WIRTSCHAFTS
UNIVERSITÄT
WIEN VIENNA
UNIVERSITY OF
ECONOMICS
AND BUSINESS

Bachelor Thesis

Open Data Hopes and Fears, Determining the barriers of Open Data

Martin Beno

Subject Area: Information Business

Student ID: h0952268

Supervisor: Prof. Dr. Axel Polleres

Date of Submission: 14.July 2016

Department of Information Systems and Operations, Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Vienna, Austria

Contents

List of Figures

List of Tables

Listings

Abstract

In recent years, the concept of Open Data has been gaining considerable attention. A steady growth in the number of published open datasets can be observed as the demand for Open Data rises. However, many potential providers are still hesitant to open their datasets and the users often face difficulties when trying to use the data. This indicates that there are still various barriers present both regarding usage and publishing of Open Data, but studies on these barriers, and mainly on their importance are rather lacking. Therefore, the main objective of this study was to gather an extensive set of information on various potential barriers of Open Data by performing a literature review and semi-structured interviews. Based on the collected information, a questionnaire was prepared in order to obtain additional detailed information on these barriers. The results of the survey are then presented in the final part of this thesis.

Acknowledgments

This thesis might not have been written without the support of a number of people. I would like to use this section of the thesis to express my gratitude towards them.

First and foremost I would like to thank my supervisor Prof. Dr. Axel Polleres and his co-supervisor Dr. Jürgen Umbrich for their continued support and giving me the opportunity to work on an exciting project. Furthermore, I would like to thank the following people for taking time from what is without a doubt a busy day and agree to being interviewed: Prof. Dr. Peter Parycek, Sonja Fischbauer, Peter Hanecak and Martin Turcek. I would then like to extend my thanks to Dr. Kathrin Figl who has provided some valuable feedback for the survey. Additional feedback was provided by Brigitte Lutz, Dr. Johann Höchtl, Martin Kaltenböck and the ADEQUATe team. Their help is greatly appreciated. Last but not least, a great thanks goes to every single participant who has completed the survey, thus providing us with valuable data for further research.

1 Introduction

The term Open Data refers to a relatively recent concept, in which machine-readable data are made freely available for anyone to reuse and redistribute. An exact definition is provided by the Open Definition:[1]

"Open Data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike."

Additionally, the Open Knowledge Foundation summarized the definition with the following 3 criteria:[2]

- Availability and access
- Reuse and redistribution
- Universal participation

Over the past few years the Open Data movement has been steadily growing, especially fueled by government and public institutions ever since the launch of the US open data portal data.gov[3] in 2009. There are several motivating factors for publishing Open Data cited in the literature. These are mainly:

- Better transparency in government resulting in more openness, reduced corruption in the government sector and improved, streamlined services [20][4];
- Enabling independent developers to develop services and applications based on published Open Data, thus creating applications with social and economic value and consequently stimulating economic growth [13];
- Promoting collaboration of citizens with the government [16][14];

Indeed, various providers including government, public and private institutions all over the world have been steadily releasing open datasets as we have observed from Table 1, which displays the number[4] of published datasets on Open Data portals in selected countries. Open Data portals are currently being used as the main distribution platform of open datasets. They can

[1] http://opendefinition.org/

[2] http://opendatahandbook.org/guide/en/what-is-open-data/

[3] http://www.data.gov/

[4] The numbers were extracted during a period of 5 months starting from December 2015 until April 2016 using the APIs of the respective Open Data portals by invoking the following command: $ 'api/3/action/package_list'

be defined as a single point of access to freely accessible and reusable data published by various public and private institutions.[5]

Table 1: Increase in the number of published datasets

	data.gv.at	data.gov.uk	data.gov.ie	data.gov.sk	govdata.de
December 2015	2308	26102	1305	874	15824
January 2016	2316	26153	1348	874	15892
February 2016	2341	27383	1382	884	17363
March 2016	2443	27966	1364	912	17700
April 2016	2478	29546	1374	913	17801

It is important to note that the Open Data portals listed in Table 1 are all governmental. While it was easy to query these portals and get the exact number of datasets, the sources of those that are published by private entities are rather fragmented, which means that specifying the exact number of the datasets for a country with non-governmental portals included would be problematic. However, efforts are being made to centralize these data. In Austria, opendataportal.at[6] is being used as a central point to publish datasets from private entities.

Although Open Data is in no way limited to government use only, most of the sources of the data sets and Open Data portals are mainly public institutions. This is also reflected in academia as only limited research regarding private entities publishing Open Data is available. This could indicate that there are still some barriers regarding publishing and using Open Data. Indeed, a recent study by Janssen et al. claims that despite numerous benefits, Open Data adoption entails a number of barriers [13]. This is further supported by a paper published by the European Commission which points out that despite the progress made to open up public data, different barriers still persist [6]. Therefore, an extensive research on these barriers is much needed. Some research on this topic has already been carried out in the recent years, such as the study by Zuiderwijk et al., which provides a systematic overview of Open Data socio-technical impediments [28]. However, much of the research done up to now has mostly been limited to merely listing the barriers, lacking a detailed analysis, which could help determine the importance and relevancy of these barriers.

[5]http://data.europa.eu/euodp/en/about
[6]https://www.opendataportal.at/

As previously mentioned, most of the research done in this field tends to focus on governmental use of Open Data. The present study however, should not be limited to Open Data barriers in government and public institutions only. It rather strives to establish a solid basis for a research on Open Data concerns and barriers, that could ultimately help mitigate any of these problems including those faced by private entities.

The remainder of this thesis is as follows: The exact research question will be stated in the following subsection. This is followed by presenting the research approach. The first findings are reported in the second chapter, where an overview of the barriers will be presented. The third chapter is concerned with the design of a survey, that was made based on the first findings of this study. The paper then concludes with the results of the survey and a discussion on mitigation strategies and possibilities for further research.

1.1 Research question

The aim of this thesis is to examine existing barriers in using and publishing Open Data and consequently asses their importance and relevancy using quantitative methods. The first research question can therefore be summarized as follows:

"What are the barriers in using and publishing Open Data?"

In order to address the research gap left by previous studies, which have already researched the barriers, using empirical research, this study seeks to obtain data which will help to answer the second research question of this paper which is:

"How important and relevant are these barriers?"

The answer to this question is essential in order to prioritize and possibly dispute the previously reported barriers.

1.2 Research approach

It was decided that the best method to adopt for this research was to obtain the information from various sources. Firstly, a systematic literature review was conducted in order to establish a solid theoretical background on the subject. Additionally, qualitative methods in the form of semi-structured interviews were used to allow a deeper insight into the barriers. These methods helped with the design of a survey, which was then sent out to a large number

of participants. Our expectation was that such quantitative measures would usefully supplement the study, by evaluating the gathered information. In summary the research approach consisted of following three parts:

- Literature review
- Semi-structured interviews
- Survey and quantitative evaluation

1.2.1 Literature review

A systematic review of the literature was the first step of this study. For this purpose, information was collected from various journals, conference papers and books. Additionally, documents on Open Data published by the European Commission including documents from the Share-PSI network were reviewed.

Open Data is closely related to a much more mature open concept, namely the Free Open Source Software (FOSS). Indeed, the criteria defined by the Open Knowledge Foundation bear many similarities to those of FOSS. Although there might be slight differences in the definition of Open Source Software based on the license used, the Free Software Foundation recognizes following criteria that need to be met [23]:

- Freedom to run the program for any purpose
- Freedom to study how the program works, and change it as you wish
- Freedom to redistribute copies
- Freedom to redistribute copies of the modified versions

Overall, these definitions imply, that there is indeed a close relation between the concept of Open Data and Open Source Software, even though one deals with data and the other one with code. Surprisingly, despite the close relation of these concepts, the study of this relation has received very little attention in the research literature so far. A recent paper, which examined the similarities and differences between these concepts, suggests that open source research and practice can be used to understand Open Data [17]. Therefore, in this thesis the barriers from Open Source Software literature will also be taken into consideration when creating the list of barriers. We will not only expand the list of barriers, but expect to discover that some of the barriers present in Open Source Software are also applicable to Open Data. This would

immensely help further studies on how to mitigate Open Data barriers, since the knowledge and experiences from Open Source Software research could potentially be used to support such future studies.

1.2.2 Interviews

With the literature review completed, the next stage of this study was to conduct interviews with participants who already have some experience in the field of Open Data. This was done in order to confirm the barriers found in the literature and to find new ones, which have previously not been mentioned in any research literature. Since the knowledge and experience in the field of Open Data varied between the participants, we decided that a semi-structured interview with many open questions was the best approach. Ultimately, four participants agreed to an interview. These interviews were conducted in the 4th quarter of 2015. The following participants were interviewed:

- Sonja Fischbauer, (OKFN Austria)
- Peter Hanecak, (COMSODE, opendata.sk[7])
- Peter Parycek, (Head of Department for E-Governance and Administration, Danube University Krems)
- Martin Turcek (Transparency international Slovakia)

The duration of the interviews varied per person, but with the exception of one case, did not exceed an hour. At the beginning of each interview, the participants were asked what their experience with Open Data was. Subsequently, some of the barriers found in the literature review were discussed, in order to determine the importance of these barriers. The participants were also proposed to talk openly about any barriers they themselves encountered during their use of Open Data. By letting the participants talk freely, we aimed at encouraging that barriers that have not yet previously been discovered would be mentioned. Lastly, the future and potentials of Open Data were discussed as well, so that additional knowledge on the challenges faced by Open Data could be obtained. Finally, the interviews were then transcribed and all mentions of Open Data barriers were documented.

One particular limitation of this study that should be pointed out is the number of interview participants. Further qualitative data collection would

[7]http://opendata.sk/liferay/web/opendata

be beneficial to further investigate potential barriers, which are currently not documented in the research literature. This limitation however, may be resolved in future revisions of this study, since many additional subjects declared their willingness to participate in an interview in the survey.

1.2.3 Survey

The final stage of the study was to design and launch an online survey. Based on the information obtained from the literature review and the interviews, a list of barriers was created. Each barrier from the list is then implemented in the survey questions and evaluated using a Likert scale. Detailed information about the survey will be further discussed in the third chapter of this paper.

2 Barriers

Having discussed the research approach of this paper, what follows is an overview of the barriers, which were reported in the research literature and obtained from the interviews. Initially, nearly a hundred potential barriers were found and documented. That list needed to be reduced before continuing with the research. The primary aim of this research was to further evaluate the reported barriers in an online survey. Obviously, designing a questionnaire, which would have measured the significance of all nearly a hundred potential Open Data barriers would not be feasible. Additionally, the survey would then take too much time to answer, which would most likely result in a lower response rate, rendering the collected quantitative data statistically not significant and therefore unusable. The barriers, that ultimately ended up in the final list are those, which have been mentioned the most in the literature and the interviews. The barriers which were not included in

The overview of the barriers has been divided into three main parts, based on whether they apply to users, providers, or both. Such distinction has already been made in previous studies: Janssen et al. found that barriers are related to either data providers (resulting in not wishing to publicize data) or data users (resulting in an inability to use the data in an easy manner) [13]. Each barrier has then been assigned to one of several categories. This may however, be inconsistent with other studies, since not all literature provide a clear categorization of the barriers and those who do often categorize the same barrier under a category with a different name. In conclusion, the barriers have been categorized in following way:

- User specific
  - Open Data portals
  - Data Quality
  - User legal constraints
- Provider specific
  - Strategic and business
  - Privacy and security
  - Provider legal constraints
- Both user and provider
  - Knowledge and experience

Each category begins with the explanation of the barriers. An overview is then documented at the end of each category in the form of a table. The table summarizes the barriers found for each category and also cites the literature source, indicates whether the barrier was also mentioned during the interviews and how many participants confirmed it. Finally the "Source" column specifies if the same barrier also exists in the Open Source Software research literature as well.

2.1 Open Data portals

It is essential, that Open Data portals fulfill certain quality standards and requirements in order to meet the demand and expectations from potential users and providers. However, there are currently no unified standards and quality enforcement rules that would ensure a certain standard of quality. This leads to various data portals providing different experiences for the users. Indeed, in their analysis of the quality of open government data portals, Martin et al. concluded that there is no single model of Open Data portals. It is rather a wide range of structures, with different processes for data search, different formats for presenting information, and a diversity of means to classify the information [2]. The heterogeneity of the portals is problematic mainly when the implemented features limit the user in some way. A comprehensive study on adoption barriers of Open Data by Janssen et al. reports several issues ranging from those that only make the experience worse, such as no information about the datasets provided by the portal, to those that heavily restrict the user. It is reported that some portals do not

allow access to data unless the user registers on their website or even worse, put the data behind a paywall [13]. Such measures go against the very principles of Open Data as defined by the Open Knowledge Foundation.[8]

However, downloading the data from the portal website directly is not the only way users can retrieve the datasets. Many users prefer the use of an API,[9] as the existence of such interface allows them to include the data directly into their application without having to first download them manually. In another major study on the limits of open data platforms, Braunschweig et al. described the existence of APIs as an obvious feature of every Open Data platform, necessary for automation [5]. Despite that, in their analysis of over fifty Open Data portals, they have found that 43% of them do not feature an API of any kind. The situation has possibly improved in the recent years, as quasi standardized platforms such as CKAN,[10] which provide an API by default unless it is disabled on purpose are becoming more prevalent. According to a portal monitoring site Open Data Portal Watch,[11] CKAN is currently the most widely used Open Data platform [22]. The existence of an API however is often not enough. The APIs must meet certain criteria to be deemed usable. To determine the current state of APIs, this study set out to benchmark all governmental Open Data portals in the European Union. Community and other non-governmental portals were ignored for this study, as the number of them grows rapidly each month, and many are run by hobbyists, who do not have the resources to provide the users with a stable and responsive server, that is able to stay online with minimal down-times.

2.1.1 API benchmark

To begin the process, a list of all available governmental portals in the EU was assembled. The second step was the filter out portals that do not meet the following two criteria: Firstly, the portal has to be associated with government and secondly, an API must be available without any heavy restrictions, such as registration required to access the API, or only limited number of queries possible. Out of the 28 member states, 23 have some form of official governmental Open Data portal. However, only 17 provide an API. Additional two portals were ruled out due to limitations of their APIs. While

[8] http://opendatahandbook.org/guide/en/what-is-open-data/

[9] http://blog.pieter.pm/2016/04/i-do-not-want-your-open-data-api-id-rather-scrape-your-website/

[10] http://ckan.org

[11] http://data.wu.ac.at/portalwatch

these numbers already look slightly better than those, reported by the Braunschweig et al. study [5], they are still somewhat disappointing as an API is a rather important feature and with the rise of standardized platforms such as CKAN, it should no longer be an issue to implement one. Many of the excluded portals however, do have their own custom website, and their own systems, making the use of these portals rather difficult. Table 2 provides an overview of the 15 portals, which were ultimately tested.

The benchmark queries various Open Data portals and prints the status of the APIs to a CSV file. The entire script is written in bash and makes use of basic GNU utilities, which are for the most part included in most Unix and Unix-like distributions with the exception of 'jq',[12] a command line utility to transform and process JSON documents. The first step[13] is to check whether the host of the API is online at all. This was initially done by pinging the host to check if it responds, however many hosts are either blocking parts of the ICM protocol immediately, or the requests are banned after several consecutive requests, which renders the ping tool unusable. This may have been done in order to prevent certain denial of service attacks which exploit this protocol. This issue was solved by using the 'netcat' tool to determine if the host responds to an http port 80 request. If the host successfully replies to the request, it can be assumed that the host and the website are online. The 'netcat' tool has various exit codes after it processes the request. It exits with a '0' if the connection succeeded. Otherwise it may respond with an exit code with a value of either '1' or '2'. Unless the exit code is '0', the exit messages are discarded to /dev/null to prevent writing log messages into the CSV file. The host is given 30 seconds to respond to the initial request. If the host takes more than 30 seconds to reply, it can be assumed that it is either offline, or too slow to be deemed usable.

Listing 1: Listing 1

```
1 #!/bin/bash
2 checkApiNc (){
3 IP=$1
4 nc -z -w30 $IP 80 2>/dev/null 1>/dev/null
```

In case the host is online and responds in a timely manner, the second step is to test the API by making a simple request. In this particular case, a list of all package names was requested from the APIs. The API is given 20 seconds window to process the query. The data is then processed and transformed

[12] https://stedolan.github.io/jq/

[13] Listing 1

Table 2: Open Data portals overview

Country	Portal	API available	Comments	Included
Austria	data.gv.at	Yes		Yes
Belgium	data.gov.be	No	The website itself is not a standard Open Data portal. It only serves as an index linking to other portals in Belgium some of which do not have APIs.	No
Bulgaria	opendata.government.bg	Yes		Yes
Croatia	data.gov.hr	Yes		Yes
Cyprus	data.gov.cy	No	Many datasets are not in machine readable format.	No
Czech Republic	opendata.praha.eu	Yes	Portal for Prague only, not the whole country.	Yes
Denmark	portal.opendata.dk	Yes		Yes
Estonia	opendata.riik.ee	Yes		Yes
Finland	avoindata,fi	Yes		Yes
France	data.gouv.fr	Yes	API is somewhat limited. The maximum number of datasets that can be queried in one request is limited.	Yes
Germany	govdata.de	Yes		Yes
Greece	data.gov.gr	Yes		Yes
Hungary	N/A	No		No
Ireland	data.gov.ie	Yes		Yes
Italy	dati.gov.it	Yes	API is limited. It is not possible to query all datasets. A "?limit=value" suffix needs to be included, otherwise the API won't respond.	No
Latvia	N/A	No		No
Lithuania	opendata.gov.lt	No	Only an index website linking to various resources. No API provided. The data is often not in machine readable format.	No
Luxembourg	N/A	No		No
Malta	data.gov.mt	No	No data portal, Only a website linking to various resources, some of which are not in machine readable format.	No
Netherlands	data.overheid.nl	Yes		Yes
Poland	N/A	No		No
Portugal	dados.gov.pt	Yes	Using their own, non-standardized version of API, which doesn't respond to the usual requests. Guide for the API was only available in Portuguese.	No
Slovakia	data.gov.sk	Yes		Yes
Slovenia	data.gov.si	No	Data is in proprietary and not machine readable formats such as pdf and docx. files.	No
Spain	datos.gob.es	No		No
Sweden	xn–ppnadata-m4a.se	Yes		Yes
United Kingdom	data.gov.uk	Yes		Yes
SUM	23/28	17/23		15/17

using the 'jq' JSON processor. This prints all the package names separated with a comma. Using the 'awk' tool, it is then possible to determine the exact number of the packages by counting the objects separated by a comma. In case the API query is not processed in the given time, the number of packages in the awk count will be '0'. Furthermore, if the API responds with an error message, it is possible that the awk tool counts some objects, as the error message may obtain several commas. It can be therefore assumed that unless the value of the processed query is less than 10, the API did not respond in time or did not return the required data. If the number of returned packages is more than 10, and the API responds in less than 20 seconds, it is assumed that the API is working correctly.

Listing 2: Listing 2

```
5  if [ "$?" = 0 ]
6    then
7    API=$(curl -s -k --max-time 20 $2 | jq '.result')
8    Print=$(echo $API | awk -F, '{print NF}')
9    if [ "$Print" -lt 10 ]
10         then
11         echo "API not responding in time, or did not
               return the required data"
12   else
13   echo "API_OK"
14   fi
15 else
16   API="Host offline"
17   Print=$(echo $API)
18   echo $Print
19 fi
20 }
```

The function is executed by passing two variables into it. The first variable is the host such as 'data.gv.at'. The second variable is the actual API request, for instance 'hostname/api/3/action/package_list'

```
21 checkApiNc $1 $2
```

It is convenient to store the retrieved data to a csv file. The initial script prints one line for each API, essentially making a column. Each new line can be translated to a comma separated value by invoking the 'tr' command. Each time the script is executed, the data is then appended to an already existing file.

Listing 3: Listing 1

```
#!/bin/bash
./portalsAPI.sh | tr '\n' ',' >> data.csv
echo
```

The APIs were benchmarked 8 times a day for the duration of 31 days, amounting to 248 requests for each portal. The intention of this test was not to stress the servers with constant requests, but to merely check their availability from time to time. However, there was a possibility that the servers would block the source of the requests, since the script sent out the same request constantly for an entire month. To prevent being blocked, the script was launched from multiple sources with different locations.

Interestingly, all of the APIs achieved nearly 100% success rate in processing the query. The failure rate percentage is in fact so low, that the very few instances where the query was not processed in a timely manner, could be attributed to a failure on the client side. In conclusion, these results suggest that if the data portal provides an API, it will be functional and able to respond in time. It could be argued that the 20 seconds in the benchmark is still too long. To address this issue, a second, albeit shorter benchmark was carried out, where the time the API and host have to respond has been reduced to 10 seconds. This implementation of the script was in operation for 14 days with 112 query requests sent to each portal respectively during this time period. The results were almost identical to those of the first pass of the benchmark with only minor, statistically not relevant differences. The results of the benchmark are presented in Table 3.

Table 3: API Benchmark results

Portal	Offline (%)	API Timeout/ Incorrect data (%)	API OK (%)
data.gv.at	0	1.2	98.8
opendata. government.bg	0	0	100
govdata.de	0	1.6	98.4
opendata.dk	0.4	1.2	98.4
opendata.riik.ee	0.4	0	99.6
opendata.praha.eu	0	0	100
avoindata.fi	0	2.4	97.6
data.gov.gr	0	0.4	99.6
data.gov.hr	0	0.4	99.6
data.gov.ie	0	0.4	99.6
data.overheid.nl	0	1.6	98.4
data.gov.ro	0	0	100
xn–ppnadata-m4a.se	0	0	100
data.gov.sk	0	0.4	99.6
data.gov.uk	0	2.4	97.6

The speed and availability of an API however, are only one of the many quality factors that an API needs to have. An interview, which was carried out as a part of an extensive study on socio-technical impediments of Open

Data by Zuiderwijk et al., reports the lack of a good API as a barrier [28]. It is important to bear in mind that the aforementioned study was carried out in 2012 and Open Data platforms such as CKAN or Socrata were not as prevalent as they are today. The reported lack of a good API could therefore also refer to the general lack of APIs available on Open Data portals in 2012, which has also been reported in the Braunschweig et al. study [5]. Nevertheless, the study implies, that APIs may have some limitations. To further investigate this matter, questions about the quality of APIs are also included in the questionnaire of the present study.

2.1.2 Summary

So far, the research has identified that there are indeed several issues with Open Data portals. A summary on all data portal barriers that will further be investigated in the survey is shown in Table 4.

Table 4: Open Data portals barriers

Barrier	References	Interview	Source
No API provided	[28], [5]	N/A	OD
API slow or unresponsive	[28]	N/A	OD
API limitations	[28], [5]	N/A	OD
Registration required before gaining access to data	[13]	N/A	OD
Difficult browsing/searching	[13]	1/4	OD
No information about the quality of the data	[13][7]	N/A	OD
No information given about the content of the dataset	[13]	1/4	OD
Language barriers	[9], [6], [2]	2/4	OD
Data is not available for download	[28], [5]	N/A	OD
Duplicate datasets	[13]	N/A	OD
Only non-value-adding data published	[13]	1/4	OD
No central portal for the data	[13], [9], [15]	2/4	OD
Restricted access	[28]	N/A	OD
Data is behind a paywall	[27], [13], [28]	N/A	OD

2.2 Data quality

Once the data is retrieved it is often found that the quality of the data is rather low. Perhaps the biggest barrier users are faced with when they download the data is that the data is not in a machine-readable format. Scanned documents in the form of a .pdf file are a common example of this particular barrier. While it could be argued that the data is technically open, the use

and reuse of such data is extremely difficult. Braunschweig et al. claim that while some might prefer to work with data in a human-readable form, such form does not allow the reuse of the underlying raw data. Furthermore, the process of transforming the data into a machine-readable form can be very challenging [5]. In the same vein, Alani et al. note that publishing data that lacks structure and semantic representation, thus making it almost impossible for machines to understand the data, inhibits its reuse [1]. Furthermore, a major report was published by the McKinsey Group in 2013. The study researched Open Data potentials, barriers and use cases across seven domains, ranging from public sector such as education, health care to institutions and consumers in private business enterprise. During their investigation of the oil and gas domain, they have found that even when data is made available, it might not be in a machine readable form [18].

Even if the data can be read and processed by a machine, users often find that the data is obsolete, inaccurate or incomplete. However, the interviews conducted for this study have shown that there are conflicting views on whether such data should be published or not. While some claimed that low quality data is a hassle and only discourages potential users and developers, others claimed that the more published data, the better, regardless of the quality of the datasets. Enforcing standards and rules could lead to less data being published. During one particular interview, a comparison to the open source code publishing platform GitHub was made. Similarly, only a small fraction of the code published is usable and would be considered high quality code, but the developers are able to find high quality projects, use the code and contribute to further development of the project. Yet, if certain quality standards were to be enforced, it could have prevented many projects from existing.

Metadata is data, used to describe the actual dataset. Unfortunately, they are often inaccurate or missing completely. Lack of agreed metadata standards is often regarded as the reason for the impracticability of metadata. A report by the European Commission states that commonly agreed metadata could considerably enhance the value for reuse of the information [6].

The summary of barriers, which were also chosen for further research by the survey are displayed in Table 5.

Table 5: Data Quality barriers

Barrier	References	Interview	Source
Data inaccuracy	[13], [28], [15]	N/A	OD
Obsolete, non-valid data	[13]	N/A	OD
Encoding problems (e.g. not UTF-8 complaint)	N/A	2/4	OD
Not machine readable	[28], [5], [18]	1/4	OD
Metadata inaccuracy	[28], [18]	1/4	OD
Data incomplete	[13], [28]	N/A	OD
No potential use of the data	[9],[13]	1/4	OD

2.3 Legal constraints

Users are often faced with various legal constraints imposed on the datasets. It might seem that by definition, understanding the legal position of the user would be simple, as Open Data is defined by the fact that the users are able to freely access, reuse and further redistribute the data. Unfortunately, this does not always reflect the real world, as many datasets are published under restrictive licenses. Furthermore, complex and hard to understand licenses make the user question the legal aspects of use and reuse of the data. Once again a comparison to Open Source software can be made here. Complex licensing situation is nothing new in the field of Open Source software. In fact, as of now the Open Source Initiative[14] currently lists 78 different licenses. The complex licensing situation is also discussed by Stol and Ali Babar in their literature review on challenges of using Open Source software. They have found multiple sources that confirm that the interpretation of these licenses is rather challenging [25]. Restrictive and complex licenses are only one part of the problem. There are cases where the published datasets do not clearly indicate the license under which they are published. This barrier was also confirmed in an interview. Both Open Source and Open Data face this issue, where code or data is often published without a license, making it unclear whether the data or code can be freely reused and further redistributed. Finally, the users often fear legal consequences resulting in a lawsuit. Talking about this issue an interviewee commented that even if the probability of winning such legal case is high, many Open Data users are independent developers or start-ups, who do not have the means and resources to fight a legal battle. To make Open Data attractive for newcomers, the legal complexity needs to be reduced. Effort should be made to always publish data

[14]https://opensource.org/licenses/alphabetical

with license included, which clearly defines the rights of the user. Table 6 below summarizes the findings on this topic.

Table 6: User legal constraints

Barrier	References	Interview	Source
Threat of lawsuits	[13], [18]	1/4	OD
Restrictive licenses	[9], [28], [15]	2/4	OD
Unclear licensing (Unclear conditions for reuse)	[15], [9], [28]	2/4	OD
Complex, hard to understand licenses	[9], [28], [25]	2/4	OD, OSS

2.4 Privacy and security

When publishing the data, providers must make sure that they will not publish any sensitive private information. Before the data is published, necessary precautions need to be made in order to ensure privacy [10]. There are also cases where disclosing information might harm the reputation of the provider. The McKinsey report found that school administrators often wish to prevent disclosure of data about poor student performance [18]. Furthermore, providers often fear that false conclusions may be drawn from the data. This topic is discussed in the Conradie and Choenni study. As a possible example they provide a situation where property value decreases if details about policy plans of new city developments would surface, especially if these were not yet finalized [8]. Additionally, by releasing the data for further reuse and redistribution, providers risk that third parties will edit the dataset and intentionally distort the data in order to gain advantage or cause harm to someone's reputation [10].

Ensuring privacy however, takes a lot of effort that many of potential providers are not willing to undertake. Overall, these barriers are summarized in Table 7:

Table 7: Privacy and security barriers

Barrier	References	Interview	Source
Privacy - Unwelcomed exposure of the data	[13], [9], [6] [15], [3], [18]	1/4	OD
Loss of control over released information	[15], [3]	N/A	OD
Fear of false conclusions drawn from the data - Data misrepresentation	[15], [10] [8], [18]	2/4	OD
Open Data may lead to corruption or falsification of the data	[3], [10]	N/A	OD
Security threats, vulnerabilities	[13]	N/A	OD
Low quality data, better not expose it	[3], [15]	1/4	OD

2.5 Strategic and business

Strategic and business decisions are often the reason why many potential providers do not publish their data. During an interview it was suggested that Open Data is simply not a priority for many companies and generates unnecessary costs with no guaranteed value in return. Additionally, some business models are dependent on keeping the data private, since opening the data would disrupt their current business strategy, which is focused on generating sales from the data they own. In the same way, many software developers do not open source their code in fear of lost sales. In their study, Morgan and Finnegan describe giving away the source code as a value-impeding characteristic. For some companies, source code enables value capture if it used for products, which they can sell [21]. Likewise, some potential data providers might use the data to develop products of their own, which they then put up for sale, thus generating revenue. Releasing the data would allow third parties to profit from the data, potentially disrupting the sales of the provider's product. Furthermore, the general lack of time, resources or simply additional expenditure are a major barrier in publishing Open Data. The costs do not end with the data being published. Similar to code, data needs to be maintained and kept up to date in order to satisfy the demand from the users. In summary, it can be argued that in some cases, Open Data requires complex strategical and business planning to determine, whether the provider gains any added value from opening the datasets, or rather makes losses. The strategical and business barriers are presented in Table 9.

Table 8: Strategic and business barriers

Barrier	References	Interview	Source
Lack of business models	[28], [25]	N/A	OD, OSS
Open Data is not a priority	[8]	1/4	OD
Resistance to change, risk-averse culture	[15], [13]	1/4	OD
No sale of data possible when an Open Data license is used	[13]	1/4	OD
Lack of strategic planning and management	[11]	N/A	OSS
Disruption of existing business model. (Such as charging money for the data)	[15], [13] [21]	N/A	OD, OSS
Uncertain economic impact	[15]	1/4	OD
Lack of resources and time	[13], [25] [11]	N/A	OD, OSS
Cost issues	[3], [25] [21], [18]	1/4	OD, OSS

2.6 Provider legal constraints

The complex licensing situation that was already discussed in chapter 2.4 applies to providers as well. Additionally, providers also face the threat of lawsuits, albeit not because of violations of the licenses, but usually when they release private or otherwise sensitive data. There is also the issue of data ownership. If the provider is not sure about the ownership of the data, they are not able to release them. In their study on barriers for local governments releasing Open Data, Conradie and Choenni argue that due to the vertical data management in the past, data sharing between departments has not always been the case. As a result departments lack a complete picture about data ownership, thus inhibiting its release [8]. The summary is displayed in Table 9 below.

Table 9: Legal barriers

Barrier	References	Interview	Source
Unclear ownership of the data, prohibiting its release	[8]	N/A	OD
Threat of lawsuits	[18]	1/4	OD
Complex, hard to understand licenses	[28]	1/4	OD

2.7 Technical

There are several barriers that might make working with data difficult for providers. As there is currently no standard, it is often expected from the provider to release data in various formats. According to one interviewee however, setting a file format standard is not feasible, since different types of data require different file formats. They further argued that not everything can be stored in a CSV file. Janssen et al. also report that having no standard software for data processing as a technical barrier [13]. The last reported barrier is the absence of metadata standards. The McKinsey report suggests that non-profit organisations could serve as a neutral organizing force to align international data standards, such as standardizing formats for metadata [18]. Some efforts towards standardizing are already in progress. The RDF specification, recommended by the World Wide Web Consortium[15] for instance, is an example of such effort.

Table 10: Technical barriers

Barrier	References	Interview	Source
Various standards and forms of machine readable data (JSON, XML, CSV ...)	N/A	1/4	OD
No standard software for processing the data (Fragmentation of software)	[13]	N/A	OD
Absence of metadata standards	[18]	N/A	OD

2.8 Knowledge and experience

Even though Open Data might be a simple concept, the actual use and publishing of the data is often difficult, especially for newcomers in the area. Open Source Software faces a similar issue. In his study on Open Source software rejections in Australia's top firms, Sigi Goode found that companies often prefer commercial, proprietary versions of software, because companies offer proper support for the software. He states that managers appeared concerned that if no equivalent to commercial support existed, they would risk having to support their software and applications with their own resources [11]. This could very well be an issue in Open Data as well. The providers usually do not provide any support on how to use their data. This is usually left for the user to figure out on their own. And while it can't be expected of providers to additionally offer a helpdesk or give them support,

[15]https://www.w3.org/RDF/

especially considering the data was free of charge, simple instructions and guidelines would be a first step in mitigating these barriers, but as of now there is a clear lack of available support.

Providers on the other hand often find that publishing the data is a rather complex task. Indeed, in their study from 2012, Zuiderwijk et al. state that opening up the data is a complex and ill-understood activity, because many barriers counteract these processes [29]. Guidelines and instructions published on data portals could help potential providers with the process. Such instructions would most likely also increase the quality of published data.

Table 11: Knowledge and experience barriers

Barrier	References	Interview	Source
No support	[13], [28] [25], [11] [21]	1/4	OD, OSS
No helpdesk	[13], [28] [25]	1/4	OD, OSS
Lack of documentation	[25], [24] [21]	1/4	OSS
No guiding principles/instructions	[29]	1/4	OD

2.9 Additional barriers

As previously mentioned, many documented barriers were not chosen for further investigation in the online survey. This was done mostly to make sure that the survey could be answered in a short time, thus increasing the response rate, but there were other reasons why some of the potential barriers were left out. Not all of the barriers can be easily evaluated using a 5-point Likert-Scale and are better suited for qualitative research, for instance in the form of an interview. Some barriers that are faced by newcomers with no previous experience with Open Data have been left out as well. The main distribution channels for the survey were mostly academic and Open Data communities, where some previous knowledge is expected. Asking such participants questions intended for newcomers would only provide us with biased answers. It should be noted, that the literature source of barriers presented in this chapter is exclusively Open Data literature.

2.9.1 Data portals - additional barriers

Returning briefly to Table 4, we have identified 15 barriers with Open Data portals. However, some additional barriers were discussed by the Martin et al. study from 2015. In their analysis, they found that 39% of the portals they have analyzed, do not support online viewing of the datasets. Perhaps even more concerning is the fact that 6% of the portals conversely only offer an online view of the data with no download option available. Lastly, they disovered that on some portals, it is not possible for external users to upload data [2]. Another noteworthy issue is the lack of communication between the data provider and the data user. In a study which set out to determine the risks of Open Data, Martin et al. claim that the relationship between the providers and user lacks communication [19]. This is consistent with the Zuiderwijk et al. study. The interviews and workshops they have conducted for their study both support the claim that no discussion is possible between the data provider and data user [28]. Even when an attempt at communication made, Janssen et al. assert that public organizations do not react to user input [13]. All documented data portal barriers, which were not included in the questionnaire are summarized in Table 12.

Table 12: Data portals - additional barriers

Barrier	References
Only online viewing possible	[2]
No online viewing possible	[2]
Heterogeneity of OD portals, no standardization	[2]
No possibility for the external users to upload data	[2]
No advanced search possibility	[28]
Lack of communication between data providers and data users	[28], [19]
Public organizations do not react to user input	[13]
Not all desired formats are available	[28]
Similar data stored in different systems yields different results	[13]
Dead links	[5]

2.9.2 Knowledge and experience - additional barriers

The number of questions about knowledge and experience in the survey is relatively limited in order to avoid biased answers. However, this does not mean that these barriers are any less significant. A report by the European Commission argues that the lack of awareness of the potential of Open Data

by administrations and businesses is holding back the development of a true market for the re-use of public data [6]. Janssen et al. suggest that potential data users also lack knowledge to use the data. The reason might be that many users do not posses the statistical knowledge necessary in order to make use of the data. Another possible explanation is that the data formats and datasets are too complex to handle and use easily [13]. While researching the potential of Open Data in education, the McKinsey study found that educational institutions often lack the skills necessary to incorporate data into their everyday workflows [18]. The rest of the barriers is summarized in Table 13.

Table 13: Knowledge and experience - additional barriers

Barrier	References
Lack of knowledge to use the data	[13], [18]
No statistical knowledge	[13], [18]
Data formats and datasets are too complex to handle and use easily	[13]
Lack of awareness about Open Data	[6]

2.9.3 Technical - additional barriers

Potential publishers often face difficulties when they have too much data. In their article, Geiger and Lucke claim that the preparation, maintenance and publishing of machine-readable data are often made difficult when there is too much information too much data to process, described as flood of information [10]. Having too much data also introduces increased costs and difficulties in storing the data and providing the necessary bandwidth in order to ensure data availability [9]. Too much data available, however, is not a barrier exclusive to publishers. In their comprehensive study on transparency and access to government information, Jaeger and Bertot argue that the volume of information and data being made available may paradoxically limit transparency by making it difficult for users to find content [12]. A summary of the barriers is presented in Table 14.

Table 14: Technical - additional barriers

Barrier	References
Flood of information to process	[10]
Storing the data	[9]
Volume of information may limit transparency	[12]

2.9.4 Data quality - additional barriers

Some issues were raised regarding data, which is being published by the government. In their risk analysis of Open Data, Martin et al. claim that the dependence of data producers on State Funding raises suspicions on the accuracy of the data. Some data can be sensitive to political pressure and therefore subject to manipulation by the state [19]. Furthermore, lack of interoperability when combining the data from various sources was also described in the report by the European Commission. The report suggests that the availability of machine-readable data with a thin layer of commonly agreed metadata could facilitate data cross-reference and interoperability and thus increasing the reuse value [6]. This barrier is also reported in the Open Data Manual, published by the Open Knowledge Foundation in which they claim that interoperability is particularly difficult when encountering non-open systems and suggest the use of vendor-neutral open systems [9]. The summary of the aforementioned and any other documented data quality barriers not included in the questionnaire is summarized in Table 15.

Table 15: Data quality - additional barriers

Barrier	References
Data is not understandable to public	[28]
Essential information is missing	[13]
Data bias	[19]
Data is not being updated	[28]
Lack of interoperability	[6], [9]

2.9.5 Other barriers

Some of the barriers, which have been documented do not fit a specific category. These barriers are presented in this chapter. Users might be unwilling to use and participate in Open Data simply because they do not see any potential or added-value in Open Data. There is no incentive for the users to use the data [13]. Providers on the other hand, often do not understand the potential added value of sharing the data. Companies are concerned about helping competitors or exposing weaknesses and are not yet convinced of the value to be gained in sharing data with consumers [18]. Lastly, there are still different legal frameworks, even within the European Union, which are making the international reuse of data difficult for the users and hinders the creation of cross-border information services [6]. These barriers are summarized in Table 16.

Table 16: My caption

Barrier	References
No incentives for the users	[13]
Lack of understanding about the added value obtained from sharing the data	[18]
Inconsistent international legal framework	[6], [9], [10]

2.10 Summary

So far, this thesis has identified a number of barriers reported in the literature, some of which were further supported by the interviews. As we have observed from the summary tables for the barriers, most of the barriers have been not been confirmed by multiple interview participants. This should not be interpreted that most of the interviewees do not think that a particular issue does present a barrier. It simply means the specific barrier was not brought up during the interview. With each interviewee having experiences in different areas of Open Data and due to the interview being quite open as opposed to a structured interview, not all barriers could be discussed. A common view shared by the participants however was that Open Data has future potential yet to be achieved.

3 Survey

Based on the information about potential barriers obtained from the literature review and the interviews, an online survey was designed. The first step was to choose the right platform for the survey, which would support all necessary features. Primary inclusion criteria for the platform were:

- Wide variety of question types such as support for array-type questions, comment fields and multiple-choice questions.
- Support for modular design, allowing the setting of conditionals
- Real-time evaluation of the results with the possibility to export to file formats supported by statistical analysis software such as R or SPSS
- Multilingual support

After careful consideration of various online services, we concluded that the free versions are limited and restricted to basic features. Therefore, it was decided that the use of Limesurvey[16] would be the best option for this research.

[16]https://www.limesurvey.org/en/

Limesurvey is a free and open source survey web application written in PHP and distributed under the GNU General Public License. It can therefore be used for any purposes free of charge. It also supports nearly all of the needed features. Some missing features, such as setting default answers in array type questions and small interface modifications were later implemented using JavaScript. The survey was then installed and launched on an university web server.[17] This also ensured that the university had full control over the collected data with no third party having access to the results.

3.1 Demographics

It should be noted that although initially designed to be a multilingual international survey, this particular instance was limited to Austrian participants only. While it might seem like a disadvantage to restrict access to a certain set of participants and thus achieving a lower response rate, it proved to be quite helpful as the survey received feedback from the participants. This feedback will be used to develop a second, international survey, where the concerns of participants are being addressed, which we plan to conduct as future work.

3.2 Questions

The questionnaire was designed for both users and providers in mind. It has a modular structure and the participants are assigned to a group depending on their answers in the introductory set of questions. All participants must answer a mandatory question in order to determine whether they have experience as an user, a provider, or both. In addition, some basic information about the participant is asked in the introductory phase of the survey such as the kind of organization the participant is associated with, which categories are of interest for them and the general motivation for publishing or using Open Data.

In order to measure the severity and importance of the barriers, most of the questions are in form of a 5-point Likert scale. Furthermore, a "No answer" choice is also available, in case the participant does not have any experience with that particular barrier, is not sure, or chooses to not answer for any other reason. It should be taken into consideration that the graphs in our study do not record the "No answer" choice. The 5 points in the scale for the user are as follows:

[17] http://odsurvey.ai.wu.ac.at/OpenDataSurveyAustria

- 1 - Not a barrier.
- 2 - Somewhat of a barrier. (It was still possible to use the data)
- 3 - Moderate Barrier. (Made it difficult to use the data)
- 4 - Serious Barrier. (Made it extremely difficult to use the data)
- 5 - Extreme Barrier. (Completely prevented me from using the data)

Correspondingly, the 5 choices for the provider set of questions are:

- 1 - Not a barrier.
- 2 - Somewhat of a barrier. (Small barrier in publishing the data)
- 3 - Moderate Barrier. (Made it difficult to publish the data)
- 4 - Serious Barrier. (Made it very difficult to publish the data)
- 5 - Extreme Barrier. (Completely hindered the publishing of the data)

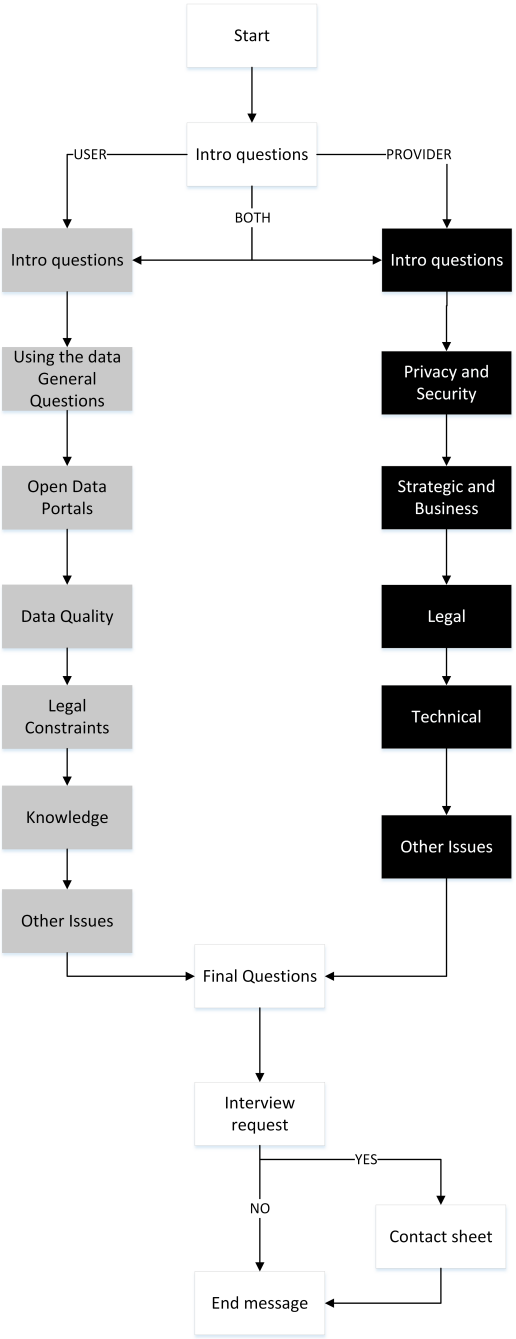
Additionally, each question category ends with an open question, where the participants are free to comment on any barriers they have encountered, which were not mentioned in the questions. By providing the participants with an option to write additional comments, new barriers that have previously not been mentioned in the research literature can be discovered. Any new barriers that have actually been discovered will be documented in the summary of the survey evaluation in chapter 4.

After the participants completed all questions regarding the barriers in either using or publishing Open Data, they were asked a set of final questions to determine their general interest and stance on Open Data. Lastly, an interview request sheet is implemented just before the end of the survey, where each participant is asked whether they would be interested in further discussing this topic in an interview. If the answer is positive, the participants are then asked to fill out a contact sheet with personal information, so that they may be later contacted for an interview. Until this point, no personal information was collected and the sheet can be left blank if the participant decides, that they would rather not disclose any personal information.

3.3 Structure

The general structure of the survey is illustrated in Figure 1 on the next page. The white rectangles depict those sets of questions that are always displayed, regardless whether the participant is an user or a provider. Grey and black rectangles represent questions for users and providers respectively. If the participant is both an user and a provider, the left branch with questions for users is executed first, followed by the right branch.

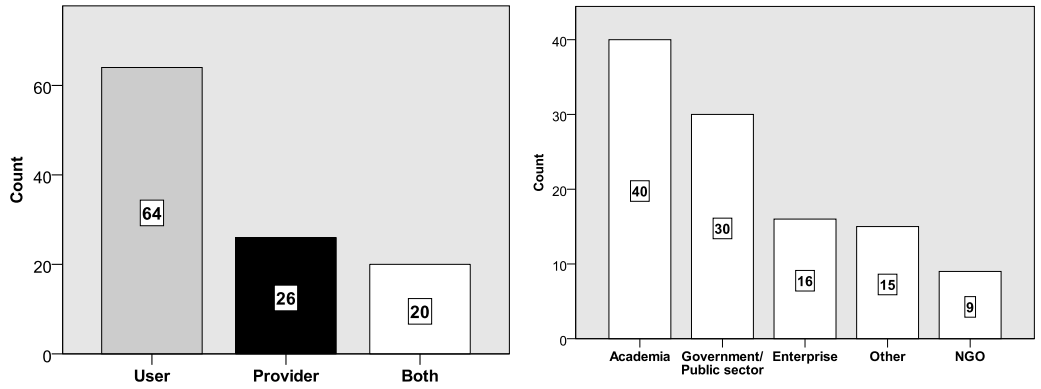
Figure 1: Modular design



4 Results of the survey

This part of the thesis discusses and analyses the results of the survey that was presented in the previous chapter. The survey was launched on the 23rd November 2015 and remained active until 17th March 2016. By the end of the period, 310 participants have launched the survey. However, more than a half only viewed the first page without answering a single question. From the remaining 149 respondents, 39 participants did not complete the questionnaire. In the analysis presented in this thesis, these responses were discarded. Finally, 110 individuals completed the entire survey.

Figure 2: Participants

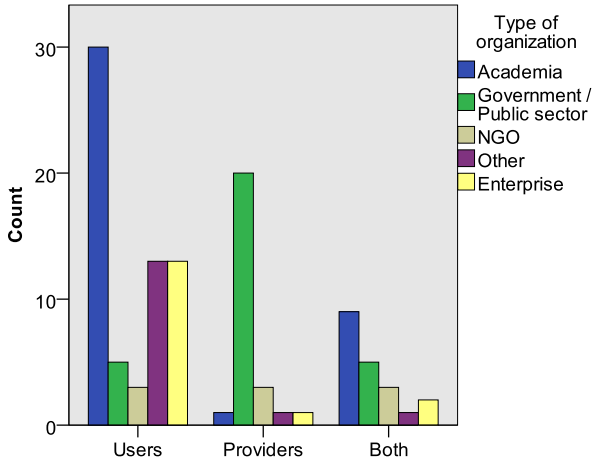


(a) The participant is

(b) Organization

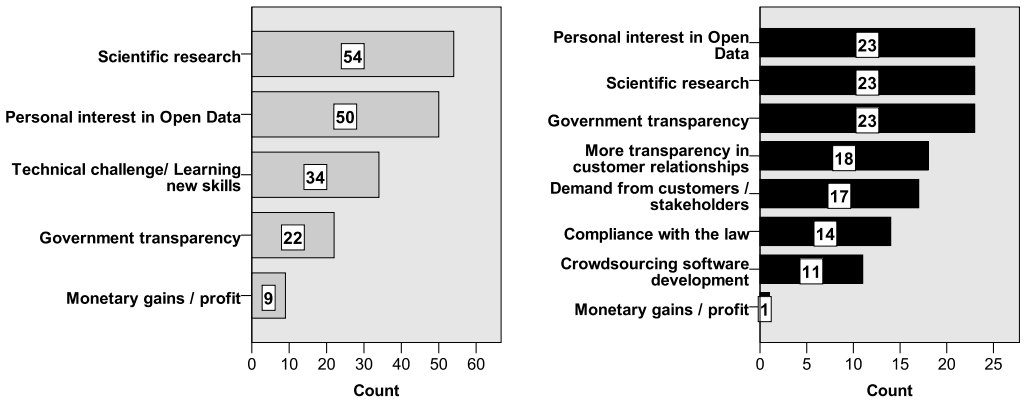
As the data in Figure 2a shows, the majority of respondents were users. The participants were also asked to specify the type of institution they are associated with. The most respondents work for public institutions such as academia or government. It could be argued, that the high number of participants from academia could be due to the way the survey was promoted. University employees and people in academia have helped to distribute the survey. It is therefore possible that the promotion of the survey was most effective among other academics. Enterprises are rather underrepresented in these results. This further supports the claim, that as of now, Open Data enjoys its popularity mainly in the public sector. To gain additional information about the participants these data were combined, to see which group has the most providers or users.

Figure 3: User is a / Type of organization



From the graph on the side, it can be observed that most of the users are from academia, while almost all providers of Open Data are from the government or public sector. A rather disappointing finding is that there are almost no providers from the private sector. In fact, only 3 participants, who are providers work in an enterprise. From the analysis of these data, combined with the fact that most users are associated with the public sector as well, it can be inferred that the private sector has yet to find its interest in Open Data. The respondents were also asked to state their motivation for the use or publishing of Open Data. The data gathered from these questions provide additional support for the claim that there is a general disinterest for Open Data in business.

Figure 4: Motivation



(a) User motivation

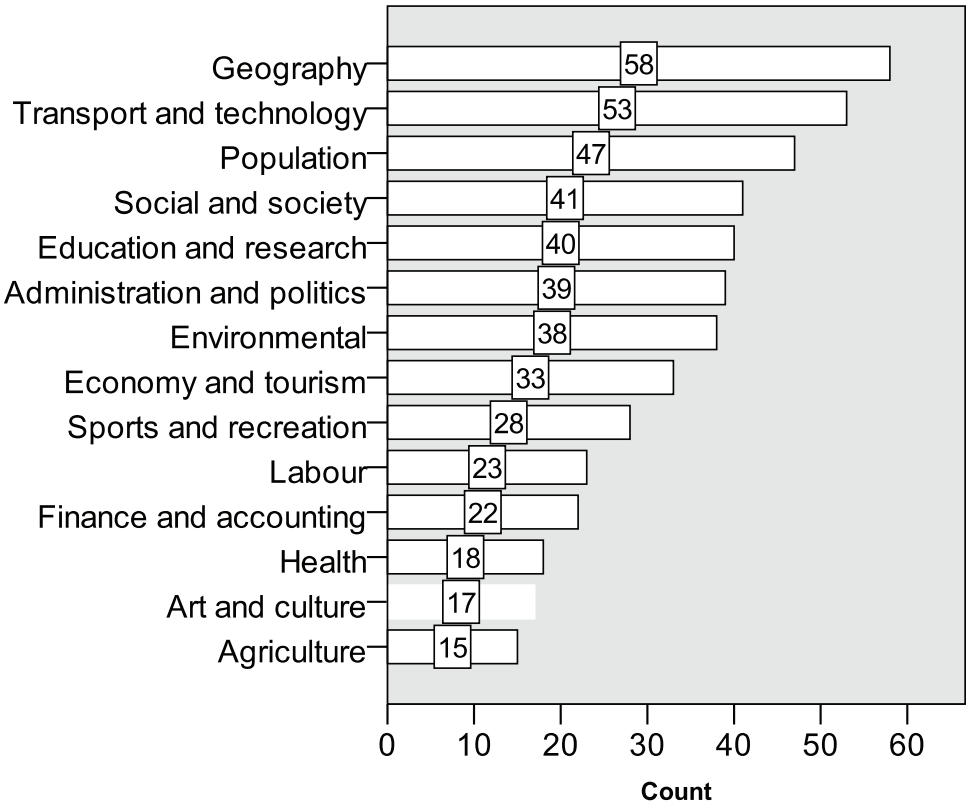
(b) Provider motivation

The graphs in Figure 4 illustrate that scientific research and personal interest in Open Data seems to be the main driving force of motivation for the use and publishing of data. Another noteworthy finding is that monetary gains and profit are least likely to be a motivation factor. There are several

possible explanations for this result. It is possible that the enterprise sector wants to achieve indirect added value from Open Data, such as increased customer satisfaction or crowdsourcing the development of their in-house software projects by allowing the public to maintain and enhance the data they use. Another possible explanation however could be that the business sector currently does not see the concept of Open Data as a strategy that would yield direct monetary profits. However, due to the low response rate from the private sector, these results should be interpreted with caution, as they might not be applicable to all businesses working with Open Data.

The following graphic shows the categories the respondents are most interested in. Since this survey was directed at Austrian participants only, the 14 categories are the same that are offered by Austrian Open Data portals data.gv.at and opendataportal.at.

Figure 5: Categories

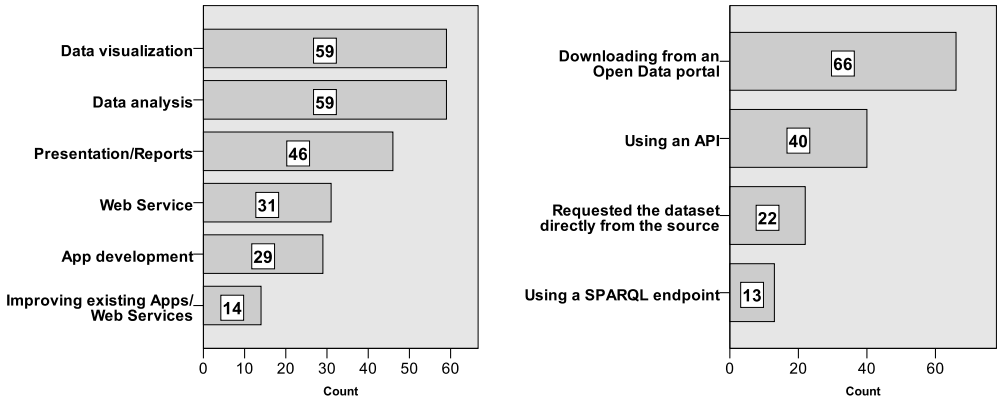


Having discussed the introduction section of the survey, the next chapter moves on to the analysis of questions, which were directed to users only.

4.1 User barriers

Before asking the users about the barriers, basic information about the use case, data retrieval methods and their motivation was collected.

Figure 6: Use cases and data source



(a) Use cases

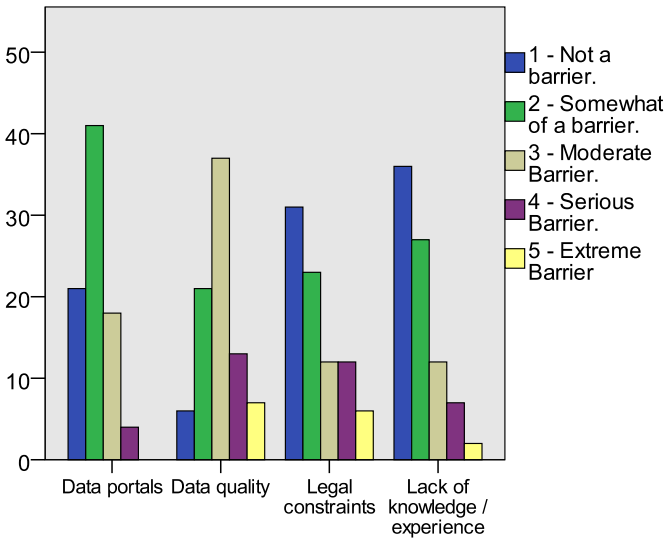
(b) Data retrieval methods

Data visualization, analysis and web service in particular are all use cases, which could take advantage of an data portal API in order to automatize obtaining up to date datasets. Figure 6b however suggests that the most used method to retrieve the data is by manually downloading them from the portals. One interesting finding is that there were 22 cases where the users requested previously unpublished data from the source directly. This could indicate that potential data providers will not publish their data on their own on various publishing platforms, but rather need to see that there is demand for the data in the first place.

4.1.1 Barriers in general

The user barriers section of the questionnaire was split into 4 categories: Data portals, data quality, legal constraints and lack of knowledge and experience. To gain an overview on which of these categories contains the strongest barriers, the respondents were asked to indicate in which of these categories they've been facing the most issues. Figure 7 below presents this overview using a stacked bar chart.

Figure 7: User barriers overview

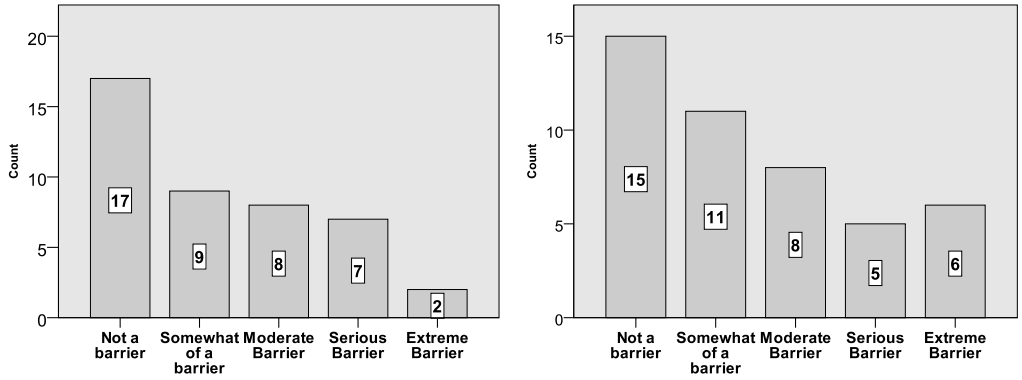


The strongest barriers are present in the data quality and data portals categories. Contrary to expectations, lack of knowledge and experience was indicated as a non-issue by most participants. This result is surprising at first, since the lack of knowledge and experience has been widely reported as a barrier both in research literature and in the interviews, which were carried out for this study. This rather contradictory result may be due to the fact, that the target group for the survey were subjects, who already have some experience with Open Data. Furthermore, taking a look back at Figure 2b, most participants are academics, so it is expected that lack of knowledge is not be an issue for them.

4.1.2 Data portals

The first set of questions was directed on the use APIs. For nearly 20 users, not having an API at all was not a barrier. The likely explanation is that those are the same users, who indicated that they use other methods of data retrieval in the previous section of the survey. Two more questions regarding the APIs were asked to determine whether slow and otherwise limited interfaces such as when only a limited number queries is possible, are a barrier for the users. The collected data however, must be interpreted with caution, as some of the respondents who indicated that they do not use an API, answered these questions. These answers were not taken into consideration to ensure the integrity of the analysis.

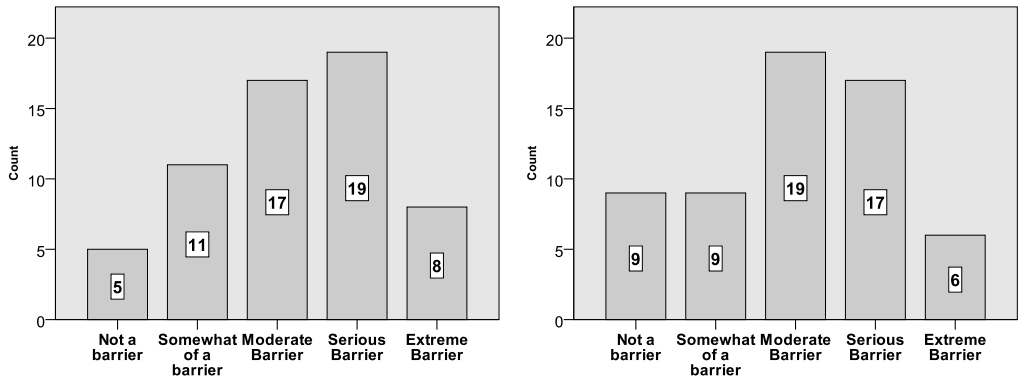
Figure 8: APIs



(a) API slow or unresponsive (b) Other limitations

Perhaps the biggest barrier, the majority of data users are reporting, is the lack of information about the content and the quality of the dataset. This can indeed be discouraging for potential developers, as they have to manually examine the data to determine whether it is of any use for them. This creates an unnecessary workload for the users. To make use of the open nature of these datasets, there should be an easy process for the users to submit any information they have found during their examination of the data back to the portal, thus saving the work for other users and improving the quality of the portal. The graphs below depict the severity of these barriers.

Figure 9: Information about the data



(a) No information about the quality (b) No information about the content

In the last question of this category, the users were asked to identify any

other barriers with Open Data portals. Some of the comments are listed below. The answers have been translated from German to English, with the original version of the answers included under each translation. In very few cases, the participants answered in English. There is no German translation provided for these comments.

- "Data.gv.at website is too slow, performance of the portal is poor"

  "data.gv.at Webseite sehr langsam, Performance des Portals schlecht"

- "Poor or no documentation of the datasets"

- "No information about the update frequency of the data. I can only develop an application if I have reliable information on when and how often the datasets updates.(e.g. "On the first day of each month")"

  "Keine Information über Update-Frequenz: ich kann eine App nur dann entwickeln, wenn ich verlässliche Information darüber habe mit welcher Frequenz die Daten aktualisiert werden (zB: "an jedem ersten Tag des Monats")."

- "Old data. OECD Data are available published on an independently maintained SPARQL endpoint. These data however, are old and do not match the data available on the OECD site. Therefore, we were not able to use the endpoint"

  "Veraltete Daten. Für OECD Daten werden von einem unabhängigen Anbieter die gleichen Daten über einen SPARQL endpoint zur Verfügung gestellt - diese Daten sind aber veraltet und stimmen nicht mit den Daten auf der OECD Seite überein, daher konnten wir den SPARQL endpoint nicht benutzen."

- "Lack of interesting datasets"

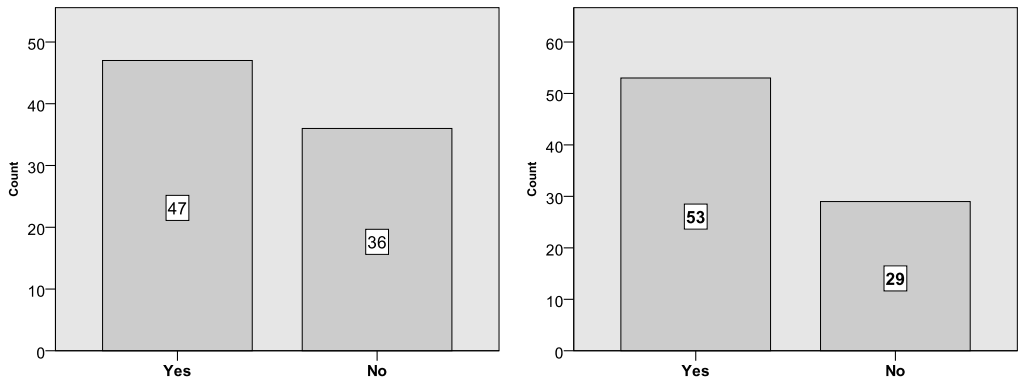
  "Mangel spannender Datensätze"

4.1.3 Data quality

As previously discussed in this thesis, there are conflicting views on whether low quality data should even be published or not. To gain a deeper insight into this topic, the participants were asked if all data regardless of quality should be published, which could be described as a "the more Open Data there is the better" approach. Furthermore, the next question asked whether

the organizations publishing Open Data should adhere to a certain set of quality standards and rules, so a certain level of quality is ensured, or if the users prefer no rules at all, as having enforced rules and standards would prevent many organization publishing their data in the first place.

Figure 10: Low data quality publishing



(a) Should the organizations open their data regardless of the data quality?

(b) Should organizations adhere to a specific set of standards and rules?

These results are consistent with the research conducted so far. There is indeed a disagreement in the Open Data community on which data should be published. There are users however, who would prefer if some quality standards and rules existed. They were asked to provide some examples on what kind of rules should be implemented. These are some of the their suggestions:

- "Topicality, completeness, primary sources, machine-readability of the data, unrestricted access, open standards, privacy, etc."

  "Aktualität, Vollständigkeit, Primärquellen, uneingeschränkter Zugang, Maschinenlesbarkeit, offene Standards, Datenschutz, etc."

- "There should not be any rules, rather recommendations, such as "use utf-8""

  "Keine Regeln, aber dringende Empfehlungen sollte es geben. Beispiel: utf-8 nutzen"

- "Minimum standard format for all data. If the data is available only in "exotic" formats, especially when these are not open formats, a version

in a minimum standard format must be available as well"

"Mindest-Format-Standard für alle Daten; Wenn Daten in "exotischen" Formaten angeboten werden (speziell, wenn es sich hierbei um nicht offene Formate handelt), müssen diese zwingend auch in den Mindest-Format-Standards angeboten werden"

- "Data should always be published in a machine-readable format. That means no PDFs for instance."

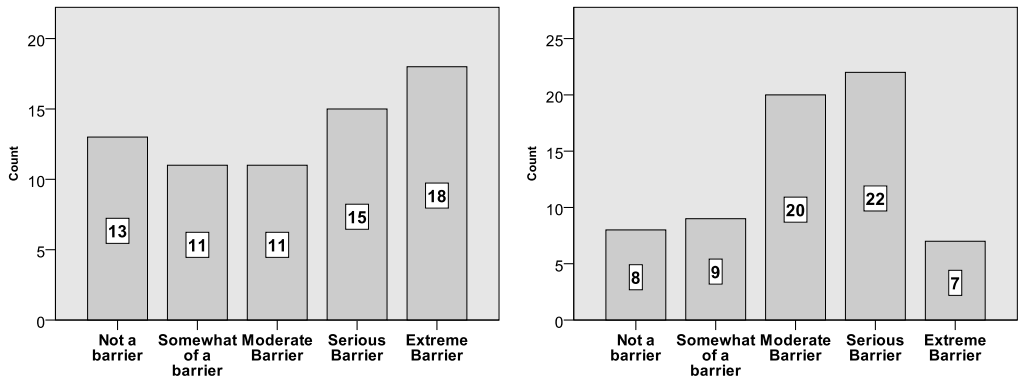
  "Daten sollten immer maschinenlesbar veröffentlicht werden, d.h. z.B keine PDFs"

- "For the moment, the concept of Open Data is not yet in full swing. Therefore I prefer quantity over quality for now"

  "Momentan ist, meiner Meinung nach, das Konzept von Open Data noch nicht völlig in Schwung gekommen. Deswegen ist für mich noch Quantität vor Qualität."

Perhaps the most obvious reported barrier is that the data is not machine-readable. As was already discussed in the second chapter of this thesis, data is often published in proprietary, human-readable only formats. Despite being reported as a moderate to extreme barrier by most, surprisingly, there are over 10 participants who indicated that non-machine readable data is not a barrier at all.

Figure 11: Machine readability and data completeness



(a) Data is not machine readable

(b) Data is incomplete

Furthermore, there were 42 participants who indicated data incompleteness as a moderate to serious barrier. These results suggest that there are currently many datasets with incomplete data being published, which drastically reduces their usability.

The respondents once again included some comments in the open comment section. They mostly confirm those barriers that were already mentioned in the questions, however some interesting points were made:

- "Some provinces publish excellent datasets on individual topics, but not all of them publish the same data. There is no Austria-wide comparability. In addition, the data is still being published in pdf format"

  "Einzelne Bundesländer veröffentlichen zwar hervorragende Datensätze zu einzelnen Themen, aber nicht alle Länder veröffentlichen dieselben Daten. Keine österreichweite Vergleichbarkeit. Außerdem werden Daten immer noch im pdf-Format veröffentlicht."

- "Metadata is often no help at all to learn more about the dataset. Either they are incomplete or missing completely"

  "Metadaten sind oft nicht hilfreich, um genaueres über die Datenqualität zu erfahren - entweder fehlen sie oder sie sind nicht komplett."

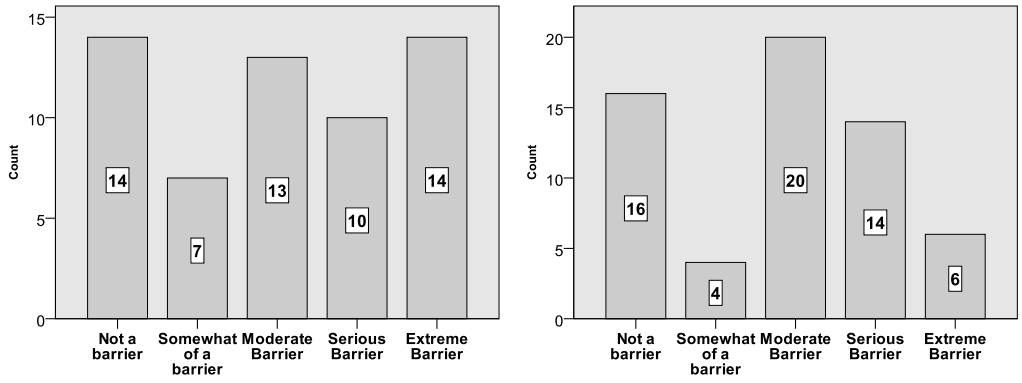
It could be argued, that if certain standards were enforced, these problems would not exist. Metadata is often being used to describe the dataset in order to provide the users with a better overview. However they are often inconsistent, wrong or simply not existing at all. The inconsistencies can be found in the published data itself as well. The fact that various governmental Open Data providers, even when working for the same state, publish different datasets, seemingly with no communication and planning with other providers, results in an inconsistent set of data available across the country. If the user finds a regional dataset on a certain topic, they have no guarantee that similar datasets exist for other regions as well.

4.1.4 Legal constraints

By definition, Open Data is freely accessible, reusable and allows for redistribution. This simple concept however, is made complicated by unclear licensing situation. As data users try to avoid any legal trouble, they are sometimes better off not using certain datasets with licenses that make the legal aspects of the free use and redistribution rather questionable. It should

not be expected from data users to understand complex legal documents. The survey researched this topic to find out how the users feel about unclear and restrictive licensing.

Figure 12: Licenses



(a) Restrictive licenses

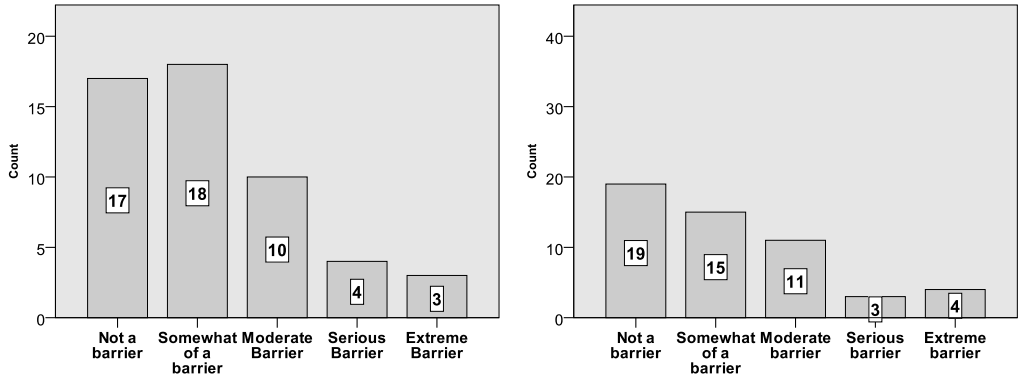
(b) Complex, hard to understand licenses

For the majority of participants, licensing is a moderate to extreme barrier. This is an important finding, as it suggests that some users were not able to use the data due to licensing issues. Potential users may then feel misled and disappointed, since Open Data has been defined by the freedom of use and reuse as one of its main concepts for them, only to later find out that there are some legal restriction after all.

4.1.5 Knowledge and experience

What was initially thought as one of the biggest barriers of Open Data, turned out to be a non-issue for most of the users. As previously explained, this is presumably due to the participants having previous Open Data experience. It is likely that with beginners in the field of Open Data, the results would change significantly. However, there is one barrier that has been inspired from the Open Source Software research literature that even experienced users are having trouble with. For more than half of the users, lack of proper documentation presents a moderate to extreme barrier. The other barriers, such as lack of support or helpdesk however, does not seem to be an issue.

Figure 13: Knowledge and experience



(a) No support or helpdesk

(b) Insufficient training

Figure 13 shows that the majority of participants in this survey do not believe that missing support or training is a barrier for them. Users were asked to state any reasons why there might be a general lack of knowledge in the public about Open Data. One user responded with:

- "According to my knowledge, no Open Data portal provides comprehensive information on the topic of Open Data. Good information is available, but it is circulating somewhere on the internet and one has to stumble upon it by luck."

  "Es gibt meines Wissens kein Portal, das einen über das Thema Open Data umfassend informiert. Gute Information gibt es, kursiert aber halt irgendwo im Internet und man muß z.T. das Glück haben, darüberzustolpern."

Indeed, a standard set of instruction and education material on Open Data could be made available on each portal. It would also need to be consistent across all data portals, as different guides would only create unnecessary confusion.

4.2 Providers

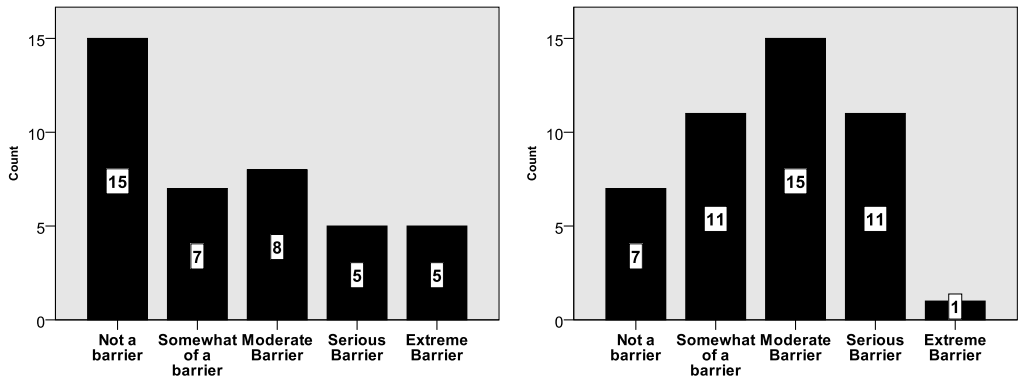
Similarly to the user part of the survey, some basic introductory questions such as motivation for publishing the data were asked before moving on to the next section of the questionnaire. The provider motivation is already discussed in chapter 4 and shown in Figure 4b. As previously stated, most

of the providers are from academia or government sector. This should be taken into consideration when analyzing the results that follow. There are some responses from providers, who are associated with the private sector, however, with such a small sample size, caution must be applied when interpreting these results, as these findings might not be statistically relevant and therefore might not represent the truth.

4.2.1 Privacy and security

As was already discussed in the second chapter, research literature and interviews confirm that many potential providers are hesitant to release their data in fear of misrepresentation, loss of control or accidentally releasing private sensitive information. For many providers however, this does not seem to be a concern as 15 providers stated that privacy and unwelcomed exposure of the data is not a barrier. Looking at Figure 14b however suggests that some providers do fear that the data they publish might be misrepresented.

Figure 14: Privacy and security



(a) Privacy, unwelcomed exposure of the data

(b) Fear of false conclusions drawn from the data

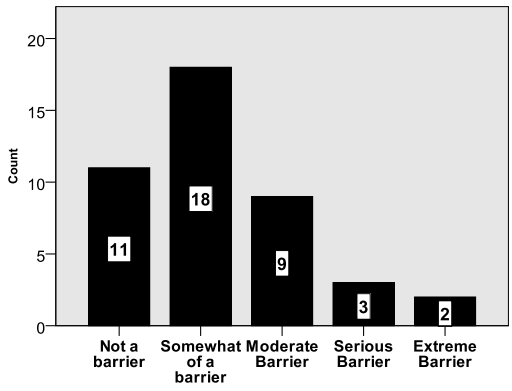
Providers might also fear that their data do not meet certain quality standards, so it is better to keep the data private. Lack of resources for the maintenance of the data seems to be the reason for this. As one particular respondent stated in the open question:

- "Keeping the data updated takes effort"

  "Aufwand der Aktuellhaltung"

This is an interesting point, especially when compared to Open Source. In Open Source, code is usually released in order to support collaboration between independent external actors and the in-house development team. This often benefits the provider of the code as they are able to crowdsource external developers who maintain and help them keep the code up to date. In their study on business value of strategic open source, Morgan and Finnegan highlight several benefits of Open Source including increased collaboration, knowledge sharing with communities, customers, and other parties. They also go on to state that Open Source facilitates joint ventures with other companies or research institutes [21]. Similarly, opening the data could facilitate collaboration between the provider and the Open Data community, thus helping the provider maintain their data.

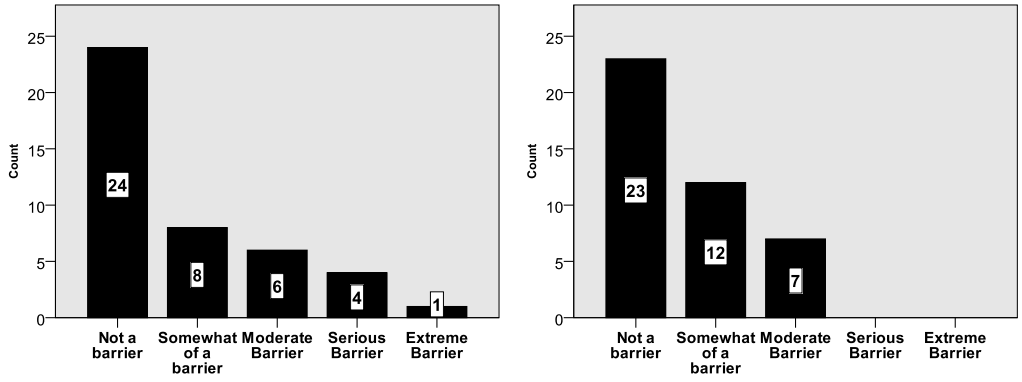
Figure 15: Poor data quality, better not expose it



4.2.2 Business and strategy

The results of this part of survey reflect that only a fraction of the participants are providers from the private sector. The fact that for the majority of providers, lost potential sales of the data do not present a barrier, suggest that the the provider is not relying on making revenue from the data. It is assumed that many private companies would be more hesitant about publishing data from which they can make revenue. Likewise, lack of Open Data business models presented a small to moderate barrier for most with very few exceptions.

Figure 16: Business and sale

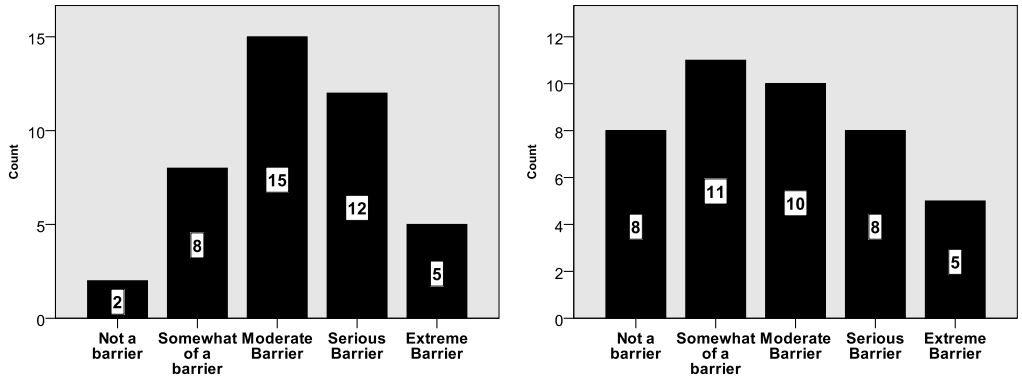


(a) Lack of Open Data business models

(b) No sale of the data possible

There were however some barriers reported in the expected areas. These results show that publishing Open Data is a process, which needs valuable resources and time. This could indicate that providers need some support and guidance on how to minimize these costs.

Figure 17: Costs and resources



(a) Lack of resources and time

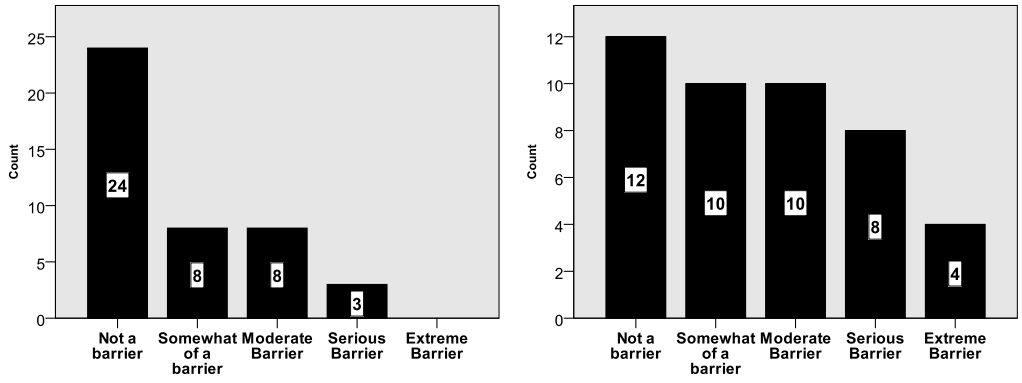
(b) Cost issues, such as hiring new talent, maintenance, training, data storage ...

4.3 Provider legal constraints

One interesting comparison to point out is that the providers do not seem to consider complex licenses as a barrier, especially compared to users. Looking back at Figure 12b, for more than half of the respondents, complex licenses

present a moderate to extreme barrier. This is however not the case with providers.

Figure 18: Licensing and ownership



(a) Complex licensing

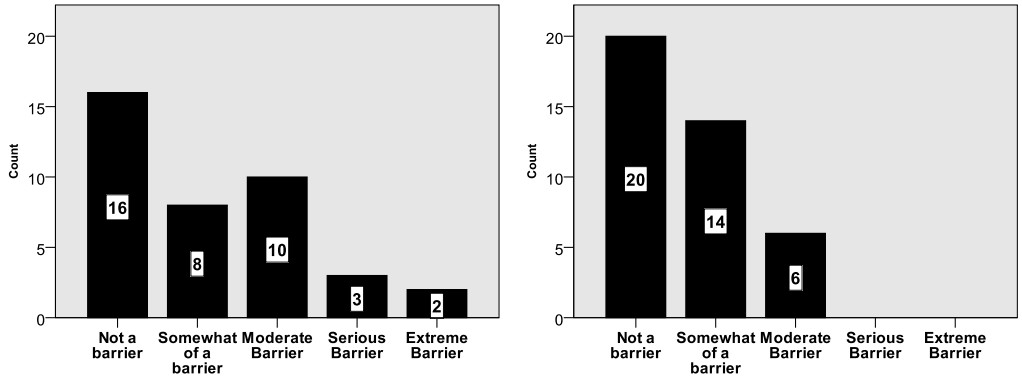
(b) Unclear ownership of the data, prohibiting its release

The participants were also asked if the the opaque ownership of the data inhibited its release. As the results show, there are indeed issues with unclear ownership leading to less data being released, as some respondents consider it an extreme barrier, completely preventing them from publishing the data.

4.4 Technical and support

Results from this category once again suggest, that many participants of this survey are experienced in the field of Open Data. Lack of documentation or support was marked as a non-barrier by the majority of respondents. Therefore it is important to bear this in mind when interpreting following charts, as these results may not apply to a beginner trying to publish Open Data.

Figure 19: Documentation and support

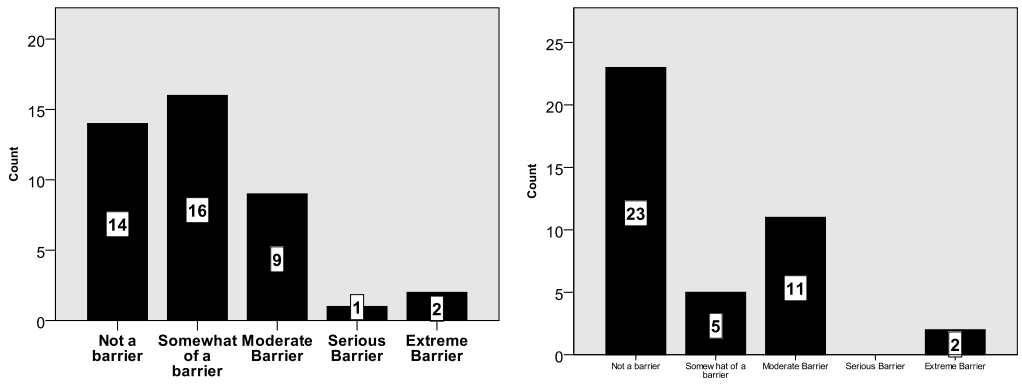


(a) Lack of guidance and documentation

(b) No support or helpdesk

The graphs that follow provide some important insight on the stance of providers on standards in formats and metadata. While some users complain about the inconsistency in the formats and metadata, it seems that providers prefer the freedom to choose any file format they prefer and write the metadata according to their own standards. This indicates that creating and enforcing an unified standard would most likely be met with resistance from the providers, thus leading to less data being published.

Figure 20: File format and metadata standards



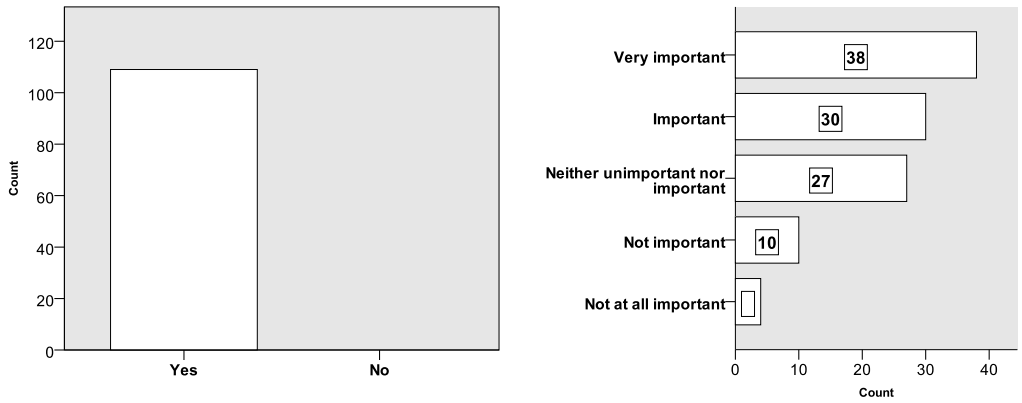
(a) Various standards and forms of machine readable data

(b) Absence of metadata standards

4.5 Open Data potentials

In the final part of the survey, participants were asked on their general opinion about Open Data, whether they will work with Open Data in the future, how important was Open Data for their project and if it was beneficial for their project or their organization as a whole. Despite many reported barriers, the results were overwhelmingly positive. From the 110 respondents, 109 suggested that they will indeed continue to work with Open Data in the future and that Open Data has been important for their project or organization. However, it is important to consider the potential bias in these results due to the fact, that most participants are assumed to have an interest in Open Data. We therefore advice a cautious interpretation of these data.

Figure 21: Open Data potential



(a) Will you be working with Open Data in the future?

(b) How important has Open Data been for your project or organization?

4.6 Summary

Most of the barriers that are widely cited in the literature have also been confirmed in our survey. We have also managed to measure the severity of these issues by presenting the results in the form of a Likert Scale. Furthermore, from the comments the participants wrote in the open questions, we have been able to identify some issues, which are to our knowledge, not yet documented in the Open Data literature. The data we have obtained from the comments however, should not be extrapolated to hold true for the entire Open Data community, as they might just represent barriers, which only specific users are faced with. The summary of these barriers is presented in Table 17.

Table 17: Discovered barriers

Data Portals
Websites are too slow
No automatic notification when a dataset is updated
No information about the update frequency of a dataset provided
Lack of data harmonisation
No way to tell whether the same data is published by other portals as well
Data Quality
Character encoding is not utf-8
Datasets have no unique identifier
Legal
Licenses are not machine-readable
Knowledge
Open Data is often misunderstood as Big Data
Technical
Pre-processing the datasets is a barrier (Formats, conversions)
Keeping the dataset updated
Strategic and business
Lack of commitment from top management
Data publishing can lead to administrative interventions

To end the survey report on a slightly negative note, it needs to be said that the results from the survey indicate that there is still much room for improvement in the Open Data concept and that the full potential has yet to be achieved. The results indicate that there are still many areas for future research, some of which are outlined in the conclusion of this thesis. It is noteworthy to add that the Open Data community enthusiastically helps with the research. The number of responses in this survey is quite impressive, especially given the fact that it was limited to Austria only. Furthermore, 27 participants expressed their interest in further discussing these topics personally.

5 Barrier mitigation strategies

Successful mitigation of these barriers is necessary if we want the Open Data concept to succeed. In the recent years Open Data movement has enjoyed a steady momentum and increased interest from analysts, statisticians, software developers, governments and even some private business entities. However, until the barriers are still present, Open Data may lose its momentum

and the interest in Open Data may start to fade.

Up to now, a number of studies have explored potential mitigation strategies on Open Data barriers. From the literature review and the survey results, we have concluded that data portals lack information about content and quality of the dataset. A recent study on benchmarking open government by Veljković et al. suggests providing rating information for datasets. Additionally, they claim recommend that it should be mandatory for data publishers to provide descriptions of the data, as such measure would significantly increase the clarity of the contained information [26]. In order to lower the cost of publishing data, Conradie and Choenni advise against releasing data just for its own sake in their study on barriers of open government data. They argue that to ensure efficient use of released data, a more focused approach is needed, where data with the highest return on investment, not only in financial sense, but in the context of the highest reuse is released. Data release based on demand, would imply lower costs [8]. Additionaly, Martin et al. propose the use of a system for measuring the popularity of the published datasets in order to identify and clarify the information needs of society [2]. A number of mitigation strategies for various barriers are proposed in another major study on Open Data risks by Martin et al. Restrictive or unclear licensing has been identified as a major barrier in our survey. To mitigate this barrier, collecting the concerns of users and modifying the licenses if they are too constraining is suggested. Data should also always be released complying with the definition of openness.[18] [19] According to Zuiderwijk et al., the use of metadata can yield significant benefits such as improving the ability to find the data, providing context for the datasets and avoiding unnecessary duplication. Additionally, it can be used to provide a link between the creator and the user of the data [30]. Many potential data providers are still hesitant to publish their data. The report on Open Data by the European Commission claims that pilot and test cases, sharing of good practices and mobilization campaigns can help the public sector in adopting a culture of Open Data [6]. Publishing Open Data however, is not an easy task. Guiding principles such as those, published by Zuiderwijk et al. can provide a systematic guidance on releasing Open Data[29]. Table 18 presents a summary of proposed strategies for Open Data barriers.

[18]http://opendefinition.org/od/2.1/en/

Table 18: Barrier mitigation strategies

Barrier	Proposed strategy	References
No information about the quality of the data	Provide rating information about the dataset	[26]
No information about the content of the dataset	Mandatory descriptions of the dataset Use of standardized metadata models	[26] [30]
Cost issues	Release data based on demand	[8]
Restrictive and unclear licensing	Collect the concerns of the users and appropriately modify the license Use license that complies with the rules as defined by the Open Definition	[19]
Difficulty in finding the data	Use of standardized metadata models	[30]
Duplicate data	Use of standardized metadata models	[30]
Lack of communication between the provider and the user	Use of standardized metadata models helps to create a link between the creator and the user	[30]
Lack of guiding principles	Make use of systematic research publications, such as the Zuiderwijk et al. study	[29]
Resistance to change, risk-averse culture	Sharing of good practices and use cases. Mobilization campaigns to promote open culture	[6]
Lack of business models	Promote networking between stakeholders	[2]
Language barriers	Publish data in multiple languages	[19]
Lack of awareness	Governments should encourage active public participation in the use of OGD portals	[2]

6 Conclusion

Open Data is a concept still very much in its early phase. And while new communities and platforms progress with a steady pace, there is still much work to be done. As the empirical findings in this thesis show, there are numerous barriers present that are making the use of Open Data difficult for both users and providers. Some work and research on strategies on how to mitigate these barriers has already been done as we have discussed in the previous chapter. Even though this thesis does not actually solve any of these barriers, we hope that it will provide a solid foundation for future research on the mitigation of Open Data barriers.

6.1 Limitations and further research

There are some limitations with this study that need to be highlighted. Firstly, this thesis discussed only a part of existing Open Data barriers. As was discussed in the earlier chapters of this thesis, more than 100 potential barriers were found in the research literature but only a part of them were analyzed in this thesis. Since this study also set out to further investigate these barriers using a survey, it was not feasible to include all of these barriers

in the questionnaire without risking a much lower response rate. The second limitation was that the survey was limited to Austria only. This makes the findings less generalisable and the reported data might not be valid in other countries, which have other Open Data portals and communities and therefore, different experiences. What is now needed is an international study on Open Data concerns and barriers. It would also be interesting to compare the results between countries.

In terms of directions for future work, we have identified following areas of research that should be carried out:

- Further work could be done by investigating the use of Open Data in the private sector. As was reporter in this thesis, the private sector is rather underrepresented in the survey participation and research literature. A future study investigating the use of Open Data in the private sector would provide valuable input in the field of Open Data research by closing this existing research gap.

- Although our survey has been limited to Austria only, it can serve as a good starting point in creating a second instance of an Open Data survey with no demographic limitations. We recommend the use The ODI Node Network [19] as it provides us with a good opportunity to reach a worldwide audience.

- This study also touched upon the close relation between Open Source and Open Data. However, it only investigated the similarities of barriers in these concepts. Another possible area for future research would be a major study on the relation and differences between Open Source and Open Data.

- In our survey, we have identified that incomplete datasets being published is a barrier. We therefore propose an extensive analysis of the datasets, which are being published on known Open Data portals, in order to determine if the data is usable on its own or presents a rather incomplete picture of the data.

- As one interviewee noted, Open Data is in need of success stories. Longitudinal case studies on successful projects using Open Data would be beneficial in analyzing strategies that work for Open Data.

[19] https://theodi.org/nodes

References

[1] Harith Alani, David Dupplaw, John Sheridan, Kieron O'Hara, John Darlington, Nigel Shadbolt, and Carol Tullo. *The Semantic Web: 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea, November 11-15, 2007. Proceedings*, chapter Unlocking the Potential of Public Sector Information with Semantic Web Technology, pages 708–721. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[2] María Del Carmen Caba Pérez Alejandro Sáez Martín, Arturo Haro De Rosario. An international analysis of the quality of open government data portals. *Social Science Computer Review*, jun 2015.

[3] José Manuel Alonso. Open government data: approaches, concerns and barriers, lessons learned, may 2011.

[4] John C Bertot, Paul T Jaeger, and Justin M Grimes. Using icts to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. *Government information quarterly*, 27(3):264–271, 2010.

[5] Katrin Braunschweig, Julian Eberius, Maik Thiele, and Wolfgang Lehner. The state of open data limits of current open data platforms. 2012.

[6] European Commision. Communication from the commission to the european parliament, the council, the european economic and social committee and the committee of the regions. open data - an engine for innovation, growth and transparent governance., dec 2011.

[7] Peter Conradie and Sunil Choenni. Exploring process barriers to release public sector information in local government. In *Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance*, ICEGOV '12, pages 5–13, New York, NY, USA, 2012. ACM.

[8] Peter Conradie and Sunil Choenni. On the barriers for local government releasing open data. *Government Information Quarterly*, 31, Supplement 1:S10 – S17, 2014.

[9] Open Knowledge Foundation. Barriers and solutions to open data, 2011.

[10] Christian P Geiger and Jörn von Lucke. Open government and (linked)(open)(government)(data). *eJournal of eDemocracy & Open Government*, 4(2), 2012.

[11] Sigi Goode. Something for nothing: management rejection of open source software in australia's top firms. *Information and Management*, 42(5):669 – 681, 2005.

[12] Paul T Jaeger and John Carlo Bertot. Transparency and technological change: Ensuring equal and sustained public access to government information. *Government Information Quarterly*, 27(4):371–376, 2010.

[13] Marijn Janssen, Yannis Charalabidis, and Anneke Zuiderwijk. Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, 29(4):258–268, 2012.

[14] Peter Johnson and Pamela Robinson. Civic hackathons: Innovation, procurement, or civic engagement? *Review of Policy Research*, 31(4):349–357, 2014.

[15] Martin Kaltenböck. *Open Government Data Weißbuch (Österreich.* Ed. Donau-Univ. Krems, Krems an der Donau, 2011.

[16] Maxat Kassen. A promising phenomenon of open data: A case study of the chicago open data project. *Government Information Quarterly*, 30(4):508–513, 2013.

[17] Juho Lindman. Similarities of open data and open source: impacts on business. *Journal of theoretical and applied electronic commerce research*, 9(3):46–70, 2014.

[18] James Manyika, Michael Chui, Diana Farrell, Steve Van Kuiken, Peter Groves, and Elizabeth Doshi. Open data: Unlocking innovation and performance with liquid information, 2013.

[19] Sébastien Martin, Muriel Foulonneau, Slim Turki, and Madjid Ihadjadene. Risk analysis to overcome barriers to open data.

[20] Patrice McDermott. Building open government. *Government Information Quarterly*, 27(4):401–413, 2010.

[21] Lorraine Morgan and Patrick Finnegan. Beyond free software: An exploration of the business value of strategic open source. *The Journal of Strategic Information Systems*, 23(3):226 – 238, 2014.

[22] Axel Polleres Sebastian Neumaier, Jürgen Umbrich. Automated quality assessment of metadata across open data portals. *ACM Journal of Data and Information quality*, Vol. V, 2016.

[23] Richard Stallman. *Free software, free society: Selected essays of Richard M. Stallman*. Lulu. com, 2002.

[24] Igor Steinmacher, Marco Aurelio Graciotto Silva, Marco Aurelio Gerosa, and David F. Redmiles. A systematic literature review on the barriers faced by newcomers to open source software projects. *Information and Software Technology*, 59:67 – 85, 2015.

[25] Klaas-Jan Stol and Muhammad Ali Babar. Challenges in using open source software in product development: A review of the literature. In *Proceedings of the 3rd International Workshop on Emerging Trends in Free/Libre/Open Source Software Research and Development*, FLOSS '10, pages 17–22, New York, NY, USA, 2010. ACM.

[26] Nataša Veljković, Sanja Bogdanović-Dinić, and Leonid Stoimenov. Benchmarking open government: An open data perspective. *Government Information Quarterly*, 31(2):278–290, 2014.

[27] Lauren Vogel. The secret's in: Open data is a foreign concept in canada. *CMAJ*, 183(7):E375–E376, Apr 2011. 21402678[pmid].

[28] Anneke Zuiderwijk, Marijn Janssen, Sunil Choenni, Ronald Meijer, and R Sheikh Alibaks. Socio-technical impediments of open data. *Electronic Journal of e-Government*, 10(2):156–172, 2012.

[29] Anneke Zuiderwijk, Marijn Janssen, Ronald Meijer, Sunil Choenni, Yannis Charalabidis, and Keith Jeffery. *Electronic Government: 11th IFIP WG 8.5 International Conference, EGOV 2012, Kristiansand, Norway, September 3-6, 2012. Proceedings*, chapter Issues and Guiding Principles for Opening Governmental Judicial Research Data, pages 90–101. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[30] Anneke Zuiderwijk, Keith Jeffery, and Marijn Janssen. The potential of metadata for linked open data and its value for users and publishers. *JeDEM-eJournal of eDemocracy and Open Government*, 4(2):222–244, 2012.

A
Transcripts of the interviews

For reference purposes we add the transcripts of the interviews where permission for recording was granted by the interviewees. These interviews were held in German.

A.1 Univ.-Prof. Mag. Dr. Peter Parycek

Zuerst möchte ich mich herzlich für ihre Hilfe bedanken. Als erstes möchte ich fragen: "Ist es in Ordnung wenn ich dieses Gespräch aufnehme?"

Kein Problem

Danke. Zuerst mal eine kleine Vorstellung. Ich bin ein Student der Wirtschaftsinformatik an der Wirtschaftsuniversität Wien. Bei einer Vorlesung an der Universität habe ich zum ersten Mal über das Konzept von "Open Data" gehört. Und da habe ich mich gefragt, wie ist es möglich, dass ich als Informatiker nicht nur als Student, sondern auch beruflich zum ersten Mal über Open Data so spät gehört habe. Da war mir klar, dass es vermutlich noch bestimmte Barrieren gibt und Open Data noch nicht so verbreitet ist wie es potentiell sein konnte. In der wissenschaftlichen Literatur habe ich circa 90 Barrieren gefunden. Die Barrieren wurden nicht nur von Open Data Literatur gesammelt sondern auch diejenigen von Open Source, die man potentiell auf Open Data anwenden könnte wurden berücksichtigt. Um das Interview zu beginnen, was denken Sie sind die größten Barrieren von Open Data. Also zum Beispiel bei den Publisher von Open Data. Wieso veröffentlichen Institutionen ihre Daten nicht?

Ja, geht es eigentlich um Verwaltung oder geht es um Verwaltung und Wirtschaft.

Beides.

Also bei der Verwaltung ist es... es gibt Verwaltungseinheiten, die wollen nicht veröffentlichen, weil sie Transparenz nicht haben wollen, sie wollen es vielleicht auch nicht machen weil es ein zusätzlicher Aufwand ist und sie daher sich auf Grund von fehlenden Ressourcen da extrem zurückhalten und manche gibt es sicher auch die es verkaufen wollen. Das würde ich sagen sind die größten Hürden. Die Mehrheit der Verwaltungseinheiten sind einfach unentschlossen. Also die wissen nicht: Habe ich was davon? Warum sollte ich es überhaupt machen? Die sind so in der Mitte. Dieses typische 20% sind dafür, 10 bis 15% sind dagegen und der Rest ist so unentschlossen. Haben genug zu tun und wissen nicht was der große Mehrwert ist und warum sie das überhaupt tun sollen. Wenn ich was mache muss ich auch wissen: was habe ich davon und generiere ich davon überhaupt einen sinnvollen Mehrwert? Weil es ist Arbeit, ich muss den Datensatz angreifen, ich muss ihn reinigen, ich muss ihn veröffentlichen, wenn ich ihn einmal veröffentlicht habe, dann sollte ich ihn wiederkehrend machen, außer es sind historische Daten, aber in der Regel sind es Daten die sich durchaus immer wieder Veränderungen unterliegen und somit ist es dann wenn ich einmal einen Datensatz veröffentliche ist es natürlich eine fortlaufende Arbeit.

Also konnte man sagen, dass es keinen klaren Business Plan gibt? Publisher sehen keinen Gewinn, keinen Profit.

In der Verwaltung muss es nicht immer gleich ein monetärer Wert sein. Aber ich sehe keinen Mehrwert. Wenn ich einen Mehrwert sehe, meine Zielgruppe für die Bürger und Bürgerinnen die durchaus meine Dienste verwenden, wenn ich einen Mehrwert sehe, kann das auch eine Motivation sein, dass ich in dem Bereich was tue. Dieser Mehrwert muss nicht immer monetär sein. Das muss nicht eine Einhahmensquelle für die Behörde sein. Also es muss einfach einen Sinn ergeben. Und wenn ich als Behörde keinen Sinn sehe, dass ich etwas veröffentliche, dann werde ich es nur unter Druck veröffentlichen.

Es ist auch möglich, dass sie Angst von einer falschen Darstellung der Daten haben.

Das ist was ich am Anfang gemeint habe. Diese Bedenken gibt es durchaus auch. Also es gibt die Bedenken, dass es eine falsche Interpretaton von Daten gibt, dass die Daten missbraucht, manipuliert werden, also diese

Bedrohungsszenarien gibt es auch, und das sind Argumente die immer wieder gennant werden. Mein persönliches Gefühl, eine subjektive Wahrnehmung aus den Diskussionen die wir gehabt haben mit der Verwaltung ist, und teilweise ist es auch belegbar durch einen größeren Workshop, dass die größere Masse einfach keinen Sinn darin sieht. Warum soll ich das tun. Warum soll ich mir den Mehraufwand antun, Daten zu reinigen und zur Verfügung zu stellen. Teilweise sehen die gar nicht, dass sie diese Daten haben.

Denken Sie, dass es dafür eine Lösung gibt? Dass man irgendwie die Institutionen überzeugen kann, dass es vielleicht doch Sinn macht?

Also ich würde auch nicht ganz generell sagen "die Institutionen". Weil was ist eine Institution. Ist die Stadt Wien eine Institution? Die Stadt Wien als Institution hat erkannt, dass es ein großes Potential hat und daher gibt es ein ganz klares politisches Commitment, es gibt doch ganz klare Aufbau und Ablauf-Organisationen, wie Daten veröffentlicht werden sollen. Das ist alles vorhanden. Und es gibt auch eine ganz... also da wurden alle Abteilungen in der Stadt Wien darauf aufmerksam gemacht, dass sie Daten veröffentlichen sollen, beziehungsweise sogar müssen. Und die Realität ist, dass es ein Paar quasi Musterschüler gibt, die veröffentlichen sehr viele Daten, hochqualitative Daten, dann gibt es welche, die nichts machen und auch nichts machen wollen aus Gründen die wir gerade genannt haben, also Transparenz, Missinterpretation und solchen Dingen und dann gibt es eine große Mehrheit, die sehen diesen Mehrwert noch nicht, daher tun sie auch nichts. Und was kann man dagegen machen? Viel Überzeugungsarbeit, viele Use Cases wo man sieht, dass ein Mehrwert entstanden ist, auch teilweise aufzeigen: Wie können überhaupt Daten entstehen. Kann ich Objekte intelligenter machen, die automatisiert Daten sammeln, die ich dann wiederum entweder selber auswerten kann oder der Allgemeine zur Verfügung stellen kann. Also das Datenthema an sich, wir stehen mit dem ja überall ganz am Anfang. Das ist nicht nur eine Sache der Verwaltung sondern auch der Wirtschaft. Also es ist jetzt Mainstream, aber es werden sicher noch ein Paar Jahre vergehen, dass man das volle Potential von Datennutzung, sowohl zu internen als auch im Open Data Bereich. Bis wir das erleben werden sicher noch einige Jahre vergehen. Und da kann man noch Überzeugungsarbeit leisten.

Sie beschäftigen sich meistens nur mit der Verwaltung oder dem

Government-Sector sozusagen...

Primär, aber wir haben auch ein anderes Projekt auch noch, wo wir Daten aus der Wirtschaft sammeln, also haben wir auch da die ersten Erfahrungen gemacht, wie die Wirtschaft darauf reagiert, wenn man konkret anfragt nach ihren Daten.

Die Leute aus der Wirtschaft mit denen ich gesprochen habe, halten es für eine Barriere, dass sich die Open Data Literatur immer wieder nur mit Government beschäftigt. Also glauben die, dass Open Data nur für Government Sector geeignet ist und nicht für sie.

Ich glaub das ist auch da ein fehlendes Bewusstsein. Ich habe ein uraltes Beispiel, das ich seit drei Jahren in meinem Foliensatz drinnen habe, eine sehr simple App "Wie komme ich von A nach B". Da brauche ich unterschiedliche Datensätze. Und da sind auch Datensätze der Wirtschaft drinnen. Zum Beispiel der Datensatz von einem Car Sharing Unternehmen. Der Datensatz von Uber. Die Datensätze von Taxis. Also wenn alle diese Unternehmen diese Datenbanken aufmachen würden, so dass wir diese echte Daten automatisiert in Applikationen verwenden könnten, könnten wir viel größere Mehrwerte entstehen lassen, in dem ich, ein simples Beispiel, von A nach B mit der bestmöglichen Variante komme. Was bestmöglich ist, definiere ich. Also möchte ich gesund sein, möchte ich einfach nur schnell von A nach B kommen. Also das ist nur eine Idee, da finden Sie sicher viele weitere Ideen und Applikationen wo unterschiedliche Daten vermischt werden. Daten, die aus dem öffentlichen Bereich kommen aber auch aus dem Privatbereich. Also das ist glaube ich eher eine vorgeschobene Ausrede zu sagen: "Ah, das hat mit mir ja gar nichts zu tun". Aber es ist auch dort natürlich die zentrale Frage... In einem Unternehmen muss ich mir immer wieder die Frage stellen: "Wofür setze ich meine Resourcen ein". Wenn ich nicht einen unmittelbaren, beziehungsweise Schrägstrich mittelbaren Mehrwert für mein Unternehmen sehe, dann mache ich es nicht. Und ich glaube, diese Mehrwerte, deren sind sich die Unternehmen noch nicht bewusst. Weil auch die Konzepte von Share-Economy und überhaupt das Konzept von "Wenn ich teile, dann kann ich auch als Unternehmen profitieren" sind noch nicht weit verbreitet und selbst dann wenn sie bewusst sind, stellt es sich immer noch die Frage, wie operationalisiere ich das jetzt in meinem Unternehmen, also wie kann ich es in

meiner Strategie berücksichtigen und wie kann ich es in weiterer Folge operationalisieren, also das sind schon noch weite Wege zu gehen. Mein Eindruck momentan ist es so, dass Unternehmen Marketing und PR daraus rausziehen und ein wenig in CSR Bereich sich rechtfertigen wollen.

Ja, kommen wir jetzt zu der Datenqualität selbst. Am Anfang haben Sie erwähnt, dass die Bearbeitung der Daten viel Arbeit ist. Daraus resultiert wahrscheinlich, dass einige Daten die doch veröffentlicht werden, von sehr niedriger Qualität sind. Also ich selbst habe viele Datensätze auf data.gv.at gefunden, die entweder korrupt waren oder falsche Daten beinhalten. Und da fragt man sich wozu überhaupt so etwas veröffentlichen, wenn ich das nicht verwenden kann.

Na ja, es gibt unterschiedliche Zugänge, also der eine, der pragmatische Zugang ist, wir sind froh, dass Daten überhaupt veröffentlicht werden, also wenn wir da mit einem zu hohen Qualitätsanspruch hineingehen würden, dann werden noch weniger Daten da. Also insofern ist es ein zweischneidiges Schwert. Also fordere ich jetzt zu hohe Qualität der Daten, dann sagt die Verwaltung: "OK, da machen wir lieber gar nichts". Ist es die Aufgabe der Verwaltung diese Daten wirklich alle zu reinigen und sie in eine höhere Qualität zu bringen? Da kommt noch die Frage Geschäftsmodell, also wenn sie das machen, dann sollten sie vielleicht dann doch beginnen Geld dafür zu verlangen. Oder ist es nicht vielleicht eh der richtige Zugang die Daten so wie sie sind zu veröffentlichen? Und da muss man sich fragen, wie entsteht so ein Datensatz? Ein Datensatz entsteht aus irgendeinem Bedürfnis heraus, der innerhalb der Verwaltung entsteht. Und das Bedürfnis war nicht den Datensatz für etwas anderes zu verwenden als ursprünglich definiert... Programmieren Sie?

Ja.

Sie haben ein ganz klares Bild, wie Sie ihren Code schreiben und wie Sie ihre Datenbank anlegen. Und da ist Ihnen wurscht, was die anderen links und rechts von Ihnen machen. Weil Sie wollen ihr Programm möglichst schnell herunterschreiben und ihre Datenbank möglichst schnell aufsetzen. Sie sprechen sich nicht mit den anderen ab. Was die Datenstruktur betrifft,

irgendwelche Formate, Schnittschtellen, sonstiges... machen Sie in der Regel nicht. Und so entstehen Datensätze. Das ursprüngliche Sinn und Zweck eines Datensatzes wie er entstanden hat, hat einen ganz spezifischen Hintergrund, eine ganz spezifische Anforderung. Und diese Anforderung mag jetzt zu einer Weiterverwendung nicht besonders geeignet sein. Die Frage ist jetzt, ist es die Aufgabe der Verwaltung diese Datensätze so zu bereinigen und so sauber hinzustellen, dass man sie quasi alles automatisiert verwenden kann. Kann man jetzt unterschiedliche Meinungen haben? Kommt wahrscheinlich auf den jeweiligen Datensatz an. Da gibt es die moralische Verpflichtung, also moralisch... aber gibt es vielleicht doch eine verwaltungstechnische Verpflichtung eine hohe Qualität zu liefern und bei anderen ist es vielleicht gut, dass der Datensatz einfach veröffentlicht worden ist. Und die Bereinigung und die Verbesserung der Daten konnte von der Wirtschaft übernommen werden und die könnten wiederum für diese Leistung Geld verlangen. Und so könnte dann in weiterer Folge auch auch Daten-Ecosystem irgendwann mal entstehen. Würde ich nicht ultimativ sehen, sondern würde ich sagen, das kommt wirklich auf die ganz spezifischen Anforderungen an, wann sollte Verwaltung hohe Qualität liefern und in welchen Fällen sind wir eigentlich nur froh, dass die rudimentären Datensätze vorhanden sind und die werden dann in weiterer Folge gereinigt und verbessert durch die Wirtschaft oder durch die Gesellschaft.

Ja, das macht Sinn, aber bestimmt sind einige Developers abgeschreckt, wenn sie mit solchen Daten arbeiten.

Ja, aber die Frage bei Developers ist natürlich auch immer: Developers, wer sind Developers? Die Developers sind... manche die spielen mit den Daten, auch schön und wichtig, aber wenn ich einen wirklichen Business Case habe, wenn ich wirklich weiß was ich will, und ich sehe, dass der Datensatz mir hilft um meinen Business Case zu erreichen, dann kann ich mir schnell ausrechnen ob sich das auszahlt den Datensatz zu bereinigen und zu nutzen.

Das ist ein ähnliches Problem wie beim Open Source. In der Open Source Literatur wird auch viel über Code Qualität geschrieben. Es ist auch viel Arbeit für andere Benutzer. Zuerst muss man den Code verstehen und oft weiter bearbeiten bevor man es verwenden kann.

Das ist eine schöne Parallele, vor allem wenn man das vergleicht mit was ich eben gesagt habe: Warum ist ein Datensatz entstanden. Genauso wie Code entsteht. Code entsteht auch aus einem spezifischen Grund. Und es mag, das jemand anderer diesen Code verwenden möchte, aber dann muss er sich viel Zeit nehmen um den Code zu lesen.

Die nächste Frage ist über Standards. Viele Leute beschweren sich, dass es keine richtigen Regeln oder Standards gibt. Deswegen gibt es so viele verschieden Formate (JSON, XML, usw...)

Das ist schnell zu beantworten. Es gibt einen Standard auf der Metadatenebene. Dort kann es einen geben. Und dort haben wir es auch in Österreich umgesetzt, weil sich alle an diesen Standard halten, und runter auf der Datenstruktur, keine Chance. Genau aus den Gründen die ich gerade gesagt habe. Keine Chance. Weil die Ursache... Also wie entsteht ein Datensatz, es gibt einen spezifischen Grund warum dieser Datensatz entstanden ist und ja, da hängen unterschiedliche Systeme dran, die können unterschiedliche Formate dann auswerfen. Teilweise geht die Forderung sehr stark alles Richtung CSV zu veröffentlichen und nicht in Microsoft Excel, persönlich habe ich das nie nachvollziehen können, weil unter Umständen ein Excel Sheet mir mehr hilft, als eine CSV Datei. Also auch da, was haben wir davon wenn wir alle dazu zwingen, alles in CSV zu veröffentlichen? Bei CSV kann ich auch viel verlieren, was ich in anderen Datenformaten vielleicht noch drinnen hätte. Also da bin ich ganz klar, das ist nicht anders machbar für die Verwaltung. Das zu vereinheitlichen, das ist Aufgabe der Wirtschaft. Die Verwaltung soll möglichst zu Nullkosten also mit Creative Commons die Daten liefern, eine Verpflichtung eingehen, dass sie die Daten neuern, aber die Qualitätssteigerung, die muss durch die Wirtschaft gemacht werden.

Weiters haben sich noch einige Leute beschwert, dass dieses ganzes Open Data Movement nicht Europaweit ist. Dass es kein zentrales Portal gibt, jedes Land hat seine eigene Regeln und Standards, und die Datensätze selbst sind auch nicht kompatibel, denn es gibt zum Beispiel Unterschiede in der Kodierung. Wäre es nicht sinnvoll zumindest für ganz Europa ein zentrales Portal zu haben?

Na ja, jetzt gibt es ein Paar zentrale Portale, die Sie sicher kennen und Sie wissen wahrscheinlich auch, dass aktuell ein Projekt läuft, ein zentrales

Data-Portal, welches 2016 gestartet werden soll und dann hätten wir dieses europäisches Portal. Vereinheitlichen kann man allerdings nur, das ist jetzt analog die gleiche Argumentationslinie, vereinheitlichen kann man nur auf der Metadatenebene, eventuell die eine oder andere API, Schnittschtellen kann man sicher auch vereinheitlichen auf Datenportalen. Also wenn man Schnittschtellen auf Datenportale vereinheitlichen, das wäre sicher ein Element, dass man sich anschauen kann. Aber bei der Datenstruktur haben wir auf der europäischen Ebene das gleiche Problem. Das simple Beispiel, wenn ich die öffentlichen Toiletten geokodiere und einer sagt WC, der nächste sagt Klo, der dritte... also allein auf der semantischen Ebene. Jetzt habe ich eine Toiletten-App, und möchte halt die anderen Städter auch einbinden. Wenn ich das machen will, muss ich in den Datensatz eingreifen, weil halt jeder anders strukturiert ist. Also das werden wir auf der europäischen Ebene nicht lösen können, ganz im Gegenteil, das wird nur noch schlimmer. Vereinheitlichen kann man nur auf der Metadatenebene und einen Gesamtüberblick über alle Daten bringen, aber darauf wird gerade gearbeitet und es gibt schon Standardisierungsmaßnahmen auch von W3C, wo der Ticketstandard propagiert wird und der auch auf vielen nationalen Portalen eingesetzt wird. Die österreichischen Metadaten sind beispielsweise Ticketkompatibel.

Bis jetzt haben Sie viele Bedenken, die in der wissenschaftlichen Literatur vorhanden sind bestätigt. Als letztes möchte ich noch fragen: Wie sehen Sie die Zukunft von Open Data, hat es wirklich eine Zukunft oder besteht immer noch die Möglichkeit, dass es scheitern wird?

Ich bin davon überzeugt, dass es nicht scheitern wird, weil die Daten eine ganz zentrale Rolle in der Wirtschaft als auch die Gesellschaft und Politik spielen werden. Durch diese enorme fortschreitende Digitalisierung, wo wir lange noch nicht am Ende sind, werden wir von Tag zu Tag mehr Daten produzieren, weil wir mehr Sensoren in Objketen einbauen werden, davon bin ich überzeugt. Damit entstehen viele Echtzeitsdatenquellen und mit denen kann ich schon eine ganze Menge anfangen, also das ist für mich nur eine Frage der Zeit bis sich das wirklich auf einer breiter Ebene durchsetzt.

A.2 Sonja Fischbauer

Vielleicht erzähle ich irgendwie so die Geschichte wenn ich für OpenData-Portal bei potentiellen Datenerstellern anklopfe und frage wie schaut es aus bei euch. Könnten wir vielleicht eure Daten haben? Dann werde ich zuerst, also oft, da gibt es zwei Möglichkeiten. Entweder kommen wir zu Technikabteilung weil es dort einen Informatiker oder Informatikerin gibt die sich schon mit dem Thema Open Data auseinandergesetzt hat und dann kommen wir zu denen oder in die Marketingabteilung. Und die fragen dann natürlich: "Was bringt uns das?"

Also wirtschaftlichen Gewinn...

Genau. Welchen wirtschaftlichen Gewinn haben wir davon. Und, daran ist natürlich gekoppelt die Frage, was wird daraus entstehen. Für uns ist es Arbeit diese Daten zu öffnen. Wir müssen in neue Infrastruktur investieren, oder wir brauchen unsere Techniker/Technikerinnen dafür, haben wir Aufwand dafür, was wird dann daraus entstehen und was bringt uns das dann sozusagen wieder zurück? Und da ist natürlich einerseits die PR-Schiene. Das bringt euch gute Publizität, weil im Sinne der Transparenz und ihr seid einfach auch noch auf einer anderen Ebene präsent. Was Unternehmen oft denn davon abhängt diesen Schritt zu machen, ist weil sie noch nicht wissen was dann damit passieren wird. Viele haben Angst davor, dass ihre Daten missbräuchlich verwendet werden

Falsche Darstellung?

Genau. Falsche Darstellung. Also zum Beispiel so, dass die negativen Teile der Information rausgeholt werden, sozusagen um zu vergleichen, zum Beispiel öffnungszeiten: "Die haben so lange Öffnungszeiten. Das ist für ihre Mitarbeiter total schlecht. Oder so irgendwas. Oder zum Biespiel die Lagerstände, solche Sachen wo die Unternehmen sehr vorsichtig sind. Und zusätzlich auch eine Komponente ist auch der Teil, dass sie sich denken, was haben wir überhaupt schon für Daten, was ist denn da überhaupt interessant. Also die Unternehmen haben oft Schwierigkeiten, oder es hat sich einfach noch nicht etabliert als Kultur, dass man nicht weiß, welche Daten es gibt und was daraus überhaupt verwendbar ist. Weil zum Beispiel das einfachste ist natürlich so Standortdaten, Öffnungszeiten oder Produktlisten. Das ist so die eine, das ist diese low-level Geschichte. Und andererseits sind es die ganz sensible Daten. So wie zum Beispiel irgendwelche Betriebsge-

heimnisdaten. Und da denken sich die Unternehmen oft. Es gibt das total Geheime, das wollen wir natürlich nicht präsentieren, weil das hilft unserer Konkurrenz und es ist marktschädigend für uns oder das was sowieso niemanden interessiert, wie zum Beispiel in der Autoindustrie: Welche Farben haben unsere Autos, wie viele Farben es gibt... Wem interessiert das? Und das sind dann so die Schwierigkeiten auf die ich stoße, dass ich den Leuten oder potentiellen Datenersteller aus der Wirtschaft dann vermitteln soll, dass es auch irgendwas dazwischen gibt. Dass es nicht nur total fade Daten sein können oder total Betriebsgeheimnisdaten sondern auch Daten, wo man auf den ersten Blick nicht weiß, was man damit machen kann. Das ist so ein Kreislauf. Die Entwickler brauchen Daten damit sie auf Ideen kommen und die Firmen brauchen entwickelte Apps damit sie Daten rausgeben. Das ist so ein "vicious circle"

Du hast erwähnt, dass Unternehmen fürchten, dass private Daten veröffentlicht werden. Ist es vielleicht möglich, dass der Begriff "Open Data" missverstanden wird? Dass sich die Leute unter dem Begriff "Data" wirklich alles vorstellen.

Genau, das ist auch absolut eine Barriere, die man angehen muss. In der Verwaltung hat sich das mit dem Open Data als Begriff schon etabliert. Das ist einfach maschinenlesbar und nicht personenbezogen. Das sind die zwei Schlagworte. Das ist die Definition und in der Open Data Community, diese zwei Wörter braucht man gar nicht zu sagen, aber wenn man auf Leute trifft, und das vergisst man oft, wenn man so außer der Community kommt, wenn man auf Leute trifft, die das Wort noch nie gehört haben, und ich hab auch vor ein Paar Jahren keine Ahnung gehabt. Als ich zum ersten Mal auf Open Data gekommen bin, habe ich mir auch gedacht: Was ist das? Ist das jetzt so: alle Daten sammeln, Facebook, privacy issues und so, aber das ist das Gegenteil davon, das wissen die meiste Leute nicht, und woher auch, weil das ein Thema ist das noch auch mehr Awareness braucht. Genau, da gebe ich dir total Recht, da stimme ich zu diesem Punkt. Unternehmen können mit dem Begriff "Open Data" noch nicht genug anfangen. Da fehlt es noch an Awareness dafür.

Hast du vielleicht eine Idee, wie man den Awareness erhöhen kann? Was kann die Community dafür tun um diese Barriere zu beseitigen?

Also die Community kann einfach... Wir brauchen mehr Maßnahmen im Punkt zu Awareness. Wir brauchen mehr wirtschaftliche Partner. Wir müssen jetzt diese 2. Revolution schaffen, dass wir da auch in der Wirtschaft Fuß fassen mit Open Data. Und die Community [...], dass sie weiterhin ihre kreativen Ideen macht, also dass Entwickler und Entwicklerrinnen weiter entwickeln, dass Leute wie ich in der Kommunikation weiter diese Botschaft nach außen tragen, aber dass wir aufpassen, dass wir auch außerhalb der Verwaltung schauen, dass wir einfach, obwohl schwierig, einfach da immer mehr die Wirtschaft [...]. Wir müssen also als Community schauen, dass wir jetzt den Fokus auf die neue Zielgruppe Wirtschaft richten und da auf darauf vorbereitet sind, dass uns andere Probleme als bei der Verwaltung entgegenkommen und dass wir die richtigen Tools aussuchen, also Use Cases, Business Modelle und neue Ideen für den wirtschaftlichen Bereich.

Eine Barriere, die ich in der Literatur gefunden habe, welche sich mit der wirtschaftlichen Perspektive beschäftigt, war, dass es keine Unterstützung gibt. Keine Guidelines, kein Help-desk. Ein ähnliches Problem gibt es auch bei Open Source, wo es bei vielen Open Source Programmen auch keine Unterstützung gibt. Es ist von der Community programmiert und es gibt keine Gewähr.

Ja, das stimmt. Und das ist ein super Punkt. So kann man auch ein Bisschen ein Business Modell schaffen, man macht eine Open Source Software, aber den Support dafür, den kann man sich dazu dann kaufen. Und solche Sachen brauchen wir auch für die Open Data in der Wirtschaft. Das wir als die Community der Wirtschaft auch helfen, dass wir die Wirtschaft bei der Hand nehmen, weil die haben natürlich anders als die Verwaltung, können die schwieriger ein Budget dafür abstellen und die müssen ihre Ausgaben auf einer anderen Art rechtfertigen. Also die haben nicht die politische Motivation, wie in der Verwaltung. Wir zum Beispiel, wir machen das von OpenDataPortal, wir haben die Ressourcen, dass wir zumindest den Datenerstellern auch technische Fragen beantworten, das wären einerseits diese Guidance geben auf der inhaltlichen Seite, also "Was ist Open Data, warum ist das wichtig, was bedeutet das für das Unternehmen?". Also die Kommunikationsseite aber dann auch die technische Infrastruktur. Und das ist etwas, was ich auch für die Zukunft ausbaufähig sehe, Dienstleistungen anbieten für die Wirtschaft, für diesen Datenöffnungsprozess. Einerseits auf der Kommunikationsebene und andererseits auf dieser technischer Ebene.

Ein Problem ist das sich die Open Data Literatur fast ausschließlich

nur mit dem Government-teil von Open Data beschäftigt. Ein Unternehmen will dann etwas veröffentlichen aber die Literatur beschäftigt sich teilweise nur mit Government.

Ja, da sind ganz andere Parameter, ganz anders Motiviert.

Du hast bestimmt mit Entwicklern gesprochen, haben vielleicht die Entwickler über Barrieren gesprochen? Ein Beispiel der in der Literatur vorkommt ist die fehlende oder langsame API, oder zum Beispiel die niedrige Qualität der Daten.

Also solche Sachen kenne ich auch, dass dieser Datenreinigungsprozess ein großer Zeitfresser ist. Das ist für die Entwickler eine Barriere. Weil solche Open Data Projekte sind für viele Entwickler auch Liebhaberprojekte, oder beginnen als solches. Und die Developers sind auch oft selbständig und haben nicht die Sicherheit eines fixen Einkommens, sondern sorgen sich selbst um ihre Projekte. Viele beginnen mit einem Open Data Projekt, weil es ihnen gefällt, ein Liebhaberprojekt. Oder sie denken sich: "Ich habe die Notwendigkeit dafür, dass es so was gibt, ich programmiere es jetzt für mich selber". Und das ist allein schon ein Aufwand und wenn dann noch dieses Datensäubern so ein großer Zeitfresser ist, dann ist es eine zusätzliche Hürde für die Entwicklerin. Es ist eigentlich eine finanzielle Hürde, weil die Zeit ist Geld, welches man in einem bezahlten Projekt verdienen könnte.

Wieso denkst du, dass solche Daten überhaupt veröffentlicht werden, warum veröffentlichen Institutionen Daten die beschädigt oder nicht maschinenlesbar sind? Könnte das vielleicht deswegen sein, weil es keine gute Standards oder Regeln gibt?

Da bin ich jetzt mehr auf dieser passiven Wissensseite, weil ich aktiv mit den Datensätzen nicht arbeite. Aber was ich mitbekomme ist, dass es schwierig ist wegen der Standards. Das sind einerseits die Lizenzen. Es gibt unterschiedliche offene, zum Beispiel im OpenDataPortal sind alle Datensätze unter der Creative Commons. Aber genau, es gibt einfach wenig Standards. Und auch für solche Datensätze die veröffentlicht werden, da gibt es natürlich Best Practices, aber da ist wirklich oft das Problem, dass sich ein Standard nicht etabliert hat. Und die einzelnen Institutionen wie die Verwaltung dann ihr eigenes Süppchen kochen und in einer Firma schaut es auch ganz anders

aus.

Man würde denken, dass es mit den Lizenzen nicht so viele Probleme geben sollte, weil die Open Definition drei klare Regeln definiert hat. (Availability and access, Reuse and redistribution, Universal participation)

Also ich bin selbst bei Lizenzen auch nicht wirklich super top. Ich kann aber ein Beispiel bringen von OpenDataPortal, da haben wir so ein Citizen Science Projekt. Da gibt es so ein Projekt, das heißt Photo Quest, wo Citizens mit ihren Smartphone Plätze fotografieren können und damit eben zum Sammeln von Science Daten beitragen können. Und die veröffentlichen dann die Daten mit einer Open Data Lizenz. Und wir haben aber auf OpenDataPortal Creative Commons Lizenz. Es ist de facto dann irgendwie dasselbe, aber es ist rechtlich gesehen eine unterschiedliche Lizenz. Wir hatten jetzt auch das Problem, das ist jetzt eine Verwaltungssache auch, zum Beispiel für unseres Projekt Gute Taten Für Gute Daten von der Open Knowledge, da machen wir auch Datensätze wie zum Beispiel Daten zu Asylwerber von Innenministerium öffentlich. Und das sind Daten die dann schon veröffentlicht werden, aber halt in PDF Form. Wir machen die Open, maschinenlesbar und setzen die Daten online, aber da ist auch die Frage: Dürfen wir das jetzt eigentlich veröffentlichen?. Die sind eigentlich als PDF allgemein zugänglich, aber das Rechtliche ist halt eine zusätzliche Komponente.

Wie siehst du die Zukunft von Open Data. Ist es ein Konzept das in der Zukunft scheitern wird?

Das hoffe ich nicht und das denke ich auch nicht. Also dafür ist schon zu viel passiert. Wir sind auf einem Punkt wo wirklich viel passiert ist, und jetzt gibt es eine neue Herausforderung mit dem Business Teil, aber die Community ist schon gestärkt, dass wir schon das Netzwerk und die Möglichkeiten haben, dass wir das auch auf die Wirtschaft drüberbringen, da bin ich zuversichtlich. Wir brauchen neue Ideen, wir müssen umdenken, aber wir haben das Netzwerk, wir haben die Ideen, wir haben die Leute.

Vielleicht braucht Open Data ein Success Story.

Genau. Sucess stories. Das ist das, was wir auch beim OpenDataPortal auf jedem Fall pushen wollen. Also einerseits die Unternehmen inspirieren,

oder anregen dazu, damit sie ihre Daten rausgeben. Mehr Use Cases und success stories brauchen wir.

B Questionnaire

The print version of the questionnaire is below beginning with the next page. Please note that no printed versions were distributed. The survey was accessible only in online form via the address odsurvey.ai.wu.ac.at/OpenDataSurveyAustria

Not all results are included in this thesis, for a complete analysis please visit http://odsurvey.ai.wu.ac.at/results

ADEQUATe Open Data Umfrage

Diese Umfrage wird im Rahmen des Forschungsprojekts ADEQUATe durchgeführt. Das Projekt wird im Rahmen des FTI - Programms „IKT der Zukunft“ durch das Bundesministerium für Verkehr, Innovation und Technologie gefördert und von der Österreichischen Forschungsförderungsgesellschaft abgewickelt [Projektnummer 849982].

Das Ziel dieses Fragebogens ist es, Informationen über Open Data Potenziale und Barrieren zu sammeln. Die Umfrage dauert etwa 5 bis 15 Minuten, je nach Ihrer Bereitschaft, auch optionale Fragen zu beantworten. Der Fragebogen ist in verschiedene Kategorien unterteilt, wobei jede Kategorie mehrere Fragen enthält. Pflichtfragen sind mit einem roten Stern gekennzeichnet. Alle anderen Fragen sind optional und können übersprungen werden, wenn Sie sich bei der Beantwortung nicht sicher sind.

Mit dem Start der Umfrage erklären Sie, dass Sie mit die Teilnahmebedingungen gelesen und verstanden haben.

Teil A: Introductory questions

A1. Ich habe Erfahrung mit Open Data als

- [ ] Benutzer (Suchen von OD, Verwenden von OD für App Entwicklung, Visualisierung, statistische Analyse)
- [ ] Anbieter (Erstellung, Sammlung und Speicherung von Open Data)
- [ ] Beides

A2. Bitte geben Sie die Art der Organisation / Institution für die Sie arbeiten an.

- [ ] Wissenschaft
- [ ] Öffentlicher Sektor/Regierung
- [ ] Nichtstaatliche Organisation (NGO)
- [ ] Unternehmen
- [ ] Sonstiges. Bitte füllen Sie das Textfeld aus

A3. Bitte geben Sie den Namen und die Webseite Ihrer Organisation an.
(*optional*)

Name

URL

A4. Wie sind Sie an Open Data beteiligt?

A5. Mit welchen Kategorien von Open Data haben Sie gearbeitet?

- Arbeit
- Bevölkerung
- Bildung und Forschung
- Finanzen und Rechnungswesen
- Geographie und Planung
- Gesellschaft und Soziales
- Gesundheit
- Kunst und Kultur
- Land und Forstwirtschaft
- Sport und Freizeit
- Umwelt
- Verkehr und Technik
- Verwaltung und Politik
- Wirtschaft und Tourismus

Sonstiges ☐

Sonstiges

A6. Arbeiten Sie mit ..... ?

	1 - Nie	2 - Mehrmals im Jahr	3 - Mehrmals im Monat	4 - Mehrmals in der Woche	5 - Mehrmals am Tag
Geodaten	☐	☐	☐	☐	☐
Statistischen Daten	☐	☐	☐	☐	☐
Wirtschaftsdaten	☐	☐	☐	☐	☐
Sonstiges (Bitte füllen Sie das Textfeld unten aus)	☐	☐	☐	☐	☐

A7. Sonstige Datentypen mit welchen Sie arbeiten:

Teil B: Introductory Questions User

B1. Als Open Data <u>Benutzer</u>, verwende ich die Daten für:

App Entwicklung ☐

Datenanalyse ☐

Datenvisualisierung ☐

Präsentation/Bericht ☐

Webdienst ☐

Verbesserung von bereits existierenden Apps/Webdiensten ☐

Sonstiges ☐

Sonstiges

B2. Verlinken Sie Ihre öffentlich verfügbaren Projekte, für welche Sie Open Data verwendet haben.

B3. Die Datensätze habe ich

- [ ] Von einem Open Data Portal herunterladen, wie z.B. data.gv.at oder opendataportal.at
- [ ] Mithilfe einer API herunterladen
- [ ] Von einem SPARQL endpoint
- [ ] Direkt von der Quelle (Regierung, Unternehmen ...) beantragt
- [ ] Sonstiges

Sonstiges

B4. Was ist Ihre Motivation als Open Data Benutzer?

- [ ] Technische Herausforderung
- [ ] Monetärer Gewinn
- [ ] Wissenschaftliche Forschung
- [ ] Regierungstransparenz
- [ ] Persönliches Interesse an Open Data
- [ ] Sonstiges

Sonstiges

Teil C: Using the data

C1. Welche der folgenden Probleme, halten Sie für eine Barriere als Open Data Benutzer? Bitte bestimmen Sie das Ausmaß der Barrieren, indem Sie die passende Antwort aus der Tabelle wählen.

	1 - Keine Barriere.	2 - Geringe Barriere. (Es war möglich die Datensätze zu verwenden)	3 - Mittelmäßige Barriere. (Es war schwierig mit den Datensätzen zu arbeiten)	4 - Starke Barriere. (Es war sehr schwierig mit den Datensätzen zu arbeiten)	5 - Außerordentliche Barriere. (Hat die Nutzung der Datensätze vollkommen verhindert)
Qualität der Datenportale (wie z.B. data.gv.at oder opendataportal.at)	☐	☐	☐	☐	☐
Datenqualitätsprobleme	☐	☐	☐	☐	☐
Rechtliche Beschränkungen	☐	☐	☐	☐	☐
Unzureichende Fähigkeiten oder Erfahrung	☐	☐	☐	☐	☐

	1 - Keine Barriere.	2 - Geringe Barriere. (Es war möglich die Datensätze zu verwenden)	3 - Mittelmäßige Barriere. (Es war schwierig mit den Datensätzen zu arbeiten)	4 - Starke Barriere. (Es war sehr schwierig mit den Datensätzen zu arbeiten)	5 - Außerordentliche Barriere. (Hat die Nutzung der Datensätze vollkommen verhindert)	Keine Antwort.
Kein API angeboten	☐	☐	☐	☐	☐	☐
API langsam oder reagiert nicht	☐	☐	☐	☐	☐	☐

	1 - Keine Barriere.	2 - Geringe Barriere. (Es war möglich die Datensätze zu verwenden)	3 - Mittelmäßige Barriere. (Es war schwierig mit den Datensätzen zu arbeiten)	4 - Starke Barriere. (Es war sehr schwierig mit den Datensätzen zu arbeiten)	5 - Außerordentliche Barriere. (Hat die Nutzung der Datensätze vollkommen verhindert)	Keine Antwort.
API Beschränkungen (wie z.B. nur eine beschränkte Anzahl von Abfragen möglich)	☐	☐	☐	☐	☐	☐
Registrierung erforderlich, bevor man Zugriff auf die Datensätze erhält	☐	☐	☐	☐	☐	☐
Kompliziertes Browsing/Suche (Kein Verzeichnis oder andere Methoden um die Suche zu erleichtern)	☐	☐	☐	☐	☐	☐
Keine Informationen über die Qualität der Daten	☐	☐	☐	☐	☐	☐
Keine Informationen über den Inhalt der Datensätze	☐	☐	☐	☐	☐	☐
Sprachbarrieren	☐	☐	☐	☐	☐	☐
Daten stehen vorübergehend nicht zum Download zur Verfügung	☐	☐	☐	☐	☐	☐
Doppelte Datensätze	☐	☐	☐	☐	☐	☐
Nur Daten die keinen Mehrwert erstellen veröffentlicht (Unbrauchbare Daten für Sie oder Ihr Projekt)	☐	☐	☐	☐	☐	☐
Kein zentrales Portal für Open Data, zersplitterte Quellen	☐	☐	☐	☐	☐	☐
Zugriff auf die Datensätze ist auf eine bestimmte Gruppe von Benutzern eingeschränkt	☐	☐	☐	☐	☐	☐
Daten sind nicht frei verfügbar (nicht kostenlos)	☐	☐	☐	☐	☐	☐

C3. Bitte geben Sie andere Barrieren mit Datenportale an, die oben nicht erwähnt sind.

C4. Als Open Data Benutzer, glauben Sie, dass Organisationen ihre Daten unabhängig von der Datenqualität veröffentlichen sollten?

Ja (Daten mit schlechter Datenqualität sollen veröffentlicht werden. Je mehr Open Data desto besser.) ☐

Nein (Daten mit schlechter Datenqualität sollen nicht veröffentlicht werden.) ☐

C5. Als Open Data Benutzer, glauben Sie, dass Organisationen Qualitätsstandards und Regeln für die Datenveröffentlichung vorgegeben werden sollten? Wenn ja, bitte geben Sie beispielhaft solche Regeln an.

Ja. (Organisationen sollten bestimmten Regeln und Standards folgen müssen.) ☐

Nein. (Das Erfordern gewisser Qualitätsstandards und Regeln würde vielen Organisationen die Datenveröffentlichung verhindern.) ☐

C6. **Wenn Sie in der ersten Frage <u>Datenqualitätsprobleme</u> als Barriere identifiziert haben, bewerten Sie bitte das Ausmaß der folgenden Barrieren von 1 - "Keine Barriere" bis zu 5 - "Außerordentliche Barriere".**

	1 - Keine Barriere.	2 - Geringe Barriere. (Es war möglich die Datensätze zu verwenden)	3 - Mittelmäßige Barriere. (Es war schwierig mit den Datensätzen zu arbeiten)	4 - Starke Barriere. (Es war sehr schwierig mit den Datensätzen zu arbeiten)	5 - Außerordentliche Barriere. (Hat die Nutzung der Datensätze vollkommen verhindert)	Keine Antwort.
Ungenauigkeit der Daten	☐	☐	☐	☐	☐	☐
Veraltete, nicht gültige Daten	☐	☐	☐	☐	☐	☐
Falsches Encoding (z.B. nicht UTF-8)	☐	☐	☐	☐	☐	☐
Datei ist nicht maschinenlesbar	☐	☐	☐	☐	☐	☐
Unrichtigkeit der Metadaten	☐	☐	☐	☐	☐	☐
Unvollständigkeit der Daten	☐	☐	☐	☐	☐	☐
Nutzlose, wertlose Daten	☐	☐	☐	☐	☐	☐

C7. **Bitte geben Sie andere <u>Datenqualitätsbarrieren</u> an, die oben nicht erwähnt sind.**

C8. **Wenn Sie in der ersten Frage die <u>rechtlichen Beschränkungen</u> als Barriere identifiziert haben, bewerten Sie bitte das Ausmaß der folgenden Barrieren von 1 - "Keine Barriere" bis zu 5 - "Außerordentliche Barriere".**

	1 - Keine Barriere.	2 - Geringe Barriere. (Es war möglich die Datensätze zu verwenden)	3 - Mittelmäßige Barriere. (Es war schwierig mit den Datensätzen zu arbeiten)	4 - Starke Barriere. (Es war sehr schwierig mit den Datensätzen zu arbeiten)	5 - Außerordentliche Barriere. (Hat die Nutzung der Datensätze vollkommen verhindert)	Keine Antwort.
Restriktive Lizenzen	☐	☐	☐	☐	☐	☐
Unklare Lizenzen (Unklare Bedingungen für die Wiederverwendung)	☐	☐	☐	☐	☐	☐

	1 - Keine Barriere.	2 - Geringe Barriere. (Es war möglich die Datensätze zu verwenden)	3 - Mittelmäßige Barriere. (Es war schwierig mit den Datensätzen zu arbeiten)	4 - Starke Barriere. (Es war sehr schwierig mit den Datensätzen zu arbeiten)	5 - Außerordentliche Barriere. (Hat die Nutzung der Datensätze vollkommen verhindert)	Keine Antwort.
Komplexe, schwerverständliche Lizenzen	☐	☐	☐	☐	☐	☐

C9. Bitte geben Sie andere Probleme mit <u>rechtlichen Beschränkungen</u> an, die oben nicht erwähnt sind.

C10. Wenn Sie in der ersten Frage die <u>unzureichenden Fähigkeiten oder Erfahrung</u> als Barriere identifiziert haben, bewerten Sie bitte das Ausmaß der folgenden Barrieren von 1 - "Keine Barriere" bis zu 5 - "Außerordentliche Barriere".

	1 - Keine Barriere.	2 - Geringe Barriere. (Es war möglich die Datensätze zu verwenden)	3 - Mittelmäßige Barriere. (Es war schwierig mit den Datensätzen zu arbeiten)	4 - Starke Barriere. (Es war sehr schwierig mit den Datensätzen zu arbeiten)	5 - Außerordentliche Barriere. (Hat die Nutzung der Datensätze vollkommen verhindert)	Keine Antwort.
Keine Unterstützung/Helpdesk	☐	☐	☐	☐	☐	☐
Fehlende Dokumentation	☐	☐	☐	☐	☐	☐
Nicht ausreichende Schulungen	☐	☐	☐	☐	☐	☐

C11. Bitte geben Sie andere mögliche Gründe für die <u>unzureichenden Fähigkeiten oder Erfahrung</u> an, die oben nicht erwähnt sind.

C12. Bitte geben Sie alle andere potentielle Barrieren an, auf die Sie als Open Data <u>Benutzer</u> gestoßen sind und die in den letzten Fragen nicht erwähnt wurden.

Teil D: Introductory questions provider specific

D1. Was ist Ihre Motivation für die Veröffentlichung der Daten als Open Data Anbieter?

- [ ] Regierungstransparenz
- [ ] Crowdsourcing der Software Entwicklung
- [ ] Nachfrage der Kunden/Stakeholder
- [ ] Wissenschaftliche Forschung
- [ ] Persönliches Interesse an Open Data
- [ ] Monetärer Gewinn
- [ ] Erfüllung von Rechtsvorschrfiten
- [ ] Mehr Transparenz in den Kundenbeziehungen
- [ ] Sonstiges

Sonstiges

Teil E: Opening the data

	1 - Keine Barriere.	2 - Geringe Barriere. (Geringe Erschwernis den Datensatz zu veröffentlichen)	3 - Mittelmäßige Barriere. (Die Veröffentlichung der Datensätze war schwierig)	4 - Starke Barriere. (Die Veröffentlichung der Datensätze war sehr schwierig)	5 - Außerordentliche Barriere. (Hat die Veröffentlichung der Datensätze vollkommen verhindert)	Keine Antwort.
Privatsphäre - Unerwünschte Aufdeckung der Daten	☐	☐	☐	☐	☐	☐
Verlust der Kontrolle über die freigegebenen Datensätze	☐	☐	☐	☐	☐	☐
Die Angst vor falschen Schlussfolgerungen aus den Daten/Falsche Darstellung	☐	☐	☐	☐	☐	☐
Open Data kann zur Verfälschung der Daten führen	☐	☐	☐	☐	☐	☐
Sicherheitsbedrohungen, Schwachstellen	☐	☐	☐	☐	☐	☐
Niedrige Qualität der Daten, besser nicht enthüllen	☐	☐	☐	☐	☐	☐

E2. Bitte geben Sie andere Datenschutz- und Sicherheits-Barrieren an, die oben nicht erwähnt sind.

E3. Als Open Data Anbieter, was sind Ihre strategischen und geschäftlichen Barrieren?

	1 - Keine Barriere.	2 - Geringe Barriere. (Geringe Erschwernis den Datensatz zu veröffentlichen)	3 - Mittelmäßige Barriere. (Die Veröffentlichung der Datensätze war schwierig)	4 - Starke Barriere. (Die Veröffentlichung der Datensätze war sehr schwierig)	5 - Außerordentliche Barriere. (Hat die Veröffentlichung der Datensätze vollkommen verhindert)	Keine Antwort.
Mangel an Open Data Geschäftsmodellen	☐	☐	☐	☐	☐	☐
Open Data ist derzeit keine Priorität	☐	☐	☐	☐	☐	☐
Widerstand gegen Veränderungen, risikoscheue Unternehmenskultur	☐	☐	☐	☐	☐	☐
Kein Verkauf der Daten möglich, wenn diese mit einer Open Data Lizenz veröffentlicht sind	☐	☐	☐	☐	☐	☐
Fehlende strategische Planung/Management mit Open Data	☐	☐	☐	☐	☐	☐
Störung des bestehenden Geschäftsmodells (Es wäre nicht mehr möglich, Bezahlung für die Daten zu verlangen)	☐	☐	☐	☐	☐	☐
Unsichere wirtschaftliche Auswirkungen	☐	☐	☐	☐	☐	☐
Mangel an Ressourcen und Zeit	☐	☐	☐	☐	☐	☐
Kostenaufwand (wie z.B. die Notwendigkeit für zusätzliche Arbeitskräfte für Wartung. Schulung/Lernkosten. Datenspeicherung, usw...)	☐	☐	☐	☐	☐	☐

E4. Bitte geben Sie andere strategische und geschäftliche Barrieren an, die oben nicht erwähnt sind.

E5. Als Open Data Anbieter, was sind Ihre rechtlichen Barrieren?

	1 - Keine Barriere.	2 - Geringe Barriere. (Geringe Erschwernis den Datensatz zu veröffentlichen)	3 - Mittelmäßige Barriere. (Die Veröffentlichung der Datensätze war schwierig)	4 - Starke Barriere. (Die Veröffentlichung der Datensätze war sehr schwierig)	5 - Außerordentliche Barriere. (Hat die Veröffentlichung der Datensätze vollkommen verhindert)	Keine Antwort.
Unklares Eigentum der Daten verbietet die Veröffentlichung.	☐	☐	☐	☐	☐	☐
Klagedrohungen	☐	☐	☐	☐	☐	☐

	1 - Keine Barriere.	2 - Geringe Barriere. (Geringe Erschwernis den Datensatz zu veröffentlichen)	3 - Mittelmäßige Barriere. (Die Veröffentlichung der Datensätze war schwierig)	4 - Starke Barriere. (Die Veröffentlichung der Datensätze war sehr schwierig)	5 - Außerordentliche Barriere. (Hat die Veröffentlichung der Datensätze vollkommen verhindert)	Keine Antwort.
Komplexe, schwerverständliche Lizenzen	☐	☐	☐	☐	☐	☐

E6. Bitte geben Sie andere <u>rechtliche</u> Barrieren an, die oben nicht erwähnt sind.

E7. Als Open Data <u>Anbieter</u>, was sind Ihre <u>technischen</u> Barrieren?

	1 - Keine Barriere.	2 - Geringe Barriere. (Geringe Erschwernis den Datensatz zu veröffentlichen)	3 - Mittelmäßige Barriere. (Die Veröffentlichung der Datensätze war schwierig)	4 - Starke Barriere. (Die Veröffentlichung der Datensätze war sehr schwierig)	5 - Außerordentliche Barriere. (Hat die Veröffentlichung der Datensätze vollkommen verhindert)	Keine Antwort.
Verschiedene Normen und Formen der maschinenlesbaren Daten (JSON, XML, CSV ...)	☐	☐	☐	☐	☐	☐
Keine Standard-Software für die Datenverarbeitung (Fragmentierung der Software)	☐	☐	☐	☐	☐	☐
Keine Metadaten-Standards	☐	☐	☐	☐	☐	☐
Fehlende Dokumentation	☐	☐	☐	☐	☐	☐
Keine Unterstützung/Helpdesk	☐	☐	☐	☐	☐	☐

E8. Bitte geben Sie andere <u>technische</u> Barrieren an, die oben nicht erwähnt sind.

E9. Bitte geben Sie andere potentielle Barrieren an, auf die Sie als Open Data <u>Anbieter</u> gestoßen sind und in den letzten Fragen nicht erwähnt wurden.

Teil F: Final questions

F1. Wie wichtig ist Open Data für Ihr Projekt oder Organisation?

	Sehr unwichtig	Eher unwichtig	Weder/noch	Eher wichtig	Sehr wichtig	Keine Antwort
Wichtigkeit	☐	☐	☐	☐	☐	☐

F2. Werden Sie mit Open Data auch in der Zukunft arbeiten?

Ja ☐

Nein ☐

F3. Stimmen Sie der Aussage "Open Data war für mein Projekt/meine Organisation vorteilhaft." zu?

1 - Stimme gar nicht zu	2	3	4	5 - Stimme völlig zu	Keine Antwort
☐	☐	☐	☐	☐	☐

Teil G: Interview

G1. Sind Sie daran interessiert, dieses Thema mit uns noch persönlich zu diskutieren?

Ja ☐

Nein ☐

Teil H: Kontakt

H1. Bitte füllen Sie dieses Formular mit Ihren Kontaktdaten aus, damit wir Sie für ein Interview einladen können.

Name

E-Mail

Organisation

Telefonnummer

Vielen Dank für die Teilnahme an der Befragung!