

Benefits, Adoption Barriers and Myths of Open Data and Open Government*

Marijn Janssen, Yannis Charalabidis & Anneke Zuiderwijk

This is an Author's Original Manuscript of an article submitted for consideration in the Information Systems Management [copyright Taylor & Francis]; Information Systems Management is available online at <http://www.tandfonline.com/doi/full/10.1080/10580530.2012.716740>.

M. Janssen, Y. Charalabidis & A. Zuiderwijk (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management (ISM)*, vol. 29, no.4, pp. 258-268.

Abstract

Although a significant number of public organizations have embraced the idea of open data, many are still reluctant to do this. One root cause is that the publicizing of data represents a shift from a closed to an open system of governance, which has a significant impact upon the relationships between public agencies and the users of open data. Yet no systematic research is available which compares the benefits of an open data with the barriers to its adoption. Based on interviews and a workshop, the benefits and adoption barriers for open data have been derived. The findings show that a gap exists between the promised benefits and barriers. They furthermore suggest that a conceptually simplistic view is often adopted with regard to open data, one which automatically correlates the publicizing of data with use and benefits. Five 'myths' are formulated promoting the use of open data and placing the expectations within a realistic perspective. Further, the recommendation is given to take a user's view and to actively govern the relationship between government and its users.

Key words: Systems theory, institutional theory, adoption, diffusion, open data, open government, governance, transformation

Introduction

The availability of open data has grown significantly, with pressure being placed on all kinds of public organizations to release their raw data. Some main motivations are that open access to publicly funded data provides greater returns from the public investment, can generate wealth through the downstream use of outputs, provides policy-makers with data needed to address complex problems (Arzberger et al., 2004) and can help to involve the citizenry in analyzing large quantities of data sets (Surowiecki, 2004). Open data is often indispensable for public policy development and service delivery, but can also be valuable for others, such as traffic information. In this research we define open data as non-privacy-restricted and non-confidential data which is produced with public money and is made available without any restrictions on its usage or distribution. Private, confidential and classified data is excluded, as this type of data is inappropriate to publicize. Data can be provided by public and private organizations, as the essence is that the data is funded by public money.

Public bodies are among the largest creators and collectors of data in many different domains (Janssen, 2011). These data domains range from traffic, weather, geographical, tourist information, statistics, business, public sector budgeting and performance levels, to all kinds of data about policies and inspection (food, safety, education quality, etc.). Oftentimes a limited number of anecdotal examples are used to provide examples of open data. These most often refer to data that is relatively safe to publicize by the government and not to data that, once released, could invoke a reaction from the public. On September 30, 2011 the Dutch National Ombudsman stated that “the Dutch government is all too often closed, instead of providing information to citizens” (Nationale_Ombudsman, 2011). The government is compared to an oyster that automatically closes up when approached. Managers and other public servants often have the tendency to avoid opening their data, as this would provide the public with new insights which might in turn result in critical questions. This is confirmed by

institutional theory, which predicts that the opening of data will reinforce existing structures instead of changing them and allowing them to fully take advantage of new developments (see for example Fountain, 2001; Kraemer & King, 2006; West, 2004).

Open data mends the traditional separation between public organizations and users. The opening of data leads to two important assumptions about government. First, it assumes the readiness of public agencies for an opening process which considers influences, discourses and exchanges as constructive and welcomes opposing views and inputs. Second, it assumes that government is to give up control, at least to some extent demanding considerable transformations of the public sector. Instead of reinforcing current processes, open data should result in *open government* in which the government acts as an open system and interacts with its environment. Not only should data be published, but actively should be sought for feedback to improve the government. The publicizing of data could have far-reaching effects on the public sector. Mechanisms for monitoring and responding to the questions asked by the public are therefore necessary and the government should be viewed as an open system interacting with its environment.

Despite the significance of open data little systematic and structured research has been conducted in this area. Most research in this area consists of conceptual papers (Bertot, Jaeger, & Grimes, 2010; McDermott, 2010), descriptions of the empirical uses of open data (Hausenblas, 2009; Napoli & Karaganis, 2010) or the design of technology and systems for harnessing the power of open data (Charalabidis, Ntanos, & Lampathaki, 2001; Kalampokis, Tambouris, & Tarabanis, 2011). None of the current research focuses on analyzing the benefits and barriers that go beyond individual projects, applications or conceptual ideas and global sketches. In this paper we analyze the benefits of and barriers to open data systems by synthesizing people's experiences with open data obtained from interviews and a group session. We start this paper by reviewing the predictions of system theory and institutional

theory on the move to opening data. Thereafter the benefits and barriers are presented. We conclude this paper by giving an overview of the myths of open data derived from the interviews. These myths are statements used to gain legitimacy for using or refusing open data but which do not have a determinable basis of fact or evidence.

Research approach

This research has an explorative nature, since open data is a recent phenomenon and is thus in an early stage of development. There is limited knowledge about this field and many research findings are fragmented. We first analyzed the literature to identify benefits of and barriers to using open data. The literature review was used as a background for organizing a group session. This session, with 9 participants, was held in June 2011 and was used to identify possible additional benefits and barriers. A group session is useful for dealing with complex, unstructured problems in which the actors have incompatible interests, diverging areas of knowledge and multiple backgrounds and is more productive than single interviews (Herik & Vreede, 2000). In a group session participants can react to each other and in this way generate more ideas than on their own. The findings were discussed and further refined based on interviews with 14 key persons. Persons representing different organizations, including Ministry of Interior, Ministry of Economic affairs, ICTU, WODC, Hack de Overheid and Municipality Den Haag. Servants at both the managerial and administrative levels were interviewed as well as users of open data to ensure a variety of responses. These interviews covered a retrospective analysis to understand the contexts of and rationales for using or not using open data. After finishing the group session, interviews were conducted. The interviews were aimed at better understanding the nature of the benefits and barriers and were used to determine the myths of open data. The myths were derived by confronting the interviewees with the benefits and barriers. The primary focus was on exposing interviewees to different

argumentations. In this context myths are considered necessary for ensuring progress but have no scientific basis.

System and institutional theories

By opening data a move from a traditionally closed to open systems is made. System theory states that these will impact the governance and feedback loops in which the government can learn from the public are needed. Institutional theory is used to predict that the opening of data will reinforce existing structures instead of changing them and transformation is needed to take advantage of open data.

Moving from closed to open systems

By publicizing data a new situation is created in which the public can use and create information through collaborative networking (Chun, Shulman, Sandoval, & Hovy, 2010). The public is outside the organizational boundaries and outside the control of the hierarchy. In fact the public becomes part of the data processing system and might process data, enrich data, combine it with other sources and might even collect their own data (for example through the use of their mobile phones). This resembles a change in the traditional boundaries between public organizations and the public in which virtually anybody in the world has access to the data. The traditional system boundaries are vanishing and the system is opened.

System theory draws attention to the important distinction between systems which are open to their environment and those which are closed (Jackson, 2003). Closed systems are much easier to manage, as they are not affected by external factors which are often unpredictable in nature. Central planning and control can be used, as there is less disruption from the environment. In contrast, the flow in open systems cannot be predefined but only guided. The opening of a system is often heralded for bringing in additional views, which has a positive

impact on its problem solving capacity (Surowiecki, 2004); and the opening of data for its use in ways that are not considered or anticipated in advance (Arzberger, et al., 2004).

The notion of *feedback* is important in open systems and refers to the situation in which activity within a system is the result of the influence of one element on another (Jackson, 2003; Wiener, 1948). The implication of the notion of feedback in systems theory is that in opening their data governments should not simply instigate one-way communication of their data but should expect or actively solicit feedback and be able to make sense of this feedback. The opening of systems provides the opportunity for creating feedback loops in which the government can learn from the public. By embedding hermeneutics the closed system is placed in the social context. The consequence is that the social context will also influence the (formerly) closed system. This implies that the relationship between a government and its environment is subject to change and that the government needs to accept that traditional planning and control instruments are no longer suitable. Opening a system typically requires a shift from mechanistic control to an evolutionary perspective which is dominated by self-organization. New governance mechanisms, capabilities and processes are necessary for dealing with these feedback loops. The nature of the response depends on the available organizational arrangements that make a response possible (Jackson, 2003).

Reinforcing and transforming institutional structures

Institutional theory analyzes the deeper and more resilient aspects of social structure by considering the processes by which structures become established (Scott, 1995). Institutional environments reward normative requirements of appropriateness and legitimacy and, in some cases, conformity to procedures, presentations, symbols and rhetoric (Scott, 1995). In institutional theory ICT is perceived, implemented and used in virtue of pre-existing institutional arrangements (sociological, cultural, legal and formal aspects) that grant stability. Stability is necessary for organizations to operate. Orlikowski (2000) argues that the

development of technology is heavily influenced by the actions (including decisions) of human agents, and that technology enacts structures. This suggests that institutions might both enable and constrain the adoption of open data. The outcomes stemming from the enactment of technology are difficult to predict because of multiple and unanticipated effects influenced by rational, social, and political logics (Orlikowski, 2000). Nevertheless, institutional theory suggests that the introduction of IT does not often change institutions but rather reinforces current work practices and organizational structures (see for example Fountain, 2001; Kraemer & King, 2006; West, 2004).

In opening data to the public, public managers (and politicians) find themselves in the midst of networks that might help them to reach advantages of open data at the expense of less control. Institutional theory argues that in open systems different steering instruments are required (Peters & Pierre, 1998). Outside the boundaries of government, command and control mechanisms cannot be used. Public managers find themselves confronted with having to deal with a variety of stakeholders (possibly unknown) that might help them to achieve the benefits of open data but might also be viewed as a threat if not properly handled. In open data the allocation of the roles of provider, processor, owner and maintainer complicates accountability issues. Which party is to blame when results of the processing of open data are incorrect? No one has an overview of what is done with the open data, and even having such an overview might violate the basic idea of open data. Whether the opening of data will unambiguously lead to a more transparent, interactive, open and hence accountable government is challenged from this perspective. Although the use of open data looks like collective accountability, it is likely that if something happens society will expect intervention from the government and will hold it responsible.

Benefits

A large number of benefits of open data were identified in the interviews. All interviewees recognized the potential benefits of open data. The basic assumption is that open data itself creates and generates more value than the selling of data sets. The benefits were clustered in 1) political and social, 2) economic, and 3) operational and technical benefits. Political and social benefits were viewed as the most important category. Political and social aspects were lumped together, as the respondents found them difficult to separate. The technical benefits were clustered together with the operational benefits. Table 1 shows the overview of the benefits. Some benefits overlap with and/or are related to each other. From the interviewees' responses it seemed that economic growth and contributing to public values (transparency and accountability) are the overarching arguments for stimulating open data.

Table 1: Overview of benefits of open data

Category	Benefits
Political and social	<ul style="list-style-type: none">More transparencyDemocratic accountabilityMore participation and self-empowerment of citizens (users)Creation of trust in governmentPublic engagementScrutinization of dataEqual access to dataNew governmental services for citizensImprovement of citizen servicesImprovement of citizen satisfactionImprovement of policy-making processesMore visibility for the data providerStimulation of knowledge developmentsCreation of new insights in the public sectorNew (innovative) social services
Economic	<ul style="list-style-type: none">Economic growth and stimulation of competitivenessStimulation of innovationContribution toward the improvement of processes, products and/or servicesDevelopment of new products and servicesUse of the wisdom of the crowds: tapping into the intelligence of the collective

	<ul style="list-style-type: none"> Creation of a new sector adding value to the economy Availability of information for investors and companies
Operational and technical	<ul style="list-style-type: none"> The ability to reuse data / not having to collect the same data again and counteracting unnecessary duplication and associated costs (also by other public institutions) Optimization of administrative processes Improvement of public policies Access to external problem solving capacity Fair decision-making by enabling comparison Easier access to data and discovery of data Creation of new data based on combining data External quality checks of data (validation) Sustainability of data (no data loss) The ability to merge, integrate and mesh public and private data

The opening of data is expected to create benefits like stimulating innovation and promoting economic growth and one interviewee even stated that the “effective use of public sector data is vital to the growth of our knowledge economy”. However, there is no way to predict and calculate the return of investment (ROI) for the issues. This is similar to many new strategies for which the ROI cannot be calculated in advance. The potential applications are hard to predict and possible ‘killer’ applications even harder. The main challenge is that open data has no value in itself; it only becomes valuable when used. One important use mentioned by the interviewees is that potential investors and companies can use open data to determine the attractiveness of potential investments.

One of the interviewees commented that “if data is used for policy-making a high level of confidence in the data and in the veracity of the interpretation is given. Therefore policy-makers should be prepared to share their data”. By opening data, users can validate and verify whether the conclusions drawn from the data are correct and justified, and they can analyze the previously collected data to sharpen the focus of policy-making. Sharing data openly and freely is often viewed as altruistic and advancing transparency and knowledge.

Most of the interviewees expect that open data can strengthen accountability, build trust and improve citizen satisfaction. The ready availability of information about what governments are doing and why is increasingly recognized as an important precondition to the meaningful exercise of democratic accountability and deliberation. One of the main benefits of opening a system is the ability to tap into the collective intelligence of the public. The key idea is that under the right circumstances, groups can generate better alternatives and make better decisions than even the smartest people can do on their own (Surowiecki, 2004).

The above list gives systematic and structured insight into the potential benefits of open data. Although sound evidence of the benefits of open data should be determined on a case by case basis, the list of potential benefits shows that the motivations for using open data are broad. Over time the benefits could change and different weight might be given to certain benefits.

Barriers

While open data can potentially provide numerous benefits, its adoption also entails a number of barriers. In the interviews, barriers were identified and categorized at the institutional level, the task complexity of handling the data, the use of open data and participation in the open data process, legislation, information quality and at the technical level, as shown in Table 2. When analyzing the barriers we found that barriers are related to either data providers (resulting in not wishing to publicize data) and data users (resulting in an inability to use the data in an easy manner). The institutional level concerns barriers from the data providers point of view, whereas task complexity and use and participation are from the user's perspective. The remaining categories (legislation, information quality and technology) can be relevant for both.

Table 2: Adoption barriers for open data

Categories	Barriers
Institutional	<ul style="list-style-type: none"> Emphasis of barriers and neglect of opportunities Unclear trade-off between public values (transparency vs. privacy values) Risk-averse culture (no entrepreneurship) No uniform policy for publicizing data Making public only non-value-adding data No resources with which to publicize data (especially small agencies) Revenue system is based on creating income from data Fostering local organizations' interests at the expense of citizen interests No process for dealing with user input Debatable quality of user input
Task complexity	<ul style="list-style-type: none"> Lack of ability to discover the appropriate data No access to the original data (only processed data) No explanation of the meaning of data No information about the quality of the open data (see category "Information Quality") Apps hiding the complexity, but also potential other use of open data Duplication of data, data available in various forms or before/after processing resulting in discussions about what the source is Difficulty in searching and browsing due to no index or other means to ensure easy search for finding the right data Even if data can be found, users might not be aware of its potential uses Data formats and data sets are too complex to handle and use easily No tooling support or help desk Focus is on making use of single data sets, whereas the real value might come from combining various data sets Contradicting outcomes based on the use of the same data Invalid conclusions
Use and participation	<ul style="list-style-type: none"> No incentives for the users Public organizations do not react to user input Frustration at there being too many data initiatives No time to delve into the details or no time at all Having to pay a fee for the data Registration required before being able to download the data Unexpected escalated costs No time to make use of the open data Lack of knowledge to make use of or to make sense of data Lack of the necessary capability to use the information No statistical knowledge or understanding of the potential and limitations of statistics Thread of lawsuits or other violations
Legislation	<ul style="list-style-type: none"> Privacy violation Security No license for using data Limited conditions for using data Dispute and litigations Prior written permission required to gain access to and reproduce data Reuse of contracts/agreements

Information Quality	Lack of information Lack of accuracy of the information Incomplete information, only part of the total picture shown or only a certain range Obsolete and non-valid data Unclear value: information may appear to be irrelevant or benign when viewed in isolation, but when linked and analyzed collectively it can result in new insights Too much information to process and not sure what to look at (Essential) Information is missing Similar data stored in different systems yields different results
Technical	Data must be in a well-defined format that is easily accessible: while the format of data is arbitrary, the format of data definitions needs to be rigorously defined Absence of standards No central portal or architecture No support for making data available Lack of meta standards No standard software for processing open data Fragmentation of software and applications Legacy systems that complicate the publicizing of data

Institutional barriers result in recalcitrance to change. One of the barriers stems from the risk-averse culture. In such a culture public accountability dominates over entrepreneurship (Bozeman, 1998). Organizations with more red tape, weak links with performance and high involvement with elected officials tend to have a risk-averse culture (ibid). Institutional theory warns for the risk of new initiatives being used for enforcing existing structures. Several interviewees suggested that this is exactly what is happening. One interviewee told us that “only data that is relatively safe is publicized and dropped in the large pile ... some of them even hope that the data will not be discovered in this pile ... there are no mechanisms for gaining any feedback about its use”. Often the existing structures are taken as a starting point and the user needs for finding, processing and using open data is neglected. It was even suggested that the reinforcement of the existing structure was strengthened due to the current budget cuts. Reinforcement also originate from the lack of systematic analysis about which type of data should be allowed to be publicized and what users expect from open data. This provides arguments for not publishing data.

Many elements were found to contribute to a higher complexity, which complicates use. Being able to use data and find patterns and trends in large amounts of data remains a significant challenge (Zurada & Karwowski, 2001). Several of these barriers express the need for having good structures and support for handling open data. Use and participation might be blocked because there might be no incentives or no added value for users to make use of open data. An underestimated subject seems to be the availability of all kinds of capabilities and knowledge levels of users for using complex and more sophisticated data. Much of the current effort is focused on how to easily use data embedded in software applications, whereas linking and combining data by users requires sophisticated knowledge. Statistical techniques are often used for the collection, analysis, interpretation and presentation of data. Yet statistical knowledge is scarce. One interviewee remarked, "Use is limited to the happy few, those who are educated and have time to explore new business opportunities". Opening data might further contribute to the digital divide, as the use of data might be limited to certain groups. Finally, the threat of lawsuits or other violations might hinder use.

The concept of information quality (IQ) has a long history, and elements determining information quality have been specified in previous research (Miller, 1996; Strong, 1997). As with most data sets, the quality of information is not automatically guaranteed and insight is needed in this before the information can be used for certain purposes. Data might be simply incorrect, but also essential information about the data sets might be missing, such as the time period in which the data was collected.

Finally, there are number of technical barriers, ranging from the unavailability of a supporting infrastructure to the lack of standards, fragmentation and legacy. These depict the need for ensuring a good infrastructure before the concept of open data will be widely accepted.

The barriers identified are often interrelated and do not stand alone. For example, the more complex the tasks that users wish to accomplish, the more barriers appear and the higher information quality should be and the more is demanded from the users. Complex tasks might be open to interpretation, which might result in a higher reluctance of public officials to publicize data. This interrelatedness makes it difficult to deal with the barriers in a straightforward manner.

Myths

In this section myths are formulated to reflect on the gap between the promises and barriers of open data. A myth is a traditional or legendary story without a determinable basis of fact or evidence. The essence of a myth is that its existence is fictional or unproven. From the conversations with the interviewees and the barriers described it can be concluded that oftentimes a conceptually simplistic look at open linked data is taken. Some interviewees stressed the potential benefits, but the 'how to' questions necessary for realizing the promised benefits were absent. This can be explained by the backgrounds of the interviewees. Those with higher level positions often focused on the benefits, whereas the persons in charge of realizing the benefits had a much more nuanced view. We argue that acknowledging the existence of this gap is important, and therefore we formulate a number of myths about open data.

Myths play an important role in policy-making, as they may inspire collective action but may also mystify and blur views on reality (Bekkers & Homburg, 2007). Bekkers and Homburg consider the concept of myth as a double-edged sword. On the one hand, myths are seductive tales containing promises and are used as a shared frame of reference that enables individuals, groups and organizations to act (ibid). On the other hand, myths are not necessarily true and not based on sound evidence (ibid). Open data is idealized by looking at the advantages that can be gained without looking at the drawbacks, one interviewee formulated the following:

“Our politicians are not aware of what is necessary. First they push to publicize data, next they complain about the transparency and ask us to react to the questions asked ... transparency seems to be desirable only for others rather than for their own activities”. This type of thinking can be viewed as myths which are used to gain legitimacy for the policy-making and the use of open data and to structure the activities without looking at the other side of the coin. Based on the interviewees we identify five myths which proved not to be supported by evidence but were often found at the heart of policy-making.

Myth 1: The publicizing of data will automatically yield benefits

Dominant in the debate about open data policies is the inescapable suggestion that publicizing data in and of itself enables or even causes the creation of competitive and other advantages. This myth avoids a focus on the barriers resulting in a lack of user actions. Nevertheless, in the long term not considering and dealing with barriers might be counterproductive. Suggesting that providing access to data is enough and not providing any means to process the data and to overcome the barriers makes the publishing of data useless. Too much emphasis is given to the data supplier and only limited attention for the user. Lowering the threshold for use should also be part of the policy.

Opening of data can be stimulated by using the ‘comply or explain’ principle, in which data is publicized unless there are arguments for not doing so. Using this principle, a culture will be created in which it is custom to publicize data. Many public organizations have jumped on the bandwagon of making data available without having a sound policy. This has even resulted in the publishing of data on central portals that were already publicly available, such as address data. The barriers listed above suggest that the creation of a huge pool of information might result in difficulty in finding the right information. The more data there is, the more difficult it is to analyze and draw meaningful conclusions (Zurada & Karwowski, 2011). Due to the

complexity, benefits like creating trust in government might not come true and bad experiences with trying to use open data might even yield the opposite.

Open data on its own has little intrinsic value; the value is created by its use. Supporting use should not be viewed as secondary to publicizing data. The publicizing of data needs to be accompanied by an infrastructure which is able to handle the data in an easy-to-use way to lower the user threshold. Hey and Trefethen (2005) argue that an e-infrastructure can enable faster, better, and different scientific research capabilities and use of data. Such an infrastructure should have facilities for the discovery, curation, provenance, analyzation and visualization of data. In a similar vein, it can be argued that an open data infrastructure can do the like for the public; however, it cannot be expected that the public has the same amount of knowledge and capabilities as researchers do. Lowering the knowledge level required for use is key to large-scale dissemination.

Myth 2: All information should be unrestrictedly publicized

Open data policies are generic and stimulate the publicizing of all data. This myth neglects a number of issues. First, data that can be traced back to the individual may not be publicized due to privacy legislation. Second, limited resources for publishing, especially if the publishing of open data does not bring benefits, can be another reason for not opening data to the public. Why spending tax payers' money on something that yields no benefits? Third, information quality might vary and be too low. Constituents expect government to account for the quality of data. Opening data that has no adequate information quality can result in discussions, confusions, less transparency and even in less trust in the government. The latter can be explained by the fact that resources are wasted and only fuzzy or even incorrect outcomes can be created when there is low information quality. The adagio "garbage in, garbage out" certainly holds true for open data. Fourth, the complexity of data structures and

difficulty to understand the data might make it less attractive to publicize it, unless the complexity is reduced and/or the use is guided. Fifth, law might prevent the publication of certain data. Information collected for a certain purpose by public organizations might not be allowed to be used for another purpose. The paradox is that regulation and policies can on the one hand enhance the publicizing of data, whereas on the other hand policies and regulations inhibit data sharing. Sixth, data sets accrue income for some public organizations. In the Netherlands some organizations' revenue model is based on the income generated by asking users' a fee for access. Indiscriminately publicizing all data will result in harming this business model. Hence, the ability to maintain exclusive control of data can significantly influence the ability to run these public organizations and can undermine their financial existence. As such, we recommend more research into the various types of open data business models. Finally, the arbitrary opening of some data might result into a biased picture of the situation. Wrong conclusions might be drawn if data providing only one view is available whereas data from opposing views or data that can be used to compliment that view is not available or used. An example provided by the interviewees is the open data about secondary schools. Data about these schools is collected and published to show the quality of the schools. But what is published is not the quality, but how well they score on arbitrary, easy-to-measure metrics. The underlying assumption is that the performance can be measured using a set of indicators, whereas these need to be interpreted with care and should be viewed in context. The risk of having an incomplete picture cannot be solved only by warning for interpretation. Data interpretation depends on the verdict of journalists and the audience and not on the public values that are hoped to be fulfilled. The fundamental question therefore is: what does the open data conceal?

Public data has a wide variety of characteristics. Information about weather, crime rates or enforcement budgets might result in different benefits and might be confronted with different

types of barriers. This data might vary in aspects such as the level of detail, quality, usefulness, intrinsic value and so on. This diverse nature of open data suggests that benefits and barriers of open data might vary according to the type of data. Furthermore, the diverse nature also suggests that different data needs to be processed in different ways. For example, criminal data needs to be processed in such a way that it cannot be traced back to the criminal, as required by privacy law, while this is not an issue with weather information.

In conclusion, consideration should be given to whether, how and which public sector data can be publicized. This is dependent on constraints from the legislative environment, the resources needed, potential risks of misuse and bias, and prospective value that can be gained from publicizing the data.

Myth 3: It is a matter of simply publishing public data

Policy-makers prefer to simply make data available. As one interviewee stated, “Preferably we can just drop the data and don’t have to worry about provenance, enriching or whatever”. This myth challenges that data can be made available without additional activities. Source data can often not immediately be used; quality assessment and the modification and processing of raw data might be needed first. In addition, data cannot be easily found if essential meta-data like the publisher, authors, timeliness and so on is not available. A key issue is not to link from the bottom up, but to also use meta-data in the linking. Meta-data is necessary to overcome barriers like searching, interpretation and so on. This is similar to searching in a library: sometimes you want to search by author (for example to see if more books by this author are available), sometimes by subject (to search for similar types of data), sometimes by date (all data published within a certain period of time). Standardization of methods and the development of robust meta-data can increase data access.

One of the underlying causes is that no feedback mechanism are available which show what is done with the open data. Governments that publish data sets face substantial criticism such as poor usability, weak application of stewardship principles, lack of data feedback and improvement mechanisms, and inadequate meta-data (Dawes & Helbig, 2010). These criticisms should be taken into account and resources made available to ensure that data sets are not just published but that they are actually user-friendly.

The barriers also show that users should become aware of the availability of the data; they might have incentives for making use of the data and may also have the knowledge and capability to do so. All these elements must come together before benefits can be accomplished. This requires an open data policy aimed at setting the right conditions and creating the right environment to stimulate open data use. With open data the publishing view is often taken, but this concerns the input views. There are no suitable metrics to evaluate whether opening data is a success (Bertot, McDermott, & Smith, 2012). Current metrics are all too often focused on the input, for example how many data sets are opened (ibid). Yet users are not interested in open data metrics; they want answers to their questions, services and other added value that can be created from open data. Processing open data is simply the hoop they have to jump through to gain answers to their questions.

Myth 4: Every constituent can make use of open data

The dream is that everyone can make use of the data that is available and that anybody can use the data directly. This might be true for relatively straightforward data or for functions for which easy-to-use software applications have already been developed. This myth assumes that open data users have the resources, expertise and capabilities to make use of the data. Reality is more stubborn. Some data requires the use of statistical techniques, a deep understanding of the underlying data and an understanding of the types of (causal) relationships. This is

knowledge that is not available to everyone and might require considerable time and effort to achieve.

In Huff's book "How to Lie with Statistics" a quote of H.G. Wells is cited: "*statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write*" (Huff, 1993, in: epigraph). This quote becomes even more realistic with the open data movement. Only persons having an understanding of statistical techniques and other knowledge needed for processing open data are able to make sense out of the data and to understand the implications.

Indeed many barriers exist to use open data. Yet also many instruments exist to lower the barriers. For example, visualization can help non-experienced users, and guidance in the potential use of data can also help constituents. This, however, requires that current efforts take the user's perspective into account and monitor the need, ultimately helping users and lowering the threshold to using open data. In general limited insight in the capabilities and needs of users is expressed. The ability to handle large volumes of data is necessary, as are automated content and topic analyses. Algorithm-learning machines can often analyze a body of data and infer rules for classifying and grouping data items. Automating the annotation and classification of data can aid to simplify searches, analyze relationships and extrapolate trends.

Myth 5: Open data will result in open government

Open government promotes transparency and engagement to allow effective oversight. This myth suggests that full, immediate and widespread disclosure of public data results in an accountable and transparent government (see for example European Commission, 2010). Yet at least two main assumptions challenge this myth. The first is that one is able to find the right data and is able to interpret and process the data in a uniform way, whereas finding the right

data might be hard, there might be a huge information overload and large differences in the way open data analysts analyze the same data and they may draw different conclusions. Furthermore open data sources might not be consistent and depict to different directions. Second, system theory provides the need for introducing feedback mechanisms to close the loop between the government and those governed. Although there are anecdotal examples the wider impact is unclear. It is easier to not publish data than to introduce mechanisms to seek feedback and discourse in a climate of decreasing budgets. Creating an open government demands considerable transformations of the public sector

The paradox is that more information does not necessarily result in better or more democratic or more rational decisions. More information can result in less understanding, more confusion and less trust (Strathern, 2000). Experience and tacit knowledge might be lacking among those who should be able to make sense of the open data. Our interviews indicate that transparency might result in better accountability, transparency and trust, but sometimes has the opposite effect. For example, publicizing data can show that the quality of the data on which important decisions are made is poor. By providing the minutes of city council meetings, people might be shocked about the discussion and lose trust in the decision-making process and the resulting decisions.

Conclusions and recommendations for further research

A comprehensive list of political and social, economic, operational and technical benefits was derived showing the potential of and the sometimes wild expectations about the impact of open data. The benefits have a generic character and do not say much about the individual data sets. The promises and potential of open data contrast sharply with the many barriers. The diverse nature of open data means that different types of results from open data have different benefits and are confronted with different barriers. Treating open data as a

homogenous topic does not sufficiently acknowledge this diversity. Hence more research into the specific benefits, barriers and value of open data is necessary. The long list of benefits presented in this paper can be used in determining the benefits on a case by case basis.

Most of what has been written about open data has focused on its benefits and positive implications. While open data potentially can provide numerous benefits, we found that it also entails a number of barriers in the field of task complexity, use, legislation, information quality and participation. These barriers are often interrelated, which adds to the overall complexity. The barriers show that one risk is that data is only publicized in name, but that the barriers create such a high threshold that the data is still private in practice. Open data has no value in itself; it only becomes valuable when used. Little is known about the conversion of public data into services of public value. Hence we strongly suggest further research in this area.

Arguments in favor of open data are based on a rather simplistic and idealized view. By confronting benefits with barriers five myths were identified. These myths are used as seductive tales of open data's importance and simplicity. The first two myths, 'the publicizing of data will automatically yield benefits' and 'all information should be unrestrictedly publicized', are used to convince data providers to open their data to the public but ignore the many barriers to and heterogeneous nature of open data. The next two myths, 'it is a matter of simply publishing public data' and 'every constituent can make use of open data', show that the user view is largely neglected and use of open data might not be easy. The last myth, 'open data will result in open government', suggests that the transformative nature of open data is more elusive than might be expected. Systems theory suggests that open data equates to less control and accountability over data and institutional theory suggest that publicizing data will reinforce existing structures instead of changing them to fully take advantage of new developments. This prediction is confirmed in the interviews and by the barriers.

Governments have to accept that they inevitably give up some level of control when opening their data to the public. New types of governance mechanisms and policies are necessary in which the more evolutionary manner of steering at arm's length is adopted.

Much of today's focus is on the suppliers of data, whereas achieving the success of open data systems depends to a large extent on the use and the quality of the data provided. One of the main criticisms of current open data initiatives is that they are largely supply-driven. Open systems require an understanding of the external world and must consider the feedback and insights of users in order to continuously improve. There is no insight into the user's perspective and users' needs are not known. This confirms that existing governance instruments are not prepared to deal with open data systems. The lack of insight into the user's perspective and the lack of appropriate governance mechanisms can explain the large gap between the promises of open data and what is actually realized. 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. In addition, incentives for stimulating and using open data, the risks of publicizing data and mitigation strategies to deal with these risks deserves more research attention. In particular principles and measures which can avoid or deal with the negative aspects of open data should receive attention.

The success of open data systems requires more than the simple provision of access to data. Also needed are the improvement of the quality of government information, the creation and institutionalization of a culture of open government, and the provision of the tools and instruments with which to use the data. This broader perspective needs to be taken by the governments that are now merely opening a portal to make data accessible. An infrastructure is necessary which helps users to make sense of data, and institutional measures are necessary to ensure public engagement. Under these conditions, open data can potentially go beyond the

current level of citizen engagement and could result in a continuous dialogue between governments and their constituents, drawing upon the collective intelligence of the public.

Acknowledgement

This paper is related to the ENGAGE FP7 Infrastructure Project (An Infrastructure for Open, Linked Governmental Data Provision Towards Research Communities and Citizens), which started in June 2011. The authors would like to thank their colleagues of the ENGAGE project for their input toward this paper. The views expressed here are those of the authors and do not necessarily reflect the views of the project.

References

Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., et al. (2004). An International Framework to Promote Access to Data. *Science*, 303(5665), 1777-1778.

Bekkers, V., & Homburg, V. (2007). The Myths of E-Government: Looking Beyond the Assumptions of a New and Better Government. *Information society*, 23(5), 373382.

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.

Bertot, J. C., McDermott, P., & Smith, T. (2012). Measurement of Open Government: Metrics and Process” in the held in , . Paper presented at the Proceedings of the 45th Annual Hawaii International Conference on Systems Sciences (HICSS2012).

Bozeman, B., & Kingsley, G. (1998). Risk Culture in Public and Private Organizations. *Public Administration Review*, 58(2), 109-118.

Charalabidis, Y., Ntanos, E., & Lampathaki, F. (2011). An architectural framework for open governmental data for researchers and citizens. Paper presented at the Electronic government and electronic participation joint proceedings of ongoing research and projects of IFIP EGOV and ePart 2011.

Chun, S. A., Shulman, S., Sandoval, R., & Hovy, E. (2010). Government 2.0: Making connections between citizens, data and government. *Information Polity*, 15(1/2), 1-9.

Dawes, S. S., & Helbig, N. (2010). Information Strategies for Open Government: Challenges and Prospects for Deriving Public Value from Government Transparency. Paper presented at the EGOV 2010.

European Commission. (2010). The European eGovernment Action Plan 2011-2015. COM (2010) 743, December 15, 2010, Brussels: European Commission.

Fountain, J. E. (2001). Building the virtual state. Information technology and institutional change. Washington DC: Brookings Institution Press.

Hausenblas, M. (2009). Exploiting Linked Data to Build Web Applications *IEEE Internet Computing*, 13(4), 68 - 73.

Herik, C. W. v. d., & Vreede, G. J. d. (2000). Experiences with Facilitating Policy Meeting with Group Support Systems. *International journal of Technology and Management*, 19(2/3/4), 246-268.

Hey, T., & Trefethen, A. E. (2005). Cyberinfrastructure for e-Science. *Science* 308(5723), 817-821.

Huff, D. (1993). How to lie with statistics. New york: W W Norton & Company.

Jackson, M. C. (2003). *Systems Thinking: Creative Holism for Managers*. Chichester: John Wiley.

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.

Kalampokis, E., Tambouris, E., & Tarabanis, K. (2011). A classification scheme for open government data: towards linking decentralised data. *International Journal of Web Engineering and Technology*, 6(3), 266–285.

Kraemer, K., & King, J. L. (2006). Information Technology and Administrative Reform: Will E-Government Be Different? *International Journal of Electronic Government Research*, 2(1), 1-20.

McDermott, P. (2010). Building open government. *Government Information Quarterly*, 27(4), 401-413.

Miller, H. (1996). The multiple dimensions of information quality. *Information Systems Management*, 13(2), 79 - 82.

Napoli, P. M., & Karaganis, J. (2010). On making public policy with publicly available data: The case of U.S. communications policymaking. *Government Information Quarterly*, 27(4), 384-391.

Nationale_Ombudsman. (2011). Nationale ombudsman: overheid zo gesloten als een oester. SC online,

Orlikowski, J. W. (2000). Using Technology and Constituting Structures: A practice Lens for Studying Technology in Organizations. *Organization Science*, 11(4), 404-428.

Peters, B. G., & Pierre, J. (1998). Governance without government? Rethinking public administration. *Journal of Public Administration and Theory*, 8(2), 223-243.

Scott, W. R. (1995). *Institutions and Organizations*. Thousand Oaks, CA: Sage.

Strathern, M. (2000). The Tyranny of Transparency. *British Educational Research Journal*, 26(3), 309-321.

Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data Quality in Context. *Communications of the ACM*, 40(5), 103-110.

Surowiecki, J. (2004). *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business Economies, Societies and Nations*: Doubleday.

West, D. M. (2004). E-Government and the Transformation of Service Delivery and Citizen Attitudes. *Public Administration Review*, 64(1), 15-27.

Wiener, N. (1948). *Cybernetics*. New York, NY: John Wiley & Sons.

Zurada, J., & Karwowski, W. (2011). Knowledge Discovery Through Experiential Learning From Business and Other Contemporary Data Sources: A Review and Reappraisal. *Information Systems Management*, 28(3), 258-274.