A Decision Model for Data Sharing

Silja Eckartz, Wout Hofman, Anne Veenstra

A Decision Model for Data Sharing

Silja Eckartz, Wout Hofman, Anne Fleur van Veenstra

Dutch National Institute of Applied Science, Technical Science Department, P.O. Box 5050, 2600 GB Delft, The Netherlands
((silja.eckartz), (wout.hofman), (anne_fleur.vanveenstra))@tno.nl

Abstract. Data-driven innovation has great potential for the development of innovative services that not only have economic value, but that help to address societal challenges. Many of these challenges can only be addressed by data sharing of public and privately owned data. These public-private data sharing collaborations require data governance rules. Data governance can address many barriers, for example by deploying a decision model to guide choices regarding data sharing resulting in interventions supported by a data sharing platform. Based on a literature review of data governance and three use cases for data sharing in the logistics sector, we have developed a data sharing decision model from the perspective of a data provider. The decision model addresses technical as well as ownership, privacy, and economical barriers to sharing publicly and privately owned data and subsequently proposes interventions to address these barriers. We found that the decision model is useful for identifying and addressing data sharing barriers as it is applicable to amongst others privacy and commercial sensitive data.

Keywords: Data Governance, Data-Driven Innovation, Public Service Innovation, Open Data, Decision Model.

1. Introduction

Data is often proclaimed to be the new oil – or the new gold – for innovation and economic growth [1]. 'Open' and 'big data' raise high expectations [2]. Open data is the provisioning of data by government organizations for free in a re-useable format [3]. Most literature takes a so-called push approach in which the data availability will contribute to public – and private sector – innovation [4]. A law such as the Freedom Of Information act in the United States, and expectations of economic growth and innovation [1] are drivers for open data. Open data aims for organizations to become more transparent and thereby accountable to citizens [5], to realize economic activity [6] or to increase organizational efficiency and effectiveness by better decisions [7]. Big data is the processing of large, (un)structured and real-time data sets for a wide variety of purposes, including the objectives of open data. Both developments are expected to not only create new economic activity, but also to contribute in addressing societal issues and challenges, such as a decrease of CO2 emissions, or a decrease of the costs incurred for health services or social welfare. Jetzek et.al [8] have

constructed and validated a model for value generation by open government data, where they have defined value from an economical and social perspective. According this model, open data has only a marginal impact on innovation.

Societal challenges can no longer be addressed by public organizations alone, they are often based on a combination of public and privately owned data. However, data sharing by private organizations may pose other challenges as those posed to public organizations [4] since private organizations have their competitive position to consider. Organizations often find the process of opening data cumbersome and many challenges and barriers occur [9]. To support the value creation with data, this paper develops, based on literature and practice, an approach to identify barriers to data sharing from the perspective of the public and the private sector, and proposes interventions to overcome these challenges and barriers.

Based on a literature review of data governance and open data, we first identify barriers to data sharing. These barriers are subsequently validated and potential interventions to overcome these barriers are identified by looking at three use cases from the logistics sector, using an interpretivist methodology [10]. Interpretivist research is "aimed at producing an understanding of the context of the information system, and the process whereby the information system influences and is influenced by the context" [11]. The use cases are investigated using on desk research and interviews with stakeholders.

In the next section, barriers to data sharing are identified from literature. Subsequently, we present three use cases of data sharing in the logistics sector that are used to validate the barriers and identify the barriers found in literature. Based on the literature and on these use cases, we present a decision model for data sharing. Subsequently, we discuss the model and present conclusions and recommendations for further research.

2. Literature review

This section identifies and analyzes potential barriers to data sharing from literature on data governance and open data.

The management of data is of raising importance for many organizations given the growing supply of structured as well as unstructured data [12]. Data governance is an emerging discipline and comprises parts of IT governance [13]. Weill and Ross [14] distinguish IT governance and IT management where governance refers to the decisions that must be made to ensure effective management and use of IT and to whom these decisions are makes, and management involves the actual decision making and implementation. According to Thomas [15], "data needs to be governed as it has neither will nor intent of its own. Tools and people shape the data and tell it where to go. Therefore, data governance is the governance of people and technology." Data governance covers aspects such as data quality, data management, metadata management, access rights, decision rights, accountability, and data policies. [16] Data governance literature shows that barriers to data sharing differ when considering

open data, which is a form of data sharing by public organizations with private organizations, and data sharing in between private organizations.

Literature on open data often takes a so-called push approach in which the data availability will contribute to public – and private sector – innovation [4]. From a data governance perspective, the most dominant open data barriers are found to focus on data quality. Data quality is specified in more detail by Batini et al. [17]. Domain-specific metadata describing the data origin, the data production date, data provenance, and for which applications the data can be used is of crucial importance. Data quality aspects that should be considered with respect to the entire data set are: accessibility, data format, semantics, conciseness, completeness, believability and reputation [18]. Data quality aspects that should be considered with respect to data elements are: validity, completeness, consistency, uniqueness, timeliness, accuracy and preciseness [19]. Other technical barriers to re-using open data are the publishing of data in a format that is not machine readable, the lack of an Application Programming Interface (API), difficulties to processing data sets, the lack of a linking or combining functionality, and difficulties in configuring data transformation [7,13,14].

Other authors take a broader perspective on open data. For example, Jetzek et al. [8] constructed and validated a model for value generation by open government data, where they have defined value from an economic and social perspective. Besides technical barriers, Zuiderwijk et al. [4], Janssen et al. [20], and Barry et al. [21] analyze potential barriers to publishing open data according to various perspectives: political, social, economic, institutional, operational, and legal. Political barriers include a lack of support, a lack of attention and a lack of knowledge about open data. Among the social barriers are a lack of interaction with users, difficulty to measure impact, cultural differences and risks and liability with respect to providing low data quality. The lack of business models is a main economic barrier to open data. Institutional barriers include a lack of standards, a lack of an open data policy, an inability to handle user requirements and a lack of guidelines. Data fragmentation, a lack of services, a lack of metadata, changing or a lack of clear semantics, and a lack of information on data quality are among the operational barriers. And the legal barriers include licensing, policy differences, lack of (detailed) policy.

In settings where data is shared with or between private organizations, most barriers to data sharing are related to privacy or to competition regarding economically sensitive data. While many authors mention privacy issues, not all of them explicitly elaborate on specific privacy problems [22]. Bizer et al. [23] elaborate the perspective of the user of data and especially the privacy issues related to combining several data sources. As long as data from several sources is viewed separately it might not involve any privacy issues but as soon as one data source is combined with another privacy threats might arise. In all settings it is important to have clear defined decision rights [14]. These are often defined by general IT governance and ownership structures. We further see that there is limited tool support and competing licenses for data sets.

Based on literature on data governance and open data we identify five main categories of barriers to data sharing: technical, data quality, ownership, privacy and economic. We consider barriers like political, institutional, and lack of or missing business models identified in open data literature as drivers to data sharing, which are

a prerequisite for analyzing data sets. Our overall model addresses these conditions, but they will not be part of the proposed decision model. Technical barriers include barriers related to re-use and precision and recall of data and/or their source [24], which can be improved by metadata as an intervention. Data quality is addressed by aspects related to individual data as well as to datasets.. Data ownership regards the question of who is allowed to use and determine re-use of data and who has decision rights. Data privacy is not mentioned as a barrier in open data literature; we can only assume that it is addressed in specifying data policies, but it is mentioned in literature on data governance. Economic barriers include interventions like billing and invoicing of data usage and address liability, which also relates to data quality. Barriers in open data literature rank liability as a social risk, but one could also consider liability from a commercial perspective among businesses.

3. Logistics use cases

This section presents three use cases in the logistics sector to validate the identified barriers as well as identifying interventions for overcoming these barriers. Logistics is considered as an application area for data sharing, since it is a fairly complex environment with many Small and Medium sized Enterprises (SMEs) and a limited number of globally operating large enterprises. Each of these enterprises is autonomous, but has to adhere to international trade agreements that also address data sharing like the Rotterdam Rules [25] that specify a clear separation of concern between transport of cargo and activities like declarations and warehousing with commercial information. Furthermore, it is generally expected that data sharing can improve decision making [7] and contribute to efficiency and effectiveness of logistics [26]. In this particular dynamic environment, the use cases considered cover i) infrastructure data provided by Rijkswaterstaat, ii) data sharing between two container terminals that serve as hubs in logistics networks and have to process large amounts of data obeying the aforementioned Rotterdam Rules, and iii) the prediction of turnaround times at one container terminal in the Rotterdam port that has to be able to load and discharge vessels and at the same time the arrival and departure of containers by road. These environments are all data intensive environments with real time data requirements. The first use case about Rijkswaterstaat is based upon a detailed web-analysis done by two researchers independently. Input for use cases two and three has been collected during two in-depth interviews with stakeholders from the terminals. Given the complex and data intensive nature of the three use cases we believe that the most important barriers have been identified and argue that these use cases are sufficient for a first validation of our decision model. In-depth case studies are needed to strengthen our validation and further develop and validate Step 4 of the decision model.

3.1. Use case 1: infrastructure and its usage

Rijkswaterstaat (RWS) is a Dutch governmental organization that is part of the Department of Infrastructure and Environment (I&M), RWS is responsible for the

design, construction, management and maintenance of the main infrastructure facilities, such as highways, water systems and waterways (www.rijkswaterstaat.nl/en/about_us/). The organization is publishing a number of data sets as open data, such as data sets of normal and current water heights and a map containing the location of objects for roads and waterways like lighting objects, painting on the surface of roads, traffic signs, locks, and bridges. The data is not only published on a map at the RWS website, but also available via the Dutch national open data portal. The map in which all this data is available is compatible with international standards of the Open Geospatial Consortium (OGC) for the exchange of geographical data.

As the organization is a public organization, they feel the obligation to publish its data. However, not all RWS data is available as open data. Roughly, the data held by the organization can be divided into static data specifying the infrastructure, which is updated yearly or a few times a year, and real-time data like traffic congestions and waiting times at locks and bridges. While most static data is published, not all real-time data sets are made available as open data. One reason is that some notifications are published as push messages (events) rather than raw data because of their urgency. It supports so-called data driven actions [27] required to improve decision making in situational awareness [7]. Also, some of the data is duplicated. Furthermore, the organization protects some data sets that contain personal data. One example is the real-time location of barges. Often, the barge operators live on their barges and therefore a barge is considered as a 'house' in terms of privacy laws. To protect the privacy of the barge operators, these data are only published anonymized.

3.2. Use case 2: sharing data between two container terminals

Lots of logistics operations occur between container terminals in the Netherlands, one for instance in the Rotterdam port and another functioning as inland port. However, little information is shared between these terminals. This use case examines the potential of information sharing for two terminals to improve their planning, be able to use the capacity of barges better and increase transshipment volumes at terminals. Shared data includes bookings, timetables, available capacity, and vessel or barge movements.

The analyzed data set includes privacy and potentially commercial sensitive information. While the location of inland barges is currently published as open data, the destination of these barges is not published. This information can be privacy sensitive as some inland shippers live on their vessels (see the previous use case). Thus, an intervention to filter out privacy sensitive information needs to be in place. Next to privacy issues, another challenge is commercial sensitivity. Sharing data on barge capacities could decreases transport prices leading to lower profits for barge operators. Sharing booking data could unveil commercial relations between logistic partners, which is considered sensitive from a competitive perspective. Intervention mechanisms are required in filtering commercial sensitive data and new business models are necessary for sharing capacity data. Thus, it is a challenge to find an economic model for sharing data between two terminals. Two options can be considered: i) a terminal sells its data to its partners and ii) the terminals make

bilateral agreements about mutual data exchange. Such agreements should also include service level agreements (SLAs) about data quality and technical formats. The terminals should further specify who is liable for the data.

3.3. Use case 3: prediction of turn around times at terminals based on open data

Time and place (location) are important aspects of logistics. Waiting times decrease the profit of carriers, since transport capacity cannot be utilized during waiting. In this respect, there is a carrier demand for predictable turn around times at drop off or pickup locations like terminals. Currently, these locations are still a 'black box' which operation is determined by its internal processes and the large number of trucks arriving and departing at its location. In the case of a terminal in the Rotterdam port, this often results in a queue at its gate. Carriers could respond to the demand for more information about turnaround times at a terminal, by collecting and sharing information about the location, the average speed the destination and activities at that destination of their trucks. This information could help to determine turn around times at different locations in order to help i) carriers to improve their planning and ii) terminals to better manage container drop off and pickup. Information could be collected using the estimated time of arrival determined by the route planner systems used in most trucks. Other information could be provided by on board units used by truck drivers to report their activities.

Sharing this information involves i) privacy issues (information about location and destination could disclose commercial relations of a carrier), ii) economic investments and collaboration structure to set up a data sharing platform (a business model needs to be thought of, e.g. subscription) and iii) technical aspects like data formats and semantics, as currently no standard format is used to describe the location of a truck.

4. A decision model for data sharing

This section introduces a decision model for organizations to decide if they want to open their data, based on the literature review and the use cases. Firstly, the overall decision model is presented and secondly, individual aspects are described with potential interventions.

4.1. The decision model

We found both data governance and open data literature to address a decision structure, albeit in a different way. Decision rights in data governance literature address the decision structure and decision processes [14], while open data literature addresses data policies from different perspectives, e.g. political on international (for instance European Union perspective) and national level, and institutional, based on a decision structure that is taken for granted. Based on the insights gained from the literature review and use cases we structure the barriers to data sharing into a decision model (figure 1). It is structured into four steps: identification of the goal of data

sharing (Step 1), identification of incentives for individual stakeholders (Step 2), identification of barriers, also referred to as constraints, to data sharing (Step 3) and the definition of the process to publish data (Step 4). Examples of goals (step 1) are for instance accountability to citizens and improved decisions (section 1). A business case and business model are examples of incentives for individual organizations (step 2). Our proposed decision model addresses detailed analysis of individual data sets (step 3), within the context of goals and incentives. This third step supports data analysis by five categories identified in literature: ownership, privacy, economic, data quality, and technical. Each of these categories contains detailed questions to support decisions regarding data sharing. The process to open date (Step 4) consists of a step-by-step guidelines covering technical aspects, such as data conversion, metadata requirements and URI strategies as well as organizational aspects around governance. A detailed description of this step is out of the scope of this paper.

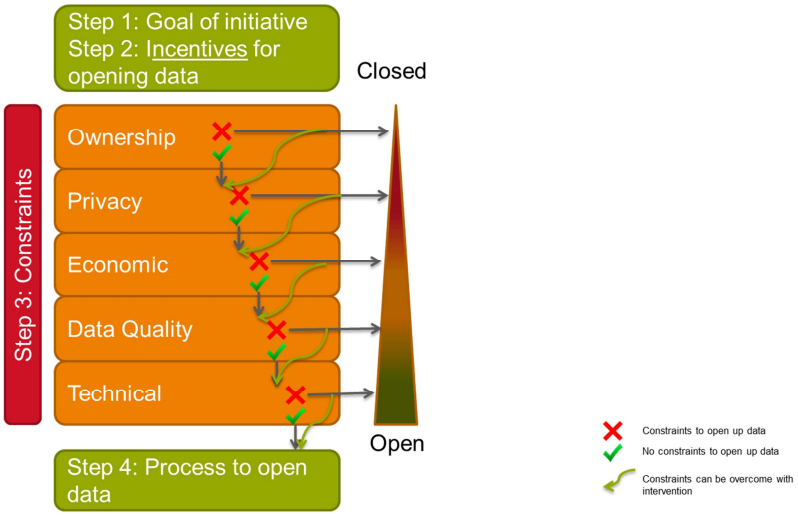

Figure 1. ***Decision model for opening up data***

The decision model works as follows. If a certain constraint to data sharing is present in a given situation, the next step is to analyze if the constraint can be overcome by an intervention. For example, when a privacy constraint occurs, anonymization by filtering or aggregation by combining a data set into a single record, are potential interventions. Interventions are usually of a technical nature, but also include organizational mechanisms. When no suitable intervention can be identified the data set cannot be shared. Hence the arrow shown on the right-hand side of figure 1. The next sections describe the five categories for analysis in more detail and introduce, where possible, interventions. The decision model should be applied both on a data set level as well as on individual data properties and even data values of a data set (see for instance [28]). It should be noted that the decision model that is presented in this section, often serves as an example rather than a definite set of issues that needs to be addressed. While the categories remain more or less the same, for every use case new issues can be added to the categories.

4.2. Ownership

Only the owner of the data can make a decision about whether to publish data or not. If more than one organization owns data, all involved organizations need to agree on opening up the data. If one or more of the data owners are negative about opening up the data one possible intervention would be to start a process that informs all skeptical data owners about the advantages of open data. Improving trust is another important intervention needed at this point in the process. Involvement and support from higher management is essential for improving trust and for establishing a culture that is positive about open data. Licensing could be an intervention describing particular rules imposed by data owners. Licensing could be under similar conditions as given for open source [29]. Licensing mechanisms reflect data policies of data owners. Security mechanisms based on identity mechanisms could serve as an intervention to share data only with trusted organizations or individuals.

A particular aspect of data ownership refers also to culture within an organization. In many cases, the data manager of a particular data set is unwilling to provide data to others, since he has no control on how his data will be used. Clear data policies (institutional perspective, [4]) and decisions structures [14] are a means to cater with these barriers.

4.3. Privacy

If a data set contains classified or privacy sensitive information that can be traced back to individual persons or companies this will constrain the data owner to publish the data. Legal constraints related to the privacy of data can also present a constraint to publishing data. Possible interventions to overcome privacy constraints are anonymization by filtering of sensitive information and aggregation of data, thus, only publishing a selection of data properties and values. Another option is the deployment of access control mechanisms combined with identity management that regulates data access. This intervention limits the openness of the data and is therefore especially applicable for company data with access restrictions and less for open data in general. In case interventions are not carried out successfully the data cannot be opened up.

4.4. Economic

Several elements related to the economic and business value aspects of a data set may hinder publication. Often data owners do not have a clear view on which data to publish as they are unaware of the potential of the different data sets. Since data is also considered to have a large economic value [1], data owners also expect to make profit with their data. It is difficult for individual data owners to assess the value of data sharing for collaboration (see the logistics cases). In the case that a data owner currently earns money by providing his data, this will constrain the publication of the data – at least openly. The data owner could consider sharing the data only among a few organizations, and assess if there are other ways to earn money with the data when it is published (even among a few organizations). This means that the data set

may be opened up but is only available for a fee (Open data business model), which is one possible intervention. Monitoring, billing, and invoicing have to be supported as intervention mechanisms, potentially with different business models.

Economic sensitivity may constrain the data to be opened up. Economic sensitivity refers to the consequences for the business processes and even the business model of a company. Opening up commercial sensitive data, such as available transport capacity, may result in a decrease of the commercial rates offered by that company and thus its profitability. Yield management mechanisms need to be in place to ensure that companies do not experience negative financial consequences when opening up data but can benefit from improved insights in consumer behavior. Another constraint to open data may be that the actual costs of publishing the data might be too high for the data owner. Only if the business case for opening data is positive the data should actually be published. To address these economic constraints it could be considered, especially when multiple organizations collaborate in a network: i) to share the costs of opening up data, ii) to define a pricing structure for data set usage and iii) to set up a separate organization to govern and maintain the data and its usage. Each of these interventions requires monitoring functions of a data sharing platform.

4.5. Data Quality

A data provider is responsible for the quality of the data that is published. Poor data quality of a data set or a selection of data properties should prevent the data from being published. Data quality could also have impact on liability in case poor data leads to accidents, incidents or increased turn around times. A related issue is that data gathered in a specific context may not be useful in another context, even though the quality 'in itself' is good. If data quality is a constraint, one needs to dive into the data to analyze if the data quality can be improved on the constraining factors to still be able to open up the data. One example is to explore whether the data set can be extended with other data to improve completeness. If data quality is too low, or if the data is not applicable in any other situation, the data set should not be opened up. In any case metadata describing the data quality should be added [27], [14]. To allow for re-use of data by others, as much context information as possible about data should be provided. Furthermore, social interaction with the data should be supported: data is often most used and most easy to interpret when a community can be built around the data platform where the data is published. Data visualization can play an important role in this.

4.6. Technical

The technical format of a data set may be a constraint to open the data [30]. If the data is unstructured it may be difficult to convert it into a machine-readable format relevant to a data user. The size of the data set, the existence of a semantic model, and identifiers are other technical issues that need to be considered. For this group of issues, many interventions are possible. Examples include: offering the data in a structured format; reusing existing vocabularies and ontologies; publish the data

according to existing data standards. Most technical interventions should be accompanied by an economic intervention.

5. Discussion

Based on barriers that organizations can apply on individual data sets, data properties, and data values, the decision model presented in the previous section provides a number of interventions. The decision model can be applied on individual data sets in the context of goals like economic growth or improved decisions (section 1) and incentives like a business case for an individual organization [4]. In the decision model, we have structured the issues in five categories that reflect a data owner's perspective, which encompasses the perspectives for open data [4]. As literature of open data considers only government data, the decision model for data sharing also considers decision structures [14], commercial sensitivity, and potential business models for data sharing amongst competitors derived from logistics use cases. The latter also refers to the institutional barrier of the inability to process data user demands, whereas in the private sector data is shared to the benefit of both a data provider and – user. Based on governance literature and the use cases, private and commercial sensitivity has been included in the decision model.

Different incentives may apply to individual organizations. On a high level, we found that a distinction can be made between legal and economic incentives. The legal incentives are centered around any regulatory measures that can be taken by the government to stimulate open data. Examples are compliance to safety or environmental regulations or compliance to directives regarding open data. The economic category includes incentives that lead to economic gains for the organizations that publish their data, such as efficiency gains, enlarging their customer base, or creating a competitive advantage. Reciprocity is another incentive as one organization is often more willing to publish data if this is matched by another organization opening their data. If neither economic nor legal incentives to publish data can be identified the process to open the data set is usually aborted as it is unlikely that a positive business case can be identified for data publication.

Analysis of data sets is not only to be done on the level of a data set, but particular data properties and data values need to be considered. It implies that data analysis requires a detailed knowledge of semantic models of a data set, including vocabularies. Analysis on a high level of detail is necessary as intervention mechanisms also have to deal with that level of detail, for instance by implementing access control at the level data properties and – values. Interventions that can be applied are for instance attribute based access control, but these require particular tool support for managing access control rules combined with identity management of partners. Although this complexity is not required in open (government) data, it will be required by the private sector and thus potentially impact public-private collaboration leading for instance to economic growth [1].

The decision model provides a number of intervention mechanisms. One intervention mechanism could potentially be used to address different barriers and an intervention mechanism can be viewed in the context of another mechanism. Security

interventions can for instance be applied in the context of a data policy. An extensive mapping of interventions to barriers in different contexts needs more attention in future research.

6. Conclusion

This paper presents a decision model for sharing publicly or privately owned data based on both a literature review and three logistics use cases. The decision model contains details addressing individual barriers in each of the five categories identified. Since we have taken use cases for data sharing in the private sector, other barriers than the ones identified for open data are introduced in our decision model, e.g. commercial sensitivity. We have also introduced a number of interventions, whereas the implementation of these interventions in for instance a data sharing platform still needs to be elaborated. Our objective is to create Web based tools supporting the decision model, in which a data owner can select particular intervention mechanisms that are supported by, for instance, a data platform. Such tools need to cater with various details of data set analysis.

Data governance literature emphasizes the importance of a decision structure [14] for data sharing in the private sector, since goals and incentives need to be clear to all stakeholders. Goals can be formulated at a macro level, e.g. (inter)national or regional, and incentives need to be formulated at organizational level, e.g. by a clear business case or a data sharing strategy supported by management. These goals and incentives are drivers for analyzing data sets as formulated by our proposed decision model. Lack of a clear decision structure, an implicit decision structure, or lack of goals and incentives for individual organizations might be a barrier to the uptake of economic growth and innovation. Organizing data use for instance from a situational awareness perspective [7] could be the basis to stimulate data sharing.

A final finding is that organizations can apply the decision model, without making a distinction to whether the data will be publically available without any restrictions (open data), or whether data will be shared in a closed community. The model introduces interventions like restricted access based on authorization and access control rules. Thus, the model can be a basis for data driven innovation for open and big data in the context of goals and incentives.

References

[1] European Commission, "Digital agenda: Turning government data into gold," European Commission, Brussels, 2011.

[2] J. Manyika et al, "Big data: the next frontier for innovatioin, competition, and productivity," McKinsey&Company, 2011.

[3] K. Janssen, "The influence of the PSI directive on open government data: An overview of recent developments," *Government Information Quarterly*, vol. 28, no. 4, pp. 446-456, 2011.

[4] P. Jaeger and J. Bertot, "Transparancy and technological change: ensuring qeual and sustained

public access to government information," *Government Information Quarterly*, vol. 27, no. 4, pp. 371-376, 2010.

[5] T. Harrison, T. Pardo and M. Cook, "Creating Open Government Ecosystems: a research and development agenda," *Future Internet*, vol. 4, no. 4, pp. 900-928, 2012.

[6] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems," *Human Factors: the journal of the human factors and ergonomics society*, vol. 37, no. 1, pp. 32-64, 1995.

[7] A. Zuiderwijk, N. Helbig, J. Gil-Garcia and M. Janssen, "Guest Editors' Introduction. Innovation through open data: a review of the state-of-the-art and an emerging research agenda," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 9, no. 2, 2014.

[8] T. Jetzek, M. Avital and N. Bjørn-Andersen, "Generating Value from Open Government Data," in *The 34th International Conference on Information Systems. ICIS 2013*, cc, 2013.

[9] M. Janssen, Y. Charalabidis and A. Zuiderwijk, "Benefits, Adoption Barriers and Myths of Open data and Open government," *Information Systems Management*, vol. 29, no. 4, pp. 258-268, 2012.

[10] H. Klein and D. Myers, "A set of principles for conducting and evaluating interpretitive field studies in information systems," *MIS Quarterly*, vol. 23, no. 1, pp. 67-93, 1999.

[11] G. Walsham, "Doing interpretive research," *European Journal on Information Systems*, vol. 15, no. 3, pp. 320-330, 2006.

[12] J. Mingers, "Combining IS research methods: towards a pluralist methodology," *Information System Research*, vol. 12, no. 3, pp. 240-259, 2001.

[13] M. Janssen and A. Zuiderwijk, "Open data and transformational government," in *tGov conference*, London, 2012.

[14] E. Barry and F. Bannister, "Barriers to open data release: a view from the top," in *2013 EGPA Annuaul Conference*, Edinburgh, 2013.

[15] P. Weill and J. Ross, IT Governance: how top performers manage IT decisions rights for superior results, Boston, MA: Harvard Business School Press, 2004.

[16] K. Weber, B. Otto and H. Osterle, "One size does not fit all - a contigency approach to data governance," *Journal of Data and Information Quality (JDIQ)*, vol. 1, no. 1, p. 4, 2009.

[17] C. Batini and M. Scannapieco, Data Quality: concepts, methodology, and techniques, Heidelberg: Springer-Verlag, 2006.

[18] S. Knight and J. Burn, "Developing a framework for assessing information quality on the World Wide Web," *Informing Science*, pp. 159-172, 2005.

[19] P. Nousak and R. Phelps, "A scorecard approach to improving data quality," 1 1 2002. [Online]. Available: http://www2.sas.com/proceedings/sugi27/p158-27.pdf. [Accessed 14 3 2014].

[20] McDonnell, "Big Data Challenges and Opportunities," 2011. [Online]. Available: http://spotfire.tibco.com/blog/?p=6793.

[21] C. Bizer et al, "Linked Data - The Story So Far," 2011.

[22] C. Batini and M. Scannapieco, Data quality: concepts, methodologies, and techniques, Heidelberg: Springer-Verlag, 2006.

[23] United Nations, "Rotterdam Rules," 2008. [Online]. Available: http://www.uncitral.org/pdf/english/texts/transport/rotterdam_rules. [Accessed 2012].

[24] S. Dalmolen, E. Cornelisse, A. Stoter, W. Hofman, H. Bastiaansen, M. Punter and F. Knoors, "Improving sustainability throuhg intelligent cargo and adaptive decision making," in *e-Freight 2012*, Delft, 2012.

[25] J. Esmeijer, T. Bakker and S. d. Munck, "Thriving and surviving in a data-driven society," TNO, Delft, 2013.

[26] W. Hofman and H. Bastiaansen, "A global IT infrastructure improving container security by data completion," in *ECITL*, Zaragoza, Spain, 2013.

[27] P. Miller, R. Styles and T. Heath, "Open data commons, a license for open data," in *LODW2008*, Beijing, 2008.

[28] T. Berners-Lee, "Linked Data - four rules," 18 June 2009. [Online]. Available: www.w3.org.DesignIssues/LinkedData.