Investigating open government data barriers: A literature review and conceptualization

Jonathan Crusoe and Ulf Melin

Tweet

LINKÖPING UNIVERSITY

Investigating Open Government Data Barriers

A Literature Review and Conceptualization

Jonathan Crusoe[1] [0000-0003-4740-1242] (✉) and Ulf Melin[2] [0000-0002-2784-863X]

[1] Division of Information Systems, Linköping University, SE-581 83 Linköping, Sweden
Jonathan.crusoe@liu.se
[2] Division of Information Systems, Linköping University, SE-581 83 Linköping, Sweden
Ulf.melin@liu.se

Abstract. When focusing on open government data (OGD) publishing and related barriers, there are several complexities present. Largely, current research is focused on publishing and usage of OGD; and we argue that there are a need to investigate and to systematise OGD barrier research in order to understand and outline an expanded scope of the phenomenon. We expand by clarifying barriers linked to the release decision and the data's organisational context. To investigate the OGD barriers, we conduct a systematic literature review, identifying 34 articles as a point of departure for our analysis. From these articles we create, present and discuss illustrations on historical development, barrier types, and different research focuses on OGD. When analysing the articles, we identify a focus on technical, organisational, and legal barrier types, while studies on open data usage and systems are less frequent. Our analysis also identifies some possible open data research barriers. In the article we also relate barriers to an expanded OGD process (Suitability, Release, Publish, Use, and Evaluation), identifying 46 barriers with possible linkages. The results is an expanded scope and a conceptual illustration of OGD barriers.

Keywords: Open data, Open government data, OGD, Barriers, Risks, Challenges, Impediments, Myths, Process, Literature review.

1   Introduction

Open government data (OGD) refer to data sets that government agencies make available for third-party usage. Putting OGD in contrast to history, where public organisations produce data enclosed in silos where its value is limited. One purpose is to make data accessible to a broader audience by stimulating the third-party development of new e-services (e.g. apps for weather forecasts and smarter travelling routes). By opening data for others to reuse, we gain OGD [17]. OGD is surrounded by many expectations of service innovation, increased transparency and inclusive government agencies [3].

As a concept and phenomenon OGD is promising, but at the same time challenging. OGD research and practice has shown several complexities when publishing data [e.g. 3, 9, 12, 17, 39, 42]. Challenges, or barriers as we conceptualise them in this paper, are described under different conceptual labels in previous research, such as; barriers [e.g.

4], impediments [e.g. 42], myths [e.g. 13, 20], and challenges [e.g. 28]. Multiple analytical perspectives are also present in previous research; focusing on barriers linked to release [e.g. 4], adoption [e.g. 20], cognitive [e.g. 37], agenda [e.g. 26], process [e.g. 9], and use [e.g. 42], use and publish [e.g. 5]. OGD is also placed in different larger context, by using, e.g. an ecosystem perspective [12]. In present OGD research, there is a tendency to study specific parts or specific phases of a larger system or process.

Another complexity is the nature of the OGD barriers and the OGD process. The barriers can e.g. be linked to different process phases, and there are several ways to divide the OGD process into phases [e.g. 3, 7, 12, 42]. Which is also further complicated when different actors can participate in one or several phases [42], and activities in one phase can create barriers in a later for someone else [39]. The complexity can be visualised as a barrier network [18], which include a reversal of the cause-and-effect; a barrier can come from a later solution. We, therefore, see earlier OGD research giving room for specialised solutions for unique and challenging problems; where solutions can cause barriers in both directions of the OGD process. There is a need to understand and systematise existing OGD barrier research with an expanded scope.

We expand the scope by taking a step back and view other actors and activities surrounding the publisher and open data publication, which includes more OGD barriers. We investigate the OGD barriers in earlier research with a systematic literature review and expand on earlier works, such as Zuiderwijk and Janssen's OGD process [38], through a systematisation. An expanded scope can serve as inspiration for a future research agenda and practice in the field. The result illustrates insights into the complexities and challenges encountered when publishing. We also outline some future OGD barrier research avenues. The following research questions are focused in this paper:

— What OGD barriers have been identified in previous research?
— Where are the OGD barriers encountered in the OGD process and how are they connected?

In the following sections, we will start by describing our research approach. Then we will discuss the identified barriers from the literature and chart illustrations, and in the next section presents the OGD barrier systematisation. We end the paper with a conclusion containing limitations, future research, and implication.

2 Research Approach

We conducted a systematic literature review [3, 10, 17] where we analysed the findings through coding [32] to make illustrations and a systematisation. The literature review [3, 10, 17] started by defining keywords, and then selected a database that was searched, results were filtered and summarised, and lastly, we analysed the findings. We created simple keywords by combining "open data" with "barrier", "risk", "challenge", or "impediment". For database, we selected Google Scholar as it has good coverage, with recall and precision that is above average for simple keyword searches [36]. For each keyword search, we looked at the first 50 results. We first searched all years, but also wanted to be sure that we caught the latest and relevant articles, we therefore explicitly

searched the years 2016 and 2017. This gave us a total of 600 articles; we discarded duplicates. We identified relevant articles by studying the title and Google Scholar summary for concepts relating to our keywords. Relevant articles were summarised with article id, authors, title, year, publish origin, topic, and important conclusions, if needed, purpose and method were included, such as the case of the systematic literature reviews. Later when analysing the articles, we discovered interesting references and snowballed. We summarised their content in the same way as above. This resulted in a final set of 34 articles. We analysed the articles to present them as data (see section 3) and to systematise OGD barriers (see section 4). For the data presentation, we decided to analyse development, types, and focuses, making a distinct coding effort for each into a representative illustration. When coding for development, we sorted the articles based on publication year. We then used them to form a graph and coded based on what and possible why. For analyzing the types, we divided articles into two groups: specific articles (containing a thematic division or typology of barriers into types), while the general articles contain barriers (but not structured by type, e.g. [5] structured by role, or [4] structured by type). We coded the first group by following their thematic division or typologies. After that, as part of categorization, we integrated the second group into the first through coding. When analysing focus we assumed that the abstracts provided an overview of the work with information about the background, objective, methods, results, and conclusions [43, 44]. We therefore initially coded [32] the abstract for research focus. If the abstract lacked the necessary information, the whole paper were explored. After codes were categorised into groups and we merged and split groups to achieve a pedagogic presentation. The product was a sketch that we as researchers discussed and analysed in a generative dialogue, and some final changes were made. The results were the three illustrations in section 3. For the systematisation, we first identified an OGD process, with different parts and activities. We then categorise coded [32] the 34 articles into it. Giving the barriers a chance to expand the OGD process by e.g. including concepts. After, we identified possible linkages between barriers following the network idea of Huang, et al. [18]. Linkages were identified when barriers shared common core phenomena or phenomena were dependent on each other. As a final step, we refined the categories by reflecting on their meaning.

3 The Identified OGD Barrier Literature

Below, we present the articles as data through illustrations following our literature analysis, in the following order: historical development of the OGD barrier discussion, barrier types, and research focuses. We note that some research only acted as *enablers* [17, 21, 35] (used to snowball) and was, therefore, not used in the later systematisation.

3.1 Historical Development

In this section, we present the historical development of the OGD barrier discussion. In Figure 1, we have divided the number of articles published 2011-2017. In 2014, we

identified an increase of published articles. We will use this peak as an important point and discuss the time-period before and after.

Before 2014, we have a total of 4 articles researching OGD barriers. In 2012, we identified what seems to be the start of a crystallisation that had its full expression in 2014. E.g. the literature review [42] focused on the user OGD barriers. The authors pull together research from different research domains. Suggesting that OGD challenges encountered covers a variety of domains and around 2012 these started to crystallise into concepts such as open data barriers, challenges, myths, risks, and impediments giving way for a collective discussion.
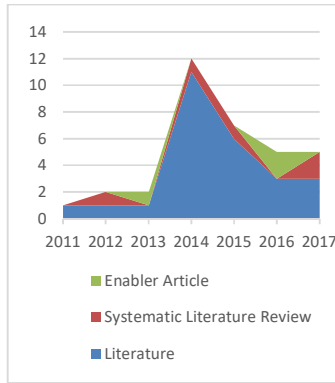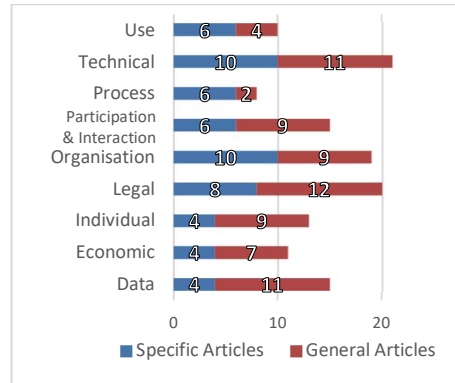

Fig. 1. Historical Development.

Fig. 2. Barrier Types found in the literature

After 2014, we identified 5 articles researching OGD barriers per year, a decrease compared to the 12 articles in 2014, conflicting with what is known in 2017: open data barriers are evident [e.g. 5, 18], something that is not reflected in the research attention. When analysing the articles, we identified a pattern. Myths, for example, [13, 20, 28] are not reused or connected to other OGD barrier research. Research has not revisited the focus on implementation and use from a multi-level perspective [26]. A process perspective [e.g. 3, 39] seems to be useful and often comes with a focus on the publisher, and therefore, lacks the barriers existing on the "outside" between publishing and feedback. Such as barriers for hackathon apps to reach the market [14]. We see a focus to understand barriers and their linked parts, leading to rather limited attempts to solve complex issues from multiple directions; from the micro to the macro perspective.

When looking at probable causes for the decline above, it can be that research is in an incubation period or have expanded. If an incubation period is present, we identify two OGD research barriers; (1) we are still searching for the abstract approach to handle barriers, e.g. like outlined in [4] and [18]. (2) We have not crystallised our knowledge about barriers, meaning that we are still working towards comprehension; one attempt has been made by [5]. Both fit with the pattern above. On the other hand, if research has expanded; it can have expanded into new domains. E.g. from government into innovation [e.g. 14]. With this, the label used to address the OGD barriers might have changed to fit the new domain. Fitting with what we identified in 2012. There is also a

possibility that the research has moved over to solutions. For both, we see two accompanying OGD research barriers; (3) keeping the knowledge body together between domains and (4) researching solutions that will not lead to barriers in other parts of the OGD process. To conclude, we identify a complexity in research and practice covering the micro technical and social (formats, metadata, skills, knowledge, etc.) and the macro technical and social (internet, ecosystem, society, etc.) making open data barriers a complex and challenging research task, as it covers multiple domains where barriers exist outside a particular actor.

3.2   Barrier Types

In Figure 2, we present barrier types identified in the 34 articles. We have coloured the two groups mentioned in section 2. When studying the different barrier types, it was common to find technical, organisational, and legal barriers (identified in 21, 19, and respectively 20 articles), something that was not surprising as open data originates in the publisher organisation's social and technical systems [3, 9, 39]. On the other hand, data barriers appeared in 15 articles, but were sometimes fused with technical barriers such as format and quality; they tended to be linked. Making data barriers implicate technical barriers. For open data, the value is generated from reuse [17] and we, therefore, expected use to be equally studied as technical, organisational, and legal barriers, it was identified in 10 articles. However, use was sometimes fused with technical barriers through access and findability. Use was also often fused with participation and seemed to be assumed to go together, as they are often integrated into the OGD processes [e.g. 3]. Making use through implicitly equally studied to technical, organisational, and legal barriers. Equally important, but less studied was the OGD process (found in 8 articles), where there were a gap between accessing data and feedback, where usage or application barriers were absent. We see that there is a lack of a wider scope when studying OGD barriers. In the OGD process for the publish-and-use relationship, we found a tendency for barriers to affect both providers and users while not stated as such. Use, participation, and process barriers need study beyond the direct interaction between providers and users. Largely, it remains unknown how organisations [e.g. 9] connect with publisher-user relation [e.g. 40] and affects the user [e.g. 6] and how this, in turn, affect participation and the organisation as a publisher. Turning our attention towards the more individual-oriented barriers, such as skills, perceived risks, knowledge, and awareness, they have been studied in 13 articles, e.g. in the role of the OGD user. These barriers differ from the above as they focus on how a person needs to adapt to open data, rather than changes in techniques and organisations. As individuals populate organisations, it can be challenging to separate the two, but we think that, based on the patterns above, it is relevant to conclude that we should aim to focus more on individual-oriented barriers in organisational settings. The final barrier type is economic, which were identified in 11 articles; we found discussions about costs, but no clear evidence that benefits would outweigh them. There was also a lack of connection between costs, organisations, and solutions (both implementation and maintenance) to barriers. As pointed out by [17] data generation and maintenance cost

money. Articles tend to focus less on the cost associated with solving barriers and challenges. However, it is not the case that costs are ignored, e.g. [5] brings up cost issues. Two major issues seem to be calculating implementation costs for overcoming barriers and maintenance costs versus benefits, e.g. how do we calculate the value of increased transparency, democratic accountability, and stimulation of innovation [20]. We see a need for more research about economic barriers in comparison to other barriers.

3.3 Research Focuses

PUBLISHING

25	IT & History 2011	38	Factors 2015
4	Senior Managers 2014	37	Perceived 2016
9	Internal 2014	29	E-Service 2017
34	Stimulate Use 2015	18	Release 2017

LEGAL & PRIVACY

| 30 | Traffic Data 2014 | 31 | Publish & Use 2014 |

USAGE

42	Impediments 2012	33	Business 2015
14	Innovation 2014	15	Assessment 2015
2	Journalism 2014	6	User Groups 2016

INTERACTION

| 21 | Engagement 2013 | 22 | Participation 2017 | 16 | Platform 2017 |

SYSTEM

27	Barriers & Risks 2013	1	Supply & Use 2014
26	Implementation & Use 2014	3	Publish & Consumption 2015
39	Socio-Technical 2014	5	Publish & Use 2017

DECISION

23	Benefits & Risks 2014
40	Dark Side 2014
41	Pro & Con 2015

ASSUMPTION

20	Five Myths 2012
13	Public Interest 2015
28	Rationalized 2016

REVIEW

| 17 | Research 2016 |
| 35 | Nordic Context 2016 |

SOCIETY

| 19 | Role In 2014 |

Fig. 3. Research focuses found in the literature (Based on abstracts).

In Figure 3, we present the identified research focuses sorted into categories. Each study is a box with citation number to the left and to the right; we have at the top focus and publishing year below. Underlined numbers refer to systematic literature reviews. We categorised the studies, which resulted in the following categorical descriptions: Assumption studies explore barriers steaming from assumptions. Decision studies approach barriers about considerations; such as the decision to publish or not. Interaction studies study barriers in the relation between actors. Legal and privacy studies how law and norms interfere with publishing. Publishing studies focus on barriers related to the publishing process. Review studies are exploring the literature. Society studies investigate open data in society. A system perspective on barriers is also present. Usage research study barriers that impede open data user reuse.

In Figure 3, we see an ongoing focus on publishing, usage, and systems over the years. The strongest focus was on publishing, while both usage and a system approach came second. Together with frequency, technical, organisational, and legal barriers are mentioned in favour overuse and participation (see Figure 2), we have identified that

challenges seem to be inside public organisations (or we need more research on use and usage). Looking at the figure, we see that interaction was studied under the years 2013 and 2017, with focuses on engagement, participation, and platform. In the earlier section, use barriers were often fused with participation and interaction barriers, where use seems to link to publishing through technical barriers. We, therefore, expected to see a balance amongst the barriers between users and publishers, with some focus on equalisation of expectations and division of labour [28], but instead found a focus on how publishers "imped" interaction where users were less problematized. For studies focusing on usage, they focused either on certain user groups and their special usage [e.g. 2, 16] or specifically on the use of provided data from a user perspective [e.g. 15, 42]. Here we found a gap in how different use barriers connect to different usages and users (see Davies [11] for examples of usages). The relative low societal focus can reflect the general maturity of the OGD field, being rather early in its lifecycle. We have, in research and practice, not yet solved issues surrounding publishing, using, and interaction. In the section above, legal barriers were mentioned second to most, but here we see that it has been studied the least. Legal and privacy barriers seem to accompany other focuses. It can be that they are solved by practical work, rather than theoretical. There is a lack of research on assumptions and publish decision barriers, unpacking this could lead to further insights about the publisher.

4   Systematization of Open Government Data Barriers

In the following section, we expand the scope surrounding the OGD process and systematise the OGD barriers. The OGD process consists of phases that involve data generation, data publication, data discovery, data usage, and user feedback [3, 39], which is the final process after data publishing. Too, understand the barriers behind publication we take a step back. Open data originates from an organisation where it starts as closed, which means that the data has an organisational context [9], before and after publication. In between data being close and open, there needs to be a decision for data release [23, 40, 41]. Meaning that we have five distinct processes containing barriers; when (1) identifying data's suitability for publishing, (2) deciding to release, (3) publishing the data, (4) someone uses the data, and (5) evaluating the impact and collect feedback. The processes are further complicated when it is known that activities in earlier phases can cause barrier further down [39] and connect with each other [18], which we argue also applies to our five distinct processes.

The relationship between the distinct processes is complex. Placing the processes beside the ecosystem [12], we find that they are connecting through ecosystem elements. We view elements as a complex network of participants with their challenges (some not related to OGD). The identification of suitable data starts when there is interest in some data. The suitable data is input into the other processes. Inside a decision network, we find decisions to release processes. Participants are the organisation's internal management and enthusiasts, but also outsiders, such as activists, researchers, and politicians [e.g. 28]. If the decision is release, the publishing process triggers inside

an organisation. The publisher (people behind data publishing) will then work to publish the data into the data market (people behind data use and products [e.g. 24]). Where users can transform the data through use into societal impact (people benefiting, affected by, or using open data outside or at the edge of the data market) [24]. Impacts could be of political and social, economic, and operational and technical nature [20]. The decision network then evaluates the impacts, which can concern the publisher, users, and citizens. Evaluation can happen through feedback, discussion, [39] and some evaluation is internal or external to the OGD process [12]. The processes form a cycle starting and ending with the decision network. For an example of the cycle; low-quality data negatively affects usage that results in an impact on the perceived usefulness of open data [17], resulting in less reason to publish. Low usefulness may lead to a revaluation of the open data's suitability. In the following subsections, we will present barriers identified for each of the five distinct processes. Bold words are barrier **categories**, while cursive are *coded*; behind both, there are citations to show quantity and origin.

4.1 Identifying Data Suitability

In the process of identifying data's suitability as open data, barriers will be encountered in the relationship between the data and organisation. Here the first thing to study is the relationship towards the *Core Task* [5, 9, 19, 25, 39, 42]. We need to explore what role open data has in the organisation and under what time-period. As been noted by Jaakkola, Mäkinen and Eteläaho; *"Some general open data related problems and challenges are…[a] guarantee of maintenance and updatedness of data is missing"* [19, p. 33]. One question is if the organisation have an interest to continually and actively collect the data. *Data* [3, 15, 20, 42] is produced by the organisation and will, therefore, have a meaning dependent on that context, with accompanying assumptions. We will have to study the data's properties (e.g. static or dynamic), context (e.g. language), and quality (e.g. completeness, fragmentation, accuracy). The production of the data happens in the *Collection Process* [3, 9, 20, 27, 42] of the organisation, which can be integrated into other activities. In it, data is collected, created, or generated. Tools, filters, data processing and methods are used, which can create barriers for later publishing and reuse. E.g. filtering before storage can result in *"No access to the original data (only processed data)"* [20, p. 262]. The *Storage System* [3, 5, 9, 18-20, 22, 25, 26, 38, 39, 42] stores data for the long term. The form and location of the storage can vary; with everything from paper-based to digital and in-house to the cloud. At the same time, the owner of the data and the storage can differ. Barriers can be an inability to convert data, improve storage system, or the share amount. Something that brings us over to *Path Dependency* [20, 25, 28, 42] as a source of barriers. E.g. municipalities seldom considered document formats when buying software, instead focusing more on specific office applications [25]. Meaning that they might have bought a system based on closed formats that are hard to open. In relation to collection, storage, and core task we have the *Internal Usage* [9, 19] that will form the data with or for a purpose. A forming that might make the data unsuitable for external reuse. The data's **Suitability** [9, 26, 38, 39]

as open data, comes from itself and relationship with its organisation. We can, therefore, talk about suitability barriers that complicate publication and reuse, where solutions can problematize the organisation's day-to-day activities.

4.2 Decisions to Release

Once suitable data have been identified, a release decision can be made. The organisation might have several datasets and no idea which to publish, which relate to **Market** barriers. Here *Supply and Demand* [26, 27, 40] can identify sought-after data; if demand is neither expressed or heard then it forms a barrier as an organisation cannot prioritise publishing. Another side is *Competition* [3, 31]. Outside the organisation there might be actors competing with or around the data. Releasing the data can dislodge their business models. On the other hand, if we turn our attention towards the publisher we have **Organizational** barriers. If publishing the data do not align with the organisation's *Objectives* [20, 34, 41], this may form a barrier as other tasks are prioritised. Prioritization can also be affected by the availability of resources. Which makes *Cost and Income* [3-5, 9, 16, 20, 26-29] interesting, as it is known that open data can lower costs, but also income [4, 5]. Meaning that open data has a risk to threaten resources used in core tasks. This also moves us to the *Division of Labour and Alternatives* [28, 29]. For Division of Labour it is a question about expectations on publisher and user; e.g., if there is a need, should publisher or user implement visualisation tools. Placing this alongside objectives and resources, it also becomes a question about alternatives. An alternative to open data within a government organisation, can, e.g. be to develop and launch e-services [e.g. 29] that are prepacked and controlled in another way that is publishing OGD for any user and any purpose. Another potential for declining to publish are **Consequence** barriers. Here *Regulation and Legislation* [3-5, 19, 20, 22, 23, 27-29, 31, 37, 40, 42] can block or problematize open data publishing. As an example, [4] identified *"Irish law can be a considerable barrier to making data available."* (p. 145). In relation to the laws we also have *External Safety* [4, 23, 29], where [23] tells us *"Detailed data about infrastructure (power plants, dams, transmitters etc.) might be misused to cause damage to the infrastructure."* (p. 35). If something does go wrong (abuse, misinformation, fraud, accidents, etc.) there will be the question of *Liability* [3, 4, 20, 23, 29, 31, 39, 40]; if this is unclear, both publisher and users might be at risk of unexpected and unwanted responsibilities. Laws and norms also make *Privacy* [1, 3, 5, 16, 18-20, 23, 26-31, 39-42] intrusions something that must be avoided (a barrier), where the solution might be to clean the data, which can also make it lose value. Another solution is to use *Embargo Periods* [39-41]. The final source of barriers is a lack of ***Skills for Release*** [26]; the organisation might lack the skills, competencies, and tools to support release decisions. If pushed, they might publish data that is a threat to external safety, useless, or reveals private information about citizens.

4.3   Publishing the Data

Publishing data is no easy task as there will be internal barriers, but also a need to consider usage. One barrier that can cause future barriers is a lack of ***Skills for Publishing*** [1, 3-5, 15, 18, 20, 22, 27, 34, 38, 39, 42], as it can e.g. result in inadequate infrastructure performance, lacking metadata descriptions, or inaccessible data. Another hurdle is **Service** barriers. *Ownership* [5, 9, 23, 27, 31, 34, 39-41], as Susha et al. [34] tells us *"For Groningen municipality, the biggest problem was to convince the data owners to publish data which they considered were not of sufficient quality."* (p. 193); can be a barrier when the organisation are unsure of service quality. Unclear data ownership can also make it hard to know what can be published. Sufficient *Quality* [1, 5, 15, 18-20, 22, 23, 26, 39-42] must also be sought and guaranteed; else it can lead to future use barriers. The data might be the wrong format, unreliable, inaccurate, and incomplete. Quality connects technical infrastructure, data's suitability, and the users' needs. For the technical infrastructure, we are interested in external and internal barrier sources. For external barriers, we must consider **Opening** barriers. Here *Access* [3, 5, 15, 20, 27, 34, 39, 41, 42], *Metadata* [3, 5, 18, 20, 26, 27, 34, 38, 39, 41, 42], and *Format* [1, 3-5, 15, 18, 20, 22, 27, 34, 38, 39, 42] can cause problems for prospecting users. Paywalls, registrations, and API only downloadable data can form access barriers. While metadata barriers can be none explained context, language, or purpose for the data, which can make it hard to identify opportunities and usages. On the other hand, formats that are badly documented or proprietary can hinder data integration and long-term sustainable usage. For the publisher selecting a format when there is no standard can be a barrier. The original storage system format can be a problem too, as exampled by Kassen [22] *"…some data are really challenging to publish in a computer-readable format due to the fact that they only exists as a paper document…"* (p. 317). In addition, converting the storage systems format to one that is open, machine-readable format can also be a barrier. For the internal technical infrastructure, we have **Administration** barriers, which likely will be experienced once data is published. The first is *Performance* [3, 4, 16, 23, 28, 29] where infrastructure might not be able to handle the external requests, which can affect the organisations day-to-day activities, but also impact users' perception of data quality (e.g. overloads that leads to system crashes). The second is *Maintenance* [5, 16, 18-20, 27, 29, 39, 42], where a possible increase in requests might require upgrading the technical infrastructure but can also lead to increased "wear and tear" (an increase in hardware replacement). As publishing is about opening closed data to third-party reuse the organisation is also opening for the public, therefore, the final is *Security* [5, 20, 37]. Where lacking security can lead to a threat against the core task, but also the external safety, as devious users gain access to private information or have the ability to damage the organisation's systems.

4.4   Using the Data

While there are user barriers outside the control of the publisher, some can be mitigated. First users must be able to find the data. **Findability** [3, 5, 16, 18-20, 23, 39, 41, 42] refers to the ability of the user to discover and identify the data, which to a degree can

be solved with an open data portal. Findability related barriers could come from different sources, such as a market flooded with similar datasets or diffuse names. Zuiderwijk and Janssen [39] tells us *"Making open data findable poses the challenge for different government organizations to collaborate."* (p. 122). Meaning that findability barriers is in the relation between publisher and user, between publishers, and both actors relation to the open data portal. Once users have found the data, they want access and read the *License* [3, 5, 14, 15, 19, 27, 31, 39, 41]. The license can cause barriers when not explained, incomprehensible, or complicated. Another part is if the license restricts usage to much or is incompatible with other licenses. There is also a question about user *Responsibility* [5, 14, 31, 40, 42] that connects back to liability. An example is *"...the UK Open Government Licence for instance requires that the re-user does not mislead others or misrepresent the Information or its source [45]..."* [31, p. 7]. The authors argue that such clauses should not be needed, but also admit that it seemed to make public agencies more comfortable with publishing. We see that an unclear license can be a barrier for users, while unclear user responsibilities can be a barrier for the publishers. At this stage, the user might want to read *Documentation* [5, 15, 39] that explains how the data can be used, accessed through an API, or see processing examples. Documentation connects back to metadata, but rather than explaining the what, it supports the user in the endeavor to use. In common for all above we have *Language* [5, 27, 39, 42] as a barrier. If the users cannot read or understand websites, licenses, documentation, or the data it will likely complicate usage. We also have the *Paywall* [5, 15, 20, 39, 42]; the user must pay for data. A paywall can stop the users, especially if there is no way to express feedback of or experiment with the data. This means that a paywall is a barrier that can worsen if there is also bad metadata or documentation. Users will also experience barriers if either license, format, IT-systems, or data content lacks *Interoperability* [19, 23, 26, 34, 42] with other data. The ability to combine data is important for open data to reach value. Interoperability can also come with collaboration barriers for publishers as standards are implemented. Another barrier is *Timely* [39, 40, 41]; the data needs to reach the user before it spoils. The barrier here is long embargo period or slow access. While mentioned barriers focuses on the experiences of the users, we can also look at user **Diversity** barriers. *Different Users* [2, 6, 14, 18, 33, 39, 42] comes with diverse needs and goals. E.g. visualise data or integrate data into an app. If not meet, users might be unable to use the data. On the other hand, publisher can find it hard to identify and meet users' needs. *Tools / Software* [3, 5, 15, 20, 34, 39, 42] can help users, but then we stand in for the question of who should implement them and how far. We also have the other side where data might need special tools to use, which can be a barrier if they are unusable or cost. Users also have their own barriers, for example *"... current challenge facing data journalism in Sweden is the lack of time to collect, analyze and present data. This may be a result of streamlined organizations as well as not having enough competence and skills to efficiently work with data."* [2, p. 404]. If data require special *Skill for Usage* [5, 6, 13, 14, 16, 18, 20, 27, 34, 39, 42], users without them are at a disadvantage and how to supply these will be a barrier for both publishers and users. This leads us to **Usability** [3, 5, 16, 18-20, 29, 34, 39] as a barrier. Data loses usability when users face some combination of the barriers above or cannot use the data, in addition high amounts of unusable data in the data market can

lower the findability of high quality data. This means that a combination of usability, awareness, and findability can result in *Low Usage* [14, 15, 34, 39].

4.5 Evaluation

Once data is published and, hopefully, in use, it is relevant to evaluate the impact and collect feedback to improve. Here we can study **Community** barriers, such as lacking *Channels* [5, 20, 22, 34, 42] used to collect and communicate and low open data *Awareness* [3, 18, 22, 31, 34]. If end-users are not aware of their open data usage, they cannot tell how it impacts. In the inverse, if users have feedback, but no channel, the process will be hard to improve (and we lose impact insight too). Users that are not aware of OGD and its possibilities will not show *Interest and Need* [1, 22, 26, 27]. Even if they have awareness and channels, they can express interest, to express need they require access to the organisation's metadata inventory. Here we have connected interest and needed to supply and demand, as without expressed interest and need it will be hard to evaluate demand for some closed data. As we argued above, different users have diverse needs, and it, therefore, becomes important to think about *Participation* [3, 13, 16, 34] in evaluation and feedback. Not involving a diverse group of users can result in a specialised solution that creates complications for other users. There is also another barrier related to participation, as Hellberg and Hedström [13] say *"We believe that not everyone is interested in using public data, even if they have the necessary resources and competences."* (p. 47). We see a possibility that low awareness and uniform participation can lead to specialised solutions, where not all with capability are interested in participation, which lowers usability by exclusion. Let us say we have participants that want to give us feedback; then we need to think about **Coordination** barriers. Here *Quality in Evaluation* [15, 20, 22] can be a challenge, as feedback can be in multitude and varied or users are without supporting documentation and tools; the user might also not know what to look for. Once we have participants with feedback, it becomes a question of *Feedback and Error Handling* [34], which connects back to administration. If errors and feedback are not handled problems will continue constraining the open data impact. Moreover, if changes are not communicated to the users, it might cause complications in their use of open data, where consequences can be hard to anticipate. Based on the identified literature, we see that evaluation and feedback need research, especially in relation to barriers and on both active and passive evaluation.

5 Conclusions and Future Work

In this paper, we have identified and systematised OGD barriers into identifying data suitability, release decisions, data publishing, data usage, and evaluation based on a systematic literature review. We have illustrated the attention given to different barriers in OGD research from a longitudinal perspective and illustrated different types of barriers and focuses in the research field, including relating 46 OGD barriers to different parts and phases of the OGD process. With the illustrations in section 3, have we positioned and contextualised our research questions, and with the systematisation (Section

4), we have integrated the research questions to contribute to a result that is valuable for researchers in the area, and in the long run for the practice. In the analysis, we moved away from a perspective on barriers as something to be "easily" reduced or managed (e.g. regarding deciding to release OGD), towards viewing barriers as a network in a cycle and part of a strategic decision in a broader context. This is a way to expand the scope of OGD, where we move away from the reported myths surrounding open data.

We have also identified a knowledge gap between use and evaluation for barriers. The government can not solve barriers here, only mitigate by, e.g. ease usage. Our study opens for research from both the individual perspective and systematic perspectives on OGD barriers and their relations. One important and unexplored dimension of this; is the responsibilities of and division of labour between the publisher and user. We have also identified a need for further insight into open data barriers and myths, especially on the economic aspects, assumptions, and the gap between use and evaluation.

This paper indicates two research implications. (1) The reviews [e.g. 3, 17, 42] are essential to identify and understand OGD barriers. To hold the knowledge body together, we need recurrent systematic literature reviews both on a general and specific level (e.g. usage or publication), where authors offer insights, research avenues, and address research barriers. One such avenue for a future systematic literature review is to investigate the focuses in previous OGD barrier studies. Hopefully, this will open for successful research and practice. (2) Current research does not support informed decisions towards a certain kind of usage; current research is rather general or even abstract. We need a thorough understanding of what barriers a publisher will encounter, but we lack the how and why. There is time to move to answer those questions and combine different conclusions to support and guide practice.

We have identified some limitations in our study. We did not use "open government data" as a keyword of the search. Instead, we used the broader "open data". The broader term usage results in that not all the literature were studies of governmental nature but may be relevant for that sub-field. We also used singulars instead of plural, because of this we did not identify [8]. We looked at the top 600 hits from Google Scholar, filtered with the year and different keywords. To avoid being too limited, we used snowballing, but there might still be important literature out there. Another limitation is that the systematisation is not yet evaluated in practice. We also acknowledge that barrier severity varies [e.g. 5]. As the systematisation is based on a broad pool of barrier research, we claim that the coverage, and to a degree the usability, of the systematisation is high, but taking the limitations into account more work can be done to evaluate and validate it.

References

1. Albano, C.S., Reinhard, N.: Open government data: facilitating and motivating factors for coping with potential barriers in the brazilian context. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 8653(June), 181–193 (2014).
2. Appelgren, E., Nygren, G.: Data journalism in sweden: Introducing new methods and genres of journalism into "old" organizations. Digital Journalism 2(3), 394–405 (2014).
3. Attard, J., Orlandi, F., Scerri, S., Auer, S.: A systematic review of open government data initiatives. Government Information Quarterly 32(4), 399–418 (2015).

4. Barry, E., Bannister, F.: Barriers to open data release: A view from the top. Information Polity 19(1-2), 129–152 (2014).

5. Beno, M., Figl, K., Umbrich, J., Polleres, A.: Open data hopes and fears: Determining the barriers of open data. 2017 Conference for E-Democracy and Open Government (CeDEM) pp. 69–81 (2017).

6. Brugger, J., Fraefel, M., Riedl, R., Fehr, H., Schöneck, D., Weissbrod, C.S.: Current barriers to open government data use and visualization by political intermediaries. Proceedings of the 6th International Conference for E-Democracy and Open Government, CeDEM 2016 pp. 219–229 (2016).

7. Charalabidis, Y., Alexopoulos, C., Loukis, E.: A taxonomy of open government data research areas and topics. Journal of Organizational Computing and Electronic Commerce 26(1- 2), 41–63 (2016).

8. Conradie, P., Choenni, S.: Exploring process barriers to release public sector information in local government, p. 5 (2012).

9. Conradie, P., Choenni, S.: On the barriers for local government releasing open data. Government Information Quarterly 31(SUPPL.1), S10–S17 (2014).

10. Cronin, P., Ryan, F., Coughlan, M.: Undertaking a literature review: a step-by-step approach. British Journal of Nursing 17(1), 38–43(2008).

11. Davies, T.: Open data , democracy and public sector. Interface pp. 1–47 (2010)

12. Dawes, S.S., Vidiasova, L., Parkhimovich, O.: Planning and de- signing open government data programs: An ecosystem approach. Government Information Quarterly 33(1), 15–27 (2016).

13. Hellberg, A.S., Hedström, K.: The story of the sixth myth of open data and open government. Transforming Government: People, Process and Policy 9(1), 35–51 (2015).

14. Hjalmarsson, A., Johannesson, P., Juell-Skielse, G., Rudmark, D.: Beyond innovation contests: A framework of barriers to open innovation of digital services. Twenty Second European Conference on Information Systems (2006), 1–15 (2014)

15. Hjalmarsson, A., Johansson, N., Rudmark, D.: Mind the gap: Exploring stakeholders' value with open data assessment, vol. 2015-March, pp. 1314– 1323 (2015).

16. Hogan, M., Ojo, A., Harney, O., Ruijer, E., Meijer, A., Andriessen, J., Pardijs, M., Boscolo, P., Palmisano, E., Satta, M., et al.: Governance, Transparency and the Collaborative Design of Open Data Collaboration Plat- forms: understanding barriers, options, and needs, pp. 299–332. Springer (2017)

17. Hossain, M.A., Dwivedi, Y.K., Rana, N.P.: State-of-the-art in open data research: Insights from existing literature and a research agenda. Journal of Organizational Computing and Electronic Commerce 26(1-2), 14–40 (2016).

18. Huang, R., Lai, T., Zhou, L.: Proposing a framework of barriers to opening government data in china. Library Hi Tech 35(3), 421–438 (2017).

19. Jaakkola, H., Mäkinen, T., Eteläaho, A.: Open data: Opportunities and challenges. Proceedings of the 15th International Conference on Computer Systems and Technologies - CompSysTech '14 pp. 25–39 (2014).

20. Janssen, M., Charalabidis, Y., Zuiderwijk, A.: Benefits, adoption barriers and myths of open data and open government. Information Systems Management 29(4), 258– 268 (2012).

21. Kassen, M.: A promising phenomenon of open da-ta: A case study of the chicago open data project. Government Information Quarterly 30(4), 508– 513 (2013).

22. Kassen, M.: Open data in kazakhstan: incentives, implementation and challenges. Information Technology & People 30(2), 301–323 (2017).

23. Kucera, J., Chlapek, D.: Benefits and risks of open government data. Journal of Systems Integration pp. 30–41 (2014).

24. Lindman, J., Kinnari, T., Rossi, M.: Business roles in the emerging open-data ecosystem. IEEE Software 33(5), 54–59(2016).
25. Lundell, B.: e-governance in public sector ict procurement: what is shaping practice in sweden? Eu-ropean Journal of ePractice pp. 66–78 (2011)
26. Martin, C.: Barriers to the open government data agenda: Taking a multi-level perspective. Policy & Internet 6(3), 217–240 (2014).
27. Martin, S., Foulonneau, M., Turki, S., Ihadjadene, M.: Open Data: Barriers, risks and op-portunities, pp. 301–309 (2013).
28. Melin, U.: Challenges and benefits in an open data initiative - a local government case study of myths and realities. 15th IFIP Electronic Government and the 8th Electronic Participation Conference (EGOV ePart 2016) 23, 111–122 (2016).
29. Melin, U., Lindgren, I.: Open government data in an e-service context - managerial and conceptual challenges. Americas Conference on Information Systems 23, 1–10 (2017)
30. Rohunen, A., Markkula, J., Heikkila, M., Heikkila, J.: Open traffic data for future service innovation - addressing the privacy challenges of driving data. Journal of Theoretical and Applied Electronic Commerce Research 9(3), 71–89 (2014).
31. de Rosnay, M.D., Janssen, K.: Legal and institution-al challenges for opening data across public sectors: Towards common policy solutions. Journal of Theoretical and Applied Elec-tronic Commerce Research 9(3), 1–14 (2014).
32. Saldaña, J.: The coding manual for qualitative researchers. Sage (2015)
33. Susha, I., Grönlund, A., Janssen, M.: Driving factors of service in- novation using open government data: An exploratory study of entrepreneurs in two countries. Information Polity 20(1), 19–34 (2015).
34. Susha, I., Grönlund, k., Janssen, M.: Organizational measures to stimulate user engagement with open data. Transforming Government: People, Process and Policy 9(2), 181–206 (2015).
35. Susha, I., Johannesson, P., Juell-skielse, G.: Open data research in the nordic region: To-wards a scandinavian approach? International Federation for Information Processing 9820, 61–73 (2016).
36. Walters, W.H.: Comparative recall and precision of simple and expert searches in google scholar and eight other databases. portal: Libraries and the Academy 11(4), 971–1006 (2011).
37. Wirtz, B.W., Piehler, R., Thom-as, M.J., Daiser, P.: Resistance of public personnel to open government: A cognitive theory view of implementation barriers towards open government data. Public Management Re-view 18(9), 1335–1364 (2016).
38. Yang, T.M., Lo, J., Shiang, J.: To open or not to open? determinants of open government data. Journal of Information Science 41(5), 596–612 (2015).
39. Zuiderwijk, A., Janssen, M.: Barriers and development directions for the publication and usage of open data: A socio-technical view, pp. 115–135 (2014).
40. Zuiderwijk, A., Janssen, M.: The Negative Effects of Open Government Data - Investigating the Dark Side of Open Data, pp. 147–152 (2014).
41. Zuiderwijk, A., Janssen, M.: Towards decision support for disclosing data: Closed or open data? Information Polity 20(2-3), 103–117 (2015).
42. Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., Alibaks, R.S.: Socio- technical imped-iments of open data. Electronic Journal of e-Government 10(2), 156–172 (2012).
43. Andrade, C. 2011. How to write a good abstract for a scientific paper or conference presen-tation. Indian Journal of Psychiatry. 53, 2 (2011), 172.
44. Jacinto, T. et al. 2014. Doing science: Writing conference abstracts. Breathe. 10, 3 (2014), 265–269.