

# Open Data Hopes and Fears

## Determining the barriers of Open Data

Martin Beno, Kathrin Figl, Jürgen Umbrich, Axel Polleres  
Department of Information Systems and Operations Management  
Vienna University of Economics and Business (WU Wien)  
Vienna, Austria  
e-mail: [firstname.lastname]@wu.ac.at

**Abstract**—In recent years, Open Data has gained considerable attention: a steady growth in the number of openly published datasets – mainly by governments and public administrations - can be observed as the demand for Open Data rises. However, many potential providers are still hesitant to open their datasets and at the same time users often face difficulties when attempting to use this data in practice. This indicates that there are still various barriers present both regarding usage and publishing of Open Data, but studies that systematically collect and assess these barriers regarding their impact are rare. Based on this observation we survey prior literature on barriers, and have developed a questionnaire aimed at both assessing the users and publishers views on obstacles regarding Open Data adoption. Using a sample of over 100 participants from Austria who completed our online survey, we draw conclusions about the relative importance of the barriers reported in the literature. The empirical findings presented in this study shall serve as a solid foundation for future research on the mitigation of Open Data barriers.

**Keywords:** *Open Data; barriers; challenges; publishing data; using data; data quality; transparency*

### I. INTRODUCTION

Over the past few years the Open Data movement has been steadily growing, especially fueled by governments and public institutions ever since the launch of the US open data portal <http://data.gov> 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 [1], [2];
- Enabling independent developers to develop *value-added services* and applications based on published Open Data, thus creating applications with social and economic value and consequently stimulating economic growth [3];
- Promoting *collaboration* of citizens with the government [4], [5].

Indeed, various providers including government, public and private institutions all over the world have been steadily releasing open datasets. As we have observed from Open Data Portal Watch project, cf. <http://data.wu.ac.at/portalwatch>, a public framework that continuously monitors a large number of public Open Data portals, the number of published open datasets is increasing at a steadily rate [6].

Although Open Data is in no way limited to government use only, most of the publishers of data sets on Open Data portals are public institutions, which is why most attention around Open Data focuses on Open Government Data. This is also reflected in academia, as only limited research regarding private entities publishing Open Data is available. The fact that private institutions do not buy in at larger scale on the promises of sharing parts of their data openly the same way that public institutions do, plus the observation that larger companies neither report back on the utility of consuming and re-using published Open Data could indicate that there are still some barriers regarding publishing and using Open Data. Indeed, a recent study by Janssen, Charalabidis and Zuiderwijk [3] claims that despite numerous benefits, Open Data adoption entails a number of barriers. This is further supported by a study by Martin, Foulonneau, Turki and Ihadjadene [7], which points out that a wider adoption of Open Data still faces significant barriers. However, a deeper understanding of the relative importance and impact of these barriers still seem to require more research, in order to effectively lower entry barriers. To this end, the aim of our study is to examine existing barriers in using and publishing Open Data and consequently assess their importance and relevancy using quantitative methods. Our research question can therefore be summarized as follows: “How important and relevant are the currently reported barriers in using and publishing Open Data?”

Firstly, we have conducted a systematic literature review in order to establish a solid theoretical background for the construction of a questionnaire, which was then sent out to a large number of participants in Austria as the focus of our study was an assessment at national level within a national research project;<sup>1</sup> we still believe that our results contain various valuable and potentially transferrable insights for Open Data efforts internationally. Our main goal was to empirically assess the importance and relevancy of Open Data barriers for users and providers in practice.

---

<sup>1</sup> ADEQUATE Open Data: Analytics & Data Enrichment to improve the QUALiTy of Open Data, cf. <http://www.adequate.at/>

Various research results on the topic of Open Data barriers have already been published in recent years. Perhaps the most complete list of barriers has been collected by Zuiderwijk et al.'s study [8] which provides a systematic overview of Open Data socio-technical impediments. While this study does present an extensive and categorized set of barriers, it lacks a detailed analysis, which could help determine the relative importance and relevancy of these barriers. This limitation has been partially lifted by a later study, in which a questionnaire was developed in order to generalize their findings [9]. However, despite their questionnaire being distributed in multiple countries, their research and results analysis was focused on Netherlands only. The study suggests comparing barriers and development directions in different countries, which we partially contribute to in the present paper, extending and adapting their approach; however, we also observe that direct comparison is difficult due to different development stages and community characteristics in different countries. Another survey on Open Data barriers was carried out as part of a case study on a semi-public organization in Netherlands [10]. This survey however, was limited to participants from the aforementioned organization and as such, those results cannot be assumed to hold true for all Open Data cases. The studies on Open Data barriers are often limited to analyzing data publishing only and the barriers faced by data users and re-users are often ignored. Indeed, Janssen, Charalabidis and Zuiderwijk claim that more research into ways of dealing with barriers and more insight into the user's perspective is necessary before open data systems will be freely adopted [3]; to this end, herein we explicitly aim at considering both the publishers' and the consumers' views. Furthermore, many studies are focused on governmental data publishing and use exclusively, not taking in account the barriers faced by the private sector. To fill these research gaps, the present study is not limited to Open Data barriers in government and public institutions only. We rather strive to establish a solid basis for a research on Open Data concerns and barriers in general, that could ultimately help mitigate any perceived problems including those faced by private entities.

In the remainder of this paper, we first present our findings from the literature review in the form of a summary of barriers of Open Data (Section II). This is followed by presenting the method and design of our online questionnaire (Section III), the results of which are analyzed in Section IV. We conclude our study with a discussion of results and limitations (Section V) and an outlook to future work as well as general conclusions in Section VI.

## II. BARRIERS TO OPEN DATA

This section gives an overview of the most relevant barriers as reported in the research literature. The overview of the barriers is divided into three main parts, based on whether they apply to data users, providers, or both. Similar distinctions have already been made in previous studies: Janssen, Charalabidis and Zuiderwijk [3] 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). Furthermore, in their risk analysis on Open Data barriers, Martin, Foulonneau, Turki and Ihadjadene claim that the challenges and constraints faced by re-users of public data differ from the ones encountered by the public data providers [7]. In the present study, each barrier has then been assigned to one of several categories. While this approach may be partially inconsistent with other studies - since not all literature provides a clear categorization of barriers and those who do, often categorized the same barrier under a category with a different name - we needed to stick to this assignment for a more digestible design of the online questionnaire, i.e., for instance not bothering pure data users with obstacles only applying to data publishers and vice versa. 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 a short introduction of the found literature barriers. We also provide an overview of each category in a table. The tables summarize the barriers found per category along with the resp. literature references and were also used as questionnaire items in our survey.

### A. Literature Search

We collected a base of articles on Open Data barriers from various online databases, such as Science Direct, ACM Digital Library and Springer Link and Google Scholar, a citation-indexing service. The term "Open Data" has been used as a basis in the search, which we combined with following keywords in our search strings: "barrier", "challenge", "impediment", "problem", "issue", "use", and "publishing". Lastly, we manually reviewed the referenced publications in the literature found in the digital libraries. The search and a manual scan for relevance performed by viewing titles and abstracts yielded 29 publications. In a next step, we excluded those articles, which have not met our selection criteria. The reasons for not accepting several publications were:

- Various issues with Open Data use and publishing described were not explicitly labeled as a barrier, challenge or an impediment.
- The study was limited to a very specific scenario, such as being heavily focused on a particular government institution.

- Open Data barriers were not the main focus of the publication and any potential barriers mentioned were merely assumptions not backed by proper research.

Ultimately, we have chosen 13 articles from which we extracted a list of barriers.

### *B. Extraction of Barriers*

Initially, nearly a hundred potential barriers were found and documented. We have eliminated duplicates and reduced the list to cover the most significant barriers. Our goal was to reach a number of barriers that would be feasible to be included in a single online questionnaire (without overburdening the participants). Our intention was to include only barriers which are relevant towards most Open Data users and publishers in our survey. For that reason, we have also excluded barriers, which are applicable to governmental organizations only. The barriers, that ultimately ended up in the final list are those, which have been either mentioned most often or have been described as significant in the literature.

### *C. User Barriers*

What follows is an overview of barriers, which make it difficult or inhibit the consumption and reuse of published open datasets. The process of Open Data use typically begins with retrieval of the data, which is mostly done by downloading them from an Open Data portal. Once retrieved, the users are often faced with corrupted, incomplete or otherwise unusable data. Furthermore, the complex licensing situation may restrict the users' rights to freely use these data in arbitrary applications, and in worst case scenarios, lead to lawsuits. Lastly, the lack of guidelines, support or a helpdesk makes the use of Open Data difficult for newcomers.

#### *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 their users. Indeed, in their analysis of the quality of open government data portals, Martín, De Rosario and Peréz [11] concluded that there is no single model of Open Data portals. It is rather a wide range of structures, with different processes and features for data search, different formats for presenting information, and a diversity of means to classify the information. Indeed we can confirm a large heterogeneity not only across but also within portals regarding formats, as well as completeness and accuracy of metadata descriptions [6]. Heterogeneity of 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, Charalabidis and Zuiderwijk [3] reports several issues ranging from barriers that only make the user experience worse, such as no information about the datasets provided by the portal, to barriers that heavily restrict the user. Some portals do not allow access to data unless the user registers on their website or even worse, put the data behind a paywall.

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, 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, Eberius, Thiele and Lehner [12] describe the existence of APIs as a feature of Open Data platforms, which is obviously necessary for automated consumption and re-use of data in other applications. Yet, in their analysis of over fifty Open Data portals in 2012, they have found that 43% of those do not feature an API of any kind. The situation has possibly improved in the recent years, as quasi-standardized platforms such as CKAN, which provide an API by default, unless disabled on purpose, are becoming more prevalent. According to the portal monitoring site Open Data Portal Watch [6], CKAN is currently the most widely used Open Data platform, followed by Socrata, OpenDataSoft, and other proprietary solutions. The existence of an API however is often not enough. APIs must also meet certain criteria to be deemed usable. An interview, which was carried out as a part of an extensive study on socio-technical impediments of Open Data by Zuiderwijk et al. [8], reports the lack of a good API as a barrier. 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 at that point of time 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, Eberius, Thiele and Lehner [12] study of the same year. Nevertheless, APIs may also have some limitations [8]. To further investigate this matter, questions about the quality of APIs are also included in the questionnaire of the present study. A summary on all reported user-side data portal barriers that we further investigated in our survey is shown in Table I.

TABLE I. OPEN DATA PORTALS BARRIERS

| <i>Barrier</i>  | <i>References</i>   |
|---|---------------------|
| No API provided                                       | [8], [12], [7]      |
| API slow or unresponsive                              | [8]                 |
| API limitations                                       | [8], [12]           |
| Registration required before gaining access to data   | [8], [7], [3]       |
| Difficult browsing/searching                          | [8], [3], [9]       |
| No information about the quality of the data          | [8], [3], [14], [9] |
| No information given about the content of the dataset | [8], [3]            |
| Language barriers                                     | [8], [7], [9], [11] |
| Data is not available for download                    | [8], [12]           |
| Duplicate datasets                                    | [8], [3]            |
| Only non-value-adding data published                  | [3]                 |
| No central portal for the data                        | [3], [9], [8]       |
| Restricted access to a certain group of users         | [8]                 |
| Data is behind a paywall                              | [8], [7], [3]       |

### 2) Data quality

After retrieval, users may find out that the quality of the obtained data is not meeting their requirements. Perhaps the biggest barrier for users who aim at reusing data in their own applications is that the data is not in a machine-readable format. Scanned documents in the form of pdf files 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, Eberius, Thiele and Lehner [12] 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. In the same vein, Alani et al. [13] note that lacking structure and semantic representation of data make it almost impossible for machines to understand the data, inhibiting its reuse. A summary of main data quality barriers from the literature, which we again selected to be included in the questionnaire, is shown in Table II.

TABLE II. DATA QUALITY BARRIERS

| <i>Barrier</i>               | <i>References</i>  |
|------------------------------|--------------------|
| Data inaccuracy              | [8], [3], [15]     |
| Obsolete, non-valid data     | [3]                |
| Not machine readable         | [8], [12]          |
| Metadata inaccuracy          | [8]                |
| Data incomplete              | [8], [7], [3], [9] |
| No potential use of the data | [3]                |

### 3) User legal constraints

Users further face various legal constraints on the datasets, imposed by terms of usage and - again heterogeneous - licenses. It might seem that by definition, understanding the legal position of the user would be simple, as Open Data is often defined by the fact that the users are able to freely access, reuse and further redistribute the data, cf. for instance the popular Open Definition at <http://opendefinition.org/>. Unfortunately, this definition 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. In fact, comparison to Open Source Software can be made here: complex licensing situation is nothing new in the field of Open Source Software. Indeed, as of now the Open Source Initiative currently lists 78 different licenses. The complex licensing situation is also discussed by Stol and Ali Babar [16] 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. 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. Similar to publishing code without selecting a specific Open Source license, publishing data without a license makes it unclear whether the data can be freely reused and further redistributed. As a consequence, the users may fear legal consequences resulting in lawsuits. Even if the probability of winning such a 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. Efforts should be made to publish data with an included license, which clearly defines the rights of the user. Table III summarizes the major legal constraints as found in the literature includes in our survey.

TABLE III. USER LEGAL BARRIERS

| <i>Barrier</i>                                   | <i>References</i> |
|--|-------------------|
| Threat of lawsuits                               | [8], [3]          |
| Restrictive licenses                             | [8], [15]         |
| Unclear licensing (Unclear conditions for reuse) | [15], [8]         |
| Complex, hard to understand licenses             | [8]               |

#### D. Provider barriers

While we have focused so far on user-side barriers that inhibit the consumption of Open Data, providers likewise face orthogonal barriers. Again, we have categorized the barriers reported in the literature for inclusion in our survey.

##### 1) Privacy and security

When publishing the data, providers must make sure that they will not publish any private or otherwise sensitive information. Before the data is published, necessary precautions need to be made in order to ensure privacy [17]. There are also cases in which disclosing information might harm the reputation of the provider. Particularly, potential publishers often fear that false conclusions may be drawn from the data. This topic is discussed in the study by Conradie and Choenni [18]. As a possible example, they provide a scenario where property value decreases if details about policy plans of new city developments would surface, especially if these were not yet finalized. 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 [17]. Ensuring privacy (for instance by ensuring an acceptable level of anonymization) however, takes a lot of effort that many of potential providers are not willing to undertake. Lastly, providers need to maintain a secure publishing infrastructure. Data integrity and authenticity must be ensured in order to avoid malicious data manipulation by a third party that could potentially damage the publisher's reputation or cause additional costs and workload for the publisher. Vulnerabilities in data publishing platforms, as well as the cost of maintenance of a secure publishing platform present an entry barrier for publishers. Overall, these publisher-side barriers are summarized in Table IV.

TABLE IV. PRIVACY AND SECURITY BARRIERS

| <i>Barrier</i>   | <i>References</i>                    |
|--|--------------------------------------|
| Privacy – Unwelcomed exposure of the data                              | [3], [15], [19], [10], [7], [8], [9] |
| Loss of control over released information                              | [15]                                 |
| Fear of false conclusions drawn from the data – Data misrepresentation | [15], [17], [18]                     |
| Open Data may lead to corruption or falsification of the data          | [17]                                 |
| Security threats, vulnerabilities                                      | [3], [19]                            |
| Low quality data, better not expose it                                 | [15]                                 |

##### 2) Strategic and business

Strategic and business decisions are common reasons why many potential providers do not publish their data. Interviews and workshops conducted by Conradie and Choenni [18] for instance clearly demonstrated that releasing Open Data was not a priority for the local governments they were investigating. Combined with the fact that data release is not part of the regular work for many data professionals, the immediate benefits of data release are not always explicit, making it an extra task, without clear incentives and value. 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, drawing again parallels to Open Source Software, many software developers do not open source their code in fear of lost sales. Morgan and Finnegan [20] for instance describe giving away the source code as value-impeding. For some companies, source code enables value capture if it is used for products, which they can sell. Likewise, some potential data providers might use the data to develop data 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 providers' product. Furthermore, the general lack of time, resources or simply additional expenditure is a major barrier in publishing Open Data. The costs do not end with the data being published. 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. Major strategical and business barriers found in the literature are again presented in Table V.

TABLE V. STRATEGIC AND BUSINESS BARRIERS

| <i>Barrier</i>   | <i>References</i>    |
|--|----------------------|
| Lack of business models  | [8], [10]            |
| Open Data is not a priority  | [18]                 |
| Resistance to change, risk-averse culture                                    | [15], [3]            |
| No sale of data possible when an Open Data license is used                   | [3]                  |
| Embedding Open Data in the strategy and work process                         | [10]                 |
| Disruption of existing business model. (Such as charging money for the data) | [15], [3], [7], [18] |
| Uncertain economic impact  | [15], [19], [7]      |
| Lack of resources and time   | [3], [7]             |
| Cost issues  | [7]                  |

### 3) Provider legal constraints

The complex data licensing situation is a barrier inhibiting data release as well on the publisher side: providers also face the threat of lawsuits, albeit not only because of violations of the licenses, but also when they release private or otherwise sensitive data. Particularly, often this includes issues of data ownership. If the provider is not sure about the ownership of the data, they are not able to release them. Conradie and Choenni [18] 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. A summary of legal constraints for providers is displayed in Table VI.

TABLE VI. PROVIDER LEGAL BARRIERS

| <i>Barrier</i>   | <i>References</i> |
|--|-------------------|
| Unclear ownership of the data, prohibiting its release | [18]              |
| Threat of lawsuits                                     | [9], [8]          |
| Complex, hard to understand licenses                   | [8]               |

### 4) Technical barriers

There are several technical barriers that might make publishing data difficult for providers. It is often expected from the provider to release data in various formats, but Martin, Foulonneau, Turki and Ihadjadene [7] state that there is currently no systematic policy of opening datasets in multiple formats. It is up to the data creators to decide on which formats to publish. Janssen, Charalabidis and Zuiderwijk [3] also report that having no standard software for data processing presents a technical barrier. Finally, the absence of metadata standards has been mentioned as a barrier [8]. It should be noted that some efforts towards standardizing are already in progress. The RDF specification, recommended by the World Wide Web Consortium, is an example of such an effort. Table VII summarizes the technical barriers faced by data publishers.

TABLE VII. TECHNICAL BARRIERS

| <i>Barrier</i>   | <i>References</i> |
|--|-------------------|
| Various standards and forms of machine readable data                     | [7], [8]          |
| No standard software for processing the data (Fragmentation of software) | [3]               |
| Absence of metadata standards  | [7], [8]          |

### E. User and provider

Some Open Data barriers, which we summarize in this third category, might affect users and publishers alike, though from different angles. The actual use and publishing of the data is often difficult, especially for newcomers in the area in terms of available support that puts cost load on providers on the one hand and the lack of inhibits use on the other hand. Open Source Software faces a similar issue. In his study on Open Source Software rejections in Australia's top firms, Goode [21] found that companies often prefer commercial, proprietary versions of software, because proper support is usually offered by the developer. 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. This could very well be an issue in Open Data as well. Indeed, the lack of support or helpdesk has been identified as a barrier for data users by Janssen, Charalabidis and Zuiderwijk [3]. However, 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 provide the users with professional support, especially considering the data was free of charge, simple instructions and guidelines would be a first step in mitigating these barriers.

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. [22] state that opening up the data is a complex and ill-understood activity, because many barriers counteract these processes. Guidelines and instructions published on data portals could help potential providers with the process. Such instructions would also most likely increase the quality of published data. Indeed, some data remains

unpublished as it is neither clear that there would be a demand from users to the providers, or the users can't communicate this demand since the availability of said data is not known to them. Table VIII presents an overview of barriers, which apply both to data users and to publishers.

TABLE VIII. KNOWLEDGE AND EXPERIENCE

| <i>Barrier</i>                     | <i>References</i>   |
|------------------------------------|---------------------|
| No support                         | [3], [8], [9]       |
| No helpdesk                        | [3], [8]            |
| Lack of documentation              | [7]                 |
| No guiding principles/instructions | [22], [8], [9], [8] |

### III. METHOD

#### A. Questionnaire construction

Based on the information about potential barriers obtained from the literature review, an online survey was designed with the open source survey web application Limesurvey, cf. <https://www.limesurvey.org/>.

The questionnaire has a modular structure with separate modules for users and providers. The modules were presented to the participants according to their indicated experience as a user, a provider, or both. In addition, basic information about the participants' background is asked in the introductory phase of the survey such as the kind of organization they are associated with, the categories of interest for them and the general motivation for publishing or using Open Data.

In order to measure the importance of the barriers, we let participants rate each identified barrier on a 5-point Likert scale:

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

Additionally, we added an open text field for each category of barriers, in which participants are free to comment on any barriers they have encountered that were not previously mentioned in the questionnaire. If comments were written in German, we translated them to English.

#### B. Participants

The survey was launched in November 2015 and remained active until March 2016. We sent an invitation to participate to through various channels in Austria, addressing both user communities, e.g. through relevant social media and "meetup" groups, through the Austrian Open Government Data cooperation (Cooperation OGD Österreich) which reaches most public bodies involved in Open Data, as well as through academic/research channels among Austrian researchers interested in Open Data topics. By the end of the period, 310 individuals have launched the survey. However, more than half of them only viewed the first page without answering a single question. Overall, 110 participants completed the entire survey.

Concerning the type of institution, most participants worked for public institutions such as in academia (36%) or government (27%). It could be argued, that the high prevalence of participants from academia was due to the way the survey was promoted, as university colleagues have helped to distribute the survey. Despite our efforts to reach the private sector, Enterprises (15%) were rather underrepresented; a possible explanation may be the high popularity Open Data enjoys in the public sector as opposed to still less adoption in private enterprises, which made it harder to reach more participants from this group. Even though this distribution of participants may be related to a bias in sample selection, it may also be viewed to provide evidence for the claim that there is still lesser interest for Open Data in business. Fig. 1 below shows that the majority of respondents were users (58%), 24% were data providers and 18% indicated that they represent both roles.

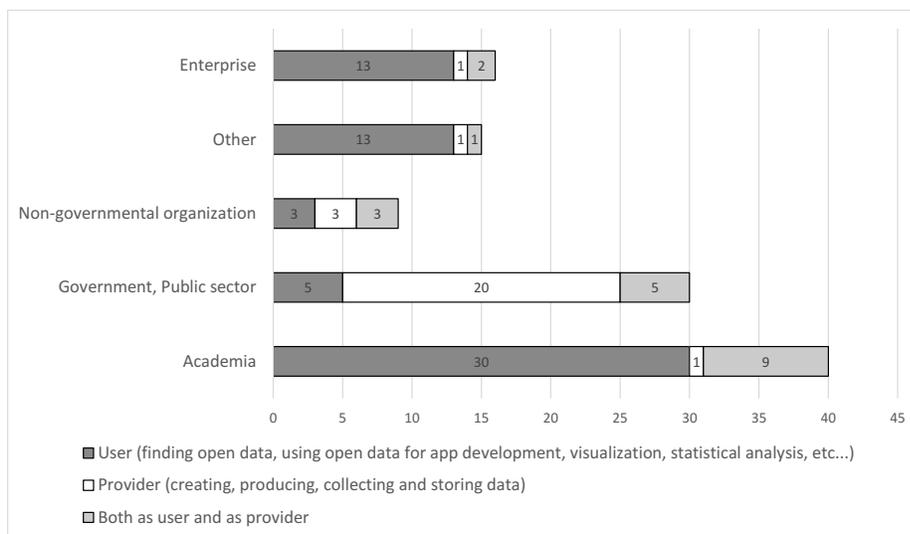


Figure 1. Types of organizations and users

## IV. RESULTS

### A. User Barriers

The barriers for users' section of the questionnaire were split into 4 categories: data portals, data quality, legal constraints and (lack of) knowledge and experience. Fig. 2 presents an overview of all user ratings, sorted per category by the mean.

#### 1) Overall Impression

To gain an overview which of the categories contains the largest barriers, the respondents were first asked to indicate in which of these categories they have been facing the most issues (see "overall impression" in Fig. 2). Overall, data quality and legal issues were rated as largest barriers, followed by data portals usability issues and lack of knowledge and experience. While lack of knowledge and experience has widely been reported as a barrier in research literature, it was rated lowest in the survey. This result may be due to the fact that the target group for the survey was subjects, who already had some experience with Open Data and most participants were academics.

#### 2) Data Quality

Regarding detailed results, we first discuss data quality issues. The most important reported barrier is that data is not machine readable ( $M=3.21$ ,  $SD=1.48$ ). In practice, data is often published in proprietary, human-readable only formats. These results suggest that there may currently be datasets with incomplete data being published, which drastically reduces their usefulness. The respondents reported some additional points in the open comments fields:

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

It could be argued that if certain standards were enforced, these problems would not exist. Metadata is often used to describe the dataset in order to provide the users with a better overview. However, metadata are often inconsistent, wrong or simply non-existent. 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, with seemingly no communication and planning with other providers, results in an inconsistent set of data available across the country. If user finds a regional dataset on a certain topic, they have no guarantee that similar datasets exist for other regions as well.

#### 3) Legal constraints

According to the widely adopted Open Definition (<http://opendefinition.org>), Open Data shall be freely accessible, reusable and permit redistribution. This simple definition however is not always followed, 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 put the legal aspects of free use and redistribution in question. It should not be expected from data users to understand complex legal documents. The survey addressed this topic to find out whether the users perceive unclear and restrictive licensing as relevant barriers to Open Data use. On average, restrictive ( $M=3.05$ ,  $SD=1.50$ ), unclear ( $M=2.98$ ,  $SD=1.38$ ) and hard to understand ( $M=2.83$ ,  $SD=1.33$ ) licenses were judged as moderate barriers. Some users may even have been not able to use the data due to licensing issues. Potential users may also feel misled and disappointed when finding out that there are legal restrictions, since Open Data has been defined by the freedom of use and reuse.

#### 4) *Data Portal Usefulness*

The lack of information about the quality ( $M=3.23$ ,  $SD=1.48$ ) and the content ( $M=3.17$ ,  $SD=1.17$ ) of a dataset were rated as biggest barriers. Such barriers can indeed be discouraging for potential users, as users have to manually examine the data to determine whether it is of any use for them, which creates an unnecessary workload for them. To make use of the open nature of 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 and effort for other users and improving the quality of the portal. Problems with APIs were ranked of medium severity. However, the collected data on APIs must be interpreted with caution, as some of the respondents who had indicated that they do not use an API, but other methods of data retrieval answered these questions on APIs as well. Problems as duplicate datasets or languages barriers were rated as less important. In the open comment field users further noted these barriers:

- “The data.gv.at website is too slow; the performance of the portal is poor”
- “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”)”
- “Old data.”
- “Lack of interesting datasets”

#### 5) *Knowledge and Experience*

On average, lack of proper documentation presents a moderate barrier ( $M=2.97$ ,  $SD=1.36$ ). The other barriers, such as lack of support or helpdesk seem to be less of an issue. 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.”

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 create unnecessary confusion.

#### B. *Opinions on Opening Data Sets*

In the literature, there are conflicting views on whether low quality data should be published or not. To gain a deeper insight into this topic, we asked participants if data should be published regardless of quality (which could be described as a “the more Open Data there is the better” approach). More than half of the participants indicated that even low quality data should be released (57%). On the other hand, more than half of participants (65%) also indicated that organizations publishing Open Data should adhere to a certain set of quality standards and rules, so a certain level of quality is ensured. Fewer participants (35%) thought enforced rules and standards would prevent organization publishing their data in the first place.

In addition, we asked users to provide some examples on what kind of rules should be implemented. These are some of their suggestions:

- “Topicality, completeness, primary sources, machine-readability of the data, unrestricted access, open standards, privacy, etc.”
- “There should not be any rules, rather recommendations, such as ‘use utf-8’”
- “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”
- “Data should always be published in a machine-readable format. That means no PDFs for instance.”
- “For the moment, the concept of Open Data is not yet in full swing. Therefore, I prefer quantity over quality for now”

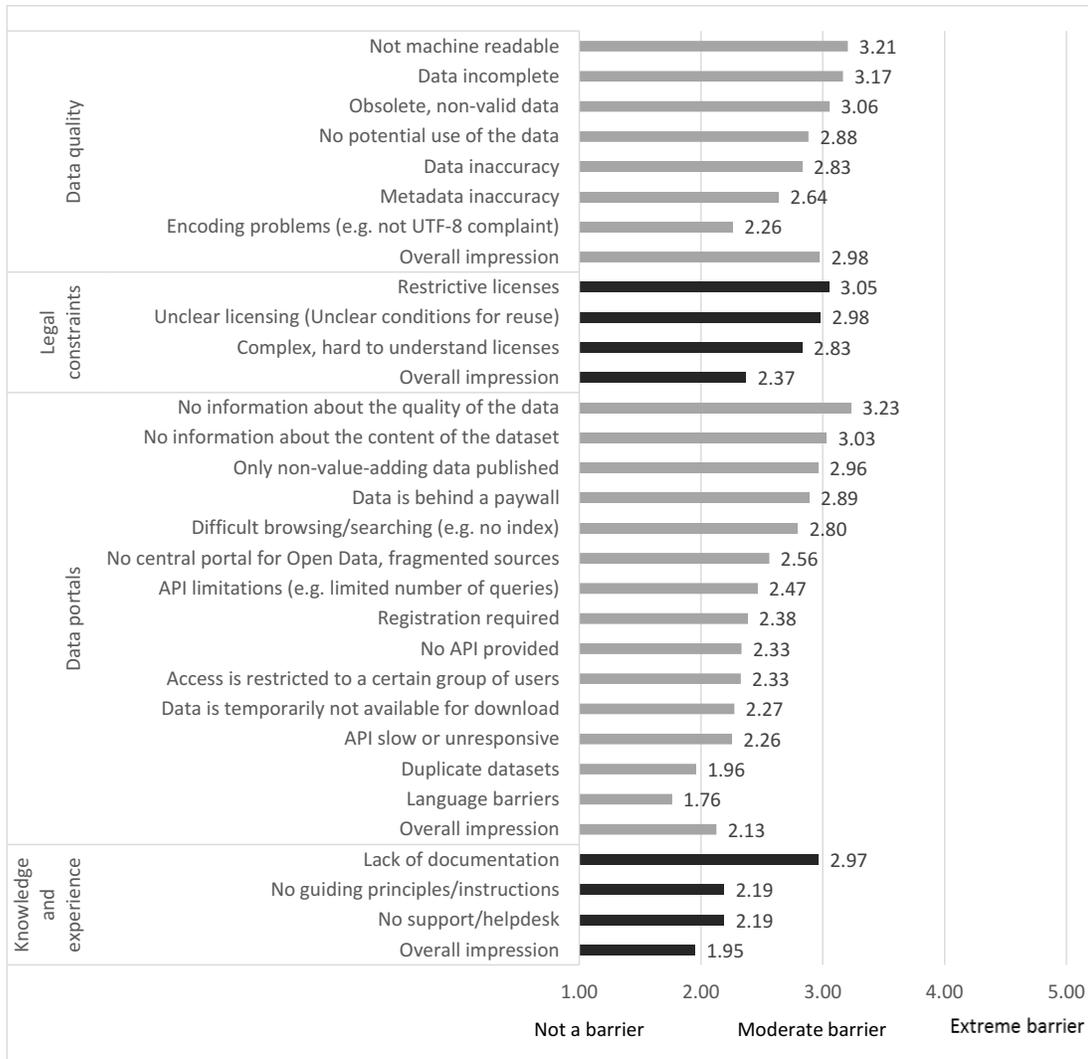


Figure 2. Barriers to using Open Data

### C. Provider Barriers

#### 1) Documentation and Support

Results from this category once again suggest that many participants of this survey are experienced in the field of Open Data: the majority of respondents marked lack of documentation or support as a non-barrier. It is important to bear the experience of the participants in mind when interpreting the low ratings of barriers of the documentation and support category as these results may not apply to beginners trying to publish Open Data.

The results also provide some 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 and write the metadata according to their own standards. Providers neither rated the absence of metadata standards nor the various standards of machine readable data as significant barriers to open their data. On basis of these results, one might argue that creating and enforcing a unified standard might be met with resistance from the providers, possibly leading to less data being published.

#### 2) Legal constraints

One interesting comparison is that providers on average rated complex, hard to understand licenses less as a barrier than users (Providers:  $M=1.77$ ,  $SD=1.00$ ; Providers:  $M=2.83$ ,  $SD=1.33$ ;  $t_{101}=-4.44$ ,  $p<0.001$ ). The providers were also asked if opaque ownership of the data inhibits its release. The results show that respondents consider unclear ownership as a weak to moderate barrier ( $M=2.59$ ,  $SD=1.32$ ).

### 3) Business and Strategy:

The results of this part of the survey reflect that only a fraction of the participants are providers from the private sector. The fact that for the majority of providers the loss of the option to sell the data due to Open Data licenses does not present a barrier ( $M=1.62$ ,  $SD=0.76$ ) suggests that the providers in this sample are not relying on making revenue from the data. We assume that many private companies would be more hesitant about publishing data from which they could make revenue. Likewise, the lack of Open Data business models presented a small to moderate barrier ( $M=1.62$ ,  $SD=0.76$ ). There were however some more severe barriers reported in the business and strategy category. On average, providers rated the lack of resources and time as largest barrier over all categories ( $M=3.24$ ,  $SD=1.05$ ). This could indicate that providers need some support and guidance on how to minimize these costs. Additionally, cost issues ( $M=2.79$ ,  $SD=1.30$ ) and resistance to change ( $M=2.57$ ,  $SD=1.21$ ) were rated as weak to moderate barriers.

### 4) Privacy and Security:

The results in this category confirm that potential providers fear misrepresentation of data. “Fear of false conclusions” was rated as moderate and received the highest average rating in this category ( $M=2.73$ ,  $SD=1.07$ ). Unwelcomed exposure of the data and loss of control over the released information were rated as weak to moderate barriers. Another weak barrier for not exposing data was low data quality ( $M=2.23$ ,  $SD=1.07$ ). Lack of resources for the maintenance of the data might be a reason for low data quality. One respondent stated in the open question: “Keeping the data up-to-date takes effort.” This is an interesting issue to be discussed in comparison with the Open Source model. 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 providers of the code as they are able to crowdsource external developers who maintain and help them keep the code up to date. Morgan and Finnegan highlight several benefits of Open Source including increased collaboration, knowledge sharing with communities, customers, and other parties and facilitation of joint ventures with other companies or research institutes [20]. Similarly, opening the data could facilitate collaboration between the provider and the Open Data community, thus helping the provider maintain their data.

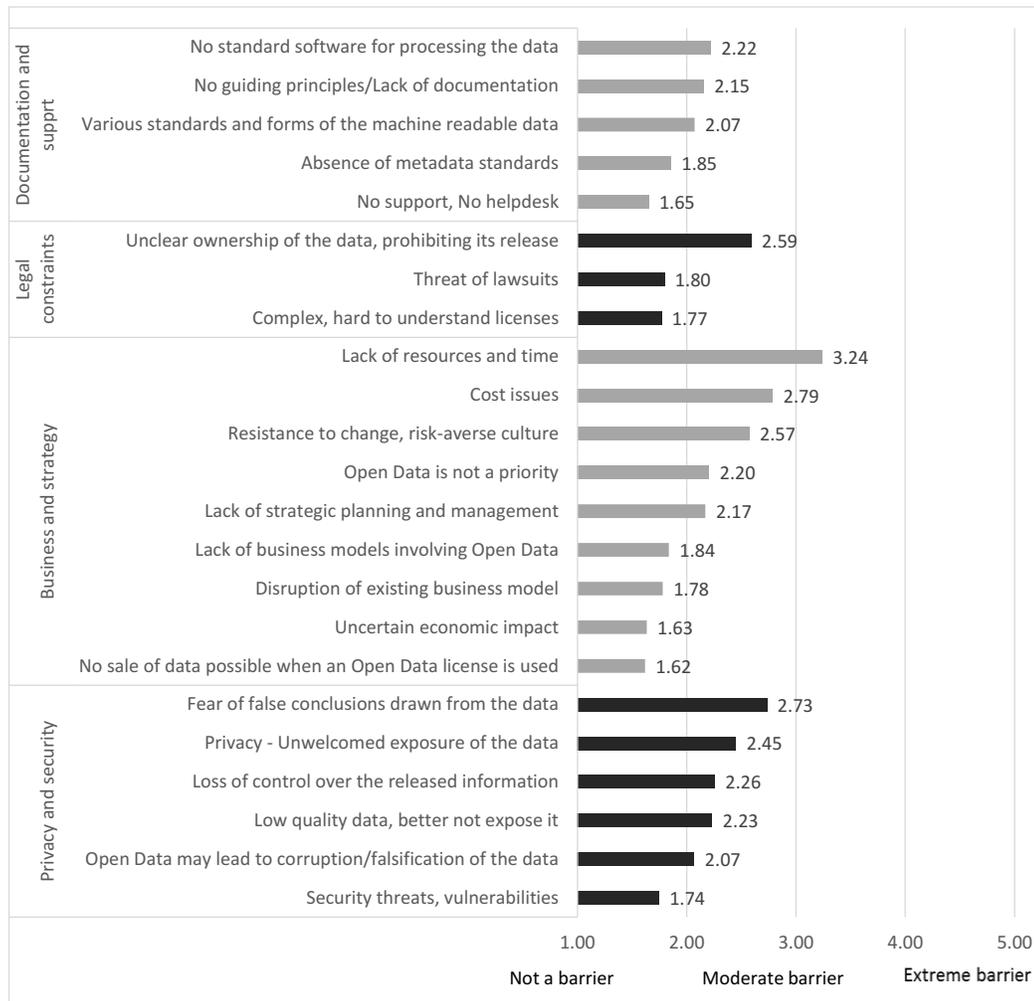


Figure 3. Barriers to opening Open Data

## V. DISCUSSION AND LIMITATIONS

The presented study has been set out with the aim of assessing the relative importance of barriers to Open Data. Most of the barriers that are widely cited in the literature have also been confirmed in our survey, but we also have extended those previous studies by evaluating and discussing the respective relevance of barriers. On the users' side of view lacking information about the quality and content of the datasets, data not machine readable, incomplete or obsolete, non-valid data as well as restrictive licenses were rated as most severe barriers. Data providers found lack of resource and time to be the most severe barrier, followed by cost issues and fear of false conclusions drawn from the data.

From the free-text comments of the participants, we could identify some issues, which to our knowledge were not yet documented in the Open Data literature. Several data portal users for instance complained about the slow responsiveness of the websites that host the portal. Indeed, poor performance can have a significant negative impact on the workflow of the user, especially when working with large datasets. Furthermore, lack of data harmonization between portals presents a barrier for the users. They are often not sure if they will be able to find a specific dataset on other portals as well. A practical example of this issue could be for instance, developing an application that lists all public libraries in Europe. While one portal may have an open dataset with the necessary information, there is no guarantee that a portal in another country will have one available as well. In the literature, various data quality issues have already been mentioned. Encoding issues, such as the data not being UTF-8 compliant is a barrier that has not yet been reported. Another survey participant identified licenses not being machine-readable as a legal barrier. Indeed, if licenses were to be published in a machine-readable format, it would help to automate working with Open Data. From the providers' perspective however, we were not able to identify any new, previously unreported barriers. It should be noted that all these findings must be interpreted with caution, as we have obtained them from the free-text comments section of the survey. As such, the aforementioned barriers were not confirmed by multiple participants and might not be relevant for the majority of Open Data users.

A number of specific limitations need to be acknowledged. One source of weakness in this study was the selection of study participants. First, there is no existing directory of Open Data users; therefore, we had to rely on personal contacts to reach a large amount of Open Data users and providers. While, we deem the sample size of 110 as large for a small country as Austria, representativeness might be biased. Specifically, we might have missed potential users who had not been able to use Open Data because of severe barriers, as they might have felt not invited or by a study on Open Data "users" or actually have been missed to be invited. Additionally, the majority of respondents were experienced in the field of Open Data as became clear from the results. Due to this fact, we have not been able to discover potential new barriers faced by newcomers or organizations that are currently not interested in publishing data in an open format. Furthermore, since we focused specifically on Austria as our main area of interest, caution must also be applied, as the findings might not be transferable to other countries. While our research questions were general in nature, countries differ according to the specific Open Data sets government institutions offer as well as in terms of maturity of their Open Data "culture" and thus, some barriers might be more pronounced for some Open Data sets than for others. It was beyond the scope of this study to examine the cross-national differences of barriers. However, we encourage fellow scholars to replicate the study in other countries to shed light on such cross-national differences in importance of barriers. This is an important issue for future research.

In terms of directions for future work, we have also identified further relevant issues for research: further work should investigate the use of Open Data in the private sector; the private sector is rather underrepresented in the survey participation and likewise in the research literature. A future study investigating the use and potential of Open Data in the private sector should provide valuable input in the field of Open Data research by closing this existing research gap.

## VI. CONCLUSION AND OUTLOOK

This paper has given an account of Open Data barriers as experienced by practitioners using and providing Open Data in Austria. Open Data is a concept still very much in its early phase in many respects: while new communities and platforms progress with a steady pace, there is still much work to be done and the group of early adopters is still relatively small and partially restricted to researchers, providers, and a handful of app developers. The relevance of addressing and lowering barriers is clearly supported by the current findings. As the empirical study showed, there are numerous barriers that are making the use of Open Data difficult for both users and providers. One of the more significant findings to emerge from this study is that not all barriers are perceived equally severe. This might provide insights for setting priorities and a further agenda to further push Open Data forward as a means for more effective public services, but also for gaining adoption in the private sector. Taken together, the current study reported here has important implications for practice and research and adds strength to a growing body of empirical work on Open Data.

## REFERENCES

- [1] P. McDermott, "Building open government," *Government Information Quarterly*, vol. 27, pp. 401-413, 2010.
- [2] J. C. Bertot, P. T. Jaeger, and J. 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*, vol. 27, pp. 264-271, 2010.

- [3] M. Janssen, Y. Charalabidis, and A. Zuiderwijk, "Benefits, adoption barriers and myths of open data and open government," *Information Systems Management*, vol. 29, pp. 258-268, 2012.
- [4] M. Kassen, "A promising phenomenon of open data: A case study of the chicago open data project," *Government Information Quarterly*, vol. 30, pp. 508-513, 2013.
- [5] P. Johnson and P. Robinson, "Civic hackathons: Innovation, procurement, or civic engagement?," *Review of Policy Research*, vol. 31, pp. 349-357, 2014.
- [6] J. Umbrich, S. Neumaier, and A. Polleres, "Quality assessment and evolution of open data portals," *Proc. Future Internet of Things and Cloud (FiCloud)*, 2015 3rd International Conference on, 2015, pp. 404-411.
- [7] S. Martin, M. Foulonneau, S. Turki, and M. Ihadjadene, "Risk analysis to overcome barriers to open data," *Electronic Journal of e-Government*, vol. 11, pp. 348-359, 2013.
- [8] A. Zuiderwijk, M. Janssen, S. Choenni, R. Meijer, R. S. Alibaks, and R. Sheikh\_Alibaks, "Socio-technical impediments of open data," *Electronic Journal of eGovernment*, vol. 10, pp. 156-172, 2012.
- [9] A. Zuiderwijk and M. Janssen, "Barriers and development directions for the publication and usage of open data: A socio-technical view," in *Open government*, ed: Springer, 2014, pp. 115-135.
- [10] A. F. van Veenstra and T. A. van den Broek, "Opening moves—drivers, enablers and barriers of open data in a semi-public organization," *Proc. International Conference on Electronic Government*, 2013, pp. 50-61.
- [11] N. Huijboom and T. Van den Broek, "Open data: An international comparison of strategies," *European journal of ePractice*, vol. 12, pp. 4-16, 2011.
- [12] A. S. Martín, A. H. De Rosario, and M. D. C. C. Pérez, "An international analysis of the quality of open government data portals," *Social Science Computer Review*, p. 0894439315585734, 2015.
- [13] K. Braunschweig, J. Eberius, M. Thiele, and W. Lehner, "The state of open data limits of current open data platforms," 2012.
- [14] H. Alani, D. Dupplaw, J. Sheridan, K. O'Hara, J. Darlington, N. Shadbolt, and C. Tullo, "Unlocking the potential of public sector information with semantic web technology," in *The semantic web*, ed: Springer, 2007, pp. 708-721.
- [15] P. Conradie and S. Choenni, "Exploring process barriers to release public sector information in local government," *Proc. Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance*, 2012, pp. 5-13.
- [16] M. Kaltenböck and T. Thurner, "Open government data weißbuch," Wien: Semantic Web Company, 2011.
- [17] K.-J. Stol and M. Ali Babar, "Challenges in using open source software in product development: A review of the literature," *Proc. Proceedings of the 3rd International Workshop on Emerging Trends in Free/Libre/Open Source Software Research and Development*, 2010, pp. 17-22.
- [18] C. P. Geiger and J. von Lucke, "Open government and (linked)(open)(government)(data)," *JeDEM-eJournal of eDemocracy and Open Government*, vol. 4, pp. 265-278, 2012.
- [19] P. Conradie and S. Choenni, "On the barriers for local government releasing open data," *Government Information Quarterly*, vol. 31, pp. S10-S17, 2014.
- [20] L. Morgan and P. Finnegan, "Beyond free software: An exploration of the business value of strategic open source," *The Journal of Strategic Information Systems*, vol. 23, pp. 226-238, 2014.
- [21] S. Goode, "Something for nothing: Management rejection of open source software in australia's top firms," *Information & Management*, vol. 42, pp. 669-681, 2005.
- [22] A. Zuiderwijk, M. Janssen, R. Meijer, S. Choenni, Y. Charalabidis, and K. Jeffery, "Issues and guiding principles for opening governmental judicial research data," *Proc. International Conference on Electronic Government*, 2012, pp. 90-101.