

Perception of Key Barriers in Using and Publishing 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), Austria; {martin.beno;kathrin.figl;jurgen.umbrich;axel.polleres}@wu.ac.at
**Complexity Science Hub Vienna, Austria;*

Abstract: There is a growing body of literature recognizing the benefits of Open Data. However, many potential data providers are unwilling to publish their data and at the same time, data users are often faced with difficulties when attempting to use Open Data in practice. Despite various barriers in using and publishing Open Data still being present, studies which systematically collect and assess these barriers are rare. Based on this observation we present a review on prior literature on barriers and the results of an empirical study aimed at assessing both the users' and publishers' views on obstacles regarding Open Data adoption. We collected data with an online survey in Austria and internationally. Using a sample of 183 participants, we draw conclusions about the relative importance of the barriers reported in the literature. In comparison to a previous conference paper presented at the conference for E-Democracy and Open Government, this article includes new additional data from participants outside Austria, reports new analyses, and substantially extends the discussion of results and of possible strategies for the mitigation of Open Data barriers.

Keywords: Open Data, barriers, data quality, questionnaire, barrier mitigation

Acknowledgement: This article was originally submitted to the International Conference for E-Democracy and Open Government 2017 (CeDEM 17).

1. Introduction

Ever since the launch of the U.S. Open Data portal <http://data.gov> in 2009, there has been an increased interest from data providers to publish their data in machine-readable format as Open Data and the demand for such data from users has been steadily growing. Federal legislations such as The Freedom of Information Act, the Memorandum for the Heads of Executive Departments and Agencies: Transparency and Open Government and the Open Government Directive have helped establish the Open Data movement in the U.S. (Kassen, 2013), and the PSI directive has been set out to achieve the same in the EU (K. Janssen, 2011). 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 (Bertot, Jaeger, & Grimes, 2010; McDermott, 2010);
- 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 (M. Janssen, Charalabidis, & Zuidervijk, 2012);
- Promoting collaboration of citizens with the government (Johnson & Robinson, 2014; Kassen, 2013).

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 (Umbrich, Neumaier, & Polleres, 2015).

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, so that 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 at larger scale buy in 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 (M. Janssen et al., 2012) claims that despite numerous benefits, Open Data adoption entails a number of barriers. This is further supported by another study (Martin, Foulonneau, Turki, & Ihadjadene, 2013), 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 relevance 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 in practice?”

In a first step, we have conducted a systematic literature review in order to establish a solid theoretical background for the construction of a questionnaire. We then sent it out to a large number of participants in Austria as the focus of our initial study was an assessment at national level within a national research project¹. The results were first presented at the Conference for E-Democracy and Open Government (CeDEM) in 2017. Limiting a survey to a national level however, lowers generalizability of our research findings, as they might not be transferable to other countries. 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

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

for some Open Data sets than for others. Therefore, in a second step, we distributed the questionnaire globally. Our main goal was to empirically assess the importance and relevance of Open Data barriers for all users and providers in practice. Thus, a new contribution of this journal article compared to the previous conference paper is that we now include more data from further countries than Austria (now 183 participants, previously 110). Furthermore, we extended our analyses and now discuss differences between the Austrian and the international sample as well as differences between users or providers working for different organization types (e.g. academia versus enterprises).

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 (2012) study, which provides a systematic overview of Open Data socio-technical impediments. Although this study presents an extensive and categorized set of barriers, it lacks a detailed analysis, which could help determine the relative importance and relevance 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 (Zuiderwijk & Janssen, 2014). However, despite their questionnaire being distributed in multiple countries, their results presentation was focused on the Netherlands only. Zuiderwijk and Janssen (2014) encourage fellow scholars to compare barriers and development directions in different countries. With our study we contribute to this stream of research, extending and adapting Zuiderwijk's and Janssen's approach (2014). However, we also observe that direct comparison of countries 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 the Netherlands (van Veenstra & van den Broek, 2013). 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 by focusing on data publishing only and the barriers faced by data users and re-users are often ignored. Indeed, Janssen, Charalabidis and Zuiderwijk (2012) 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; to this end, we explicitly aim at considering both the publishers' and the consumers' views in the present study. 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, we do not limit our study to Open Data barriers in government and public institutions only, but include more categories of organizations as e.g. academia. We strive to establish a solid basis for research on Open Data concerns and barriers in general that could ultimately help to mitigate any perceived problems including those faced by private entities.

In the remainder of this article, we first present our findings from the literature review in the form of a summary of barriers of Open Data (Section 2). Section 3 describes the method and the design of our online questionnaire (Section 3), the results of which are analyzed in Section 4. In Section 5, we propose several mitigation strategies for various Open Data barriers. We conclude our article with a discussion of results and limitations in Section 6 and an outlook to future work as well as general conclusions in Section 7.

2. 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 (2012) found that barriers are related to either data providers (resulting in not intending to publish 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 (2013) claim that the challenges and constraints faced by re-users of public data differ from the ones encountered by the public data providers. In the present study, we assign each barrier to a specific category. 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 - it was important for us to use an assignment of barriers to categories for a more digestible design of the online questionnaire, i.e., for instance not bothering pure data users with obstacles applying only to data publishers and vice versa. We categorized the barriers in the 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)

We describe each category below, starting with a short introduction of the barriers found in the literature. We also provide an overview of each category in a table. The tables summarize the barriers found per category along with the respective literature references and the short labels of the barriers in the left columns were also used as questionnaire items in our survey.

2.1. Literature Search

We collected a base of articles on Open Data barriers from various online databases, such as Science Direct, ACM Digital Library, Springer Link, and Google Scholar, a citation-indexing service. The term "Open Data" has been used as a central keyword in the search, which we combined with the following keywords in our search strings: "barrier", "challenge", "impediment", "problem", "issue", "use", and "publishing". 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 in the article, neither was there any other indication that the authors considered them to be such. As an example, Tammisto and Lindman (2012) acknowledge the confusion regarding what the term "Open Data" actually means, which could be considered a barrier, especially for the newcomers. However, they do not label the potential confusion regarding the term "Open Data" a barrier and even go on to claim that the ambiguity of the term may have positive effects on the Open Data movement.
- The article 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 up by proper research.

Ultimately, we have chosen 13 articles from which we extracted a list of barriers. The majority of research on Open Data was published from 2009 onwards and is still ongoing. The publication years of our selection of literature on Open Data barriers had a peak in 2012. Indeed, 6 of the 13 selected publications are from this particular year. As mentioned in the first chapter of this article, the Open Data movement emerged after the launch of the US Open Data portal in 2009, which might explain the lack of literature on Open Data from before that point in time. We have however, included one article from 2007 (Alani et al.). Even though the aforementioned article does not focus on Open Data specifically, it discusses the retrieval of public sector information using semantic web technologies, a topic, which we feel is closely related to Open Data.

2.2. 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 was low enough to be included in a single online questionnaire (without overburdening the participants). Our intention was to focus on those 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.

2.3. User Barriers

In this section we present 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. In addition, the lack of guidelines, support or a helpdesk makes the use of Open Data difficult for newcomers.

2.3.1. 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 Pérez (2015) 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 (Umbrich et al., 2015). 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 (2012) 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 application programming interface (API), as the existence of such an interface allows them to include the data directly into their application without having to first download the data manually. In another major study on the limits of Open Data platforms, Braunschweig, Eberius, Thiele and Lehner (2012) 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 (Umbrich et al., 2015), 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. (2012), 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 et al. (2012) study of the same year. Nevertheless, APIs may also have some limitations (Zuiderwijk, Janssen, Choenni, et al., 2012). 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 1.

Table 1: Open Data Portals Barriers (Source: Authors)

Barrier	References
No API provided	(Martín et al., 2015; Martin et al., 2013; Zuiderwijk, Janssen, Choenni, et al., 2012)
API slow or unresponsive	(Zuiderwijk, Janssen, Choenni, et al., 2012)
API limitations	(Martín et al., 2015; Zuiderwijk, Janssen, Choenni, et al., 2012)
Registration required before gaining access to data	(M. Janssen et al., 2012; Martin et al., 2013; Zuiderwijk, Janssen, Choenni, et al., 2012)
Difficult browsing/searching	(M. Janssen et al., 2012; Zuiderwijk & Janssen, 2014; Zuiderwijk, Janssen, Choenni, et al., 2012)
No information about the quality of the data	(Alani et al., 2007; M. Janssen et al., 2012; Zuiderwijk & Janssen, 2014; Zuiderwijk, Janssen, Choenni, et al., 2012)
No information given about the content of the dataset	(M. Janssen et al., 2012; Zuiderwijk, Janssen, Choenni, et al., 2012)
Language barriers	(Huijboom & Van den Broek, 2011; Martin et al., 2013; Zuiderwijk & Janssen, 2014; Zuiderwijk, Janssen, Choenni, et al., 2012)
Data is not available for download	(Martín et al., 2015; Zuiderwijk, Janssen, Choenni, et al., 2012)
Duplicate datasets	(M. Janssen et al., 2012; Zuiderwijk, Janssen, Choenni, et al., 2012)
Only non-value-adding data published	(M. Janssen et al., 2012)
No central portal for the data	(M. Janssen et al., 2012; Zuiderwijk & Janssen, 2014; Zuiderwijk, Janssen, Choenni, et al., 2012)
Restricted access to a certain group of users	(Zuiderwijk, Janssen, Choenni, et al., 2012)
Data is behind a paywall	(M. Janssen et al., 2012; Martin et al., 2013; Zuiderwijk, Janssen, Choenni, et al., 2012)

2.3.2. Data Quality

After retrieval, users may find out that the quality of the obtained data is not meeting their requirements. Perhaps the strongest 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 (2012) 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. (2007) 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 2.

Table 2: Data Quality Barriers (Source: Authors)

Barrier	References
Data inaccuracy	(Conradie & Choenni, 2012; M. Janssen et al., 2012; Zuiderwijk, Janssen, Choenni, et al., 2012)
Obsolete, non-valid data	(M. Janssen et al., 2012)
Not machine readable	(Martín et al., 2015; Zuiderwijk, Janssen, Choenni, et al., 2012)
Metadata inaccuracy	(M. Janssen et al., 2012; Martin et al., 2013; Zuiderwijk & Janssen, 2014; Zuiderwijk, Janssen, Choenni, et al., 2012)
No potential use of the data	(M. Janssen et al., 2012)

2.3.3. 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 (OSS) can be made here: complex licensing situation is nothing new in the field of OSS. 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 (2010) in their literature review on challenges of using OSS. 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 3 summarizes the major legal constraints as found in the literature includes in our survey.

Table 3: User Legal Barriers (Source: Authors)

Barrier	References
Threat of lawsuits	(M. Janssen et al., 2012; Zuiderwijk, Janssen, Choenni, et al., 2012)
Restrictive licenses	(Conradie & Choenni, 2012; Zuiderwijk, Janssen, Choenni, et al., 2012)
Unclear licensing (Unclear conditions for reuse)	(Conradie & Choenni, 2012; Zuiderwijk, Janssen, Choenni, et al., 2012)
Complex, hard to understand licenses	(Zuiderwijk, Janssen, Choenni, et al., 2012)

2.4. 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.

2.4.1. 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 (Geiger & von Lucke, 2012). 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 (2014). 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 (Geiger & von Lucke, 2012). 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 4.

Table 4: Privacy and Security Barriers (Source: Authors)

Barrier	References
Privacy – Unwelcomed exposure of the data	(Conradie & Choenni, 2012, 2014; M. Janssen et al., 2012; Martin et al., 2013; van Veenstra & van den Broek, 2013; Zuiderwijk & Janssen, 2014; Zuiderwijk, Janssen, Choenni, et al., 2012)
Loss of control over released information	(Conradie & Choenni, 2012)
Fear of false conclusions drawn from the data – Data misrepresentation	(Conradie & Choenni, 2012; Geiger & von Lucke, 2012)
Open Data may lead to corruption or falsification of the data	(Geiger & von Lucke, 2012)
Security threats, vulnerabilities	(Conradie & Choenni, 2014; M. Janssen et al., 2012)
Low quality data, better not expose it	(Conradie & Choenni, 2012)

2.4.2. 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 (2014) 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 a current business strategy, which is focused on generating sales from the data owned. In the same way, drawing again parallels to OSS, many software developers do not open source their code in fear of lost sales. Morgan and Finnegan (2014) 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 user demand. 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 5.

Table 5: Strategic and Business barriers (Source: Authors)

Barrier	References
Lack of business models	(van Veenstra & van den Broek, 2013; Zuiderwijk, Janssen, Choenni, et al., 2012)
Open Data is not a priority	(Geiger & von Lucke, 2012)
Resistance to change, risk-averse culture	(Conradie & Choenni, 2012; M. Janssen et al., 2012)
No sale of data possible when an Open Data license is used	(M. Janssen et al., 2012)
Embedding Open Data in the strategy and work process	(van Veenstra & van den Broek, 2013)
Disruption of existing business model. (Such as charging money for the data)	(Conradie & Choenni, 2012; Geiger & von Lucke, 2012; M. Janssen et al., 2012; Martin et al., 2013)
Uncertain economic impact	(Conradie & Choenni, 2012, 2014; Martin et al., 2013)
Lack of resources and time	(M. Janssen et al., 2012; Martin et al., 2013)
Cost issues	(Martin et al., 2013)

2.4.3. Provider Legal Constraints

The complex data licensing situation is a barrier also inhibiting data release on the publisher side: providers also face the threat of lawsuits, albeit not only because of violations of the licenses, but when they release private or otherwise sensitive data. Particularly, this includes issues of data ownership. If the provider is not sure about the ownership of the data, he or she is not able to release them. Conradie and Choenni (2014) 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 6.

Table 6: Provider Legal Barriers (Source: Authors)

Barrier	References
Unclear ownership of the data, prohibiting its release	(Geiger & von Lucke, 2012)
Threat of lawsuits	(Zuiderwijk & Janssen, 2014; Zuiderwijk, Janssen, Choenni, et al., 2012)
Complex, hard to understand licenses	(Zuiderwijk, Janssen, Choenni, et al., 2012)

2.4.4. Technical Barriers

There are several technical barriers that might make publishing data difficult for providers. It is often expected from providers to release data in various formats, but Martin, Foulonneau, Turki and Ihadjadene (2013) 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 (2012) 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 in the literature (Zuiderwijk, Janssen, Choenni, et al., 2012). It should be noted that some efforts towards standardizing are already in progress. The Resource Description Framework specification, recommended by the World Wide Web Consortium, is an example of such an effort. Table 7 summarizes the technical barriers faced by data publishers.

Table 7: Provider Legal Barriers (Source: Authors)

Barrier	References
Various standards and forms of machine readable data	(Martin et al., 2013; Zuiderwijk, Janssen, Choenni, et al., 2012)
No standard software for processing the data (Fragmentation of software)	(M. Janssen et al., 2012)
Absence of metadata standards	(Martin et al., 2013; Zuiderwijk, Janssen, Choenni, et al., 2012)

2.5. User and Provider: Knowledge and Experience

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 OSS rejections in Australia's top firms, Goode (2005) found that companies often prefer commercial, proprietary versions of software, because in that case 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 (2012). However, the providers usually do not offer 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 if 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, Zuiderwijk et al. (2012) 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 for providers that there would be a demand from users, or the users can't communicate this demand since the availability of said data is not known to them. Table 8 presents an overview of barriers, which apply both to data users and to publishers.

Table 8: Knowledge and Experience (Source: Authors)

Barrier	References
No support	(M. Janssen et al., 2012; Zuiderwijk & Janssen, 2014; Zuiderwijk, Janssen, Choenni, et al., 2012)
No helpdesk	(M. Janssen et al., 2012; Zuiderwijk, Janssen, Choenni, et al., 2012)
Lack of documentation	(Martin et al., 2013)
No guiding principles/instructions	(Zuiderwijk & Janssen, 2014; Zuiderwijk, Janssen, Choenni, et al., 2012; Zuiderwijk, Janssen, Meijer, et al., 2012)

3. Method

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 had a modular structure with separate modules for users and providers. The modules were presented to the participants according to whether they had indicated experience as a user, a provider, or both. In addition, basic information about the participants' background was asked in the introductory section of the survey such as the kind of organization they were associated with, the categories of data 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 (It was impossible to use/publish the data.)

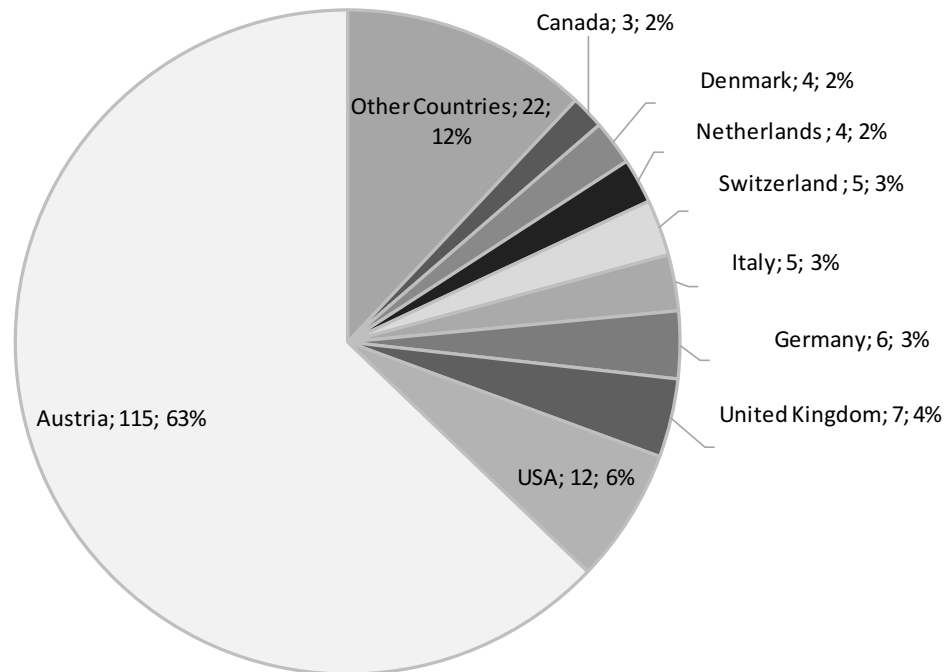
Additionally, we added an open text field for each category of barriers, in which participants were free to comment on any barriers they have encountered, which were not already mentioned in the questionnaire.

3.1. Participants

The national version of 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 had 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.

The second, international version of the survey was launched in July 2017 and remained active until the end of September 2017. In contrast to the first questionnaire, which was in German, the second one was in English. Both questionnaires were identical, however, the English version additionally asked the respondents to report their country of residence. We promoted the questionnaire using social media and various mailing lists. Furthermore, we have harvested the publicly available contact information found in the datasets’ metadata published on various Open Data portals and sent an e-mail invitation to publishers of datasets. Ultimately, 276 participants have started the survey, 73 have fully completed the survey. Of those 73 participants, five had indicated that their country of residence was Austria. In total, we therefore had 115 participants from Austria (68%) and 68 from other countries (37%). Figure 1 depicts the distribution of the countries of residence of the participants.

Figure 1. Countries of Residence of Participants (Other countries with only 1-2 participants each were: AGO, ALB, ATA, CZE, ESP, FRA, HUN, LTU, LUX, NOR, PHL, URY, AUS, BEL, BGR, IRL, RUS)
(Source: Authors)



4. Results

In the following, we present results for the total sample of 183 participants. We checked with t-tests for independent samples whether the Austrian sample had rated barriers differently than the international sample and report significant differences in respective sections. To screen for possible differences between ratings of users or providers from various institutions, we conducted ANOVAs (analyses of variance) for each barrier with “type of organization” [government and public sector, academia, other, enterprise and non-governmental organization] as independent factor. In case of a significant overall result, we performed post-hoc analyses (Fisher’s LSD tests) to determine which types of organization significantly differ from each another.

Figure 2 shows the distribution of the type of institution in which participants were employed. Most participants worked for public institutions such as government (36%) or academia (30%). The prevalence of participants from academia was higher in the Austrian sample (42; 37% of participants) than in the international version (12; 18% of participants). A possible reason for the high rate of participants from academia in the Austrian sample might be the way the survey was promoted, as university colleagues have helped to distribute the survey. The highest share of participants in the international sample stem from government and the public sector (34; 50% of participants). Despite our efforts to reach the private sector, enterprises (11%) were rather underrepresented in both samples. A possible explanation may be the high popularity Open Data enjoys in the public sector as opposed to still lower adoption in companies, which makes it harder

to reach more participants from this group. Even though the distribution of participants may be induced by a sample selection bias, it may also be viewed to represent the real distribution of users and providers and thus, provide evidence for the claim that companies are rather disinterested in Open Data.

Figure 2: Types of Organizations in Austrian and International Sample (Source: Authors)

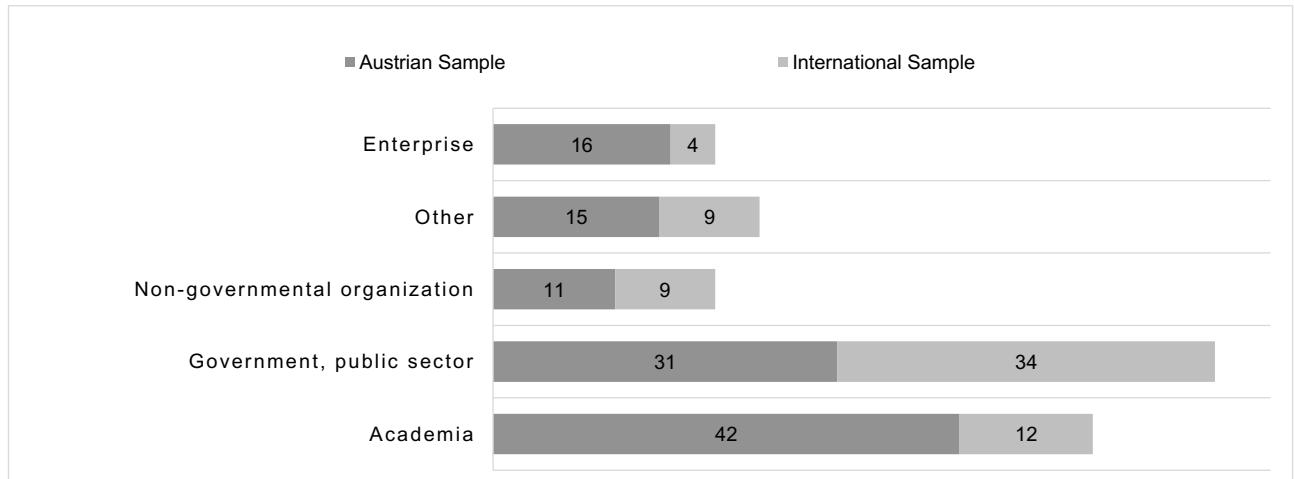


Fig. 3 below shows that the largest group of respondents were users without experience as providers (44%), 28% were data providers and 28% indicated that they represent both roles. The distribution of user types was different in the Austrian versus the international sample. While 59% (68 out of 115) of participants of the Austrian sample were users, only 18% (12 out of 68) of the international sample were users. This difference might be related to the types of organization from which we could motivate participants to fill out the questionnaire. As Fig. 4 shows, 61% (33 out of 54) of participants from academia were users only, while a similar percentage of participants from governmental organizations and the public sectors (58%, 38 out of 65) were providers only.

Figure 3: Types of Users in the Austrian and the International Sample (Source: Authors)

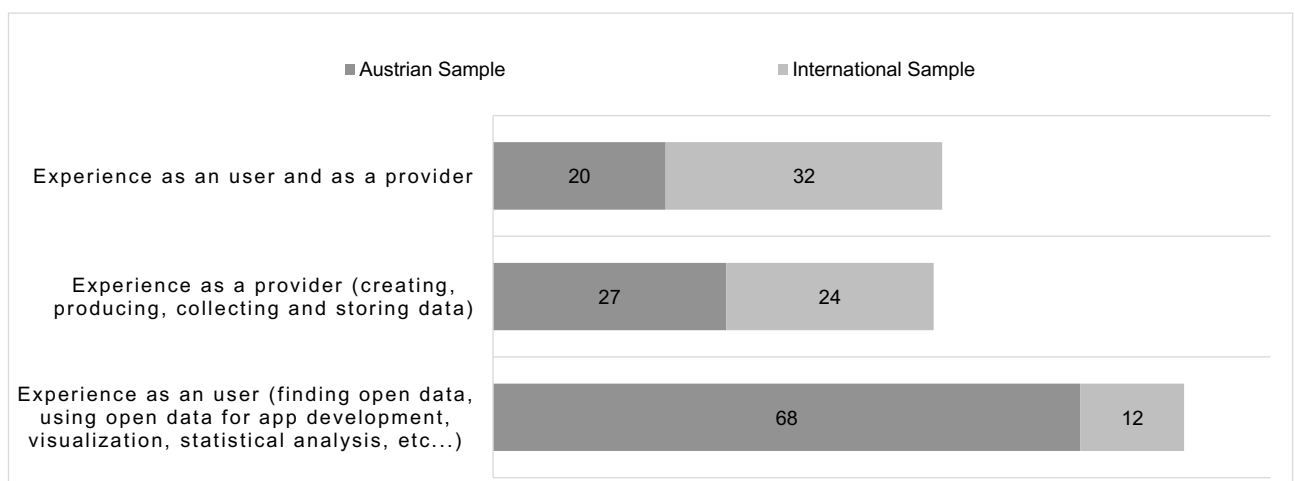
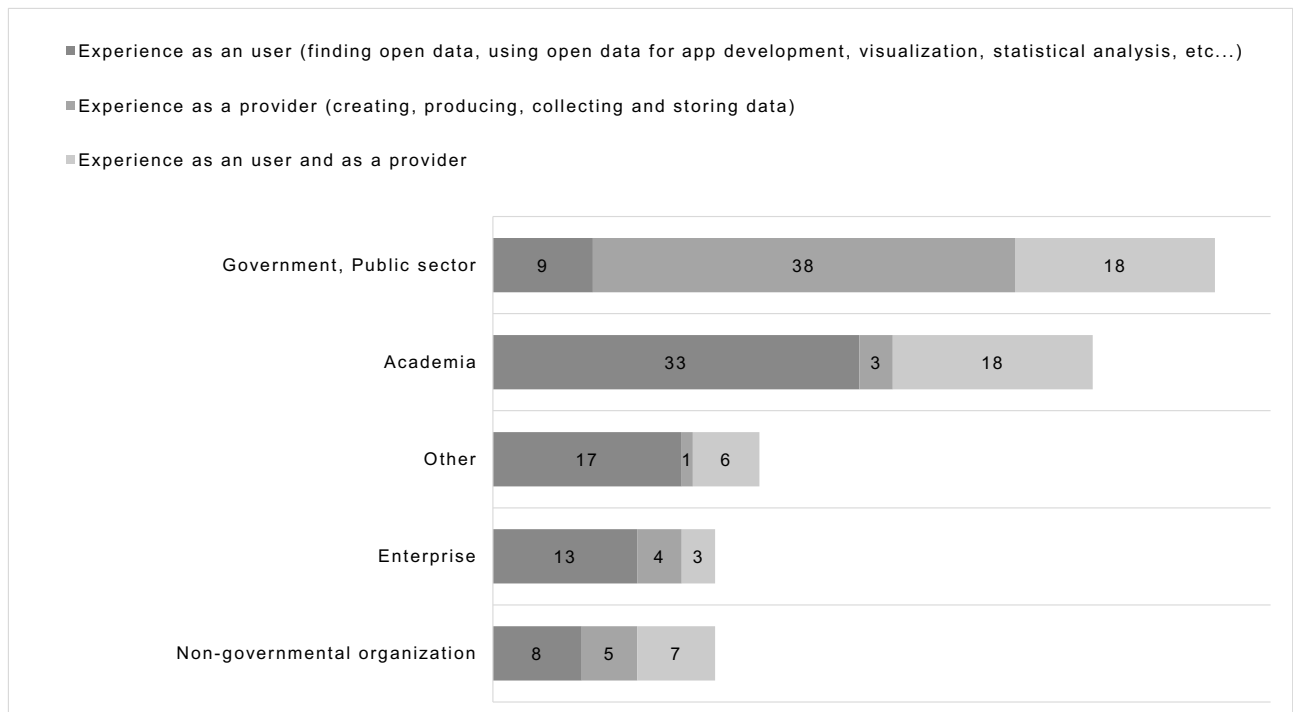


Figure 4: Types of Users and Types of Organizations (Source: Authors)



4.1. User Barriers

The questionnaire section on barriers for users was split into 4 categories: data portals, data quality, legal constraints and (lack of) knowledge and experience. Fig. 5 presents an overview of all user ratings, sorted per category based on the mean.

4.1.1. Data Quality

We first discuss data quality issues. The most important reported barrier is that data is not machine readable ($M=3.33$, $SD=1.47$). This result seems to be consistent with other research, which found that in practice, data is often published in proprietary, only human-readable formats. Indeed, in their analysis of 260 Open Data Portals, Neumaier, Polleres and Umbrich (2016) found that 23% of the published data were PDF files. Furthermore, our results suggest that there may currently be datasets with incomplete data being published, which drastically reduces their usefulness. The respondents also reported some additional points in the open comments fields:

- "Old data could be even worse than no data."
- "Datasets are not updated regularly."
- "Most importantly a lack of adherence to standard classifications so that data can be linked in a meaningful way. I.e. health with level of education etc."

By not updating the published datasets regularly, publishers risk their data becoming obsolete and no longer accurate or of any potential use. Users in our sample rated obsolete, non-valid data as second most relevant data quality barrier. To ease the process of linking data, metadata is often

used to describe the datasets in order to provide the users with a better overview. However, metadata are often incomplete (Martin et al., 2013), difficult to interpret or simply non-existent (Zuiderwijk, Janssen, Choenni, et al., 2012). It could be argued that if certain publishing standards were implemented, these problems would not exist. By publishing data in a machine-readable format with accurate metadata, many quality issues could be avoided (Attard, Orlandi, Scerri, & Auer, 2015).

4.1.2. 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 and complicated by an unclear licensing situation. As data users try to avoid any legal trouble, they are sometimes better off not using certain datasets with licenses which 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 users perceive unclear and restrictive licensing as relevant barriers to Open Data use. On average, restrictive ($M=3.05$, $SD=1.43$), unclear ($M=3.01$, $SD=1.29$) and hard to understand ($M=2.82$, $SD=1.24$) licenses were judged as moderate barriers. Some users may not have been able to use the data at all 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. Indeed, the comments below illustrate that there is a sense of frustration amongst the data users caused by the various legal constraints:

- “In the case of administrative boundary information (GADM), it has an unclear license. It really makes it hard to know if we can use it.”
- “Licenses for Open Data should all adhere to CC-BY only. There should be no qualifications of this or any other charging or restrictive conditions. The data production is funded by tax payers and should be available for all to use (subject to appropriate confidentiality restrictions). Governments should not try to use data as sources of revenue, it is counterproductive to the overall economy.”
- “This area is a complete mess. They should converge on a set of licenses like the open source community has done.”

Interestingly, the last comment seems to suggest that Open Source Software does not suffer from licensing issues. However, as we have discussed in the second section of this article, complex and restrictive licenses are very much present in OSS as well. There is no single unifying OSS license, and the existence of multiple licenses reflects a divergence of philosophies within the OSS community (Yoo, 2016). While some licenses generally do not restrict the user in any way, the GPL for instance, requires any code combined with GPL-licensed code to be published as GPL, arguing that it is crucial that developers give the same freedom to others as they themselves have been given. The viral aspect of GPL makes it impossible to merge code to certain OSS projects with incompatible licenses. We once again draw a parallel to Open Data, where combining data from various datasets may not be possible due to licensing issues. While we agree that OSS may have a slightly better licensing situation, possibly due to it being a much more mature concept, we do not see copying the OSS licensing model as a suitable solution to mitigate the licensing barriers of Open Data.

User evaluations of the barrier “Unclear licensing (Unclear conditions for reuse)” significantly differed between the types of organization users worked for ($F(4,84)=3.15$, $p=0.018$). Users working for non-governmental organizations (NGOs) rated that barrier more severe than users from other organizations (mean difference=0.99-1.80, $SD=0.45-0.53$, $p\leq 0.03$). In addition, academic users rated this barrier more severe than users employed in an enterprise (mean difference=0.82, $SD=0.39$, $p=0.04$). One possible interpretation could be that academic users who are interested in Open Data might reflect more deeply on the various types of license models, another explanation might be that some license models allow free use for research and non-commercial use.

In summary, these results show the need for a clear, less restrictive licensing situation. The existing legal issues present an unnecessary barrier that could easily be resolved by agreeing to use licenses that do not restrict the data users in any way, which is the fundamental idea behind Open Data.

4.1.3. Data Portal Usefulness

The lack of information about the quality ($M=3.12$, $SD=1.25$) and the content ($M=3.00$, $SD=1.20$) of a dataset were rated as strongest 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 (e.g. an updated dataset or revised metadata), thus saving the work and effort for other users and improving the quality of the portal.

Problems with APIs were ranked of medium severity. International participants rated it more problematic than Austrian participants that data is temporarily not available ($M_{Int}=2.93$, $SD_{Int}=1.46$; $M_{Aus}=2.25$, $SD_{Aus}=1.22$; $t(81)=-2.29$, $p=0.02$) and that access to datasets is restricted to certain user groups ($M_{Int}=3.21$, $SD_{Int}=1.52$; $M_{Aus}=2.35$, $SD_{Aus}=1.51$; $t(81)=-2.43$, $p=0.02$). The reason for this is not clear, but a possible explanation might be that the overall quality of Austrian Open Data portals is higher than the average. 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 the Open Data users were once again asked to describe any other issues with Open Data portals:

- “Usability of the data portal - should be user friendly and easy to use.”
- “Inflexibility and slowness of responses received to questions/requests from public data owners.”
- “Complexity of queries. Necessity to understand how the data is organized in a very detailed way before being able to query it. More work is needed to make it possible for the non-programmer to access the data.”
- “Possible duplication of data in varying formats and timeliness.”

A recurrent theme in the comment fields are the complaints about data not being kept up-to-date. It is interesting to note that this particular barrier was also identified by the participants in the “Data Quality” section of the questionnaire:

- “Unclear long-term commitment of publishers.”
- “No governmental data in Ireland is ever updated after initial upload. Ever.”

It could be argued that the publishers do not update the data, as they are not sure whether the data is being used at all and keeping it up-to-date would be considered a wasted effort. However, the following comments suggest that even after the users express demand, their requests are often ignored:

- “I have not received feedback for requests to update a dataset.”
- “Inflexibility and slowness of responses received to questions/requests from public data owners”

Keeping the data up-to-date takes effort. This is an interesting issue to be discussed in comparison with the Open Source Software model. In OSS, 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 OSS including increased collaboration, knowledge sharing with communities, customers, and other parties and facilitation of joint ventures with other companies or research institutes (Morgan & Finnegan, 2014). Similarly, opening the data could facilitate collaboration between the provider and the Open Data community, thus helping the provider maintain their data.

4.1.4. Knowledge and Experience

On average, lack of proper documentation presents a moderate barrier ($M=2.97$, $SD=1.22$). The other barriers, such as lack of support or helpdesk seem to be less of an issue. It is important to bear in mind the possible bias in these responses, resulting from the survey’s target audience being already users of Open Data. It can therefore be assumed, that the respondents are at least somewhat experienced, and consider the lack of support or a helpdesk rather insignificant.

In the open comment field, one user gave an interesting suggestion: “I think more examples with code etc. should be made available for each data entry/API etc. so that users can pick these up and modify as required”. Indeed, education material such as instructions on how to use the data could be made available on each portal, which could help to mitigate the entry barriers, especially for the newcomers to Open Data. It would be beneficial if consistent guidelines could be used across data portals, as different guidelines would create unnecessary confusion. A further examination of the barrier “No guiding principles/instructions” reveals that it was rated most severe by users working for the government or the public sector, differing significantly from academic and enterprise users and other organizations ($F(4,77)=2.66$, $p=0.039$; mean difference= $0.79-1.13$, $SD=0.32-0.45$, $p\leq 0.05$). A possible explanation for these differences is that government agencies are often forced to work with Open Data under new directives (K. Janssen, 2011; Kassen, 2013) and may therefore not be as experienced as those users who use Open Data of their own volition such as academics, or private entities interested in exploiting the benefits of Open Data.

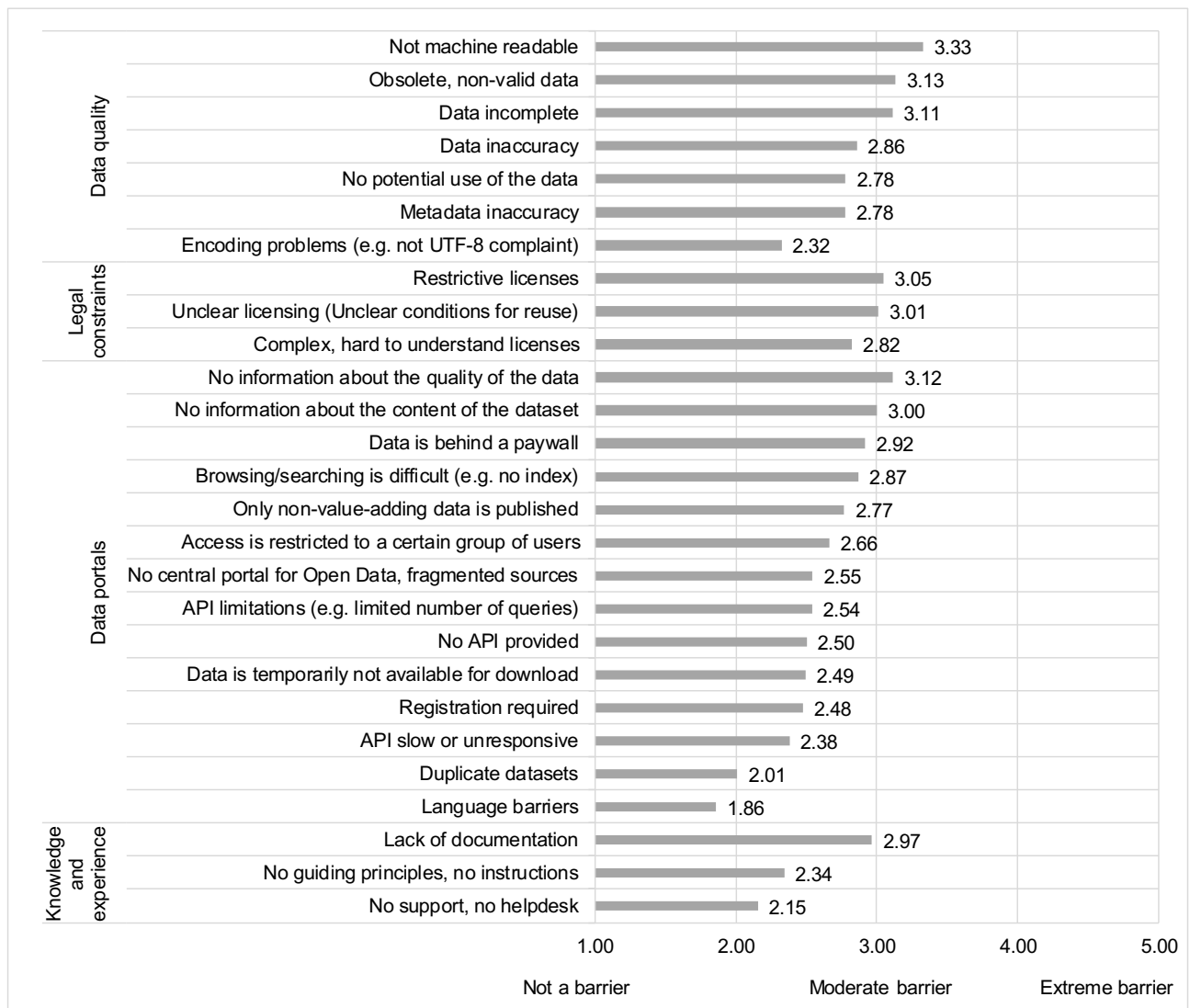
4.2. Opinions on Opening Data Sets

In a further item, we asked the 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 (61%). On the other hand, more than half of participants (62%) 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 (38%) 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:

- “You need a minimum quality standard, and hopefully some practical standards. For example: 1) Always provide metadata on quality, how and when the data was collected, etc. 2) Share the data in open machine readable formats, such as csv, shapefiles, etc.
- “Universal system for indicating data quality needed. Also, for data gathered through a survey, copy of the actual survey available.”
- “Quality rules should be recommendations, the target to reach, not a barrier to entry. However, to keep users from wasting time on low quality datasets, a public quality ranking would be useful.”
- “Follow at least community-wide metadata standards for content and completeness.”
- “Support Http Content Negotiation according to Standards, so the most common RDF libraries can directly resolve the URIs”

Figure 5: Barriers to Using Open Data (Source: Authors)



4.3. Provider Barriers - 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 low 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.

4.3.1. Legal constraints

One interesting comparison is that providers on average rated complex, hard to understand licenses less as a barrier than users (Providers: $M=1.91$, $SD=1.01$; Users: $M=2.82$, $SD=1.24$; $t_{101}=-5.39$, $p<0.001$). It is difficult to explain this result, but the reason might be that data providers either have more experience in the legal aspects of working with data than the users, or in case of bigger institutions, have an in-house legal department advising them on such matters.

Concerning legal barriers, we could identify differences between providers from different organizations for the ratings of three items:

- 1) "Unclear ownership of the data, prohibiting its release" ($F(4,88)=3.62$, $p\leq 0.01$). Providers from government organizations rate this barrier less severe than providers from academia and non-governmental organizations (mean difference= $0.88-1.05$, $SD=0.33-0.41$, $p\leq 0.03$). Conradie and Choenni (2014) claim that due to the vertical data management in the past, data sharing between departments has not always been the case which resulted in a situation within organizations, where departments lack a complete picture about which data the others control. It could be argued that a lack of proper data organization is more prevalent in academia and non-governmental organizations, which would explain this finding.
- 2) "Threat of lawsuits" ($F(4,86)=3.13$, $p=0.02$). Providers employed by the government rate the barrier lower than providers from academia (mean difference= 0.70 , $SD=0.27$, $p=0.01$), furthermore, providers from the "other" category of organization rate this threat lower than all organizations (mean difference= $1.18-1.42$, $SD=0.51-0.61$, $p\leq 0.03$). This might be due to the variety of participants who had chosen this category e.g. a journalist or a member of a church who are in a less likely position to be sued.
- 3) "Complex, hard to understand licenses" was also rated as more severe barrier by providers from academia than from government ($F(4,84)=2.81$, $p=0.03$, mean difference= 0.82 , $SD=0.27$, $p=0.003$). As we have already mentioned, organizations with their own legal department do not rate the complex licensing situation as a strong barrier. Academics are more likely to publish their data without having to deal with legal departments first and thus, are also left on their own to interpret the meaning of the licenses.

Lastly, 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 ($M=2.57$, $SD=1.29$).

4.3.2. 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 ($M=1.67$, $SD=0.92$) 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 ($M=2.09$, $SD=1.23$).

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 ($M=3.30$, $SD=1.08$). This could indicate that providers need some support and guidance on how to minimize these costs. Additionally, cost issues ($M=2.74$, $SD=1.24$) and resistance to change ($M=2.84$, $SD=1.24$) were rated as weak to moderate barriers.

International participants more often than Austrian participants report that Open Data is not a priority ($M_{Int}=2.87$, $SD_{Int}=1.19$; $M_{Aus}=2.24$, $SD_{Aus}=1.25$; $t(81)=-2.45$, $p=0.02$). This difference could be caused by regional differences between the participant countries, but also by a change of perception of Open Data in the one and a half years that have passed between the two surveys. A detailed analysis is unfortunately not possible from the data available, since there are no questions about the perception of Open Data over time in our questionnaire. However, hints about the different perception could be based on recent political changes, such as the new US administration for instance. The current Whitehouse administration has, as one of its first acts, removed all datasets from <http://data.whitehouse.gov> without restoring or replacing any Open Data since². In contrast, the previous administration under former President Barack Obama, has issued a Memorandum on Transparency and Open Government, essentially launching the open government movement (Veljković, Bogdanović-Dinić, & Stoimenov, 2014). It is possible that changes in the political climate have an influence on transparency, openness, and therefore the perception of Open Data as well.

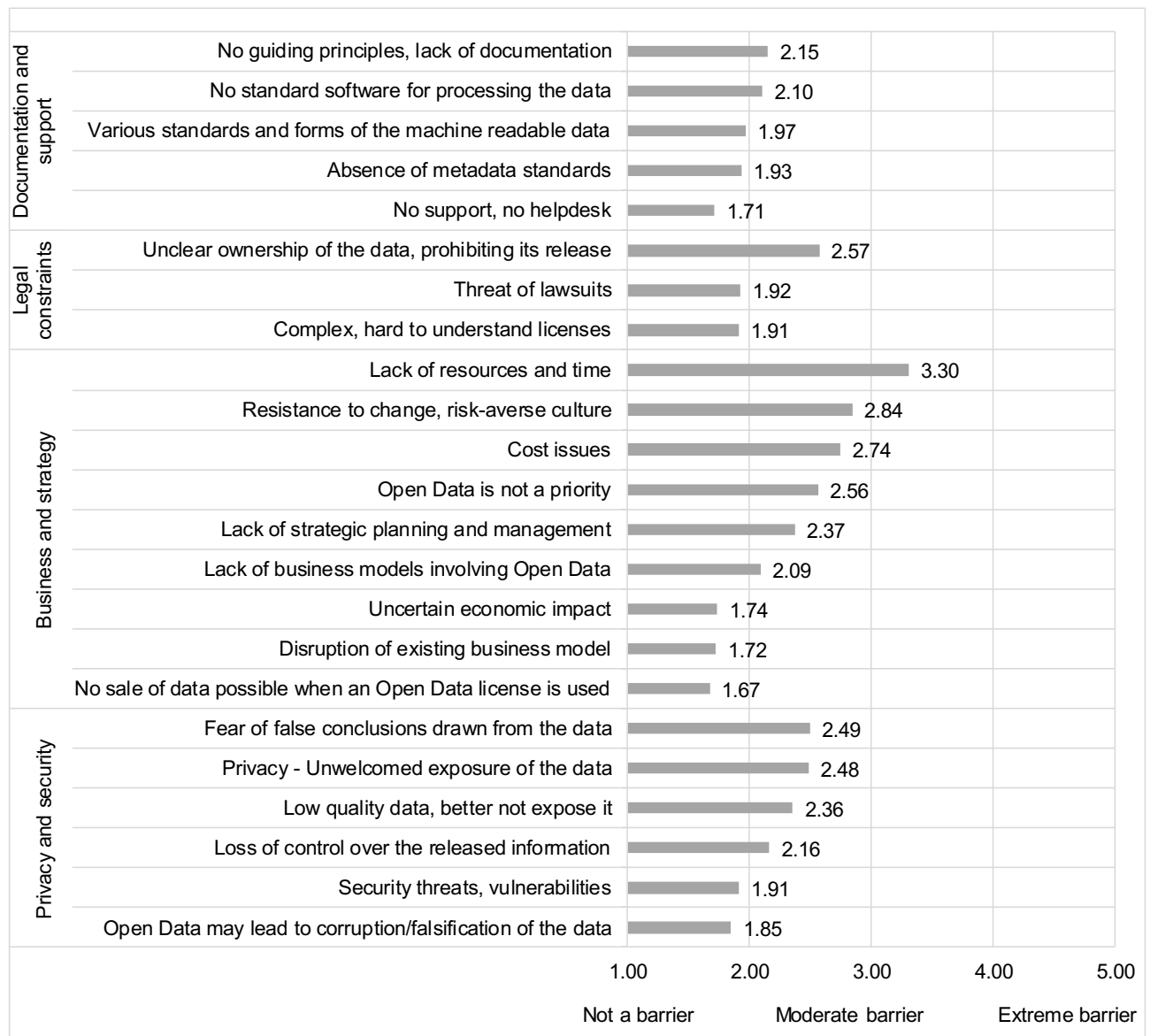
One interesting finding is that providers' evaluations of the barrier "Resistance to change, risk-averse culture" significantly differed between the types of organization users worked for ($F(4,83)=2.66$, $p=0.04$). Providers from enterprises rated this barrier more severe than all others types of organizations (mean difference= $1.24-1.80$, $SD=0.56-0.76$, $p\leq 0.03$). The reason for this might once again be the fact, that government and public institutions are often required to publish Open Data under certain directives, as we have already pointed out in an earlier section of this article. The risk of an unknown payoff usually does not apply to public institutions as they often do not expect a payoff in the first place. Private entities on the other hand more often avoid investing in strategies without a guaranteed return on investment.

4.3.3. 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 ($M=2.49$, $SD=1.17$). 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 ($M=2.36$, $SD=1.10$). Lack of resources for the maintenance of the data might be a reason for low data quality. International participants are less afraid than Austrian participants that false conclusions may be drawn from the data ($M_{Int}=2.26$, $SD_{Int}=1.22$; $M_{Aus}=2.74$, $SD_{Aus}=1.06$; $t(81)=-2.03$, $p=0.05$) and that Open Data may lead to corruption/falsification of data ($M_{Int}=1.65$, $SD_{Int}=0.96$; $M_{Aus}=2.07$, $SD_{Aus}=0.99$; $t(81)=-2.09$, $p=0.04$).

² cf. http://data.wu.ac.at/portalwatch/portal/open_whitehouse_gov/1738/evolution

Figure 6: Barriers to Opening Open Data (Source: Authors)



5. Barrier Mitigation Strategies

So far this article has focused on the identification and analysis of Open Data barriers. Based on these insights, we now propose mitigation strategies, which could improve the experience for both Open Data users and publishers. In recent years, the Open Data movement has gained momentum and enjoyed an increased interest from analysts, statisticians, software developers, governments and private business entities. However, as long as severe barriers persist, Open Data may lose its momentum and the interest in it may start to fade. While it would be out of scope for this article to develop mitigation strategies for each single barrier that has been discussed, we want to focus on most relevant potential mitigation strategies of Open Data barriers.

According to the literature review and the survey results, data portals sometimes lack information about content and quality of the dataset. A recent study on benchmarking open

government (Veljković et al., 2014) suggests to provide rating information for datasets. The authors argue that this measure could significantly increase the users' trust in the published data. Additionally, they recommend that it should be mandatory for data publishers to provide description of the data, in order to increase the clarity of the contained information.

In order to lower the cost of publishing data, Conradie and Choenni (2014) advise against releasing data just for its own sake in their study on barriers of open government data. They claim that to ensure efficient use of released data, a more focused approach is needed, according to which data with the highest return on investment is released, not only in financial sense, but to achieve the highest reuse. Demand-based data release would imply lower costs. Additionally, Martin et al. (2013) propose to use a system to measure the popularity of the published datasets in order to identify and clarify the information needs of society.

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 by literature (Martin et al., 2013). Furthermore, released data should comply with the definition of openness as defined by the Open Definition (<http://opendefinition.org>). This view is also held by Toots, McBride, Kalvet and Krimmer (2017), who advise to review data licenses to ensure their compatibility with Open Data goals.

According to Zuiderwijk, Jeffery and Janssen (2012), the use of metadata can yield significant benefits such as improving the ability to find the data, providing the context for the datasets and avoiding unnecessary duplication. Additionally, metadata can be used to establish a link between the creator and the user of the data and to establish a communication channel, enabling the users to give valuable feedback to the publishers.

Many potential data providers are still hesitant to publish their data. A report on Open Data by the European Commission (2011) 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. Publishing Open Data however, is not an easy task. Guiding principles such as those published by Zuiderwijk et al. (2012) can provide a systematic guidance on releasing Open Data. Furthermore, in their study on Open Data barriers and drivers, Toots, McBride, Kalvet and Krimmer (2017), recommend a number of strategies to improve the general lack of knowledge and experience. They suggest to implement training programs for public sector officials in order to build data-related knowledge and skills.

Table 9: Barrier Mitigation Strategies (Source: Authors)

Identified Barrier	Proposed mitigation strategy	References
No information about the quality of the data	Provide rating information about the dataset	(Veljković et al., 2014)
No information about the content of the data	<ul style="list-style-type: none"> • Mandatory descriptions of the dataset • Use of standardized metadata models 	(Veljković et al., 2014) (Zuiderwijk, Jeffery, et al., 2012)
Cost issues	Release data based on demand	(Conradie & Choenni, 2014)
Restrictive and unclear licensing	<ul style="list-style-type: none"> • Collect the concerns of the users and appropriately modify the license • Use a license that complies with the rules defined by the Open Definition 	(Martin et al., 2013; Toots et al., 2017)
Difficulty in finding the data, Duplicate data, Lack of communication between the provider and the user	Use of standardized metadata models	(Zuiderwijk, Jeffery, et al., 2012)
Lack of guiding principles	Make use of systematic research publications on guiding principles	(Zuiderwijk, Janssen, Meijer, et al., 2012)
Resistance to change, risk-averse culture	<ul style="list-style-type: none"> • Sharing of good practices and use cases • Mobilization campaigns to promote open culture 	(Commision, 2011)
Lack of business models	Promote networking between stakeholders	(Martín et al., 2015)
Language barriers	<ul style="list-style-type: none"> • Publish data in multiple languages (If not possible, publish at least metadata in multiple languages) 	(Martín et al., 2015; Zuiderwijk, Jeffery, et al., 2012)
Lack of knowledge and experience	Training and skills development	(Toots et al., 2017)

6. Discussion and Limitations

The presented study has been set out with the aim of assessing the relative importance of barriers for using and publishing Open Data. Most of the barriers that are widely cited in the literature have also been confirmed in our survey. Furthermore, we contribute to existing research 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 resources 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. 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. Another survey participant identified licenses not being machine-readable as a legal barrier, a barrier that has to our knowledge not yet been discussed in the literature. 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.

The generalizability of our findings is subject to certain limitations. For instance, our respondents were at least somewhat experienced in the field of Open Data; discovering potential new barriers faced by newcomers or organizations that are currently not interested in publishing or using data in an open format were out of scope of the current study. Another limitation was the language of the questionnaire versions; first we used a German questionnaire for collecting data in Austria, then an English questionnaire to collect data from users and providers from other countries. As a consequence, users and providers who did not speak or were not comfortable with the language used were unlikely to participate in the questionnaire which might have introduced a sample selection bias.

From a statistical point of view, we note that there are limits to the conclusion validity. To be able to compare evaluations of different user types and evaluations of the international with the Austrian version of the questionnaire, we had to perform a high number of single, separate analyses. As we reported the results without Bonferroni alpha level adjustments, the reader should take into consideration that when performing multiple tests the likelihood of incorrectly rejecting a null hypothesis is increased (cumulative type-1 error). Furthermore, when comparing different organization types, for some combinations (e.g. providers from enterprises) the number of responses was low. In conclusion, we do advise the reader to bear in mind that we report preliminary, not necessarily statistically conclusive findings.

Based on the fact that we used a convenience sample and the distribution of users types in the Austrian and international sample were different, the generalizability of the findings is limited. However, we do believe that inviting participants via mailing lists was consistent with the goals of the study and that the sample size was large in relation to the low basic population of highly specialized users and providers of Open data we were looking for.

The private sector is rather underrepresented in the literature in the research field as well as in our survey sample. A future study investigating the potential of Open Data use in the private sector could provide valuable new insights.

7. Conclusion and Outlook

This article has given an account of Open Data barriers as experienced by practitioners using and providing Open Data. 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, data providers, and a handful of app developers. The relevance of addressing and lowering barriers is clearly supported by the current findings. As the results of the empirical study showed, there are numerous barriers 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. Such results might provide insights for setting priorities and an 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

- Alani, H., Dupplaw, D., Sheridan, J., O'Hara, K., Darlington, J., Shadbolt, N., & Tullo, C. (2007). Unlocking the potential of public sector information with semantic web technology *The semantic web* (pp. 708-721): Springer.
- Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly*, 32(4), 399-418.
- Bertot, J. C., Jaeger, P. T., & Grimes, J. M. (2010). 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.
- Braunschweig, K., Eberius, J., Thiele, M., & Lehner, W. (2012). The State of Open Data Limits of Current Open Data Platforms.
- Commision, E. E. (2011). *Open data. An engine for innovation, growth and transparent governance*. Retrieved from
- Conradie, P., & Choenni, S. (2012). *Exploring process barriers to release public sector information in local government*. Paper presented at the Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance.

- Conradie, P., & Choenni, S. (2014). On the barriers for local government releasing open data. *Government Information Quarterly*, 31, S10-S17.
- Geiger, C. P., & von Lucke, J. (2012). Open government and (linked)(open)(government)(data). *JeDEM- eJournal of eDemocracy and Open Government*, 4(2), 265-278.
- Goode, S. (2005). Something for nothing: management rejection of open source software in Australia's top firms. *Information & Management*, 42(5), 669-681.
- Huijboom, N., & Van den Broek, T. (2011). Open data: an international comparison of strategies. *European journal of ePractice*, 12(1), 4-16.
- Janssen, K. (2011). The influence of the PSI directive on open government data: An overview of recent developments. *Government Information Quarterly*, 28(4), 446-456.
- Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, 29(4), 258-268.
- Johnson, P., & Robinson, P. (2014). Civic Hackathons: Innovation, Procurement, or Civic Engagement? *Review of Policy Research*, 31(4), 349-357.
- Kassen, M. (2013). A promising phenomenon of open data: A case study of the Chicago open data project. *Government Information Quarterly*, 30(4), 508-513.
- Martín, A. S., De Rosario, A. H., & Pérez, M. D. C. C. (2015). An International Analysis of the Quality of Open Government Data Portals. *Social Science Computer Review*, 0894439315585734.
- Martin, S., Foulonneau, M., Turki, S., & Ihadjadene, M. (2013). Risk analysis to overcome barriers to open data. *Electronic Journal of e-Government*, 11(1), 348-359.
- McDermott, P. (2010). Building open government. *Government Information Quarterly*, 27(4), 401-413.
- Morgan, L., & Finnegan, P. (2014). Beyond free software: An exploration of the business value of strategic open source. *The Journal of Strategic Information Systems*, 23(3), 226-238.
- Neumaier, S., Umbrich, J., & Polleres, A. (2016). Automated quality assessment of metadata across open data portals. *Journal of Data and Information Quality (JDIQ)*, 8(1), 2.
- Stol, K.-J., & Ali Babar, M. (2010). *Challenges in using open source software in product development: a review of the literature*. Paper presented at the Proceedings of the 3rd International Workshop on Emerging Trends in Free/Libre/Open Source Software Research and Development.
- Tammisto, Y., & Lindman, J. (2012). Definition of open data services in software business. *Software Business*, 297-303.
- Toots, M., McBride, K., Kalvet, T., & Krimmer, R. (2017). *Open data as enabler of public service co-creation: exploring the drivers and barriers*. Paper presented at the E-Democracy and Open Government (CeDEM), 2017 Conference for.
- Umbrich, J., Neumaier, S., & Polleres, A. (2015). *Quality Assessment and Evolution of Open Data Portals*. Paper presented at the Future Internet of Things and Cloud (FiCloud), 2015 3rd International Conference on.
- van Veenstra, A. F., & van den Broek, T. A. (2013). *Opening moves—drivers, enablers and barriers of open data in a semi-public organization*. Paper presented at the International Conference on Electronic Government.

Veljković, N., Bogdanović-Dinić, S., & Stoimenov, L. (2014). Benchmarking open government: An open data perspective. *Government Information Quarterly*, 31(2), 278-290.

Yoo, C. S. (2016). Open Source, Modular Platforms, and the Challenge of Fragmentation.

Zuiderwijk, A., & Janssen, M. (2014). Barriers and development directions for the publication and usage of open data: A socio-technical view *Open Government* (pp. 115-135): Springer.

Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., Alibaks, R. S., & Sheikh_Alibaks, R. (2012). Socio-technical impediments of open data. *Electronic Journal of eGovernment*, 10(2), 156-172.

Zuiderwijk, A., Janssen, M., Meijer, R., Choenni, S., Charalabidis, Y., & Jeffery, K. (2012). *Issues and guiding principles for opening governmental judicial research data*. Paper presented at the International Conference on Electronic Government.

Zuiderwijk, A., Jeffery, K., & Janssen, M. (2012). The potential of metadata for linked open data and its value for users and publishers. *JeDEM-e-Journal of e-Democracy and Open Government*, 4 (2) 2012.

About the Authors

Martin Beno

Martin Beno is a research assistant at the Institute for Information Business at the Vienna University of Economics in Austria. He joined the research group while pursuing his degree in Business Information Systems. His current research focuses on Open Data, Semantic Web technologies, and the adoption of the Open-Source model in public institutions.

Kathrin Figl

Dr. Kathrin Figl is an Assistant Professor at the Institute for Information Systems and New Media at the Vienna University of Economics. She received her Doctoral (awarded with the Dr. Maria Schaumayer Award) and two Master's in Information Systems and Psychology, both with honours, from the University of Vienna. Her research focuses on human-centric development and design of information systems and she has published more than 70 research papers and articles, among others in the Journal of the Association for Information Systems, Decision Support Systems, Information & Management, Business & Information Systems Engineering, Requirements Engineering, Information and Software Technology and the International Journal of Human-Computer Studies.

Jürgen Umbrich

Dr. Jürgen Umbrich graduated in Business Engineering from the University of Karlsruhe in Germany (Dipl.-Inf.) in 2008 and holds a PhD degree from the Digital Enterprise Research Institute at the National University Ireland, Galway since 2012. His PhD topic was centred around efficient SPARQL query processing over evolving RDF data on the Web, partially through lightweight RDFS reasoning and a novel hybrid execution query framework. Before he joined the WU, he worked one year as a post-doctoral researcher at Fujitsu Ireland in Galway exploiting the benefits of Linked Data for enterprise applications.

Axel Polleres

Axel Polleres works as the chair for Data and Knowledge Engineering and Head of the Institute for Information Business at Vienna University of Economics in Austria, which he joined in Sept 2013.. Since January 2017 he is also a member of the Complexity Science Hub Vienna Faculty. He obtained his Ph.D. and habilitation from Vienna University of Technology and worked at University of Innsbruck, Austria, Universidad Rey Juan Carlos, Madrid, Spain, the Digital Enterprise Research Institute (DERI) at the National University of Ireland, Galway, and for Siemens AG's Corporate Technology Research division before joining

WU Wien. His research focuses on querying and reasoning about ontologies, rules languages, logic programming, Semantic Web technologies, Web services, knowledge management, as well as Linked and Open Data. Moreover, he actively contributed to international standardisation efforts within the World Wide Web Consortium (W3C) where he co-chaired the W3C SPARQL working group.