector-an Overview of its Benefits, Barriers, Drivers, and Enablers. *HICSS*, 2533-2542. http://hdl.handle.net/10125/79648

Khandkar, S. H. (2009). Open coding. *University of Calgary*, *23*, 2009. http://pages.cpsc.ucalgary.ca/saul/wiki/uploads/CPSC681/open-coding.pdf

Kim, J. H., & Eom, S. J. (2019). The managerial dimension of open data success: Focusing on the open data initiatives in Korean local governments. *Sustainability*, *11*(23), 6758. https://doi.org/10.3390/su11236758

Kim, S., Kim, H. J., & Lee, H. (2009). An institutional analysis of an e-government system for anti-corruption: The case of OPEN. *Government Information Quarterly*, *26*(1), 42–50. https://doi.org/10.1016/j.giq.2008.09.002

Kučera, J. (2017). Analysis of barriers to publishing and re-use of open government data. *IDIMT 2017: Digitalization in Management, Society and Economy - 25th Interdisciplinary Information Management Talks*, 305-314.

Luna-Reyes, L. F. (2018). US open data policy: Advances and recommendations. *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, 1–10. https://doi.org/10.1145/3209281.3209308

Máchová, R., & Lnenicka, M. (2017). Evaluating the quality of open data portals on the national level. *Journal of Theoretical and Applied Electronic Commerce Research*, *12*(1), 21–41. https://doi.org/10.4067/S0718-18762017000100003

Martin, C. (2014). Barriers to the open government data agenda: Taking a multi-level perspective: barriers to the open government data agenda. *Policy & Internet*, *6*(3), 217–240. https://doi.org/10.1002/1944-2866.POI367

Martin, S. (2013). Risk analysis to overcome barriers to open data. *Electronic Journal of e-Government*, *11*(2), 348-359.

McIntosh, M. J., & Morse, J. M. (2015). Situating and constructing diversity in semi-structured interviews. *Global Qualitative Nursing Research*, *2*, 233339361559767. https://doi.org/10.1177/2333393615597674

Melin, U. (2016, August). Challenges and benefits in an open data initiative-a local government case study of myths and realities. In *15th IFIP Electronic Government and the 8th Electronic Participation Conference (EGOV ePart 2016) 23*, 111-122. https://doi.org/10.3233/978-1-61499-670-5-111

Merriam-Webster. (n.d.-a). *Benchmark.* In Merriam-Webster.com dictionary. Retrieved January 26, 2022, from https://www.merriam-webster.com/dictionary/benchmark

Merriam-Webster. (n.d.-b). *Community.* In Merriam-Webster.com dictionary. Retrieved January 26, 2022, from https://www.merriam-webster.com/dictionary/community

Merriam-Webster. (n.d.-c). *Metadata.* In Merriam-Webster.com dictionary. Retrieved August 15, 2022, from https://www.merriam-webster.com/dictionary/metadata

Ministry of the Interior and Kingdom Relations (2015). *Kamerbrief Nationale open data agenda 2016*. https://open.overheid.nl/repository/ronl-archief-51c13163-825d-48c0-b240-b2d0f6d33f61/1/pdf/kamerbrief-over-nationale-open-data-agenda-2016-noda.pdf

Ministry of the Interior and Kingdom Relations (2020). *Open Government, Open Democracy - The Netherlands Open Government Action Plan 2020-2022.* https://www.informatiehuishouding.nl/binaries/informatiehuishouding/documenten/instrumenten/2020/12/29/index/2020-12-23+The+Netherlands+Open+Government+Action+Plan+2020-2022.pdf

Morelli, N. (2017). Open data: creating communities and practices for a new common. *International Conference on Internet Science*, *10750.* 126-136. Springer International Publishing. https://doi.org/10.1007/978-3-319-77547-0_10

OECD (2008). *Recommendation of the council for enhanced access and more effective use of public sector information.* https://www.oecd.org/sti/40826024.pdf

O'Leary-Kelly, S. W., & J. Vokurka, R. (1998). The empirical assessment of construct validity. *Journal of Operations Management*, *16*(4), 387–405. https://doi.org/10.1016/S0272-6963(98)00020-5

Open Government Partnership (n.d.-a). *Approach*. Retrieved from: https://www.opengovpartnership.org/about/approach/

Open Government Partnership (n.d.-b). *Pilot open data communities concept (NL0048).* https://www.opengovpartnership.org/members/netherlands/commitments/NL0048/

Ostrom, E. (2008). Tragedy of the commons. *The new palgrave dictionary of economics*, *2*. Palgrave Macmillan, New York. https://hdl.handle.net/10535/5887

Parkes, E., Karger-Lerchl, T., Wells, P., Hardinges, J., & Vasileva, R. (2018). Using open data to deliver public services. *Report for Open Data Institute*. https://theodi.org/article/using-open-data-for-public-services-report-2/

Patton, M. Q. (1990). *Qualitative research & evaluation methods: Integrating theory and practice*. SAGE Publications, Inc.

Purwanto, A., Zuiderwijk, A. and Janssen, M. (2020). Citizen engagement with open government data: Lessons learned from Indonesia's presidential election. *Transforming Government: People, Process and Policy*, *14*(1), 1-30. https://doi.org/10.1108/TG-06-2019-0051

Reis, H. T., & Judd, C. M. (Red.). (2000). *Handbook of research methods in social and personality psychology*. Cambridge University Press.

Ritchie, J., Lewis, J., Nicholls, C. M., & Ormston, R. (Red.). (2013). *Qualitative research practice: A guide for social science students and researchers*. SAGE

Rose, C. (1986). The comedy of the commons: custom, commerce, and inherently public property. *The University of Chicago Law Review*, *53*(3), 711-781. https://doi.org/10.2307/1599583

Saxena, S. (2017). "Usage by stakeholders" as the objective of "transparency-by-design" in open government data: Case study of Sri Lanka's open data initiative. *Information and Learning Science*, *118*(7/8), 420–432. https://doi.org/10.1108/ILS-05-2017-0034

Saxena, S. (2018). Drivers and barriers to re-use Open Government Data (Ogd): A case study of open data initiative in Philippines. *Digital Policy, Regulation and Governance*, *20*(4), 358–368. https://doi.org/10.1108/DPRG-08-2017-0045

Sayogo, D. S., Pardo, T. A., & Cook, M. (2014). A framework for benchmarking open government data efforts. *2014 47th Hawaii International Conference on System Sciences*, 1896–1905. https://doi.org/10.1109/HICSS.2014.240

Scott, W. R. (2005). Institutional theory: Contributing to a theoretical research program. *Great minds in management: The process of theory development*, 37(2), 460-484.

Scott, W. R. (2013). *Institutions and organizations: Ideas, interests, and identities* (Fourth edition). SAGE Publications, Inc.

Sekaran, U., & Bougie, R. (2016). *Research methods for business: A skill-building approach* (Seventh edition). Wiley.

Shadbolt, N., & O'Hara, K. (2013). Linked data in government. *IEEE Internet Computing*, *17*(4), 72–77. https://doi.org/10.1109/MIC.2013.72

Shao, D. S., & Saxena, S. (2019). Barriers to open government data (OGD) initiative in Tanzania: Stakeholders' perspectives. *Growth and Change*, *50*(1), 470-485. https://doi.org/10.1111/grow.12282

Shepherd, E., Bunn, J., Flinn, A., Lomas, E., Sexton, A., Brimble, S., Chorley, K., Harrison, E., Lowry, J., & Page, J. (2019). Open government data: Critical information management perspectives. *Records Management Journal*, *29*(1/2), 152–167. https://doi.org/10.1108/RMJ-08-2018-0023

da Silva Craveiro, G., & Albano, C. S. (2015). Budgetary data (In an open format) benefits, advantages, obstacles and inhibitory factors in the view of the intermediaries of this system: A study in latin american countries. *Open and Big Data Management and Innovation*, *9373*, 223–235. Springer International Publishing. https://doi.org/10.1007/978-3-319-25013-7_18

Smith, G., & Sandberg, J. (2018). Barriers to innovating with open government data: Exploring experiences across service phases and user types. *Information Polity*, *23*(3), 249–265. https://doi.org/10.3233/IP-170045

Steckler, A., & McLeroy, K. R. (2008). The importance of external validity. *American Journal of Public Health*, *98*(1), 9–10. https://doi.org/10.2105/AJPH.2007.126847

The World Bank (n.d.). *Starting an Open Data Initiative: Benefits of Open Data*. http://opendatatoolkit.worldbank.org/en/starting.html

TU Delft (n.d.). *MSc Complex Systems Engineering and Management.* https://www.tudelft.nl/onderwijs/opleidingen/masters/cosem/msc-complex-systemsengineering-and-management/

Ubaldi, B. (2013). Open government data: Towards empirical analysis of open government data initiatives. *OECD Working Papers on Public Governance*, *22*. OECD Publishing, Paris. https://doi.org/10.1787/5k46bj4f03s7-en

Unisys (n.d.). *Digital government services.* Retrieved February 8, 2022, from: https://www.unisys.com/glossary/digital-government-services/

Van Thiel, S. (2015). *Bestuurskundig onderzoek: Een methodologische inleiding (3e ed.)*. Coutinho

Van Veenstra, A. F., & Van Den Broek, T. A. (2013). Opening moves–drivers, enablers and barriers of open data in a semi-public organization. *International Conference on Electronic Government*, *8074,* 50-61. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-40358-3_5

Weiss, R. S. (1994). *Learning from strangers: The art and method of qualitative interview studies*. Free Press, New York.

Wieczorkowski, J. (2019). Barriers to using open government data. *Proceedings of the 2019 3rd International Conference on E-Commerce, E-Business and E-Government- ICEEG 2019*, 15–20. https://doi.org/10.1145/3340017.3340022

Williams, M., & Moser, T. (2019). The art of coding and thematic exploration in qualitative research. *International Management Review*, *15*(1), 45-55.

Yin, R. K. (2015). *Qualitative research from start to finish*. Guilford publications.

Yin, R. K. (2018). *Case Study Research and Applications*. 6th ed. SAGE

Zeleti, F. A., Ojo, A., & Curry, E. (2016). Exploring the economic value of open government data. *Government Information Quarterly*, *33*(3), 535-551. https://doi.org/10.1016/j.giq.2016.01.008

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.

Zuiderwijk, A., Pirannejad, A., & Susha, I. (2021). Comparing open data benchmarks: Which metrics and methodologies determine countries' positions in the ranking lists? *Telematics and Informatics*, *62*, 101634. https://doi.org/10.1016/j.tele.2021.101634

Zuiderwijk, A., Shinde, R., & Janssen, M. (2019). Investigating the attainment of open government data objectives: Is there a mismatch between objectives and results?. *International Review of Administrative Sciences*, *85*(4), 645-672. https://doi.org/10.1177/0020852317739115

Zuiderwijk-van Eijk, A. M. G., & Reuver, M. D. (2021). Why open government data initiatives fail to achieve their objectives: categorizing and prioritizing barriers through a global survey. *Transforming Government: People, Process and Policy, 15*(4), 377-395. https://doi.org/10.1108/TG-09-2020-0271

# 8 Appendices

## Appendix 1: Case study protocol

**1. Introduction**

This case study protocol gives an overview of all the relevant information regarding the case study executed for the thesis. Some of the information of the case study protocol, is part of the thesis as well. To avoid duplication, the duplex information is removed from the appendix and instead a reference to the corresponding section in the main text is added.

**2. Previous research on the topic**

The Open Data Institute (ODI) has developed a framework that exists out of six skills sets, one of those skill sets is management. One of the skills that is part of the management set is building communities. Furthermore, one of their online courses (Measuring success for open data) states “open data communities need to be built and success stories communicated”. The institute does not elaborate on the specific skills.

Morelli (2020) discusses the creation of communities around open data, arguing that it is one of the two elements that would make open data a common. However, the work of Morelli is not focused on OGD but entails a broader perspective.

Attard et al. (2015) have inventoried evaluated aspects in open government initiative evaluations. The following aspects were considered: Data, functionality, features, stakeholder participation, initiative maturity, stakeholder feedback. Only two of the 25 evaluations included the stakeholder participation. One of the two evaluations does not specifically mention data communities, the other one (Sayogo, Pardo & Cook, 2014) does.

Sayogo et al. (2014) use ‘Engagement Capability’ as one of the three factors describing the stage of open government data portal development and one of two factors to review the progress of open data portal (together with Data Manipulation Capability). The OGD portals are divided into three categories: No Features, Participative and Collaborative. One of the conclusions is that (2014, p. 1904) “The types of user engagement provided in the OGD portal might correlate to the level of user’s engagement”. However, it is not clear what the effects of data communities are on national governments' open data platforms.

Other papers on open data found in this literature review mention communities, but considering the context in which the terminology is used, they are not referring to open data communities. For instance the work of Dawes et al. (2016) contains an ecosystem design of Open

Government Data Platforms, also mentioning community characteristics. In the work of Dawes et al. (2016) involved (existing) communities related to the data-sets are discussed, but these are not open data communities. Since Sayogo et al. (2014) proposes that the engagement capabilities can be used describe the progress of an Open Data Portal, the question rises if the presence of communities can fetch away barriers for open (government) data and even broader what the exact effects are of those communities on the Open Data platforms.

**Main research question**

Since there is little existing research on data communities and the concept is still relatively new, this research will have an exploratory, qualitative nature. The thesis will investigate the possible link between data communities and the Open Data platforms. The next chapter will further elaborate on the thesis approach, formulate sub questions and illustrate the conceptual model.

*"What are the potential effects of open data communities on open data benefits and barriers?"*

**Additional research questions**

Based on the literature review and the information collected to write the introduction, a set of sub questions were formulated. Each question is briefly explained in Chapter 1, section 3.

**3. Design**

**Design choices**

A multiple-case study design will be used, since the evidence from multiple cases is often considered more compelling. Furthermore, the study is regarded as being more robust (Yin, 2018). The cases are analysed holistic, since their won't be a combination of both qualitative and quantitative data collection.

**Object of study**

The object of study is a data community, which meets the following requirements: the data community is set up by a national government (1) and is focusing on open data (2).

**4. Case Selection**

According to Yin (2015), cases can either be a literal replication or a theoretical replication. Open data communities that are currently used are limited. Furthermore: The prediction is that the selected cases will deliver similar results.

Therefore, literal replication will be more suitable. Three or four cases are optimal for literal replication.

**Criteria for project selection**

Different techniques can be used for selecting cases (Patton, 1990). For this research intensity sampling will be used. The sampling strategy can be defined as "The case is information rich but not an extreme case" (Patton, 1990, pp. 182-183). This sampling strategy is comparable to extreme case sampling because it uses the same logic: a lot of interesting information is available on the extremities of the distribution range. However, when choosing for intensity sampling the focus lies less on the extremes. There are three criteria formulated for selecting the cases:

1. Case study factor criterion:
2. Case study quality factor criterion:
3. Case study performance information criterion:

The full explanation of the three criteria can be found in Chapter 3.

**Informant selection**

After the case study is selected according to the paragraph above, relevant managers and users of the selected open data community will need to be interviewed. I will seek to speak with at least two interviewees per role described below and a total minimum of 8 participants. This should be enough to satisfy the research objectives and goals. Potential interviewees include the following:

1. Community managers, specifically managing one (or more) open data communities;
2. Community contributors, who are members of communities within the selected case;
3. Open data researchers.

**Case Study Data and Delegation tables**

This section shows examples of the tables that will be used to collect case study information, contacts, and documents. The tables are not numbered since they do not contain any actual information. The personal information is collected and stored according to the data collection plan (see section 6 of the case study protocol).

| # | Case study name | Contact details (such as e-mail address) | Online/offline interview | Interview date | Transcript sharing date |
|---|---|---|---|---|---|
| 1 | | | | | |
| 2 | | | | | |
| 3 | | | | | |

| # | Case study name | Materials/documents received |
|---|---|---|
| 1 | | |
| 2 | | |
| 3 | | |

**5. Case Study Procedures and Roles**

**Field procedures**

This paragraph will focus on the case study procedures. For each case, multiple interviews will be conducted. For the Dutch case, a direct contact is available. For the other cases, gaining access to key organisations or interviewees may be a challenge. The snowballing effect will be used to gather new contacts (for possible interviews).

To conduct the fieldwork, resources (besides time) are required:

- A personal computer, with MS Teams
- A software program to make notes
- A safe (virtual) storage location for the interview data
- A TU Delft e-mail address to approach possible interviewees
- A voice recorder (when teams fails to record)
- Paper and a pen to make extra notes

These resources forms the basis of the data collection environment. However, there are more practical things to think of when doing fieldwork, such as: Online interviews should be conducted from a quiet location, meaning there is no diversion and no other people around.

**Communication**

See the appendix (Appendix 1.1 and 1.2) for interview communication per e-mail and for interview communication.

**Team members**

The case study is executed as part of a master's thesis and therefore individually. It is however supervised.

- **Thesis author**
  Tom Schuurmans
  Technical University of Delft, MSc. Student:
  Complex Systems Engineering & Management
  Jaffalaan 5, 2628 BX Delft
  +31 65 59 85421
  t.f.schuurmans@student.tudelft.nl

- **First Supervisor**
  Dr. A.M.G. (Anneke) Zuiderwijk- van Eijk
  Assistant Professor
  +31 15 27 86471
  A.M.G.Zuiderwijk-vanEijk@tudelft.nl
  Faculty of Technology, Policy and Management
  Jaffalaan 5, 2628 BX Delft

- **Second Supervisor, Chair**
  Prof dr. M.E. (Martijn) Warnier
  Professor
  +31 15 27 82232
  m.e.warnier@tudelft.nl
  Jaffalaan 5, 2628 BX Delft

**6. Data Collection**

**Requested documents**

Documents will be needed to execute the document analysis. Therefore, documents should be searched online or requested when they cannot be found. Despite the fact that as many documents as possible need to be collected, the success of the study is not dependent on fully obtaining the list below. Some of the documents will be available for some of the selected cases, but not for all of them.

- Project overview (either technical, financial, capacity wise)
- Policy documents (regarding goals, implementation or maintenance)
- Quality Management Plan of the case community
- General Agency/Company Information

**Questionnaire**

The same set interview questions were asked to all interviewees, however based on their background some general questions differed. The questions regarding benefits, barriers and institutional instruments were all identical. This allows for easier comparison and analysis between the responses of interviewees. Each interview is unique however and the input depends upon the answers study participants.

**Data collection plan**

A Data Management Plan (DMP) has been drafted in order to provide an overview of the research data that were collected, how they were collected and how the data were used during and after the research. This section provides a brief overview of these factors.

Both the DMP and the Informed consent form (attached to the e-mails and mentioned in the interview protocol) were reviewed by the Human Research Ethics Committee (HREC).

The table below (Table 15) includes the type of data that is collected, the data format, the method of collecting the data, the purpose of collecting the data, the location where the data is stored and who has access to the data.

*Table 15: Overview of collected case study data*

| Type of data | File format(s) | How will data be collected (for re-used data: source and terms of use)? | Purpose of processing | Storage location | Who will have access to the data |
|---|---|---|---|---|---|
| Interview recordings | .MP4 | During the interviews | Capturing experts opinions on data communities | TU Delft OneDrive | Tom Schuurmans - and my supervisor Anneke Zuiderwijk- van Eijk |
| Interview transcripts | .docx | Produced from the interview recording | Capturing experts opinions on data communities | TU Delft | (Same as row above) |
| Participant's list | .xlsx | Overview of interviewed experts | Finding domain specific experts | TU Delft OneDrive | (Same as row above) |
| Anonymized transcripts / interview Summary | .docx | Produced from the transcripts | Provided a privacy preserving version of the data for long term archival | TU Delft OneDrive | Publically available, will be attached as thesis appendix and included in TU Delft Repository |
| Report | .pdf | Serves as record of the process as well as documentation | Long term documentation | TU Delft OneDrive | (Same as row above) |

The research data that can be directly traced back to the participants will be destroyed after the interviews are transcribed. Pseudonymised or aggregated data will be stored on the systems of the TU Delft. This (pseudonymised) data will eventually be stored on a research data repository, which makes it accessible for future usage. This is in accordance with the TU Delft Research Data Framework Policy.

**7. Analysis**

The interview questions either focus on the background of the interviewee or question a single element of the theoretical framework. This includes the benefits, barriers and institutional instruments that were identified as potentially relevant for open data communities.

The interview transcripts will be labelled according to the code tree. The table can be found in the thesis (Table 10). By using the labelled answers, an overview can be given of the (potential) contributions of the open data community to open data benefits and barriers. Furthermore,

institutional instruments that stimulate value creation within the open data community can be identified similarly.

The results of the analysis can be found in Chapter 5 of the thesis.

**8. Plan Validity**

Based on the work of Yin (2018), four types of relevant plan validity have been identified:

- Construct validity
- Internal validity
- External validity
- Reliability

In order to ensure construct validity, existing literature is used to identify barriers and benefits of open data. There is no causal validity between results and an intervention. The case study findings will be used to formulate (explorative) best practices. The domain to which study findings can be generalised is limited to open government data communities on a national level.

The main text of the thesis (see Chapter 3) discusses the validity of the research in more detail.

**9. Study Limitations**

There is not much information found on the characteristics of open data communities. Scientific literature is very limited, and only little policy documents were found. Data communities are rare and some of the successful communities seemed to be inactive or even shut down. There are no conflicts of interest.

The main text of the thesis (see Chapter 6) discusses the limitations of the research in more detail.

**10. Reporting**

The nature of the thesis is exploratory and should therefore focus on providing information for future research. Therefore, the target audience can be defined as Open (Government) Data researchers and academia. The relation of this thesis to larger studies can be seen as follows: The subject of open data is researched a lot. Open government data initiatives are benchmarked, but communities seem to be absent in most of those benchmarks. This may relate to the lack of knowledge on how those communities effect open data benefits and barriers. This research will form a base for exploring the contributions of the communities towards open data efforts and provide insights in their functioning.

**11. Planning**

The table below (Table 16) shows the overall planning for the case study.

*Table 16: Case study planning*

| ***Task*** | **Execution weeks** |
|---|---|
| *Planning* | Week 10 |
| *Preparation* | Week 11 - 13 |
| *Data collection* | Week 14 - 16 |
| *Data analysis* | Week 17 - 20 |
| *Reporting* | Week 21 - 23 |

**Appendix 1.1: E-mail template 1 (first invitation)**

TO: Survey Participant

FROM: Tom Schuurmans, Thesis Author

SUBJECT: Open Data communities - Case Study

Dear **[name]**,

I would like to get in touch with you for a research project I am doing on the value creation of the Dutch open data community. I am very curious about the role you play within this community and your own experiences with the Open Data community.

If you are open to contributing, I would like to invite you for an interview. The interview will be used together with a series of other interviews for a master's thesis. I would be happy to send you more information.

Best regards,

Tom Schuurmans

**E-mail attachment(s):**

**[Basic research introduction]**

**Appendix 1.2: E-mail template 2 (confirmation and RFI)**

TO: Survey Participant

FROM: Tom Schuurmans, Thesis Author

SUBJECT: Open Data communities - Case Study

Dear **[name]**,

Thank you for agreeing to participate in the Open data communities case study concerning national open data communities on Open Government Data Platforms (OGPDs). I have attached relevant background information regarding the research I'm doing. This includes the objectives, goals, and methods of the research. We made an appointment to conduct the interview on **[insert day/month]** at **[insert time]**. The interview will be **[online/on a specific location]**. If for any reason this is no longer a suitable timeslot, please contact me via the contact details below.

I've also attached the Informed Consent form. I want to ask you to read the form and send it back signed before the interview. If you have questions about it, we can also go through the form at the start of the interview. Please keep in mind that you can withdraw from participating in the research at any moment.

If you have any questions, please use the following information to contact me.

By email, at t.f.schuurmans@student.tudelft.nl.

By telephone, at +316 xxxx xxxx.

Regards,

Tom Schuurmans

**E-mail attachment(s):**

**[List of objectives/research interests]**

**[Informed Consent Form]**

## Appendix 2: Interview protocols

### Appendix 2.1: Interview protocol – interviews with open data community managers

**Interview Introduction**

Nice to meet you! My name is Tom Schuurmans and as part of my graduation research on open data communities, I would like to ask you some questions. My research focuses on the contribution of 'data communities' to open data value creation.

We will go through a list of questions together. You may have already seen the questions, if requested. I want to ask you to answer the questions into as much detail as possible. It would be very helpful for me if you can go into more detail than just a yes or a no.

If you are unable to answer a question, don't worry. Some questions may not be applicable to you or your role or experience with the open data community. It would be helpful if you could refer me to someone else who could be able to answer the specific question. Also, I would welcome any appropriate documentation or other information sources related to the open data community that you think might be valuable for my research.

**Recording and informed consent**

Before we can start the actual interview we will discuss information regarding your participation in this research study. I will provide you with the necessary information regarding your rights, but also explain what you can expect from me and my research supervisor. As stated in the Informed Consent form, your participation is completely voluntary. Furthermore you have the right to withdraw from participating in this research at any point.

In this study, I collect the following personal information: your name, title, e-mail address, current function and the recording of this interview. I will collect this information because it may provide context for the obtained research findings. Only the people on my research project team, being myself and my supervisors (Dr. Anneke Zuiderwijk and Prof. Dr. Martijn Warnier) will have access to this recording as well as its transcript.

I will be making notes of the interview. However I would prefer to use the MS Teams transcript function, to help me with my note-taking and working out the interview transcript. Do you agree with me recording our conversation?

**[If the respondent agrees with the use of the MS Teams recording, move to the next section of the interview. If there are any questions left, indicate that the recording is not mandatory but that the transcript recording will be kept confidential. It will only be used to make the verbatim transcript of the conversation.]**

I will make a transcript of the interview. The interview recording will be deleted within two months after the research and the transcript will be kept in a secure place. The anonymised (non-personal) information you provide in this interview will be used in my master thesis report. This report will be openly shared on the research repository of the TU Delft. An anonymised summary of the interview will also be published along with the master thesis report in the TU Delft repository. The anonymised information may also be used for scientific publications and presentations. My supervisors and I will always treat your data confidentially.

Again, I would like to stress that your participation in this study is completely voluntary. If you do not wish to answer one or more questions, don't hesitate. Furthermore, you can also withdraw from the study at any time. You do not need to specify a reason to do so. If you have questions about this study, please contact me or my first supervisor, whose information you can find on the informed consent form.

**[Check again whether the respondent has any questions left. If not, check if the respondent has either signed the human subjects consent form or indicates verbally that he/she agrees to this. If the respondent does not want to sign the form and cannot clearly indicate that he or she agrees to the informed consent form; thank him or her for the time and attending the interview and then end the interview here.]**

**[When agreed upon: Switch on MS teams recording and transcript functions, make sure that it is recording.]**

**1. Background Information**

In the first section of the interview, I will ask you to elaborate on your background.

Q1. Where are you currently employed and what is your current function?

Q2. What is your role in the open data community?

Q3. When did you start to interact with/work on/use the data community?

**2. The open data community**

This section concerns the [case country] open data community itself.

Q4. In which country/countries is/are the data community active? (Are there international contacts?)

Q5. When was the data community implemented?

Q6. What was the initial goal of the data community?

Q7. How active is the data community currently?
(How many members does the community have and how many of those are active?)

Q8. Who is active in the data community currently?

Q9. How is the data community implemented?

Q10. Which (technical) platform is used in the data community?

Q11. Which other interaction mechanisms are currently used in the community (think of e-mail, message system, forum)?

Q12. Is additional information/ documentation on the data community publicly available and where can it be found?

Q13. Are there other characteristics of the data community relevant for this research?

**3. Open data impact**

I would now like to ask you some questions about open data benefits in relation to data communities.

For each of the following benefits I will ask you to what extent the open data community (that you participate in) contributes to the utilization of the benefit.

Q14. To what extent does the open data community contribute to increased social control (of society on the government)?

Q15. To what extent does the open data community contribute to increased civic participation and public engagement?

Q16. To what extent does the open data community contribute to more informed citizens?

Q17. To what extent does the open data community contribute to increased accountability (of the government)?

Q18. To what extent does the open data community contribute to gained access to external capacity and resources for solving problems?

Q19. To what extent does the open data community contribute to increased intra-governmental collaboration?

Q20. To what extent does the open data community contribute to increased problem-solving capacity?

Q21. To what extent does the open data community contribute to the use of collective intelligence to solve public problems?

Q22. Does the open data community that you participate in also experience other benefits than the ones above?

**Extra explanation [only to be used when asked for]:**

| Benefit | Explanation |
|---|---|
| Increased social control | A set of rules, laws, standards in a society (in this case the country of the open data community, which regulate human behaviour. |
| Increased civic participation and public engagement | The involvement of specialists who listen to, their understanding of, and interaction with non-specialists. |
| More informed citizens | Refers to citizens having more knowledge or information about something. |
| Increased accountability | The willingness to accept the responsibility for in this case the government (of the open data communities' country). |
| Gained access to external capacity and resources for solving problems | External capacity refers to capacity outside of the data uploader. Problem solving resources can be human resources, but also money and/or other resources that help you identify and solve problems effectively and efficiently. |
| Increased intra-governmental collaboration | Intra-government collaboration is collaboration between different governments. Can be on a national (e.g., ministries) or local level (e.g., municipalities). |
| Increased problem-solving capacity | Capacity that helps you identify and solve problems effectively and efficiently. |
| Use of collective intelligence to solve public problems | Collective intelligence refers to the ability of a certain group of people to perform a broad spectrum of different tasks. |

**4. Open data challenges**

Open data communities may potentially contribute to mitigating various barriers for open data provision and use. For each of the following barriers I will ask you to what extent the open data community (you participate in) contributes to the mitigation of the barrier.

Q23. To what extent does the open data community contribute to the governmental organisation (responsible for or owning the data) being interested in using open data?

Q24. To what extent does the open data community contribute to the governmental organisation (responsible for or owning the data) having the organisational capabilities, routines, and processes to use open data?

Q25. To what extent does the open data community contribute to higher engagement of public managers with the use of open data?

Q26. To what extent does the open data community contribute to the increase of human resources with the knowledge, skills and/or capabilities to use open data?

Q27. To what extent does the open data community contribute to the data provider paying attention to the data users' requests and suggestions?

Q28. To what extent does the open data community contribute to data portal users discovering and/or locating data?

Q29. To what extent does the open data community contribute to combining and connecting data?

Q30. To what extent does the open data community contribute to data portal users interaction with the data provider?

Q31. Does the open data community that you participate in also mitigate other barriers than the ones above?

**Extra explanation [only to be used when asked for]:**

| **Barrier** | **To what extent does the open data community you participate in contribute to mitigating each of the following barriers?** |
|---|---|
| The organisation is not interested in using open data | In this case the organisation is the (governmental) party uploading the data, in this case it can be ministries or other national bodies. |
| The organisation does not have the organisational capabilities, routines, and processes to use open data | In this case the organisation is the (governmental) party uploading the data, in this case it can be ministries or other national bodies. Capabilities: the ability to do something, routines: a sequence of actions that is regularly followed, processes: a set of steps taken to achieve a particular goal. |
| Low engagement of public managers with the use of open data | Public managers in the organisation of the data uploading party having little interest in using open data. |
| There are no human resources with the knowledge, skills and/or capabilities to use open data | Persons in the organisation of the data uploader don't have any or multiple of the following: knowledge: the theoretical or practical understanding of a subject, skills: a learned power of doing something competently, capabilities: the ability to do something. |
| Data provider ignores my requests and suggestions | 'My' refers in this case to the user of the open data platform. |
| Difficulty in discovering/locating data | Refers to the end user having difficulties with discovering/locating data. |
| Data cannot be combined and connected | Refers to the end user having difficulties with combining and connecting data. |
| Difficulties to interact with the data provider | Refers to the end user having difficulties to interact with the data provider. |

**5. Institutionalization and instruments**

We will now discuss some institutional instruments and forces that may be relevant for data communities.

Q32. To what extent do policies, regulations, laws and other formal institutional instruments contribute to the value creation of the open data community?

Q33. To what extent do norms, cultural characteristic, pressure and influence from certain actors, and other informal institutional instruments contribute to the value creation of the open data community?

Q34. To what extent do rewards, sanctions and other enforcing institutional instruments contribute to the value creation of the open data community?

Q35. What are managerial challenges and difficulties that data community can be faced with?

Q36. How can reactions, scepticism or enthusiasm influence the development of an open data community?

**6. Final questions**

There are some final questions that I want to ask you.

Q37. Which recommendations on maintenance and continuity of data communities can you give, based on your research and experience with open data?

Q38. Do you have recommendations for other data communities in general?

Q39. Do you have any documentation on the data community, for instance, policy documents, architecture overviews, that you can share with me?

We have now reached the end of this interview. Thank you very much for participating in the interview and making time. I will create a transcript of our conversation and send you the answers to the questions. You will be able to review them and have sufficient time to add or remove anything you want. After the thesis is finalised, I will share the results with you.

Q40. Do you have any questions left for me? Or do you want to say anything about this interview?

**[Stop the MS transcript recording]**

**Appendix 2.2: Interview protocol – interviews with open data community users**

**Interview Introduction**

Nice to meet you! My name is Tom Schuurmans and as part of my graduation research on open data communities, I would like to ask you some questions. My research focuses on the contribution of 'data communities' to open data value creation.

We will go through a list of questions together. You may have already seen the questions, if requested. I want to ask you to answer the questions into as much detail as possible. It would be very helpful for me if you can go into more detail than just a yes or a no.

If you are unable to answer a question, don't worry. Some questions may not be applicable to you or your role or experience with the open data community. It would be helpful if you could refer me to someone else who could be able to answer the specific question. Also, I would welcome any appropriate documentation or other information sources related to the open data community that you think might be valuable for my research.

**Recording and informed consent**

Before we can start the actual interview we will discuss information regarding your participation in this research study. I will provide you with the necessary information regarding your rights, but also explain what you can expect from me and my research supervisor. As stated in the Informed Consent form, your participation is completely voluntary. Furthermore you have the right to withdraw from participating in this research at any point.

In this study, I collect the following personal information: your name, title, e-mail address, current function and the recording of this interview. I will collect this information because it may provide context for the obtained research findings. Only the people on my research project team, being myself and my supervisors (Dr. Anneke Zuiderwijk and Prof. Dr. Martijn Warnier) will have access to this recording as well as its transcript.

I will be making notes of the interview. However I would prefer to use the MS Teams transcript function, to help me with my note-taking and working out the interview transcript. Do you agree with me recording our conversation?

**[If the respondent agrees with the use of the MS Teams recording, move to the next section of the interview. If there are any questions left, indicate that the recording is not mandatory but that the transcript recording will be kept confidential. It will only be used to make the verbatim transcript of the conversation.]**

I will make a transcript of the interview. The interview recording will be deleted within two months after the research and the transcript will be kept in a secure place. The anonymised (non-personal) information you provide in this interview will be used in my master thesis report. This report will be openly shared on the research repository of the TU Delft. An anonymised summary of the interview will also be published along with the master thesis report in the TU Delft repository. The anonymised information may also be used for scientific publications and presentations. My supervisors and I will always treat your data confidentially.

Again, I would like to stress that your participation in this study is completely voluntary. If you do not wish to answer one or more questions, don't hesitate. Furthermore, you can also withdraw from the study at any time. You do not need to specify a reason to do so. If you have questions about this study, please contact me or my first supervisor, whose information you can find on the informed consent form.

**[Check again whether the respondent has any questions left. If not, check if the respondent has either signed the human subjects consent form or indicates verbally that he/she agrees to this. If the respondent does not want to sign the form and cannot clearly indicate that he or she agrees to the informed consent form; thank him or her for the time and end the interview here.]**

**[When agreed upon: Switch on MS teams recording and transcript functions, make sure that it is recording.]**

**1. Background Information**

In the first section of the interview, I will ask you to elaborate on your background.

Q1. Where are you currently employed and what is your current function?

Q2. What is your role in the open data community?

Q3. When did you start to interact with/work on/use the data community?

**2. The open data community**

This section concerns the [case country] open data community itself.

Q4. How did you get in touch with the community>

Q5. How often are you logging in?

Q6. In which way are you using the community? Are you asking questions, and are you answering questions?

Q7. How do you think the software of community is working?

Q8. How does the community contribute to the goals of your organisation?

Q9. Do you meet new people in the community who are relevant for your network?

Q10. Which other interaction mechanism are you using, following from your activity on the community platform?

Q11. Are more colleagues within your organisation using the community?

Q12. Do you also use data.overheid.nl and if yes, in which way?

Q13. Are there other characteristics of the data community relevant for this research?

**3. Open data impact**

I would now like to ask you some questions about open data benefits in relation to data communities. For each of the following benefits I will ask you to what extent the open data community (that you participate in) contributes to the utilization of the benefit.

Q14. To what extent does the open data community contribute to increased social control (of society on the government)?

Q15. To what extent does the open data community contribute to increased civic participation and public engagement?

Q16. To what extent does the open data community contribute to more informed citizens?

Q17. To what extent does the open data community contribute to increased accountability (of the government)?

Q18. To what extent does the open data community contribute to gained access to external capacity and resources for solving problems?

Q19. To what extent does the open data community contribute to increased intra-governmental collaboration?

Q20. To what extent does the open data community contribute to increased problem-solving capacity?

Q21. To what extent does the open data community contribute to the use of collective intelligence to solve public problems?

Q22. Does the open data community that you participate in also experience other benefits than the ones above?

**Extra explanation [only to be used when asked for]:**

| **Benefit** | **Explanation** |
|---|---|
| Increased social control | A set of rules, laws, standards in a society (in this case the country of the open data community, which regulate human behaviour. |
| Increased civic participation and public engagement | The involvement of specialists who listen to, their understanding of, and interaction with non-specialists. |
| More informed citizens | Refers to citizens having more knowledge or information about something. |
| Increased accountability | The willingness to accept the responsibility for in this case the government (of the open data communities' country). |
| Gained access to external capacity and resources for solving problems | External capacity refers to capacity outside of the data uploader. Problem solving resources can be human resources, but also money and/or other resources that help you identify and solve problems effectively and efficiently. |
| Increased intra-governmental collaboration | Intra-government collaboration is collaboration between different governments. Can be on a national (e.g., ministries) or local level (e.g., municipalities). |
| Increased problem-solving capacity | Capacity that helps you identify and solve problems effectively and efficiently. |
| Use of collective intelligence to solve public problems | Collective intelligence refers to the ability of a certain group of people to perform a broad spectrum of different tasks. |

**4. Open data challenges**

Open data communities may potentially contribute to mitigating various barriers for open data provision and use. For each of the following barriers I will ask you to what extent the open data community (you participate in) contributes to the mitigation of the barrier.

Q23. To what extent does the open data community contribute to the governmental organisation (responsible for or owning the data) being interested in using open data?

Q24. To what extent does the open data community contribute to the governmental organisation (responsible for or owning the data) having the organisational capabilities, routines, and processes to use open data?

Q25. To what extent does the open data community contribute to higher engagement of public managers with the use of open data?

Q26. To what extent does the open data community contribute to the increase of human resources with the knowledge, skills and/or capabilities to use open data?

Q27. To what extent does the open data community contribute to the data provider paying attention to the data users' requests and suggestions?

Q28. To what extent does the open data community contribute to data portal users discovering and/or locating data?

Q29. To what extent does the open data community contribute to combining and connecting data?

Q30. To what extent does the open data community contribute to data portal users interaction with the data provider?

Q31. Does the open data community that you participate in also mitigate other barriers than the ones above?

**Extra explanation [only to be used when asked for]:**

| **Barrier** | **To what extent does the open data community you participate in contribute to mitigating each of the following barriers?** |
|---|---|
| The organisation is not interested in using open data | In this case the organisation is the (governmental) party uploading the data, in this case it can be ministries or other national bodies. |
| The organisation does not have the organisational capabilities, routines, and processes to use open data | In this case the organisation is the (governmental) party uploading the data, in this case it can be ministries or other national bodies. Capabilities: the ability to do something, routines: a sequence of actions that is regularly followed, processes: a set of steps taken to achieve a particular goal. |
| Low engagement of public managers with the use of open data | Public managers in the organisation of the data uploading party having little interest in using open data. |
| There are no human resources with the knowledge, skills and/or capabilities to use open data | Persons in the organisation of the data uploader don't have any or multiple of the following: knowledge: the theoretical or practical understanding of a subject, skills: a learned power of doing something competently, capabilities: the ability to do something. |
| Data provider ignores my requests and suggestions | 'My' refers in this case to the user of the open data platform. |
| Difficulty in discovering/locating data | Refers to the end user having difficulties with discovering/locating data. |
| Data cannot be combined and connected | Refers to the end user having difficulties with combining and connecting data. |
| Difficulties to interact with the data provider | Refers to the end user having difficulties to interact with the data provider. |

**5. Institutionalization and instruments**

We will now discuss some institutional instruments and forces that may be relevant for data communities.

Q32. To what extent do policies, regulations, laws and other formal institutional instruments contribute to the value creation of the open data community?

Q33. To what extent do norms, cultural characteristic, pressure and influence from certain actors, and other informal institutional instruments contribute to the value creation of the open data community?

Q34. To what extent do rewards, sanctions and other enforcing institutional instruments contribute to the value creation of the open data community?

Q35. What are managerial challenges and difficulties that data community can be faced with?

Q36. How can reactions, scepticism or enthusiasm influence the development of an open data community?

**6. Final questions**

There are some final questions that I want to ask you.

Q37. Which recommendations on maintenance and continuity of data communities can you give, based on your research and experience with open data?

Q38. Do you have recommendations for other data communities in general?

Q39. Do you have any documentation on the data community, for instance, policy documents, architecture overviews, that you can share with me?

We have now reached the end of this interview. Thank you very much for participating in the interview and making time. I will create a transcript of our conversation and send you the answers to the questions. You will be able to review them and have sufficient time to add or remove anything you want. After the thesis is finalised, I will share the results with you.

Q40. Do you have any questions left for me? Or do you want to say anything about this interview?

**[Stop the MS transcript recording]**

**Appendix 2.3: Interview protocol – interviews with open data-community researchers**

**Interview Introduction**

Nice to meet you! My name is Tom Schuurmans and as part of my graduation research on open data communities, I would like to ask you some questions. My research focuses on the contribution of 'data communities' to open data value creation.

We will go through a list of questions together. You may have already seen the questions, if requested. I want to ask you to answer the questions into as much detail as possible. It would be very helpful for me if you can go into more detail than just a yes or a no.

If you are unable to answer a question, don't worry. Some questions may not be applicable to you or your role or experience with the open data community. It would be helpful if you could refer me to someone else who could be able to answer the specific question. Also, I would welcome any appropriate documentation or other information sources related to the open data community that you think might be valuable for my research.

**Recording and informed consent**

Before we can start the actual interview we will discuss information regarding your participation in this research study. I will provide you with the necessary information regarding your rights, but also explain what you can expect from me and my research supervisor. As stated in the Informed Consent form, your participation is completely voluntary. Furthermore you have the right to withdraw from participating in this research at any point.

In this study, I collect the following personal information: your name, title, e-mail address, current function and the recording of this interview. I will collect this information because it may provide context for the obtained research findings. Only the people on my research project team, being myself and my supervisors (Dr. Anneke Zuiderwijk and Prof. Dr. Martijn Warnier) will have access to this recording as well as its transcript.

I will be making notes of the interview. However I would prefer to use the MS Teams transcript function, to help me with my note-taking and working out the interview transcript. Do you agree with me recording our conversation?

**[If the respondent agrees with the use of the MS Teams recording, move to the next section of the interview. If there are any questions left, indicate that the recording is not mandatory but that the transcript recording will be kept confidential. It will only be used to make the verbatim transcript of the conversation.]**

I will make a transcript of the interview. The interview recording will be deleted within two months after the research and the transcript will be kept in a secure place. The anonymised (non-personal) information you provide in this interview will be used in my master thesis report. This report will be openly shared on the research repository of the TU Delft. An anonymised summary of the interview will also be published along with the master thesis report in the TU Delft repository. The anonymised information may also be used for scientific publications and presentations. My supervisors and I will always treat your data confidentially.

Again, I would like to stress that your participation in this study is completely voluntary. If you do not wish to answer one or more questions, don't hesitate. Furthermore, you can also withdraw from the study at any time. You do not need to specify a reason to do so. If you have questions about this study, please contact me or my first supervisor, whose information you can find on the informed consent form.

**[Check again whether the respondent has any questions left. If not, check if the respondent has either signed the human subjects consent form or indicates verbally that he/she agrees to this. If the respondent does not want to sign the form and cannot clearly indicate that he or she agrees to the informed consent form; thank him or her for the time and end the interview here.]**

**[When agreed upon: Switch on MS teams recording and transcript functions, make sure that it is recording.]**

**1. Background Information**

In the first section of the interview, I will ask you to elaborate on your background.

Q1. Where are you currently employed and what is your current function?

Q2. What is your field of expertise within the open data sector?

Q3. When did you start your research into open data?

**2. The open data community**

This section concerns the open data community itself.

Q4. Does your research include the interaction of different parties with open data? Or open data engagement?

Q5. Are you familiar with the concept of open data communities?

Q6. Are you familiar with data.overheid.nl? Do you use it?

Q7. Have you ever visited the Dutch open data community?

Q8. Are you aware of other open data communities on a national level?

Q9. What do you know about the (technical) platforms used in open data communities?

Q10. Do you additional research information/ documentation on the data communities?

Q11. Are there other characteristics of the data community relevant for this research?

**3. Open data impact**

I would now like to ask you some questions about open data benefits in relation to data communities. For each of the following benefits I will ask you to what extent the open data community (that you participate in) contributes to the utilization of the benefit.

Q12. To what extent does the open data community contribute to increased social control (of society on the government)?

Q13. To what extent does the open data community contribute to increased civic participation and public engagement?

Q14. To what extent does the open data community contribute to more informed citizens?

Q15. To what extent does the open data community contribute to increased accountability (of the government)?

Q16. To what extent does the open data community contribute to gained access to external capacity and resources for solving problems?

Q17. To what extent does the open data community contribute to increased intra-governmental collaboration?

Q18. To what extent does the open data community contribute to increased problem-solving capacity?

Q19. To what extent does the open data community contribute to the use of collective intelligence to solve public problems?

Q20. Does the open data community that you participate in also experience other benefits than the ones above?

**Extra explanation [only to be used when asked for]:**

| Benefit | Explanation |
|---|---|
| Increased social control | A set of rules, laws, standards in a society (in this case the country of the open data community, which regulate human behaviour. |
| Increased civic participation and public engagement | The involvement of specialists who listen to, their understanding of, and interaction with non-specialists. |
| More informed citizens | Refers to citizens having more knowledge or information about something. |
| Increased accountability | The willingness to accept the responsibility for in this case the government (of the open data communities' country). |
| Gained access to external capacity and resources for solving problems | External capacity refers to capacity outside of the data uploader. Problem solving resources can be human resources, but also money and/or other resources that help you identify and solve problems effectively and efficiently. |
| Increased intra-governmental collaboration | Intra-government collaboration is collaboration between different governments. Can be on a national (e.g., ministries) or local level (e.g., municipalities). |
| Increased problem-solving capacity | Capacity that helps you identify and solve problems effectively and efficiently. |
| Use of collective intelligence to solve public problems | Collective intelligence refers to the ability of a certain group of people to perform a broad spectrum of different tasks. |

**4. Open data challenges**

Open data communities may potentially contribute to mitigating various barriers for open data provision and use. For each of the following barriers I will ask you to what extent the open data community (you participate in) contributes to the mitigation of the barrier.

Q21. To what extent does the open data community contribute to the governmental organisation (responsible for or owning the data) being interested in using open data?

Q22. To what extent does the open data community contribute to the governmental organisation (responsible for or owning the data) having the organisational capabilities, routines, and processes to use open data?

Q23. To what extent does the open data community contribute to higher engagement of public managers with the use of open data?

Q24. To what extent does the open data community contribute to the increase of human resources with the knowledge, skills and/or capabilities to use open data?

Q25. To what extent does the open data community contribute to the data provider paying attention to the data users' requests and suggestions?

Q26. To what extent does the open data community contribute to data portal users discovering and/or locating data?

Q27. To what extent does the open data community contribute to combining and connecting data?

Q28. To what extent does the open data community contribute to data portal users interaction with the data provider?

Q29. Does the open data community that you participate in also mitigate other barriers than the ones above?

**Extra explanation [only to be used when asked for]:**

| **Barrier** | **To what extent does the open data community you participate in contribute to mitigating each of the following barriers?** |
|---|---|
| The organisation is not interested in using open data | In this case the organisation is the (governmental) party uploading the data, in this case it can be ministries or other national bodies. |
| The organisation does not have the organisational capabilities, routines, and processes to use open data | In this case the organisation is the (governmental) party uploading the data, in this case it can be ministries or other national bodies. Capabilities: the ability to do something, routines: a sequence of actions that is regularly followed, processes: a set of steps taken to achieve a particular goal. |
| Low engagement of public managers with the use of open data | Public managers in the organisation of the data uploading party having little interest in using open data. |
| There are no human resources with the knowledge, skills and/or capabilities to use open data | Persons in the organisation of the data uploader don't have any or multiple of the following: knowledge: the theoretical or practical understanding of a subject, skills: a learned power of doing something competently, capabilities: the ability to do something. |
| Data provider ignores my requests and suggestions | 'My' refers in this case to the user of the open data platform. |
| Difficulty in discovering/locating data | Refers to the end user having difficulties with discovering/locating data. |
| Data cannot be combined and connected | Refers to the end user having difficulties with combining and connecting data. |
| Difficulties to interact with the data provider | Refers to the end user having difficulties to interact with the data provider. |

**5. Institutionalisation and instruments**

We will now discuss some institutional instruments and forces that may be relevant for data communities.

Q30. To what extent do policies, regulations, laws and other formal institutional instruments contribute to the value creation of the open data community?

Q31. To what extent do norms, cultural characteristic, pressure and influence from certain actors, and other informal institutional instruments contribute to the value creation of the open data community?

Q32. To what extent do rewards, sanctions and other enforcing institutional instruments contribute to the value creation of the open data community?

Q33. What are managerial challenges and difficulties that data community can be faced with?

Q34. How can reactions, scepticism or enthusiasm influence the development of an open data community?

**6. Final questions**

There are some final questions that I want to ask you.

Q35. Which recommendations on maintenance and continuity of data communities can you give, based on your research and experience with open data?

Q36. Do you have recommendations for other data communities in general?

Q37. Do you have any documentation on the data community, for instance, policy documents, architecture overviews, that you can share with me?

We have now reached the end of this interview. Thank you very much for participating in the interview and making time. I will create a transcript of our conversation and send you the answers to the questions. You will be able to review them and have sufficient time to add or remove anything you want. After the thesis is finalised, I will share the results with you.

Q38. Do you have any questions left for me? Or do you want to say anything about this interview?

**[Stop the MS transcript recording]**

## Appendix 3: Interview transcripts

A total of eight interviews took place with nine interviewees. Verbatim transcripts were made of all of these interviews. As a last step, the Dutch interview-transcripts were translated into English.

The transcripts can be retrieved directly from the research team upon request. To do so, please contact the author of this thesis via e-mail: tomschuurmans@live.nl. The members of the Graduation Committee also have the transcripts of the interviews. Their contact information is listed in the Colophon.

Overview of the interview transcripts available:

- Interview 1 (with interviewees 1 and 2, open data community managers)
- Interview 2 (with interviewee 3, open data community user)
- Interview 3 (with interviewee 4, open data community user)
- Interview 4 (with interviewee 5, open data community user)
- Interview 5 (with interviewee 6, open data community user)
- Interview 6 (with interviewee 7, open data community user)
- Interview 7 (with interviewee 8, open data researcher)
- Interview 8 (with interviewee 9, open data researcher)