The open data landscape

Charalabidis, Yannis; Zuiderwijk, Anneke; Alexopoulos, Charalampos; Janssen, Marijn; Lampoltshammer, Thomas; Ferro, Enrico

DOI
10.1007/978-3-319-90850-2_1

Publication date
2018

Document Version
Final published version

Published in
Public Administration and Information Technology

Citation (APA)
Charalabidis, Y., Zuiderwijk, A., Alexopoulos, C., Janssen, M., Lampoltshammer, T., & Ferro, E. (2018). The open data landscape. In *Public Administration and Information Technology* (pp. 1-9). (Public Administration and Information Technology; Vol. 28). Springer. https://doi.org/10.1007/978-3-319-90850-2_1

Chapter 1
The Open Data Landscape

"Open data has many different aspects: objectives and benefits for a variety of stakeholders, but open data also has a dark side."

1.1 Creating a New World of Open Data

The opening of data has grown tremendously over the past decade. More and more datasets have been opened to the public, application programming interfaces (APIs) gave been design for enabling the public to make use of real-time data and new apps based on this data have been developed. Data about policy-making, software code (open sources), documents, minutes, financial data and so on has been opened resulting in a large repository of government data that can be on open data portals and government websites. Nevertheless the potential is even higher, as most of the data is still closed and is not directly accessible to the public. Furthermore, more and more data is collected and can be share in nowadays words driven by The Internet of Things (IoT). The IoT consist of devices that are able to collect data such as GPS (geographical location), compass, temperature, movement, pollution and so on. Devices collecting data combined with data analytics are expected to transform the government and society. This can provide insight into the energy consumption of smart cities (https://amsterdamsmartcity.com/projects/energy-atlas) or the pollution (http://airindex.eea.europa.eu/). These initiatives are all driven by the opening of data and extended by user-friendly apps to enable large use by the public.

Over the course of the past few decades, many governments have imitated all kinds of projects to open their data to the public. This practice have been followed by private organizations that started also opening some of their data resulting in the creation of business value (Zuiderwijk, Janssen, Poulis, & van de Kaa, 2015). The availability of open government data has grown significantly, with pressure being placed on all kinds of public organizations to release their raw data for the good.

The movement of opening data resembles a move from a closed to an open system (Janssen, Charalabidis, & Zuiderwijk, 2012). Open systems are encountered with uncertainties from the environment and are less predictable and therefore not

Y. Charalabidis et al., *The World of Open Data*, Public Administration and Information Technology 28, https://doi.org/10.1007/978-3-319-90850-2_1

easy to manage. By opening some data, also insight into the functioning of the government is revealed. This might be viewed as a risk by some public servants, whereas others views this as a way to strengthen the democratic system by creating transparency and accountability. The public is empowered by giving it the data and the means for making sense out of the data. Also businesses can benefit from the opening of data and enrich their existing products or develop new products (Zuiderwijk, Janssen, Van de Kaa, & Poulis, 2016).

Open data and open government are related. Open governments objectives relate to creating transparency, accountability and engagement to strengthen the governance and empower citizens. The opening of data is a means for this, but not sufficient as also institutional measures might be necessary. This includes steps to take measures when corruption or fraud is detected using open data. Open can include Open Government Data (OGD), but also Open Business Data (OBD) or Citizen-Generated Data (CGD). The latter is data collected by citizens, which can be done by using IoT devices.

Also the public can become part of the policy-making processes. Ordinary people can become part of the policy-making and might collect data, process data, combine it with other sources to create new insight to help policy-makers. In this way, new opportunities for involving the public in policy-making processes become available. Also citizens 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).

Open data can be looked at in various ways and there are various definitions available. Instead of giving another formal definition we prefer to look at the characteristics of what makes data really open. The Sebastopol principles elaborate on what makes data "open data" (Malamud et al., 2013). Open data should be primary data, published in a timely manner and allowing diverse groups with different interests to take advantage of this. This includes the following aspects.

- Data must be complete
- Data must be primary
- Data must be timely
- Data must be accessible
- Data must be machine processable and made online in persistent archives
- Access must be non-discriminatory
- Data formats must be non-proprietary
- Data license must be unrestricted and bear no usage costs
- Also data should be as accurate as possible.

Indeed most of the data will not meet this list of requirements. Nevertheless, data is only truly open if most of these criteria are met. In this book the 5 stars model of Tim Berners-Lee will be discussed which provides insight into the maturity of the data, where each additional star means that the data meets the criteria of the previous steps (http://5stardata.info/en/).

1.2 Historical Developments

The opening of data by the government has already a long history. Traditionally data was only opened upon request by the public. The right to have access to data is central to the Freedom of Information (FOI) Act. Although many countries already had a FOI act before, the FOI is linked to the article 19 of the 1948 Universal declaration of basic human rights of freedom of expression (http://www.un.org/en/universal-declaration-human-rights/). Many countries have Freedom of Information Acts (FOIAs) in place in which citizens can ask for information (Petticrew & Roberts, 2008). FOIA allows the public to ask for (partial) disclosure of information and data is not released yet. The amount of FOI varies overtime and the requests are often coming from the same stakeholders who have the opportunity and time to ask for this data. Governments have developed procedures and processes to receive FOI requests, process them and give answers. Some people have misused this act to ask many questions requiring many resources of the government. Yet, the asking for information cannot be used by companies for innovating their products or developing new value propositions. Also following FOI is a cumbersome and sometimes lengthy procedure which makes it less suitable for certain applications.

Whereas, FOI is based on the 'based upon request' principle, the proactive provision of data to the public is based on the 'open by default' principle. The pro-active opening of data streams is initiated by Obama's Memorandum on 'Transparency and Open Government' published in 2009 (McDermott, 2010). Obama's Memorandum encourage the active disclosure of public data, instead of waiting for requests. This Memorandum resulted in the development of open data portals (see for example www.opendata.gov) in which open data is released to the public. Policies stimulating the opening of data were developed and public organizations were asked to start with the release of their datasets. The USA example served as a sources of inspiration for many other governments. For example, the EU Public Sector Information (PSI) directive, which is focused on making public sector data available and ensuring a level playing field (European_Parliament_and_Council, 2003).

The Open Government Partnership (OGP) is a partnership launched in 2011 to stimulate open government by empowering citizens, fighting corruption, and harness new technologies to strengthen governance (https://www.opengovpartnership.org/). The opening of data is an important means for this. Opening up government data is a voluntary initiative that countries can join and is aimed at securing and taking actions to strengthen governance.

1.3 Objectives of Open Data

The objectives of open data relate to coming closer to an open government, stimulating and enabling private sector innovation, and stimulating engagement and participation of stakeholders like citizens and companies. The three areas are visualized in the figure below. Government should become transparent and accountable by

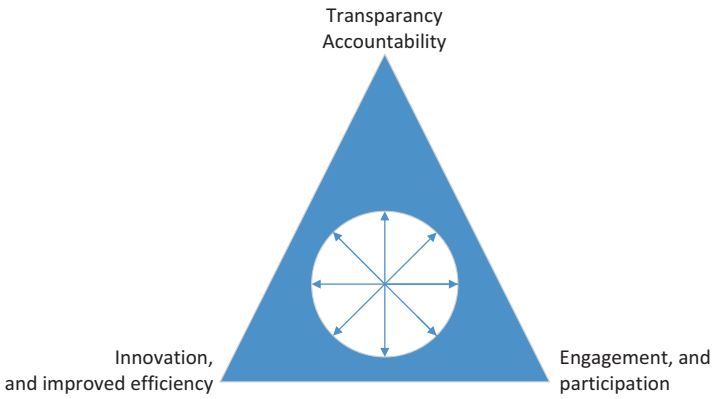

Fig. 1.1 Objectives of open government data

promoting the public right of access to information (McDermott, 2010). This can even be viewed as a requirement of a democratic system and concerns the opening of data about the functioning of the government and their decision-making.

The second one is has economic motives to encourage the opening of government data which can be used by companies and society to create value. The government has a lot of data that, when opened, can be used to create new entrepreneurial activities, to add value to existing services offerings, or to create new insights which enable to improve business (Fig. 1.1).

The third area of open data objectives concerns the stimulation of engagement and participation. Open government data gives governments a new means to communicate their activities to citizens and other stakeholders and to invite various actors to give feedback on government activities and participate in them.

1.4 The Stakeholder Landscape

There are often many stakeholders involved in the opening of data. Often the actor that is sharing the information is not necessarily the organization that collected the data or processed the data. Many more organizations and departments might be involved. Some organizations might support the opening of data like software vendors, whereas other stakeholders are directly involved. The stakeholder landscape adds to the complexity of open data as responsibilities for opening data might not be clear, the ownership of data cannot be defined easily and many parties should collaborate for opening data (Table 1.1).

1.5 Open Data and Big Data: A World Apart?

The field of open data consist of the many areas referred to the term ‘data’ in general as shown in the figure below. The origin of the data can be the government, businesses or citizens. Open data refers to the situation that data is made available

Table 1.1 Overview of main stakeholders

Stakeholder name	Simplified Stakeholder descriptions
Politicians	Often in charge of creating open data policies and assigning resources and budget for realizing open data policies.
Data collector	These organizations are collecting data. These are often public organizations, but can also be citizens who collected data or businesses.
Data processor	Data needs to be analyzed and processed. Many organizations have data analysts who are able to make sense of data.
Data publishers	The organizations that publish the data. Often in open data portals, but also using APIs and other means.
Infrastructure providers	Data needs to be communicated over a secure and reliable network infrastructure.
Companies and businesses	Business might use open data to add value to their existing products and services or analyze data for their own advantages. They might even use open data for empowering their lobbyist.
Infomediaries	Infomediares (information + intermediaries) are organizations that collect and process open data. They have more resources and can lower the threshold for citizens to use the data
Citizens	Citizens are often the intended users of data. The idea is that they will engage with the government, use open data to scrutinize decision-making in the government to create transparency and held them accountable. Ultimately this should result in a higher trust in the government, however, this is challenged by many researchers.
Software vendors	Often many software providers are involved to collect, process or to publish data. There are specialized organizations that are able to anonymize documents for publications. There is software that can be used to analyzed and visualize data, and there is separate software to make data available.
Regulators and ombudsman	Regulators have two different roles on the one hand can regulators be in charge of ensuring compliance with the data regulations and data protection act. On the other hand they can use the data for investigating governments and to look at what is happening.

outside the own organization for use by others. Ideally to everybody without any restrictions for further use. Yet, licenses might limit what can be done with the data. Often data might not be used for commercial use which limits the use for businesses to make profit from data.

Big data is commonly characterized by several Vs, including Volume, Velocity, Variety (McAfee & Brynjolfsson, 2012). Gandomi and Haider (2015) add another three Vs to this list; Value, Variability and Veracity. The essence of big data is that this concerns data that cannot be handled in traditional ways (Elgendy & Elragal, 2014a). Big data is closely related to Big Data Analytics (BDA) which are needed to create value of the data (Elgendy & Elragal, 2014a; Holsapple, Lee-Post, & Pakath, 2014). Although big data and open data are closely related, yet they are not the same as, big data is characterized by its size and open data by its availability (Janssen, Matheus, & Zuiderwijk, 2015).

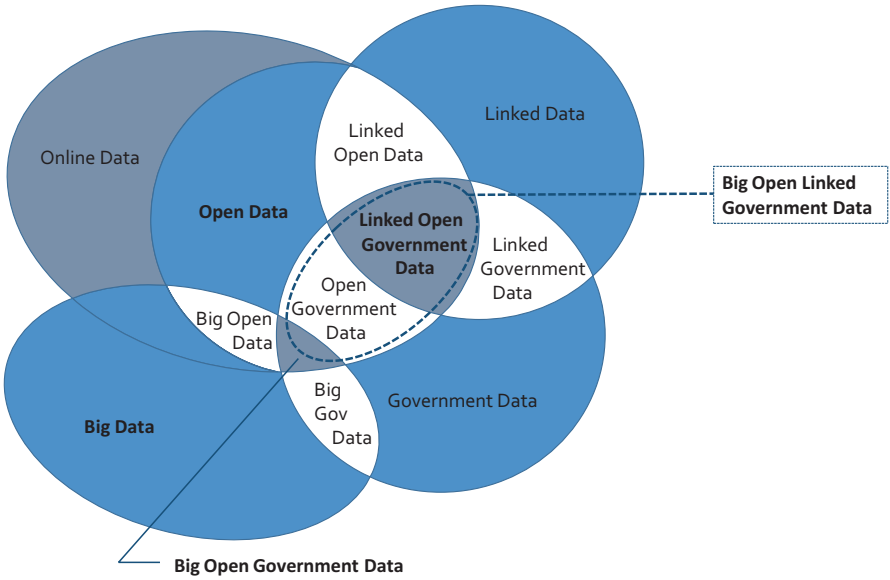

Fig. 1.2 Overview of the field of open data

Data often originates from many sources which are often beyond the control of a single actor like social media and devices. Therefore there is a need to link data to created 'linked data'. Linked data is about relating structured data into machine-readable format that can be semantically queried (Bizer, Heath, & Berners-Lee, 2009). This enables the searching for the data, but also to combine different datasets to create value from them. The creation of value from data requires combining large datasets originating from different and heterogeneous data sources (Janssen, Estevez, & Janowski, 2014). *Big Open and Linked Data* (BOLD) is an acronym often used for depicting to the use of data in the digital age referring to the changing nature of data (Janssen et al., 2015) (Fig. 1.2).

1.6 Benefits of Open Data

There are many benefits can be accomplished with the opening of data that range from political to technical benefits (Janssen et al., 2012) as listed in the table below. The benefits are not mutually exclusive, but they are a good starting point for making the case for opening data (Table 1.2).

Table 1.2 Overview of benefits of open data

Category	Benefits	Description
Political and democratic	More transparency	The creation of more insight into the functioning of the government
	Democratic accountability towards citizens (users)	The answering for the actions taken to the public. The ability for other organizations to scrutinize the government ad to check their actions
	Trust in government	The creation of more trust in the government by acting in a transparency and accountable way.
	More participation and self-empowerment public engagement	The ability of citizens and business to participate better in democratic decision-making and have a deeper understanding of the issues.
	Empowering the public	Providing the means to understand and participate in decision-making.
	Equal access to data	Leveling the play field by ensuring that the public have the same data as policy-makers have.
	New governmental services for citizens/improvement of citizen services	The creation of new data-driven services that can be more customer-centric and deal with societal problems.
	Improvement of policy-making processes	By providing information policy-makers can tap into the wisdom of the crowds and in this way improve the policies.
Organizational	More visibility for the data provider	Organizations might be viewed as innovative and gain more visibility
	Improvement of citizen satisfaction	The opening of data can result in more satisfied citizens.
	Use of the wisdom of the crowds:	By opening of the data organizations can tap into the intelligence of the collective
Innovation	Stimulation of knowledge developments	The opening of data can result in new innovative applications with the data by others, which in turn can stimulate innovation within the government.
	Creation of new insights in the public sector	Opening of data can create new insights. Also government can start to use each other data and create new insights.
	New (innovative) social services	New services can be developed for the public
Economic	Stimulation of competitiveness and innovation	Data can be used to create new businesses, for the development of new products and services or to extend current service offering
	Economic growth	The creation of a data economy. Data is fueling economic growth. Creation of a new sector adding value to the economy
	Availability of information for investors and companies	Companies can use open data to determine where to invest and where to locate their businesses.

(continued)

Table 1.2 (continued)

Category	Benefits	Description
Operational and technical	Reuse of data	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)
	Improve administrative processes and policies.	The opening of data and feedback gained can be used to optimize administrative processes and policies.
	Improving the quality of data	External quality checks of data (validation) and the public can help to improve the quality of data.
	New data	The ability to merge, integrate and mesh public and private data. Creation of new data based on combining data.

Based on Janssen, Charalabidis, and Zuiderwijk (2012)

1.7 The Dark Side of Open Data

All too often the focus of politicians is on the benefits and the possibilities of open data, whereas the public administration is afraid of the risks of opening data. The opening of data might require considerable resources, however, the opening might not result in any public value at all. Resources might be wasted on releasing data that are not used or even not relevant. Zuiderwijk and Janssen (2014a, 2014b) found the following issues that might hinder the opening of data, although there are many mechanism that can be used to overcome them. For example, privacy-enhancement mechanisms (PEM) are often used to comply with the data protection act (Table 1.3).

The risks might result in inertia and the avoidance of the opening of data. Nevertheless most of the issues can be dealt with, however, the costs needed to deal with them often hinder the opening of data. Budgets are tight and many organizations have no or limited budget for opening data.

1.8 Developments

Whereas much focus is still on opening data, there are developments to have 'openness by default' and "transparency-by-design". These concepts refer to the situation where software is designed in such a way that when data is collected the data is collected in such a way that the opening of data is possible (Janssen, Matheus, Longo, & Weerakkody, 2017).

Data is fragmented, described in different formats by different organizations. In many portals data is opened, but not well-described which makes searching for data and the interpretation of the usefulness of datasets difficult. Semantic descriptions,

Table 1.3 Overview of risks of open data

Category	Risk	Description
Legislation	Non-compliant	All kinds of legislation might be applicable from different domains. There might be unawareness of which legislation might be applicable.
	Privacy	The data protection act poses strict requirement on what can be published and what cannot be published. Although there are PEM, the privacy of persons can be violated unintentionally or data might be used for another purpose than what it was collected for.
Governance	Responsibilities	Difficulties with data ownership and stewardship. Unclear responsibility and accountability.
	Maintaining quality	Unclear if data is updated and what the data quality is.
Interpretation	Bias	Published data can be biased
	Ambiguity misinterpretation	Misinterpretation and misuse
Data quality	Poor data quality	Data might have different qualities (completeness, accuracy, timeliness). Decisions can be made on poor information quality or the wrong insights can be created.
	Timeliness	The most recent data might not be available. Embargo period prohibits the publication of recent data

Based on Zuiderwijk and Janssen (2014a; 2014b)

adding metadata and linking the data improves the use of the data. In addition, meta-search engines, have become available which have indexed many data portals. Also there are data standardization working groups that are developing comprehensive meta-data models for describing open data like CERIF (Jeffery, Houssos, Jörg, & Asserson, 2014).

Also automatic annotation and retrieval software has been developed. Data range from structured to unstructured data and data might not be used easily. Unstructured data can be transferred into structured data by annotating the data. For example, this happens when somebody adds the persons in a picture on Facebook. More and more automatic tools can be used to automatically annotate unstructured information.

Also in the field of -statistical data and visualization there are initiatives to make the collection, linking and analysis of Linked Open Statistical Data (LOSD) easier (Kalampokis, Tambouris, & Tarabanis, 2017). In the ideal situation no knowledge of software is needed and by drag and drop applications statistical data can be combined and visualized.