

Design Principles for Improving the Process of Publishing Open data

Anneke Zuiderwijk, a.m.g.zuiderwijk-vaneijk@tudelft.nl, Delft University of Technology, Faculty of Technology, Policy and Management, Delft, The Netherlands

Marijn Janssen, m.f.w.h.a.janssen@tudelft.nl, Delft University of Technology, Faculty of Technology, Policy and Management, Delft, The Netherlands

Sunil Choenni, r.choenni@minvenj.nl, Research and Documentation Centre – Ministry of Security and Justice, Den Haag, The Netherlands; r.choenni@hr.nl, Rotterdam University of Applied Sciences – Creating O10, Rotterdam, The Netherlands

Ronald Meijer, r.f.meijer@minvenj.nl, Research and Documentation Centre – Ministry of Security and Justice, Den Haag, The Netherlands

Abstract

- Purpose: Governments create large amounts of data. However, the publication of open data is often cumbersome and there are no standard procedures and processes for opening data. This blocks the easy publication of government data. The purpose of this paper is to derive design principles for improving the open data publishing process of public organizations.
- Design/methodology/approach: Action Design Research (ADR) was employed to derive design principles. The literature was used as a foundation, and discussion sessions with civil servants were used to evaluate the usefulness of the principles.
- Findings: Barriers preventing easy and low-cost publication of open data were identified and connected to design principles, which can be used to guide the design of an open data publishing process. Five new principles are 1) start thinking about the opening of data at the beginning of the process, 2) develop guidelines, especially about privacy and policy sensitivity of data, 3) provide decision support by integrating insight in the activities of other actors involved in the publishing process, 4) make data publication an integral, well-defined and standardized part of daily procedures and routines, 5) monitor how the published data are reused.
- Research limitations/implications: The principles are derived using ADR in a single case. A next step can be to investigate multiple comparative case studies and detail the principles further. We recommend using these principles to develop a reference architecture.
- Practical implications: The design principles can be used by public organizations to improve their open data publishing processes. The design principles are derived from practice and discussed with practitioners. The discussions showed that the principles could improve the publication process.
- Social implications: Decreasing the barriers for publishing open government data could result in the publication of more open data. These open data can then be used and stimulate various public values, such as transparency, accountability, innovation, economic growth and informed decision and policy-making.
- Originality/value: Publishing data by public organizations is a complex and ill-understood activity. The lack of suitable business processes and the unclear division of responsibilities blocks publication of open data. This paper contributes to the literature by presenting design principles which can be used to improve the open data publishing process of public sector organizations.

Keywords: open data, e-government, publishing process, principles, business process reengineering, action design research

Acknowledgement: This research is related to the ENGAGE FP7 Infrastructure Project (An Infrastructure for Open, Linked Governmental Data Provision Towards Research Communities and Citizens), see www.engage-project.eu and www.engagedata.eu. The authors would like to thank their colleagues of the ENGAGE project for their input for this research although the views expressed in this paper are the views of the authors and not necessarily of the project.

1. Introduction

Governmental organizations produce considerable amounts of data while performing their common tasks and activities. They can decide to publish these data on the internet, and we then refer to these data as open data. Open data have received considerable attention recently. Many studies claim the potential of open data to create considerable advantages (for example, Bertot and Jaeger, 2006, Janssen, 2011, McDermott, 2010, Zhang et al., 2005). After data have been published by public organizations, these open data can be reused and the reuse of open data is expected to increase transparency, innovation, economic growth and to improve informed decision and policy-making (for instance, Janssen, 2011, Napoli and Karaganis, 2010, Neuron et al., 2013). Open data are also praised for their power to increase participation, interaction, self-empowerment and social inclusion of users of open data (e.g. citizens) and providers of open data (Bertot et al., 2010, European Commission, 2012, Janssen, 2011, Klischewski, 2012, Meijer, 2007). In this way, an open government can use information technologies to generate a participatory, collaborative dialogue between policymakers and citizens (Evans and Campos, 2013, Maier-Rabler and Huber, 2011).

Even though opening government data has the potential to create many advantages, opening data is a complex and ill-understood activity. Ideally datasets are published by default, however, there are many serious considerations influencing the decision to publish requiring well-defined processes and procedures before opening data (Zuiderwijk et al., 2012a). For instance, public organizations see the threat of privacy violations by releasing data and of being legally liable when opened data are misused (Kalidien et al., 2010, Kulk and Loenen, 2012). Accidentally publicizing policy or privacy sensitive data could damage the reputation of an organization. Furthermore, the publication of data may be restricted because the data are owned or created by different organizations, so that the interests of all these organizations should be taken into account (Faerman et al., 2001). Moreover, changes to different organizational layers are required to be able to systematically publish open data (van_Veenstra and van_den_Broek, 2013). There is a lack of insight in how the publishing process could be modified to improve it (Zuiderwijk and Janssen, 2013). Public values such as privacy, trust and security are important (Rana et al., 2013), but they complicate publishing data (Meijer et al., 2013). The process in which open data are created, collected and published, the so-called open data publishing process, needs to take these aspects into account, but it is unclear how this should be done.

Because of these complexities, governments find it difficult to open their data. They do not have appropriate processes and procedures for opening the data which they have collected or created (Zuiderwijk et al., 2012b). Often the publication of data is not an integral part of the data collection or creation process and it is usually viewed as a separate activity which is not integrated in the daily procedures and routines (Zuiderwijk and Janssen, 2013). Although some helpful guidelines for releasing

government data were published in the past (for instance, Pollard, 2011, Kalampokis et al., 2011), these cannot be used to improve the open data publishing process of public organizations. None of these provides detailed guidelines, foundations or principles for the design of the *open data publishing process*, whereas this is vital for publishing open data. Insufficient attention is given to the problem of *how* data can be published and fuelled in open data portals. Moreover, the literature often neglects principles that could guide process design (Aier et al., 2011).

These barriers are also present in the Netherlands. In the last decades, a number of letters were sent to the Dutch Lower Chamber in which several ministers emphasized the importance of an open government and open government data (Tweede_Kamer_1998-1999, 1998, Tweede_Kamer_1999-2000, 2000, Tweede_Kamer_2010-2011, 2011, Verhagen, 2011), showing the attention of Dutch politicians for open data. It was stated that each public body should open up as much data about the preparation and execution of policy-making as possible, in an accessible and understandable way (Tweede_Kamer, 2011). This should stimulate the use of these data and result in valuable new insights. In addition to these letters, the Netherlands has endorsed the declaration of the Open Government Partnership in 2012, which means that the Netherlands has embraced a high-level Open Government Declaration and delivered an Open Government Action Plan (Rijksoverheid, 2013). The declaration also includes the commitment to increase the availability of information about governmental activities (Open_Government_Partnership, 2011). Despite all these initiatives to stimulate the publication of open data, various other countries world-wide seem to be more advanced in publishing open data than the Netherlands (Nugroho, 2013). The various complexities and barriers of publishing open data are still present in many Dutch public organisations.

Because of these impediments the open data publishing process needs to be redesigned and improved. The aim of this paper is to derive process design principles for designing and improving the open data publishing process of public organizations. Action Design Research (ADR) is performed at a public organization in the Netherlands. The approach for attaining this objective is described into more detail in the following section. Subsequently, a brief overview is provided of the literature about business process design principles. Then the results of the performed ADR are presented by showing the existing challenges in the current publishing process and by showing the principles which could improve the process of publishing open data. This section also reports on discussions with employees of the public organization to see whether the proposed design and the design principles can be applied and whether additional design principles are needed. Finally, the paper ends with conclusions about the applicability of the proposed design principles.

2. Research Approach

The aim of this paper is to derive principles for designing and improving the administrative processes used for publishing open data. The ADR perspective was used to achieve this objective (see Sein et al., 2011). ADR deals with two challenges: “(1) addressing a problem situation encountered in a specific organizational setting by intervening and evaluating; and (2) constructing and evaluating an IT artefact that addresses the class of problems typified by the encountered situation” (Sein et al., 2011). More specifically, knowledge is generated by 1) formulating the problem (i.e. the challenges in the publishing process), 2) building, intervening and evaluating by applying organizational intervention, 3) reflecting and learning, and 4) formalizing of learning. We adopted the ADR-approach, as it assists in making theoretical contributions as well as in solving current and anticipated problems of practitioners and it

“reflects the premise that IT artefacts are ensembles shaped by the organizational context during development and use” (Sein et al., 2011, p. 37).

Figure 1 shows the ADR approach of this research. First, the literature was examined to obtain background information about the open data publishing process and about the key propositions of design principles. This was done by searching for journal and conference papers, books, reports and other information in various databases, including Science Direct, Scopus, ACM Digital Library and Google Scholar. During this search, combinations of the terms open data, open government data, publishing process, principles, business process reengineering and re-design were used. The relevance of the hits was determined by the search machines and by scanning the titles and abstracts of the documents. Most of the obtained documents appeared not to be useful because they did not relate to the open data publishing process or business process reengineering. Finally, 57 publications were selected that were relevant for this research.

The background information that was obtained with this literature review formed the basis for the performance of the ADR. ADR was used to analyse the current publishing process and to create an overview of the shortcomings and problems in the publishing process. Subsequently, the design principles from the literature were used for improving the open data publishing process in the ADR. This resulted in the refinement of some principles, removing some non-relevant principles, and in the identification of new principles. Finally, the design and the identified design principles were discussed with employees of the public organization to see whether it is feasible to apply the principles and whether additional design principles are needed.

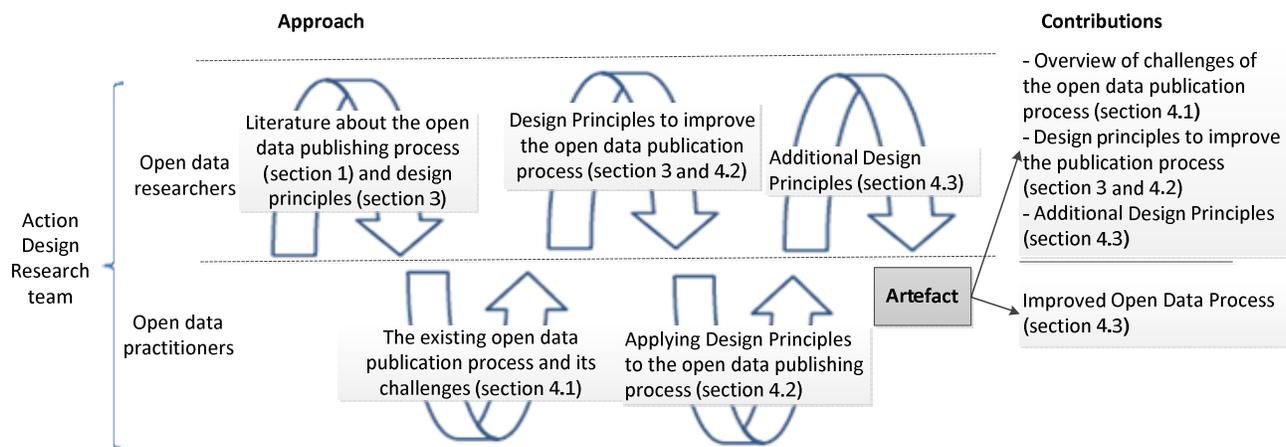


Figure 1: Action Design Research (ADR) approach (adapted from Sein et al., 2011).

ADR is performed at the Dutch Research and Documentation Centre (Wetenschappelijk Onderzoek en Documentatiecentrum; WODC). The WODC is a criminal justice knowledge centre that is part of the Dutch Ministry of Security and Justice. The main tasks of the WODC are as follows.

- Conducting research and letting other organisations conduct research, including the evaluation of policy and policy programs;
- Advising about intended policy and policy programs;
- Developing and maintaining data and making data accessible;
- Disseminating knowledge that is available within the organization;

- Documenting scientific publications in the field of Security and Justice (Staatscourant, 2011, nr. 22848).

In the context of the tasks and activities that are described above, the WODC systematically collects stores and enhances data. The WODC has been selected for this study because the WODC already gives access to data and publishes data on the internet for many years. The WODC aims to improve the publishing process to cope with the challenges of open data such as privacy protection and prevention of misuse and misinterpretation. The organisation aims to give access to data and publish data in a more systematic way, to better achieve the advantages of publishing open data, such as internal and external transparency, knowledge dissemination, reuse of data and better archiving of datasets.

The WODC does not only provide access to data by publishing them on the internet, but also by providing researchers with data and information on their request. The WODC was already providing access to individual researchers for many years. In this paper we focus on the active publication of data. Although information requests are part of the WODC's knowledge dissemination, just like open data does, we do not consider information requests as open data, as these two ways of providing data access require different processes.

The researchers who performed research at the WODC were part of the development team and participated in the discussions, which is typical for an ADR approach. During the discussions, which took place approximately every other week between November 2011 and July 2013, the current publishing process was discussed, as well as the possibilities to improve this process by using the design principles derived from the literature. The publishing process was revised on the basis of the discussions and evaluated during the discussions. The implementation of the improved publishing process is not realized yet. As such the evaluation of the implemented artefact is outside the scope of this study.

3. Business Process Design Principles

The use of principles is common when designing artefacts (for example, Bharosa et al., 2013, van_Bommel et al., 2006, Richardson et al., 1990). Principles can be defined as “general rules and guidelines, intended to be enduring and seldom amended, that inform and support the way in which an organization sets about fulfilling its mission” (TOGAF, 2011, <http://pubs.opengroup.org/architecture/togaf9-doc/arch/index.html>). Principles can be used as generic prescriptions for the design and implementation of information systems (Housel et al., 1986, Markus et al., 2002).

Principle-based design has a long history (Richardson et al., 1990) and can be used in various areas, including Business Process Reengineering (BPR) (Hammer, 1990). Hammer and Champy (1993) defined ‘reengineering’ as “the fundamental rethinking and radical redesign of business processes to achieve dramatic improvements in critical, contemporary measures of performance, such as cost, quality, service, and speed” (p. 46). Although there is no consensus as to what constitutes BPR, there is general agreement that Information Technologies (IT) are powerful enablers and can result in a radical improvement by causing people to rethink business process, which in turn can lead to organizational changes in which a new strategic vision is central, rather than technology (Hammer & Champy, 1993). This radical approach has often been criticized, as this often resulted in high failure rates of BPR projects (Weerakkody et al., 2011). Transformation processes are complex, as governments have to deal with organizational and technical issues (Janssen and Klievink, 2010). In public organizations radical reinvention is fraught with difficulty, as traditional public organizations are characterized by bureaucracy

(Weerakkody et al., 2011). For this reason, Weerakkody et al. (2011) show that “e-government-induced change requires a plan for a radical improvement which, in contrast to BPR, is obtained by incremental steps and has a high level of participation” (p. 320). This change should be embedded in the existing processes.

BPR is aimed at redesigning business processes, whereas principles are often used for designing system architectures. In the private sector, the BPR movement has shown mixed results for transforming processes, while the academic and research community have barely used BPR to examine the way that transformation could be achieved in the public sector (Weerakkody et al., 2011). Nevertheless, although the lessons learned from applying BPR in the private sector should not be copied on a one-to-one basis to the public sector, research has shown that the lessons learned from applying BPR in the private sector could be used in achieving transformational change in the public sector (Jurisch et al., 2012, Weerakkody et al., 2011, Weerakkody and Dhillon, 2008). The key propositions of BPR could be realised by putting a number of principles into practice, derived from Hammer (1990). The first principle emphasizes to organize around outcomes and focus on the final objective instead of focussing on single tasks. One person oversees and coordinates the whole process. Second, those who use the output of the process perform the process. Third, information-processing work should be subsumed into the real work that produces the information. Fourth, geographically dispersed resources should be treated as though they were centralized, and fifth, parallel activities should be linked instead of integrating their results. Sixth, put the decision point where the work is performed, and build control in the process, and, finally, capture information once and at the source (Hammer, 1990). These principles are often not followed in business processes, but may be used to solve some of the challenges that the open data publishing process deals with. For instance, the principle that information should be captured once and at the source could be applied by redesigning the open data publishing process in such a way that the person who created the data assesses whether the data can be published on the internet and provides all the right information to make the data publishable, such as the appropriate metadata.

In their work published in 1993, Hammer and Champy added the principle that reengineered processes must be kept simple to meet requirements of quality, service, flexibility and low costs. As a consequence, reengineering is often characterized by the integration of several jobs into one, while workers make decisions themselves that used to be made by their managers. Additionally, the work is performed where it makes most sense, which could also be in another organisation. Furthermore, the reengineered processes have multiple versions to satisfy the diverse market and user demands and the processes are performed simultaneously rather than in a linear sequence. Moreover, checks and controls are reduced, reconciliation is minimized by reducing the number of external contact points that a process has, and centralized and decentralized operations are hybridized. Finally, one case manager is assigned to provide a single point of contact. The case manager has an overview of the complex processes and communicates with the customer (Hammer and Champy, 1993).

Some of these principles may also be applied to the open data publishing process. For example, Hammer and Champy (1993) demonstrate that the reengineered process should have multiple versions to satisfy market and user demands. Applied to the open data publishing process, one could argue that different types of data, such as privacy-sensitive, policy-sensitive and non-sensitive data require different approaches and decisions. As a consequence, various types of data may require multiple versions of processes. In the following section we examine whether BPR can be used to deal with the challenges of the open data publishing process.

4. Action Design Research

Section 3 described how principles for the publication of open data can be developed. In the following sections, we present the steps of the ADR approach. First the challenges for publishing open data that the principles need to deal with are derived from practice (section 4.1). Subsequently, the derived principles from the literature are applied to practice (section 4.2). Then, in the final ADR step we derive additional design principles from the steps that were taken before (section 4.3).

4.1 Analysis of the existing open data publication process: The challenges

The publishing process that the WODC used until the end of 2012 only limitedly resulted in the aimed advantages, such as knowledge dissemination and reuse of data. The discussions with employees of the WODC provided the following reasons for this.

1. Employees were informed about the rules for publishing data when they started working at the WODC (Wetenschappelijk_Onderzoek_en_Documentatie_Centrum_(WODC), 2013). However, many employees only read the document containing these rules once and did not think about them, did not remember them or for some other reason did not take them into account when they were performing their research and producing data. As a result most employees started thinking about data publication only after the data had already been collected and the final dataset had been created. As a consequence, the individual researchers had used and stored the data in such a way that they were useful for their purposes, but this may not have been the best way to support data reuse. For instance, the researcher may not have added much metadata (i.e. data about the data (Jeffery, 2000)) to a dataset, because he or she knows these metadata by heart. If the data were published and someone wants to reuse them, this potential data user may not have any knowledge about these metadata and may not be able to interpret and use the data in an adequate manner because of this lack of metadata (Zuiderwijk et al., 2012c). If the researcher still wanted the dataset to be reusable after the research, additional effort was required to prepare and process the datasets to enable publication and data reuse.
2. In the old process, a few times per year an employee of the statistical division sent an e-mail to the managers of each department asking whether they could provide the statistical division with datasets that could be published. The division managers could then consult the employees of each division to see which datasets were appropriate for publication. As only very few guidelines were provided to researchers about how datasets can be prepared for publication, many researchers were uncertain about whether their datasets were appropriate for publication. Preparing a dataset for publication is very complex, as there are many issues concerning privacy sensitivity of data and the fear of misinterpretation of data (Conradie and Choenni, 2012, Zuiderwijk et al., 2012b). This challenge is widely acknowledged in international research (Zhang et al., 2005, Dawes, 1996, Kulk and Loenen, 2012, Sayogo and Pardo, 2012, van_Veenstra and van_den_Broek, 2013). It was also argued that the privacy act does not provide sufficient guiding principles for opening up data. This resulted in differences and inconsistencies between the ways that datasets were constructed by researchers or, for many researchers, it became automatism not to open the data because of the risk avoiding norms. Changing this automatism is very difficult. In addition, employees of the WODC expressed that publishing data could be risky when decisions about

opening or not opening data are based only on random single datasets, without taking a broader framework into account. As a consequence, only a few advocates of publishing data sent their datasets to the statistics division to consider publication.

3. Although the statistics division aimed to keep track of datasets that could be published, this division did not have a complete overview of all the research that was finished and all the datasets that were produced with this research. In the old process, the statistics division did not have much insight into how and why employees decided whether a dataset was appropriate for publication. Furthermore, the statistics division was not informed about when a certain dataset was under construction, was being prepared for publication or was waiting for further processing. For this reason the statistics division did not know how they could support data publication and which additional guidelines were needed for data preparation and publication. In the old data publication process there was a lack of insight in the activities performed by other actors involved in the process, as there was no single manager who had a complete overview of the publishing process and the datasets that were produced within this process. This challenge is explained into detail by Zuiderwijk and Janssen (2013).
4. The document that informed employees about the rules for publishing data (Wetenschappelijk_Onderzoek_en_Documentatie_Centrum_(WODC), 2013) also described rules for archiving research. Nevertheless, the divisions worked in different ways, as datasets were saved and registered at different places in different ways. Some divisions stored the datasets themselves in addition to the storage by the statistics division. Some divisions already checked whether datasets could be published in addition to the check that the statistics division performed.
5. The open data process was not organized around outcomes, but on the single task to publish datasets on the internet. The publication process did not consider how WODC datasets could be reused and what the WODC could learn from this reuse. The process did not take into account how the WODC could profit from the wisdom of the crowd. This challenge was also found by Pollock (2011). Public organizations that are publishing data should make a distinction between the provision of data on the one hand and the interaction that this provision aims to result in on the other hand (Robinson et al., 2009). Van Veenstra and Van den Broek (2013) also write that organisations are more internally focused when they just start to publish open data, although their focus may shift towards factors related to the reuse of open data later on.

To summarize, the data that were created were not systematically checked and prepared for publication because of several problems. The main problems concerned privacy sensitivity and the fear of misuse and misinterpretation. As a result of these problems, only a low number of datasets has been published. Between 1982 and 2000 the WODC has published 21 datasets. In 2001, the Dutch Data Protection Act (Wet Bescherming Persoonsgegevens; WBP) was introduced, which aims to guarantee citizens the right to privacy protection (Winter et al., 2008). In connection with this new act and an increase in attention for privacy protection in society, the WODC changed its open data policy in 2007. No WODC-datasets have been released between 2000 and 2008. From 2008 up to and including 2012, data that are considered by WODC-researchers to be qualified for public opening have partly been collected and stored in a so-called digital 'research data safe'. Over these 5 years, 54 datasets have been stored in the safe. Seven of these datasets have been published.

4.2 Applying Design Principles to the Process of Publishing Open Data

The previous section reported on the challenges for the publication of open data in our ADR. In this section we relate these challenges to the design principles of Hammer (1990) and Hammer and Champy (1993), to examine whether they could contribute to the improvement of the publication of open data. Five challenges were identified, namely: 1) late involvement, 2) lack of guidelines for publishing data (e.g. concerning privacy-sensitivity of data), 3) lack of insight in the activities of other actors involved in the publishing process, 4) different approaches and 5) lack of focus on outcomes. These challenges and the design principles from the literature were discussed with employees of the WODC. The discussions showed that some of the BPR principles could be used to improve the publishing process of the Research and Documentation Centre. A number of BPR principles could be translated to the WODC on a one-to-one basis, such as the principle to organize around outcomes and the principle to subsume information-processing work into the real work that produces the information. Table 1 provides an overview of the extent to which the BPR principles can be applied to the open data publishing process of the WODC. The table shows both the principles mentioned by Hammer (1990) and the principles mentioned by Hammer and Champy (1993). As the work done by Hammer and Champy (1993) elaborates on the work done by Hammer (1990), there may be some overlap between the principles that were developed in these publications. For instance, principle two of Hammer and Champy (1993) partly seems to resemble the sixth principle of Hammer (1990). Nevertheless, the business processes derived from both Hammer (1990) and Hammer and Champy (1993) are presented in this paper, as there are also several differences between the principles. Table 1 uses the common structure for describing principles, including their explanation, rationale and implications (TOGAF, 2011).

Table 1: BPR principles to improve the publication of open data (based on Hammer, 1990, Hammer and Champy, 1993)

	BPR Principle	Explanation and rationale	Implications
Hammer, 1990	1. Organize around outcomes, not tasks.	If the open data publishing process would be organised around outcomes rather than tasks, all tasks would be aimed at the final objective of sending datasets to the archiving organization and publishing them.	Following this principle can help creating datasets in such a way that they may be more useful for external data reuse than they would be without outcome orientation.
	2. Have those who use the output of the process perform the process.	This principle is not applicable, because the output of the process is used by external organizations and persons. It is not possible to let open data users conduct the open data publishing process.	-
	3. Subsume information-processing work into the real work that produces the information.	This principle indicates that the employee who produced the dataset should also prepare the dataset's publication, as this employee has most knowledge about how the data were produced and thus about how they can be processed.	This principle can help in creating a more efficient open data publishing process.
	4. Treat geographically dispersed resources as though they were centralized.	One of the most important resources in the open data publishing process is the data. Datasets can be treated as though they were centralized by using a central system in which projects and decisions on forthcoming datasets are registered.	Using a centralized registration system for datasets can help in maintaining a clear overview of the decisions taken for each created and published dataset.
	5. Link parallel activities instead of integrating their results.	There are a few parallel activities in the open data publishing process, such as the simultaneous creation of datasets and the simultaneous preparation for the publication of datasets. These parallel activities can be linked by using a central registration system for all datasets.	This principle can help in creating a more efficient open data publishing process. Furthermore, using a centralized registration system for datasets can help in maintaining a clear overview of the decisions taken for each created and published dataset.
	6. Put the decision point where the work is performed, and build control in the process.	This principle indicates that the employees who create the dataset decide whether the dataset could be published and how it should be prepared for publication, as the employee who created the dataset has most knowledge about how the data were produced and thus about how they can be processed. Control can be built in by having one coordinator who monitors the progress of preparing and publishing the dataset several times in the process. Moreover, each division could have one data manager who oversees the process for his/her division and who corresponds with the coordinator.	This principle can help in creating a more efficient open data publishing process, as it helps in giving each employee a clear role in the open data publishing process.
	7. Capture information once and at the source.	Following this principle, the coordinator should capture the data once and at the individual researcher(s) who collected the data (source).	This principle can help in creating a more efficient open data publishing process.

	BPR Principle	Explanation and rationale	Implications
Hammer and Champy, 1993	1. Several jobs are integrated into one.	This principle is not applicable, because the aim of the changes in the publishing process of the WODC is not to reduce the number of jobs.	-
	2. Workers make decisions themselves that used to be made by their managers.	This principle indicates that the employee who produced the dataset should also make decisions about data preparation and data publication, as he/she has most knowledge about how the data were produced and thus about how they can be processed.	This principle can help in creating a more efficient open data publishing process.
	3. The process steps are performed in a natural order: the reengineered processes should be performed simultaneously rather than in a linear sequence.	There are a few activities in the open data publishing process that can be performed simultaneously, such as the simultaneous creation of datasets and the simultaneous preparation of the publication of datasets.	This principle can help in creating a more efficient open data publishing process.
	4. Processes have multiple versions, so that organizations are able to satisfy the diverse and changing market and user demands.	Multiple versions of processes can be required for different types of data, such as policy-sensitive data, privacy-sensitive data and non-sensitive data, or when data are collected by an external organisation instead of the WODC itself.	This principle can help to deal with market and user demands, but also to comply with legislation (e.g. privacy-legislation).
	5. The work is performed in the part of the organization where it makes most sense, which could be across organization boundaries.	This principle indicates that the employee who produced the dataset should also prepare its publication, as this employee has most knowledge about how the data were produced and thus about how they can be processed.	This principle can help in creating a more efficient open data publishing process.
	6. Checks and controls are reduced and should only be used to the extent that they are economically important.	In line with this principle, checks and controls need to be reduced in the open data publishing process, and should only be used to the extent that they are economically important.	This principle can help in creating a more efficient open data publishing process, but may have the risk that wrongful data (e.g. privacy-sensitive data) are published, as these might not be checked sufficiently.
	7. Reconciliation is minimized by reducing the number of external contact points that a process has.	Each sub process of the open data publishing process should have as few external contact points as possible.	This principle can help in creating a more efficient open data publishing process.
	8. One case manager is assigned to provide a single point of contact.	This principle indicates that the employee who produced the dataset is assigned as a case manager. This employee provides a single point of contact to the coordinator of the open data publishing process.	This principle can help in creating a more efficient open data publishing process.
	9. The centralized and decentralized operations are hybridized and prevalent in the reengineered processes.	Following this principle, the centralised and decentralised operations in the open data publishing process are hybridized and both prevalent. If appropriate, decisions are made on a central level; otherwise they are made on a decentralised level.	This principle can help in creating a more efficient open data publishing process.

Table 1 confirms that the publishing process of the WODC could be reengineered according to Hammer's and Champy's BPR principles, as many of these principles could have a positive effect on the open data publishing process, for instance, by creating more efficiency. Yet, from the discussions with WODC-employees it became clear that not all challenges are addressed by the BPR principles. Challenges two and four are not addressed by the BPR principles. Furthermore, the principles that can be applied do not totally solve the challenges. Additional principles are needed to solve the challenges that the publishing process has to deal with. There is a need for principles which are specifically focused on open data and which could be used additionally to the BPR principles. The application of the BPR principles to the open data publishing process shows the need for more specifically applied principles to improve this process. The following section addresses this issue.

4.3 Towards New Principles to Improve the Process of Publishing Open Data

The results of applying BPR principles to the open data publishing process showed that additional principles are needed specifically for the open data publishing process. Brainstorm sessions were held with employees of the WODC to find out which additional principles could be developed. These brainstorm sessions resulted in a list of five new principles (see Table 2). The principles refer to the whole open data publishing process, while the BPR principles are more focused on the level of single tasks.

Table 2: New principles to improve the publication of open data.

Challenge	New Principle	Explanation and rationale	Implications
1. Late involvement	Start thinking about the opening of data at the beginning of the process	Most employees were involved late in the publication process, as they started thinking about data publication only after the data had already been collected and the final dataset had been created. If employees are involved early in the publication process, they are expected to consider the publication of the data while creating and collecting datasets. Employees regularly need to be pointed at their early involvement in publishing open data.	Following this principle can help creating datasets in such a way that they may be more useful for external data reuse than they would be without early involvement.
2. Lack of guidelines for publishing open data	Develop guidelines, especially about privacy and policy sensitivity of data	Only few guidelines concerning privacy and policy sensitivity, decisions support and defined processes and procedures were provided to employees about how datasets can be prepared for publication. As a consequence, many employees were uncertain about whether their datasets were appropriate for publication and how they should publish them, which resulted in limited publication of datasets. Providing guidelines could help employees with assessing under which conditions datasets can be published. Guidelines are needed especially with regard to privacy sensitiveness of the data and avoiding the risk of misinterpretation of data.	Guidelines can help in 1) assuring that all employees are familiar with the way the publishing process should be performed and 2) assessing whether certain datasets are appropriate for publication.
3. Lack of insight in activities of other actors	Provide decision support by building in insight in the activities of other actors involved in the publishing process	The lack of insight in the activities that other actors within the organization perform in the publishing process has led to a lack of clarity about how employees decide which datasets are appropriate for publication and about which barriers they face. There was no overview of all the research that was performed and finished and which datasets were produced.	Insight and control can help in clarifying the responsibilities of different employees involved in the process and clarifying difficulties and problems that these employees face.
4. Different approaches	Make data publication an integral, well-defined and standardized part of daily procedures and routines	Each division worked in a different way. Standardizing the publishing process could help in making it more transparent, efficient, effective, and reliable and make people become more familiar with how the process could be performed. If the publication process is performed in the same way by different departments, the employees know what they can count on and how they could take this into account in performing their other activities.	Following this principle may help in performing the publishing process in a standardized way in different divisions, so that each employee knows what he or she can count on.
5. Lack of focus on outcomes (e.g. data use)	Monitor how the published data are reused	The old publication process did not include mechanisms to find out how WODC datasets were used and what the WODC could learn from this reuse. Feedback loops can help in learning more about the way that datasets are used and about what can be learned from this (Zuidervijk and Janssen, 2013). E-government initiatives like open data could then have an impact on, for instance, policy-making (Srivastava, 2011)	This principle could help in making use of the wisdom of the crowd. Knowledge about data reuse could also give more motivation to publish reusable data in the future.

The ADR-approach of our Dutch case resulted in the development of five new principles, namely 1) start thinking about the opening of data at the beginning of the process, 2) develop guidelines, especially about privacy and policy sensitivity of data, 3) provide decision support by building in insight in the activities of other actors involved in the publishing process, 4) make data publication an integral, well-defined and standardized part of daily procedures and routines and 5) monitor how the published data are reused. Discussions with the employees of the WODC showed that these five principles could improve the open data publishing process.

Figure 2 shows the effect of the new principles by showing both the old and the new publishing process. As there were differences in how the WODC departments conducted the old publishing process, Figure 2 shows how the old process was usually performed, although there were some exceptions. The five new design principles are represented in the figure with numbers in the redesigned process. The figure reveals that in the redesigned publishing process, the coordinator gets involved earlier than in the old publishing process, in this way dealing with the first challenge of the publishing process of the WODC. The coordinator provides the individual researchers with guidelines for publishing data and with a metadata template, in this way dealing with the second and the fourth challenge. Subsequently, more insight in the activities of other actors is created by using a central registration system (challenge three). This registration system is explained more into detail below. Finally, challenge five is addressed by investigating how feedback can be received on the published data.

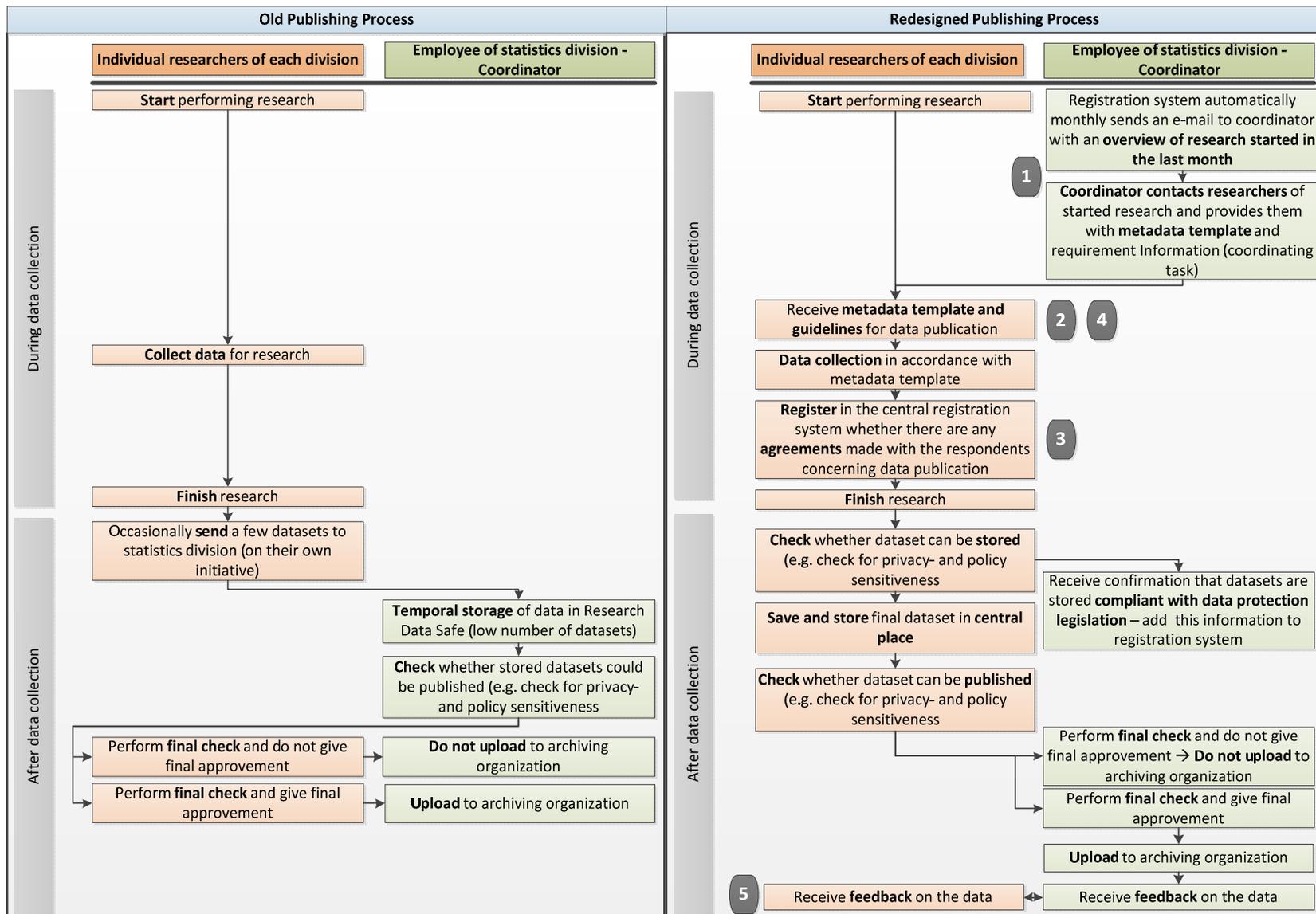


Figure 2: A comparison of the old publishing process and the redesigned publishing process, showing the effect of the new design principles.

Although the new principles were developed in a Dutch case, we believe that at least some of these principles can also be applied to open data publishing processes in other countries, as other countries partly deal with the same challenges for publishing data. Examples of the way in which these principles could help in solving issues and improving the publishing process are discussed below.

First, the principles of early involvement (1), guidelines (2), building insight (3) and standardization (4) could profit from ICT support in which employees register projects and decisions on datasets. Projects are already registered before the datasets are created, which improves early involvement. The registration system could point at the steps that need to be taken to publish open data, and in this way this system could provide guidelines. The registration system could keep track of the status of datasets and of all the decisions that need to be made before it can be published. Various employees can be authorised to get insight in the system, which could improve building insight in each other's activities in the publishing process. As the system points at the same process steps for publishing data, the system could help in improving standardization. The principles can also be used to develop a reference architecture for open data publication, which could address authentication, preservation of data, updating of the information quality, ensuring privacy and support transfer of ownership. It could exist of a repository of experiences, components, tasks and services for designing the open data publishing process.

Second, although this is not in line with the BPR principles which give an important role to teams rather than managers (Hammer and Champy, 1993), the discussions showed that insight in the activities of other actors involved in the publishing process may also be improved by appointing one employee as the manager of the publishing process. Moreover, one contact could be appointed for each division to communicate with the manager. These measures could help in obtaining a more complete overview of the publishing process and the datasets that are produced within this process. The manager monitors the progress of preparing and publishing the dataset regularly in the process.

Third, guidelines can help in clearly defining 'open data' and ensuring that all employees use this definition consistently. In addition, guidelines could help in describing how the publishing process could be performed, who is responsible for which tasks, which data are privacy sensitive and policy sensitive and who is available for dataset specific questions. Finally, experts could perform a final privacy-sensitivity and misinterpretation check on the data. These could be external experts, such as employees working for an archiving organization.

Fourth, the principles that are proposed by the BPR movement refer to a radical reinvention of corporations (Hammer and Champy, 1993). In line with Choenni, Bakker and Baets (2003), the discussions with WODC-employees showed that the successful transformation of processes of public organizations should be implemented gradually. The implementation should start small and in the beginning principles should be applied case by case. Radical redesign of the open data publishing process at public organizations is not expected to be successful.

5. Conclusions

The publication of data by public organizations is a complex and ill-understood activity, which suffers from many impediments, such as the threat of privacy violation and of being legally liable when opened data are misused. Because of these impediments the open data publishing process needs to be improved. The aim of this paper was to derive process design principles for improving the open data publishing

process of public organizations. Action Design Research (ADR) was performed at a public organization in the Netherlands.

Business process design principles were derived from the literature and linked to the open data publishing process of a public organization. Five main challenges were identified in the open data publishing process, namely 1) late involvement, 2) lack of guidelines for publishing data (e.g. concerning privacy-sensitivity of data), 3) lack of insight in the activities of other actors involved in the publishing process, 4) different approaches and 5) lack of focus on outcomes. Several of these challenges were also identified in open data research in other countries, in this way demonstrating the importance of this research and its applicability to a wider context. The linkage of Business Process Reengineering (BPR) principles to the challenges of the open data publishing process showed that although the publishing process of the WODC could be reengineered according to the BPR principles, additional principles are needed to solve the challenges that the publishing process has to deal with. There is a need for principles which are specifically focused on open data and which could be used additionally to the BPR principles. Five additional principles were developed, namely:

- 1) Start thinking about the opening of data at the beginning of the process;
- 2) Develop guidelines, especially about privacy and policy sensitivity of data;
- 3) Provide decision support by building in insight in the activities of other actors involved in the publishing process;
- 4) Make data publication an integral, well-defined and standardized part of daily procedures and routines;
- 5) Monitor how the published data are reused.

Although these principles have not been fully implemented yet, discussions with employees of the WODC indicated that these five principles could improve the open data publishing process. Examples of the way that these principles could help in solving issues were discussed. For instance, the principles can be used to develop a reference architecture for open data publishing and the principles could profit from ICT support in which employees register projects and decisions on datasets.

ADR appeared to be a useful approach, as it assisted both in making theoretical contributions as well as in solving current and anticipated problems of practitioners. In line with the ADR approach, this paper contributes to the literature in four ways, 1) BPR design principles were applied to the open data publication process for improvement (research contribution), 2) additional design principles were developed, specifically appropriate for the open data publishing process (research contribution), 3) the open data publishing process was redesigned (practical contribution) and 4) the redesigned open data process could decrease the barriers for publishing open government data and could result in the publication and use of more open data, in this way stimulating various public values, such as transparency, accountability, innovation, economic growth and informed decision and policy-making (social contribution). Further research should focus on applying the design principles to other cases, which will likely result in further refinement of the principles.

References

- AIER, S., FISCHER, C. & WINTER, R. 2011. Construction and Evaluation of a Meta-Model for Enterprise Architecture Design Principles. *10th International Conference on Wirtschaftsinformatik*. Zurich, Switzerland.
- BERTOT, J. C. & JAEGER, P. T. 2006. User-centered e-government: Challenges and benefits for government Web sites. *Government Information Quarterly*, 23, 163–168.

- 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, 264–271.
- BHAROSA, N., JANSSEN, M. & BAJNATH, S. 2013. Deriving Principles for Guiding Service Encounters: A Participative Design Research Approach. *International Journal of Information Systems in the Service Sector*, 5, 1-16.
- CHOENNI, S., BAKKER, R. & BAETS, W. 2003. On the Evaluation of Workflow Systems in Business Processes. *Electronic Journal on Information Systems Evaluation (EJISE)*, 6, 33-44.
- CONRADIE, P. & CHOENNI, S. Exploring process barriers to release public sector information in local government. 6th international conference on theory and practice of electronic governance (ICEGOV), October 22-25, 2013 2012 Albany, New York, United States of America.
- DAWES, S. 1996. Interagency information sharing: Expected benefits, manageable risks. *Journal of Policy Analysis and Management*, 15, 377-394.
- EUROPEAN_COMMISSION. 2012. *Communication from the commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Towards better access to scientific information: Boosting the benefits of public investments in research* [Online]. Brussels: European Commission. [Accessed October 6 2013].
- EVANS, A. M. & CAMPOS, A. 2013. Open Government Initiatives: Challenges of Citizen Participation. *Journal of Policy Analysis and Management*, 32, 172-185.
- FAERMAN, S. R., MCCAFFREY, D. P. & SLYKE, D. M. V. 2001. Understanding interorganizational cooperation: public-private collaboration in regulating financial market innovation. *Organization Science*, 12, 372-388.
- HAMMER, M. 1990. Reengineering work: don't automate, obliterate. *Harvard Business Review*, 68, 104-112.
- HAMMER, M. & CHAMPY, J. 1993. *Reengineering the corporation. A manifesto for business revolution*, New York, HarperBusiness.
- HOUSEL, T. J., SAWY, O. A. E. & DONOVAN, P. F. 1986. Information Systems for Crisis Management: Lessons from Southern California Edison. *MIS Quarterly*, 10, 389-400.
- JANSSEN, K. 2011. The influence of the PSI directive on open government data: An overview of recent developments. *Government Information Quarterly*, 28, 446-456.
- JANSSEN, M. & KLIEVINK, B. 2010. Gaming and simulation for transforming and reengineering government: Towards a research agenda. *Transforming Government: People, Process and Policy*, 4, 132-137.
- JEFFERY, K. G. 2000. Metadata: The future of information systems. In: BRINKKEMPER, J., LINDENCRONA, E. & SØLVBERG, A. (eds.) *Information Systems Engineering: State of the art and research themes*. London: Springer Verlag.
- JURISCH, M., IKAS, C., PALKA, W., WOLF, P. & KREMAR, H. 2012. A review of success factors and challenges of public sector BPR implementations. *45th Hawaii International Conference on System Sciences (HICSS)*. Hawaii: IEEE Computer Society.
- KALAMPOKIS, E., TAMBOURIS, E. & TARABANIS, K. 2011. Open government data: A stage model. In: JANSSEN, M., SCHOLL, H. J., WIMMER, M. A. & TAN, Y.-H. (eds.) *The 10th Conference on Electronic Government (EGOV 2011)*. Delft, the Netherlands: LNCS.
- KALIDIEN, S., CHOENNI, R. & MEIJER, R. F. 2010. Crime statistics online: potentials and challenges. In: CHUN, S. A., SANDOVAL, R. & PHILPOT, A. (eds.) *Proceedings of the 11th Annual International Digital Government Research Conference on Public Administration Online: Challenges and Opportunities*. Puebla, Mexico: Digital Government Society of North America.
- KLISCHEWSKI, R. 2012. Identifying Informational Needs for Open Government: The case of Egypt. *45th Hawaii International Conference on System Sciences*. Grand Wailea, Maui, Hawaii: IEEE Computer Society.
- KULK, S. & LOENEN, B. V. 2012. Brave New Open Data World? *International Journal of Spatial Data Infrastructures Research*, 7, 196-206.
- MAIER-RABLER, U. & HUBER, S. 2011. „Open“: the changing relations between citizens, public administration and political authority. *eJournal of eDemocracy & Open Government (JeDEM)*, 3, 48-58.
- MARKUS, M. L., MAJCHRZAK, A. & GASSER, L. 2002. A Design Theory for Systems that Support Emergent Knowledge Processes. *MIS Quarterly*, 26, 179-212.
- MCDERMOTT, P. 2010. Building open government. *Government Information Quarterly*, 27, 401-413.
- MEIJER, A. 2007. Publishing public performance results on the Internet. Do stakeholders use the Internet to hold Dutch public service organizations to account? *Government Information Quarterly*, 24, 165–185.

- MEIJER, R., CHOENNI, S., SHEIKH_ALIBAKS, R. & CONRADIE, P. 2013. Bridging the Contradictions of Open Data. In: CASTELNOVO, W. & FERRARI, E. (eds.) *13th European Conference on eGovernment*. University of Insubria, Como, Italy: Academic Conferences and Publishing International Limited, UK.
- NAPOLI, P. M. & KARAGANIS, J. 2010. On making public policy with publicly available data: The case of U.S. communications policy making. *Government Information Quarterly*, 27, 384-391.
- NEURONI, A. C., RIEDL, R. & BRUGGER, J. 2013. Swiss Executive Authorities on Open Government Data – Policy Making beyond Transparency and Participation. *46th Hawaii International Conference on System Sciences*. Grand Wailea, Maui, Hawaii: IEEE Computer Society.
- NUGROHO, R. P. 2013. A comparison of open data policies in different countries. Lessons learned for an open data policy in Indonesia. *Master thesis, Delft University of Technology, The Netherlands*. Delft.
- OPEN_GOVERNMENT_PARTNERSHIP. 2011. *Open Government Declaration* [Online]. Available: <http://www.opengovpartnership.org/open-government-declaration> [Accessed October 4 2013].
- POLLARD, P. 2011. Opening up government data: Making the case. European Public Sector Information Platform.
- POLLOCK, R. 2011. *Building the (Open) Data Ecosystem* [Online]. Available: <http://blog.okfn.org/2011/03/31/building-the-open-data-ecosystem/> [Accessed September 13 2013].
- RANA, N. P., DWIVEDI, Y. K. & WILLIAMS, M. D. 2013. Analysing challenges, barriers and CSF of egov adoption. *Transforming Government: People, Process and Policy*, 7, 177-198.
- RICHARDSON, G. L., JACKSON, B. M. & DICKSON, G. W. 1990. A Principles-Based Enterprise Architecture: Lessons From Texaco and Star Enterprise. *MIS Quarterly*, 14, 385-403.
- RIJKSOVERHEID. 2013. *Actieplan Open Government Partnership* [Online]. Data.overheid.nl. Available: <https://data.overheid.nl/actieplan-open-overheid> [Accessed October 4 2013].
- ROBINSON, D. G., YU, H., ZELLER, W. P. & FELTEN, E. W. 2009. Government Data and the Invisible Hand. *Yale Journal of Law & Technology*, 11, 160-175.
- SAYOGO, D. S. & PARDO, T. A. Exploring the Motive for Data Publication in Open Data Initiative: Linking Intention to Action. 45th Hawaii International Conference on System Sciences, 2012 Hawaii
- SEIN, M. K., HENFRIDSSON, O., PURAO, S., ROSSI, M. & LINDGREN, R. 2011. Action Design Research. *MIS Quarterly*, 35, 37-56.
- SRIVASTAVA, S. C. 2011. Is e-government providing the promised returns?: A value framework for assessing e-government impact. *Transforming Government: People, Process and Policy*, 5, 107-113.
- STAATSCOURANT 2011. Organisatieregeling_Ministerie_van_Veiligheid_en_Justitie.
- TOGAF 2011. Architecture Principles. In: THE_OPEN_GROUP (ed.) *TOGAF Version 9.1*.
- TWEEDE_KAMER 2011. Toepassing van de Wet openbaarheid van bestuur.
- TWEEDE_KAMER_1998-1999 1998. Actieprogramma Elektronische Overheid. .
- TWEEDE_KAMER_1999-2000 2000. Actieprogramma Elektronische Overheid.
- TWEEDE_KAMER_2010-2011 2011. Hergebruik en Open Data: naar betere vindbaarheid en herbruikbaarheid van overheidsinformatie.
- VAN_BOMMEL, P., HOPPENBROUWERS, S. J. B. A., PROPER, H. A. & VAN_DER_WEIDE, T. P. Giving Meaning to Enterprise Architectures; Architecture Principles with ORM and ORC. In: HUTCHISON, D., KANADE, T., KITTLER, J., KLEINBERG, J. M., MATTERN, F. & MITCHELL, J. C., eds. *International Conference of On the Move to Meaningful Internet Systems*, 2006. 1139-1147.
- VAN_VEENSTRA, A. F. & VAN_DEN_BROEK, T. A. Opening Moves – Drivers, Enablers and Barriers of Open Data in a Semi-public Organization. In: WIMMER, M. A., JANSSEN, M. & SCHOLL, H. J., eds. *Electronic Government 2013*, 2013 Koblenz, Germany. *Lecture Notes in Computer Science (LNCS)*, 50–61.
- VERHAGEN, M. 2011. *De digitale agenda voor slimme ondernemers. Toespraak minister van Economische Zaken, Landbouw en Innovatie, Maxime Verhagen, bij de presentatie van de Digitale Agenda op 17 mei 2011 in Den Haag* [Online]. Available: <http://www.rijksoverheid.nl/documenten-en-publicaties/toespraken/2011/05/17/de-digitale-agenda-voor-slimme-ondernemers.html> [Accessed November 18 2011].
- WEERAKKODY, V. & DHILLON, G. 2008. Moving from E-Government to T-Government: A Study of Process Reengineering Challenges in a UK Local Authority Context. *International Journal of Electronic Government Research*, 4, 1-16.
- WEERAKKODY, V., JANSSEN, M. & DWIVEDI, Y. 2011. Transformational change and business process reengineering (BPR): Lessons from the British and Dutch public sector. *Government Information Quarterly*, 28, 320-328.

- WETENSCHAPPELIJK_ONDERZOEK_EN_DOCUMENTATIE_CENTRUM_(WODC) 2013.
Informatiebeveiliging en gegevensbeheersplan WODC. Den Haag.
- WINTER, H. B., JONG, P. O. D., SIBMA, A., VISSER, F. W., HERWEIJER, M., KLINGENBERG, A. M. & PRAKKEN, H. 2008. *Wat niet weet, wat niet deert. Een evaluatieonderzoek naar de werking van de Wet bescherming persoonsgegevens in de praktijk*, Groningen, Pro Facto, Rijksuniversiteit Groningen, De Jong Beleidsadvies, WODC.
- ZHANG, J., DAWES, S. & SARKIS, J. 2005. Exploring stakeholders' expectations of the benefits and barriers of e-government knowledge sharing. *Journal of Enterprise Information Management*, 18 548-567.
- ZUIDERWIJK, A. & JANSSEN, M. 2013. A coordination theory perspective to improve the use of open data in policy-making. In: WIMMER, M. A., JANSSEN, M. & SCHOLL, H. J. (eds.) *12th conference on Electronic Government (EGOV)*. Koblenz, Germany: IFIP International Federation for Information Processing 2013.
- ZUIDERWIJK, A., JANSSEN, M., CHOENNI, S., MEIJER, R. & SHEIKH_ALIBAKS, R. 2012a. Socio-technical impediments of open data. *Electronic Journal of eGovernment*, 10, 156 - 172.
- ZUIDERWIJK, A., JANSSEN, M., MEIJER, R., CHOENNI, S., CHARALABIDIS, Y. & JEFFERY, K. 2012b. Issues and guiding principles for opening governmental judicial research data. In: SCHOLL, H. J., JANSSEN, M. F. W. H. A., WIMMER, M. A., MOE, C. E. & FLAK, L. S. (eds.) *11th conference on Electronic Government*. Kristiansand, Norway: IFIP International Federation for Information Processing 2012.
- ZUIDERWIJK, A., JEFFERY, K. & JANSSEN, M. 2012c. The potential of metadata for linked open data and its value for users and publishers. *Journal of e-Democracy and Open Government*, 4, 222-244.