MANAGING THE PRIVACY RISKS OF OPEN DATA

How do municipalities manage the privacy risks when publishing open government data?

By

Eva Peeters Weem

S1878131

Submitted in partial fulfillment of the requirements for the degree of Master of Science, program Public Administration, University of Twente

03-07-2018

Supervisors:

Dr. L.C.P. Broos, Privacy law

Prof. Dr. M.A. Heldeweg LL.M, Law governance & technology

UNIVERSITEIT TWENTE.

Table of content

Summary

Both the national and local government have committed to publishing more public records as open data as part of a broader open government strategy. Open data is expected to have appositive effect on transparency and democracy as a whole. On the other hand, there might be privacy risks connected to the publication of open data. Municipalities have large amounts of public records that could potentially be released open data. This thesis is focussed on the policies and procedures that Dutch municipalities use when they publish open data. These policies and procedures are described and compared to prevalent risk management methodology. The main research question is formulated as follows:

How have Dutch Municipalities, that are noteworthy regarding the publication of open data, designed and implemented comprehensive open data policies and procedures to protect citizen's privacy when they publish open data? And to what extent does this design and implantation integrate prevalent risk management framework methodology?

The research was based on desk research, document analysis and interviews. Firstly, it needed to be determined how many municipalities have been publishing open data and which are noteworthy. This has been be determined through quantitative desk research. Secondly, document analysis was conducted on full scope of documents that constructed the open data policies and procedures. Thirdly, the implemented and applied risk management procedures have been compared to prevalent risk management methodology by analysing the data from the interviews combined with desk research.

A proper open data policy is in all three cases part of a larger data management policy. The municipal organization needs data management: data is used to develop and substantiate public policy and employees wanting to share data in a sustainable way (Gemeente Utrecht, 2014a; Gemeente Haarlem, 2017a; Gemeente Den Haag, 2011). Procedures for publishing open data are aimed at removing personal data from datasets. The municipalities are aware of the privacy risks of open data publication. However, the risks are not formally identified and qualified. Unlawful publication of personal data, in other words personal data breaches, is one risk that is formally identified and qualified by two out of three participating municipalities. The risk of re-identification through combination of datasets is known, however, this risk is difficult to qualify. It is difficult to determine the likelihood, impact and to what extent municipalities are responsible in these situations (Gemeente Utrecht, personal communication, December 13, 2017; Gemeente Utrecht, personal communication, January 15, 2018; Gemeente Den Haag, personal communication, January 31, 2018; Gemeente Haarlem, personal communication, February 7, 2018).

1 Introduction

This chapter will introduce the topic and research question of this thesis. Firstly, the broad topic of open data will be discussed. Secondly, the purpose of the research will be discussed. The purpose of the research is followed by the main research question and the sub questions. Lastly, the scientific, societal and economic relevance will be explained.

1.1 Managing the privacy risks of open data

1.1.1 Open government

The Dutch national government has adopted an active policy regarding digitalisation. The national government has formulated a national digital agenda, parallel to the digital agenda of the European Union and is a member of the global Open Government Partnership (OGP) (Ministerie Van Binnenlandse Zaken, 2015). An important part of the open government agenda is publication of open data. Open data is an important instrument to achieve the objectives of open government policies (Hardy & Maurushat, 2017). Open data is the active publication of data that is free for anybody to access, use, modify and share for any purpose (Viale Pereira, Macadar, Luciano, & Testa, 2016). Open data can be published by anyone. (Ministerie van Economische Zaken, 2016; Plasterk, 2015). The contributions of the Dutch national government towards open government are showing results. The Netherlands are internationally one of the frontrunners on open government development (Ministerie van Binnenlandse Zaken, 2015). The Dutch government cites the benefits of open data such astransparency, accountability and economic stimulation to advance with open government and open data initiatives (Ministerie van Binnenlandse Zaken, 2015).

The publication of open data, or open government data (OGD) when published by government sources, is expected to achieve transparency, empower citizens and increase accountability. Open data is an essential tool to make government ‘open’ and transparent (Janssen, Charalabidis & Zuiderwijk, 2012). However, there is also awareness of the risks and limitations to opening up data. There are privacy risks involved with the publication of open data. Separate datasets without personal data can be combined with other datasets to deanonymize and identify individuals in datasets (Ministerie van Economische Zaken, 2016).

1.1.2 Open data of municipalities

Municipalities collect considerable amounts of data about citizens and their direct living environment. Some of that data could possibly be published as open data at a later time (Ministerie van Economische zaken, 2016). The nationally formulated strategy for open government and open data publication leaves discretionary room for municipalities to structure the publication of open data. Municipalities are individually responsible to design and manage open data policies. Consequently, municipalities are also responsible for the management of the privacy risks that arise from the publication of open data (Ministerie van Binnenlandse zaken, 2015).

The Dutch government mainly focusses on the economic and societal benefits of open data (Ministerie van Economische Zaken, 2016). However, there are ways in which open data can be used in an unlawful way or undesirable way. Datasets that separately do not contain personal identifiable information, can be combined to unmask personal data of citizens and results in the infringement of privacy of citizens (Zuiderwijk, Janssen, Choenni & Meijer, 2014). The risk of re-identification or unmasking of personal data makes publication of open data a complex activity. Municipalities might not always fully recognize the complexity of open data publication (Zuiderwijk, Janssen, Choenni & Meijer, 2014). One of the main goals of open data is to improve transparency of the workings of government and thereby improving citizens trust in government. However, the privacy risks of open data might undermine the benefits of transparency (Meijer, Conradie & Choenni, 2014). Citizens need to trust local government to protect their privacy as municipalities collect their (sensitive) personal data (Meijer, Conradie & Choenni, 2014).

1.1.3 Managing the risks of open data

The processing of citizens personal data is regulated by national and EU level privacy law. The General data protection regulation (Gdpr) prohibits the publication of open data with identifiable personal data if there is no justification to do so (regulation (EU)2016/679, p. 36). Assuming that municipalities abide by these regulations and remove personal data from open data, there might still be privacy risks. One pseudonymized dataset might not pose privacy risks, however, combining multiple datasets may re-identify individuals in the dataset. The privacy risks may vary per dataset and on all the other available open data (Janssen, Charalabidis & Zuiderwijk, 2012).

Municipalities need to implement policies and procedures to manage the privacy risks in order to comply with the Gdpr (regulation (EU)2016/679, p.47; Meijer, Conradie & Choenni, 2014). The impact of privacy risks can be weighted and mitigated. However, this depends on policies and procedures the municipality put in place to identify, weigh and mitigate the risks that related to the publication of open data (Wieczorek-Kosmala, 2014). The Gdpr requires municipalities in their role as controller of personal data, to implement protocols to assess and mitigate the privacy risks when they process personal data (regulation (EU)2016/679, p. 47-48).

Municipalities can mitigate risks by formulating and structurally implementing procedures to set a context to-, assess- and mitigating measures. These measures together form a risk management framework (ISO, 2009). Risk management enables an organization to consistently and methodologically identify, assess and mitigate risks. The success of structural mitigation of privacy risks is dependent on the risk management of the municipality (Brooks, Garcia, Lefkovitz, Lightman, Nadeau, 2017).

1.1.4 Purpose of the research

This thesis focusses on the question how Dutch municipalities manage the privacy risks that are related to the publication of open data. The purpose of this thesis project is to examine how municipalities have formulated policies and procedures to protect the privacy of citizens when open data is published. This research project will help to determine to what extent privacy risks are identified, assessed and mitigate the privacy risks and to what extent these policies and procedures implement prevalent risk management methodology. Based on these finding recommendations can be made on how the privacy of citizens can be better protected. Protecting citizens privacy is necessary as infractions of citizens' privacy might undermine the benefits of open data and open government (Janssen, Charalabidis & Zuiderwijk, 2012). Aside from the privacy risks, municipalities can find themselves in violation of national and Gdpr when risk management procedures and policies are not sufficiently formulated and applied. This can result in legal liability or fines (regulation (EU)2016/679, p. 82).

1.2 Research questions

The central research question of this thesis is as follows:

How have Dutch Municipalities, that are noteworthy regarding the publication of open data, designed and implemented comprehensive open data policies and procedures to protect citizen's privacy when they publish open data? And to what extent does this design and implantation integrate prevalent risk management framework methodology?

In order to answer the main research question three sub questions have been formulated:

1. Which Dutch municipalities are noteworthy regarding the publication of open data?
2. How are comprehensive open data policies and procedures applied?
3. To what extent do comprehensive open data policies and procedures apply and implement prevalent risk management methodology?

1.3 Relevance of the research

1.3.1 Scientific relevance

This thesis project is scientifically relevant as it will contribute to the body of knowledge on the measures aimed at protecting the privacy of citizens in the context of open data. More specifically this will provide more insight in the quality of privacy risk management of municipalities on the topic of open data. This research will provide more knowledge on what measures need to be taken to successfully implement open data policies. Examining the implementation of risk management methodology is one of the most important indicators on how privacy risks are assessed and mitigated. Examining to what extent formulated policies and procedures implement prevalent risk management methodology, makes it is possible to assess the quality of open data policies and procedures.

1.3.2 Societal relevance

Open data has the potential to bring about transparency and citizen empowerment that strengthen democracy (Janssen, Charalabidis & Zuiderwijk, 2012; Viale Pereira, Macadar, Luciano, & Testa, 2016; Attard, Orlandi, Scerri & Auer, 2015; European Commission, 2011; Open Government Partnership, 2017; Rijksoverheid, 2017). However, the benefits of open data might be undermined when the publication of open data results in infringements of citizen’s privacy. Privacy infringements may lead to citizens losing trust in governments to protect their personal data. By mitigating the privacy risks the undermining of the societal benefits can be prevented (Janssen, Charalabidis & Zuiderwijk, 2012). Knowledge on mitigating privacy risks is relevant for the national government, municipalities, citizens and regulatory bodies in the field of privacy. The results of this research will provide more transparency on how municipalities protect citizen’s personal data. The achieved transparency will coincide with the objective of transparency, that is part of the broader open government policies (Ministerie van Economische Zaken, 2016).

1.3.3 Economic relevance

It is expected that open data will bring about economic benefits aside from the societal and scientific benefits. The European Commission (2010) estimates that the economic benefits may add up to €40 billion a year in the EU. Some other authors disagree with estimates, however, these authors agree that opening up data will stimulate different types of innovations and yield economic benefits (Kuk & Davis, 2011). Researching how municipalities protect the privacy of citizens will provide information on how municipalities can reduce the privacy risks related to the publication of open data. This knowledge will contribute to the successful implementation of open data policies. The publication of municipal open data will help with achieving the expected economic benefits (Kuk & Davis, 2011). However, there are also financial risks of open data in the form of fines or other legal procedures. For example, in the Netherlands processors of personal data can be fined in case of a personal data breaches. Fines or other legal procedures can impede the realization of expected benefits of open data (Autoriteit Persoonsgegevens, 2017b).

1.4 Conclusion

This chapter introduced the topic, the purpose and relevance of this research. Open data is a part of the Dutch national Digital Agenda. Open data is expected to be beneficial to democracy, the economy and be a positive influence for innovation. However, there are possible negative side-effects of open data, for example: accidentally exposing personal data of citizens. In order to ensure the expected benefits, open data publication needs to be managed properly (Janssen, Charalabidis & Zuiderwijk, 2012). Municipalities are responsible for open data publication on the local level (Ministerie van Binnenlandse zaken, 2015). This thesis will focus on how municipalities manage the publication of open data. The purpose of this research is to protect the privacy of citizens by evaluating if municipalities take sufficient steps to manage the risks regarding open data publication. (Janssen, Charalabidis & Zuiderwijk, 2012).

2 Theoretical framework

This chapter creates a context for the research in this thesis and consists of three parts. The first part is a literature review on the concept of open data. The literature review will include a definition of the concept of open data and open government data, the benefits and the possible risks of open data. The second part is the conceptual framework that further clarifies the concepts of privacy and risk management. The third part is the legal framework that summarizes EU-level legislation on personal data protection.

2.1 Literature review

2.1.1 Open data

Open data refers to data that is free for anybody to use, modify and share for any purpose (Viale Pereira, Macadar, Luciano, & Testa, 2016). Open data can refer to various types of data. Open data can be primary or secondary. Ideally it is primary data however, it is not always possible to publish primary data. Data can be in real-time, location-based, generic documentation, pictures, video, reports, maps and so forth (Alamgir Hossain, Dwivedi & Rana, 2016). The most comprehensive definition is based on ten principles formulated by the Sunlight Foundations (2010). In order for open data to be considered 'open' it needs to be: complete, primary, accessible, machine processable, non-discriminatory, non-proprietary, permanent, licence free and free of change.

The concept of open data in this thesis is specified to open government data (OGD). This is a sub type of open data that originates from the fulfilment of public tasks that adheres to the standard principles of open data. (Attard, Orlandi, Scerri & Auer, 2015). This thesis exclusively focusses on open government data published by Dutch municipalities. The Dutch Ministry of Internal Affairs has published data principles that state date open data must: be accessible unless otherwise decided, collected as part of public task, free, non-proprietary, accessible without registration, machine readable-, processable, include meta data, as close to the primary source as possible and findable (Ministerie van Binnenlandse Zaken, 2017). The definition by the Dutch Ministry of Internal Affairs takes into account that some primary data is not appropriate to publish as-is. Therefore, their definition of open data allows modified datasets to be considered open data (Ministerie van Binnenlandse Zaken, 2017). Data is considered readable if it is structured in rows and columns and published formats in CSV-, XML, or JSON- format. Other types of formats for example Pdf. are not readable (Algemene Rekenkamer, 2016).

The definition of open data makes it possible to differentiate open data from 'normal' public data that can be found online. A lot of normal public data is freely accessible to the public however, some public data can be more difficult to find and cannot be reused in the same way open data is supposed to be readable and reusable (Ruijer, 2017). Open data can also be referred to as proactive data. Proactive data is all data that is made public by a government without having to request it being released. Based on this definition, all open data is considered as proactive data. However, not all proactive data is open data as proactive data also includes press releases or other government documentation that does not meet all principles of open data (Ruijer, 2017).

2.1.2 Benefits of open data

Janssen, Charalabidis and Zuiderwijk (2012) have published the most comprehensive and structured summary of all potential benefits of open data as they specially set out to analyse all benefits and risks of open data. They identify a total of 31 expected benefits of open data clustered in three types of benefits: political & social, economic & operational and technical. At the top of their list and one of the most repeated benefits of open data and open government is more transparency and democratic accountability (Viale Pereira, Macadar, Luciano, & Testa, 2016; Attard, Orlandi, Scerri & Auer, 2015; European Commission, 2011; Open Government Partnership, 2017; Rijksoverheid,2017; Janssen, Charalabidis & Zuiderwijk, 2012; Weerakkody, Irani, Kapoor, Sivarajah & Dwivedi, 2016; Welle Donker & Van Loenen, 2016). Transparency and democratic accountability are part of the political and social benefits. This cluster of benefits includes, more participation of citizens, increase of trust in government, improving policy making, better and more equal access to government data and new and better services for citizens (Janssen, Charalabidis & Zuiderwijk, 2012).

The economic benefits are clustered in the second category. These benefits include a stimulation of competitiveness through better availability of information, stimulation of innovation, improvement of products and services, development of new product and services and making use of the intelligence of society (Janssen, Charalabidis & Zuiderwijk, 2012). The European Commission expected the yearly benefits to add up to €40 billion per year in the European Union (European Commission, 2010). However, other researchers expect that these expectations are overestimated and that possible benefits would be smaller (Kuk & Davis, 2011).

The third category are the operational and technical benefits of open data. These benefits include: the ability to easily reuse data, optimization of administrative processes, improvement of public policies, enabling of comparison during decision-making, easy access to data, creation of new data through combing of datasets, validation of data, better preservation of data and the integration of public and private data (Janssen, Charalabidis & Zuiderwijk, 2012).

2.1.3 Privacy risks of open data

Although most reports assume that the benefits of open data outweigh the risks it is important to identify the risks related to the publication of open data (Zuiderwijk &Janssen, 2014). Open data is expected to bring about benefits regarding public values such as transparency, trust, security and privacy. However, the risks related to open data may outweigh the benefits of open data. This might lead to contradicting results on the public values that initially promoted open data (Meijer, Conradie & Choenni, 2014). Privacy risks are the most important risks within the scope of this thesis.

Open data is by law generally prohibited from including data that directly identifies individuals. The first risk of open data is unlawful publication of personal data. Mistakes with properly filtering out personal data can be made. Such a mistake was made in New York City where one dataset included the personal email addresses of members of the New York City Commission on Women's Issues (Keenan, 2012). These situations of unintended publication are legally referred to as personal data breaches in the Regulation protecting personal data (Wbp) and the Gdpr. Personal data breaches can result in large fines for the controller (Autoriteit Persoonsgegevens, 2017b).

The second risk of open data is re-identification of individuals in datasets. Datasets should not be considered as isolated silo's. To extract (predictive) knowledge datasets are often combined and analysed using big data methods (Mantelero, 2017). Re-identification is the process identifying the subjects in the dataset. Re-identification is usually done by using subject patterns found in other (public) datasets. By combining information from multiple datasets individuals can be identified in datasets that separately do not identify individuals (Lavrenovs & Podins, 2016; Mantelero, 2017). The risk of re-identification is always present (Meijer, Conradie & Choenni, 2014). Re-identification is especially easy when individual patterns are known (Lavrenovs & Podins, 2016). There are multiple examples of re-identification. In the United Kingdom students were able to deanonymize data on re-offenders from the Ministry of Justice (Keenan, 2012). By using Internet Movie Database (IMDB) Narayanan and Shmatikov (2008) were able to uniquely identify 95% of the users in a 500,000-user's database published by Netflix.

The impact of privacy breaches is difficult to predict as this depends on how data is combined and how the unmasked data is used. This also makes it difficult to formulate a list of all variables that might pose privacy risks. Without a properly considerating the privacy risks can become significant. However, some broadly formulated negative effects can be identified (Zuiderwijk & Janssen, 2014). The risk of privacy breaches effects society as a whole. Privacy breaches might reduce the collective trust in public organizations. The privacy risks of open government data may have a contractionary effect on transparency as citizens may lose trust in their government to protect their data (Bargh, Choenni & Meijer, 2017; Meijer, Conradie & Choenni, 2014).

Privacy breaches may also have negative effects on the individuals. Individuals affected by privacy breaches may have their identity stolen, be publicly embarrassed, face discrimination, lose confidence in professional secrecy meant to protect their personal data, unauthorised re-identification, lose employment or lose business opportunities (Bargh, Choenni & Meijer, 2017; European Union, 2016). An example of how re-identification could reveal sensitive personal details on individuals took place in Riga. The city of Riga published open data containing ride registration from the city's public transportation. By identifying ride patterns, assumptions could be made on someone's religion, political opinions, sexual orientation or membership to a specific community (Lavrenovs & Podins, 2016).

On the other hand, removing privacy sensitive sections may undermine the usability of a specific open dataset (Meijer, Conradie & Choenni, 2014). With the focus on open data these two fundamental principles of a democracy come in conflict with each other. Open data will contribute to more transparency however, by releasing more open data governments can actually harm citizens' privacy. Both full transparency and perfect privacy do not exist, rather these concepts should be considered as relative concepts. Open data requires the continuous weighing of the principles of transparency to the principles of privacy (Janssen &Van den Hoven, 2015; Green et al, 2017).

2.1.4 General risks

There are more general risks to open data aside from the privacy risks. Public organizations tend to underestimate the complexity of the process of publishing open data. Zuiderwijk, Janssen, Choenni and Meijer (2014) identified five challenges that public organizations face when open data is published. The challenges in the publishing process may result in problems such as privacy violations, illegal publication or the misuse of open data. The five challenges are: late involvement, lack of guidelines or protocols for the publication of open data, lack of understanding of activities of other actors in the publication process, differing approaches between actors and lack of focus on the outcomes.

The first risk is unintended publication of inappropriate data. Not all data is appropriate to publish as open data due to privacy concerns, policy sensitive content, the level of security, ownership of data by multiple actors and compliance with different laws. Publishing inappropriate data might be unlawful, harm the reputation of the organization or lead to reduces trust in the organizations (Zuiderwijk, Janssen, 2014; Kucera & Chlapek, 2014). Open data might reveal trade secrets, security secrets or infrastructure details that could be misused to damage the publisher. Data that on its own could be harmless, could become a threat if it is combined with other datasets to cause damage to security or infrastructure (Kucera & Chlapek, 2014).

The second risk is biased data. The selection process that determines what data can be published may lead to publication of datasets with certain arguments or biases. Certain (sensitive) data is not always published due to a higher and possible harmful trade-off for the publisher (Zuiderwijk & Janssen, 2014). This risk refers to the publication of data that is not illegal to publish but might result in negative publicity or attitudes towards the publishers (Kucera & Chlapek, 2014). Either more narrow selection or more broad selection, might lead to the publishing of data harmful to the publishers (Zuiderwijk, Janssen, 2014; Kucera & Chlapek, 2014).

The third risk is publishing complex open data that is misinterpreted or misused. Opening up data to everyone also means opening up data to people who do not, or only partially, have the capabilities to properly use and interpret the data. This might lead to the wrong conclusions being drawn from the data. The misuse and misinterpretation may lead to incorrect information being spread and the reputation of the publisher being harmed. Misuse and misinterpretation may occur by accident or purposely (Zuiderwijk, Janssen, 2014; Kucera & Chlapek, 2014).

The fourth risk is related to data quality. Data quality refers both to the accuracy of the information and the usability of datasets. Publishers of open data regularly do not use systems that assess and manage the accuracy and usability of open data (Zuiderwijk & Janssen, 2014). The lack of data management might lead to datasets containing inaccurate data. Poor data management might lead to the publication of datasets that overlap and create an overload of data (Kucera & Chlapek, 2014).

All the identified general risks can result in a lack of transparency and trust in government. This will contradict the expected benefits of open data as a contributor to trust and transparency. Open data might also create a situation where there is too much information. An overload of information can hinder the use of available data. Users maybe are no longer able to find the right information thereby hindering transparency instead of creating it (Meijer, Conradie & Choenni, 2014). A higher quantity of available information does not necessarily improve the quality of use (Zuiderwijk & Janssen, 2014).

2.2 Conceptual framework

2.2.1 Privacy

Privacy risks are often acknowledged in studies on the effects of open data. However, a conceptualization of privacy is generally absent. The concept of privacy is often discussed however it can be difficult to define. Government employees who will be or are already confronted with open government policies are shown to have difficulties with describing the concept of privacy, personal data and classifying information as personal data (Badrul, Parslow, Lundqvist & Williams, 2016). The discussion about the concept of privacy usually concerns a distinction between a private and public sphere and how much control an individual has on the information gathered about them (Fuchs, 2011).

One classic definition of privacy is “The right to be left alone” (Warren and Brandeis, 1890, p. 193). The concept of privacy can refer to several definitions ranging from three to six according to the author (Fuchs, 2011). The six definitions by Solove (2008) are arguable most complete as they cover all definitions formulated by other authors. The six definitions are:

- The right to be left alone,
- Limiting the access to the self,
- Secrecy,
- Control over personal information,
- Personhood,
- Intimacy (Solove, 2002, p. 1092).

The conceptualization of privacy can be difficult as the definitions are arbitrary and might depend on the scope of a particular situation (Fuchs, 2011). The concept of privacy within the scope of this thesis is limited to the control over personal information. One of the main considerations of this regulation is that natural persons get proper protection of their personal data and that they should have the control over their own data (Regulation (EU)2016/679, p. 1). Control over personal data is contrary to the concept of open data. Open data is meant to be used and distributed with only a few limitations (Viale Pereira, Macadar, Luciano, & Testa, 2016).

2.2.2 Risk Management

2.2.2.1 General risk management

The privacy risks of re-identification or accidental publication of personal data cannot be fully eliminated (Zuiderwijk, Janssen, Choenni & Meijer, 2014). However, by implementing risk management policies and procedures that are proven to be effective, enable an organization to mitigate the risk of open data publication. Risk management (RM) refers to coordinated activities aimed at directing and controlling risks within an organization (ISO, 2009). There are differences between the RM of public and private organisations. Public organisations often are subject of political decisions, public interests and publicly funded. These differences produce different types of risks that are mainly non-financial risks or political risks, contrary to the private sector risks that are mainly financial (Oulasvirta & Anttiroiko, 2017).

Risk in the context of this thesis refers to "*an uncertain event or condition that, if it occurs, has either positive or negative effects on project objectives"* (Hillson & Simon, 2007; Project Management Institute, 2008). There are several different models for risk management. However, generally these models include a similar set of activities. The first activity is setting a context for risk management focussed on the strategic objectives of the organization. The second activity is to assess the risks that the organization might encounter during its operations. Risks are assessed on their likelihood and their impact. The third step is developing mitigating measures to the identified risks and implementing these measures. This process is ongoing to constantly improve risk management (Wieczorek-Kosmala, 2014). Open data needs continuous management to mitigate the risks (Simperl, O'Hara & Gomer, 2016) In order to continuously mitigate the risk, the publisher should implement a system of several steps:

4. Be aware and describe the data situation;
5. Know what data is in the datasets and scan for personal data and all other types of data that; might be (legally) restricted from open publication;
6. Understand how the datasets can be used. Publishers should consider in what ways the datasets could be combined and used to re-identify individuals and reveal personal data. Publishers should anticipate users combining data;
7. Understand the legal and governance issues that can be anticipated before publishing the data;
8. Be aware of consent and ethical issues that can be anticipated before publishing the data;
9. Have a proper risk management process that assesses the risks of publication;
10. Formulate a plan for what happens after the data is published;

11. Publishers need employ a system to manage published datasets. The publisher needs to be able to remove or redact datasets as part of control measures;
12. Publishers should employ a system through which users can provide feedback and express privacy concerns (Keenan, 2012; Lavrenovs & Podins, 2016; Simperl, O'Hara & Gomer, 2016).

Sufficiently applied risk management methodology will enable a municipality to process personal data in a repeatable and measurable way (Alashwal, Abdul-Rahman & Asef, 2017). Therefore, the extent to which risk management methodology is applied can be used as an indicator on the quality of the protection of citizens privacy by municipalities in the context of this thesis.

2.2.2.2 Enterprise risk management

A prevalent model for RM in enterprises is the Enterprise Risk Management (ERM) framework (Wieczorek-Kosmala, 2014). This framework is developed by the Committee of Sponsoring Organisations of the Treadway Commission (COSO). The ERM framework is an integrated framework designed for the private sector. However, COSO is of the opinion that the ERM framework is also applicable to other types of organisations (COSO, 2004). The ERM framework can best be described as a process and implemented as a strategy aimed at identifying and managing risks that might have a negative impact on the organisation and its operations. In total, there are several principles that form the ERM framework: an organisation wide process, by people on all levels of the organisation, applying as a strategy, identifying risks and managing these within the risk appetite, providing assurance to board of directors and achieving objectives of the organisation (COSO, 2004).

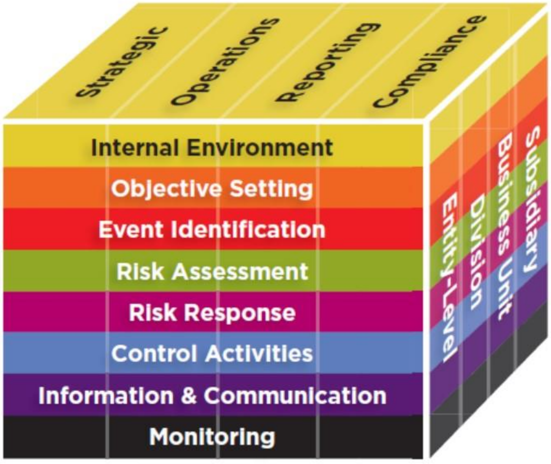


Figure 2,COSO ERM Framework ©, 2004

The three sides of the framework represent the concepts on which the ERM framework is build. The top side is dedicated to the achievement of goals on four levels: strategic, operational, reporting and compliance. The right side of the framework is dedicated to the entity levels on which the framework needs to be implemented in order to secure an organisation-wide strategy. The front side of the framework represents the eight components of the ERM framework. These components together produce a comprehensive RM strategy and include the concepts such as risk identification, response to risks and management of the risks (COSO, 2004).

The three steps of Wieczorek-Kosmala (2014) can also be identified in the COSO (2004) framework and the risk management measures by Simperl, O'Hara and Gomer (2016), Keenan, (2012) and Lavrenovs and Podins (2016). The COSO framework uses eight components that could be grouped within the three broader steps of context, assessment and mitigation. The COSO framework uses more detailed components to manage risks (COSO, 2004). The measures by Simperl, O'Hara and Gomer (2016), Keenan, (2012) and Lavrenovs and Podins (2016) add risk management measures that are similar to COSO (2004) more detailed components. For example, COSO's (2004) internal environment and objective setting matches with the steps described above 'awareness' and 'describing the data situation' (Keenan, 2012; Lavrenovs & Podins, 2016; Simperl, O'Hara & Gomer, 2016). However, these specific measures are more specific to the topic of open data compared to the more abstract ERM framework. These measures translate the more abstract COSO components to workable measures for publishers of open data (Keenan, 2012; Lavrenovs & Podins, 2016; Simperl, O'Hara & Gomer, 2016).

2.3 Legal framework

2.3.1 General data protection regulation

The legal scope of this thesis is limited to the Council Regulation (EU)2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data [2016] OJ L 119/1, the General data protection regulation (Gdpr). The Gdpr has replaced the Dutch Law protecting personal data (Wbp) and the previous Directive 95/46/EC as from May 25th of 2018 (Autoriteit Persoonsgegevens, 2017a). The main legal framework of this thesis is the new Gdpr.

The Gdpr is only applicable to open data publications that might contain personal data. Personal data is "any information concerning an identified or identifiable natural person" (regulation (EU) 2016/679, p. 33). These identifiers include but are not limited to: names, identification numbers, specific information on physique, genetics, economic, cultural or social identity (Regulation (EU) 2016/679, p.33). If masked personal data that can be re-identified by addition of other data should also be considered as information on an identifiable person. Based on article 4 Gdpr, a municipality can be described as a controller of personal data. Publishing open data is a form of processing when a dataset contains personal data. The recipients of open data are all those who view and use open datasets, both natural and legal persons (regulation (EU)2016/679, p. 33). When municipalities publish open data that unlawfully contain personal data this is referred to as a personal data breach.

Publishing open data is a form of processing when a dataset contains personal data. The recipients of open data are all those who view and use open datasets, both natural and legal persons. An anonymized dataset is legally referred to as pseudonymized. When municipalities publish open data that unlawfully contain personal data this is referred to as a personal data breach.

2.3.2 Risk mitigating measures

Controllers need to be able to demonstrate how they assess and mitigate the privacy risks of processing personal data. The controller needs to adopt policies that embed risk assessment and risk mitigation in work processes (Van Dijk, Gellert & Rommetveit, 2016). The Gdpr includes several measures that are used to protect personal data including: data protection impact assessments (DPIA's), privacy be design, privacy by default, data protection officers, notifications of personal data breaches and (Van Dijk, Gellert & Rommetveit, 2016).

The first measure is data protection impact assessments (DPIA's). The DPIA is an assessment of the impact of the processing on the protection of personal data. This assessment is mandatory for controllers who intend to process personal data in a way that is likely to result in high risks to the rights and freedoms to the data subject. The DPIA needs to be conducted before the processing of personal data. The DPIA needs to contain a description of the processing, assessment of the risks and mitigating measures (regulation (EU)2016/679, p.53). The second measure is the notification of personal data breaches. The Netherlands already implemented a duty to report personal data breaches in the Dutch Wbp in January 2016 (Wet bescherming persoonsgegevens, 2016). The Gdpr includes an article implementing the same duty to report personal data breaches in the whole European Union. The controller needs to report the personal data breach within 72 hours of becoming aware of the breach.

The third measure is the installation of a data protection officers (DPO). The Gdpr requires certain organizations to install a DPO. Public authorities such as municipalities are one type of organization that is required to appoint a DPO. The Gdpr introduced the function of a DPO who has expert knowledge on data protection law and assists the controller with complying to applicable data protection regulation. The DPO may be an employee or external. The DPO needs to be able to perform the role in an independent fashion (regulation (EU)2016/679, p.55).

The fourth and fifth measures are privacy-by-design and privacy-by-default, two linked concepts but two distinct methods for data protection. As they are often mentioned together, the concepts of privacy-by-design and default refer to different applications of data protection (regulation (EU)2016/679 p.48). Privacy-by-design refers to the embedding of privacy-protecting strategies in the design stage and onwards instead of integrating privacy protecting measures in the last stages of the development. Based on the privacy-by-design strategy privacy-protection becomes part of the design process and will lead to new technologies with minimal data collection (Hader et al, 2017). Privacy-by-default refers to technologies applying a default setting that collects only the minimal amount of data for the technology to function. It is allowed for a technology to have different data settings that collect more personal data than is essential, however these settings require the explicit consent of the user before they are applied. This creates an opt-in situation where users have to give permission before more personal data is collected (regulation (EU)2016/679, p. 48).

2.4 Conclusion

The Dutch national government has formulated polices to promote open data in all layers of government. By promoting open data publication, the Dutch national government hopes to achieve a range of benefits. First, societal benefits among others more transparency, increased accountability and more democratic participation. Second, opening up government data is expected to benefit the economy by stimulating innovation and new business opportunities. Third, opening up government data and making data more accessible is expected to stimulate scientific research (Ministerie van Binnenlandse Zaken, 2015; Zuiderwijk, Janssen, Choenni & Meijer, 2014). On the other hand, there are possible risks to publishing open government data. These risks include the publication of inappropriate data, re-identification of individuals and the misuse of data (Zuiderwijk, Janssen, Choenni & Meijer, 2014).

Municipalities can adopt risk management methods to structurally assess and mitigate all the risks that are related to the publication of open data. A properly formulated and applied risk management will help to protect the municipality from unexpected events and be prepared in case of incidents (Wieczorek-Kosmala, 2014). One of the leading risk management frameworks is the Enterprise Risk Management (ERM) from COSO. A fully adopted framework is integrated in all the layers and all departments of the organization (COSO, 2004).

The Gdpr is the main legal framework applicable to the context of this thesis. The new regulation is mainly focussed on giving natural persons more control over their personal data, data minimalization and risk management. This links to the necessity for a formal risk management framework. The Gdpr requires controllers of personal data to provide evidence on how they protect the personal data under their control (Van Dijk, Gellert & Rommetveit, 2016).

3 Methodology

This chapter will describe the methods of research used in this thesis. First, the broad strategy and design will be explained. This will provide a general overview of the research. This will be followed by a more detailed description of the data collection methods. The third part describes the methods for data analysis. The fourth part of this chapter will describe the operationalization of the concepts used in the study.

3.1 Strategy and design

The goal of this research project is to gain empirical knowledge on how municipalities that are noteworthy regarding the publication of open data, have formulated and applied measures to mitigate the privacy risks connected to open data publication. This thesis focussed on the set up of privacy protecting policies and procedures. This knowledge is to be gathered through desk research, interviews and document analysis on policy and risk management documentation.

Noteworthy in the context of this thesis, refers to municipalities that are actively publishing open data multiple datasets and stand out because of their publishing activities. Noteworthy municipalities have published a large number of datasets compared to other municipalities. The term noteworthy also includes other factors for example the use of a municipal data platform and/ or recognized reputation as noteworthy. Given limited time and resources only a few municipalities can be researched. The selection of municipalities will be accompanied by an explanation for every chosen municipality.

The complete sets of privacy protecting policies and procedures of the selected municipalities are the units of analysis in this thesis. Recommendations will be made based on the final results. These are both recommendations for the municipalities that participated in the research and recommendations for other municipalities that are working with open data. Because of the use of qualitative research methods, it is not possible to generalize the results of the results for all Dutch municipalities. However, by analysing the privacy risk management framework of the noteworthy municipalities regarding open data the results can provide general guidance for the municipalities their (and possibly expending) open data activities.

Firstly, it needs to be determined how many municipalities are publishing open data. This is determined through desk research. This part of the research collects general data on the publishing activities of all 388 Dutch municipalities. This research is limited to whether or not municipalities have published open data and if so, on which portals. Partly based on these results and partly based on other sources, the noteworthy municipalities selected for the qualitative data analysis. A maximum of three noteworthy municipalities is selected as cases for this thesis. Secondly, document analysis is conducted on the full scope of documents that construct the complete set of privacy protecting policies and procedures. After the document analysis one or possible more interviews are conducted. Thirdly, implemented privacy protecting policies and procedures will be compared to prevalent risk management methodology through desk research.

3.2 Operationalisation

Open data

There are two conditions for the operationalization of open data in the context of this thesis. Frist, there are several different definitions of open data. This thesis uses the definition formulated by the Dutch national government. Open data is defined as: public, non-proprietary and licence free, the data has been paid for from public funds, preferably machine processable and accessible (Data.overheid, 2017). When proactive data does not meet all of these principles it will not be referred to as open data. This operationalisation will guarantee that counted open datasets are comparable. Second, the published open datasets need to be redundant to data that municipalities are required to provide to the Central Statistics Bureau (CBS). Municipalities are required to provide certain data to the CBS (CBS, 2017). This thesis is focussed on open data that is published on the initiative of the individual municipalities, voluntary publication of open data. When municipalities provide data to the CBS they do not have to consider whether or not to publish these datasets. This is already determined by the CBS. Therefore, this data falls outside of the full authority of the municipality itself and outside the scope of the research.

Open data portals

Two types of data portals are identified in this research: individual portals and shared portals. The term individual portal refers to portals that are created and managed by one municipality and contain open data relevant to the managing municipality. The term shared portal refers to national or regional portal on which multiple municipalities or other government organizations can publish open datasets. These shared portals are operated by an overarching- or third party. An example of a shared portal is Dataplatform.nl, this website was created by Civity (Dataplatform, 2018a). Some municipalities use both an individual platform and a shared portal.

3.3 Data collection

3.3.1 First sub-question

Information on the data portals of Dutch municipalities will be gathered through internet research. This stage of the research will provide quantitative data on the publishing activities and the data portals on which open data is published. This data will be collected through Google searches with the search term: open data [name of the municipality]. This search will be conducted for all Dutch municipalities as they existed in 2017. The search term is limited to open data published by the municipality and excludes open data published by other organizations. The internet search will indicate whether the municipality has published open data, on which platform it has been published, if this is an individual portal or a shared portal and the number of datasets that are available. The completed dataset also includes the number of inhabitants per municipality. All the collected data will be gathered in one spreadsheet.

Alongside the internet search for publishing activities, a second internet search for more information on open data and Dutch municipalities is conducted. This internet search is aimed at gathering extra data on which municipalities are noteworthy. As part of this research the Association of Dutch Municipalities (Vereniging van Nederlandse gemeenten, VNG) was contacted. This association generally has a broad overview on various topics relevant to municipalities. A few questions on which municipalities they perceive to be noteworthy were sent to the VNG in order to gather more data. Based on all the collected data requests for cooperation was sent to multiple municipalities. It was expected that some municipalities would not want to cooperate. Therefore, initial requests for cooperation were sent to seven municipalities.

3.3.2 Second sub-question

The data used to answer to the second research question on how are comprehensive open data policies and procedures are applied, is gathered through document analysis and interviews. The documents that have been analysed, were gathered through a combination of desk research and an inquiry to receive documentation. The combination ensured that the full range of documentation constructing the comprehensive privacy risk mitigating policies and procedures were gathered and analysed. The initial requests for information and documentation were addressed to the Data Protection Officer. Based on the legal function description it is likely that the DPO will have the most information and expertise on how the municipality manages the privacy risks when publishing open data. However, depending on the way a municipality has allocated roles and tasks, officials other than the DPO were contacted. Documents are analysed by using an analysis framework.

After the initial request for participation further communications were undertaken with municipalities that are willing to participate. During these communications, information about data collection were shared and arrangements were made regarding sharing documentation and interviews. The goal of these communications is primarily to prepare for carrying out research. However, these communications were processed in such a way that relevant information could be used as research data. E-mails including relevant data have been documented and personal meetings were audio recorded and transcribed.

The interview will take place after the document analysis is completed. The results of the documents analysis will function as a baseline for the interview. The interview will be conducted with the same official that provided the documentation. These interviews were aimed to ask follow-up questions and collect addition data. The additional data concerns the context of the documents, how open data policies and procedures are applied in practice and gather information about work methods that might not be documented.

3.3.3 Third sub-question

The third sub question on the extent to which applied open data policies and procedures implement prevalent risk management methodology, is aimed at evaluating to what extent the privacy risk mitigating policies and procedures implement prevalent risk management methodology. The policies and procedures used for the publication of open data were compared to prevalent risk methodology. Research data on procedures and policies of the municipality have been gathered through document analysis and interviews. Information regarding prevalent risk management methodology is gathered through desk research.

3.4 Data analysis

3.4.1 First sub-question

The data from the internet searches and the outreach to the VNG was combined to identify several municipalities that could be possible subjects for this study. The main indicator was the number of published datasets. The second factor to be considered is the use of an individual data platform. Lastly, the input from the VNG was combines with the other two factors. The information gathered from the VNG was mainly used to include municipalities that might not have stood out, based only on the other two factors.

3.4.2 Second-sub-question

Documentation analysis determined if the all the components of a risk management framework were present in policies and procedures. The COSO framework (COSO, 2004) and the risk mitigating theory of Simperl, O'Hara and Gomer (2016) have been used to build a general analysis framework. This framework is made up of ten broad risk management components that result in 32 variables. These variables were scored from 0 to 3. This score indicates to what extent the variable is present in the documentation and appears to be applied in practise. See table 1 for a description of the scores. These scores are a simple tool to determine set up and existence of policies and procedures.

Scoretabel		
Item is not part of open data activities	Indicates no action taken on the item	0
Policy, procedure or information on item that is part of open data activities is not available in documentation	Indicates no documented policy on the item	1
Policy, procedures or information partially available in policy or documentation	Indicates partial set up	2
Information comprehensively included policy or documentation	Indicates set up	3

Table 1

The initial scoring was done based on the documents provided by the municipalities. These initial results will be the starting point for the interview(s). There was at least one interview per municipality. The interview(s) were used to discuss these results and to gain more information on the application of the described policies and procedures. The data from the interviews was used to determine the final scores in the analysis framework.

3.4.3 Third sub-question

The results of the second sub question were the base for the answer to the third research question. These results were compared to prevalent risk management methodologies to evaluate which aspects were implemented. This evaluation was focused on how all the separate policies and procedures constitute a comprehensive framework and how this compares to prevalent risk management methodologies.

3.5 Conclusion

This research project is based on desk research, document analysis and interviews. The first and third sub questions were answered based on desk research and the second sub question will be based on document analysis and interviews. First, it needed to be determined how many municipalities are publishing open data and which municipalities are noteworthy. This was determined through desk research. Secondly, document analysis was conducted on full scope of documents that construct open data publication process. After the document analysis one or possible interviews have been conducted. Thirdly, the implemented and applied risk mitigating policies and procedures were compared to prevalent risk management methodology through desk research. Based on the results of this research, recommendations will be made to improve risk management procedures. These results will also include exemplary policies procedures.

4 Results

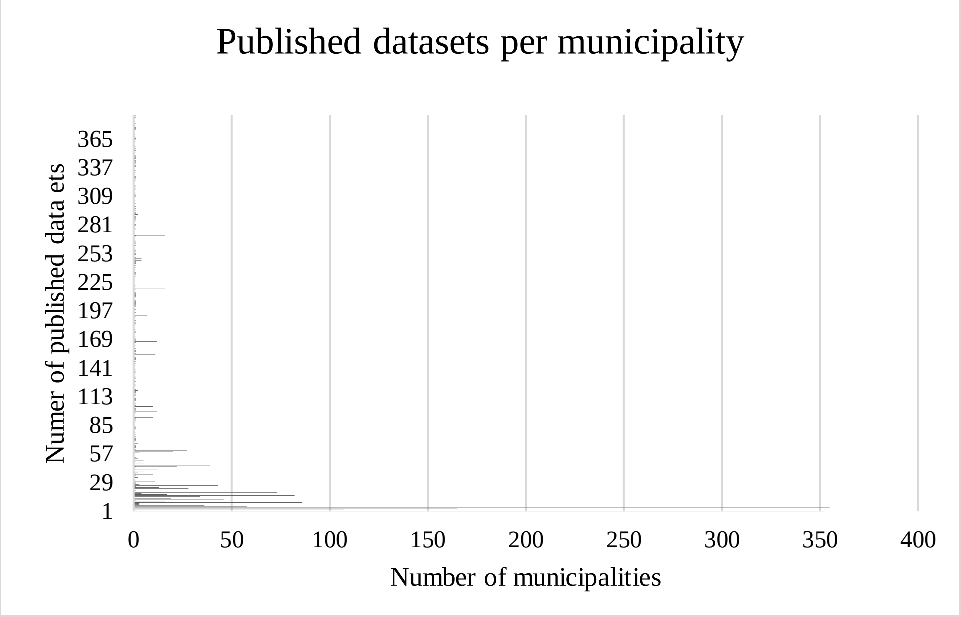
4.1 Sub-question 1: Which Dutch municipalities are noteworthy regarding the publication of open data?

4.1.1 Publishing municipalities

The selection of noteworthy municipalities is predominantly based on the active publication of open data. The active open data publication was researched and mapped through internet searches regarding active open data publication of all Dutch municipalities. A Google search was conducted for all Dutch municipalities with the following search term: open data [name of the municipality]. The hits on the first page would generally provide sufficient links towards open data portals or other open data activities. A second search for "open data" would be conducted on the website of a specific municipality when no direct hits would come up in the initial search. The total number of datasets per municipality could generally be found on the open data portal. The results of the internet searches were documented in one large spreadsheet that included the name of the municipality, the number of citizens, whether the municipality used their own portal or a shared portal, a link to the used portal, number of datasets, and the date of the internet search. All searched where conducted between October 5th and October 26th of 2017. The collected data is presented here. All 388 Dutch municipalities where researched, two thirds of Dutch municipalities have published at least one open dataset.

Publishing open data	Count	%
Have published open data	252	64,95
Have not published open data	136	35,05
Total	388	100

Table 1



Most of the municipalities have only published one dataset. This single published dataset almost always concerns data on placement and types of public lighting. Due to the structure of ten portals it was difficult to properly determine the exact number of datasets. These have been counted as ‘Not able to determine total amount of datasets’. Two municipalities stand out among the rest. These are the city of Amsterdam (352) and the city of Utrecht (355). The city of The Hague (165) and the City of Rotterdam (105) are the third and fourth largest publishers.

Most municipalities that have published open data use a shared data portal. Only 41 municipalities use an individual portal that only contains open data on their municipality. The other 211 municipalities use a shared portal. Some municipalities have published open data on both individual and shared portals. The most used portal is dataplatform.nl.

Portal	Count	%
Dataplatform.nl	175	69,44
Data.overheid.nl	12	4,76
Dataplatform.nl & data.overheid.nl	24	9,52
Own portal	33	13,09
Own portal & shared portal	8	3,17
Total	252	100

Table 2

4.1.2 Determining noteworthy municipalities

Based on the number of published datasets and the use of their own open data portals the municipalities Amsterdam, Rotterdam, Utrecht and The Hague, Eindhoven (58) and Leeuwarden (43) stand out as noteworthy. The VNG was contacted in order to get more insight in which municipalities could be determined noteworthy but that would not stand out based mainly based on the number of published datasets. Three municipalities indicated to be noteworthy by the VNG were: Tilburg (36), Eindhoven and Haarlem (46) (VNG, personal communication, November 21, 2017).

Based on a combination of the collected data and the answers received from the VNG the following municipalities were requested to participate: Amsterdam, Utrecht, Rotterdam, The Hague, Haarlem, Eindhoven and Leeuwarden. Based on lower number of published datasets Tilburg was not included in the initial requests for participation. A response from Eindhoven and Leeuwarden was never received. Three municipalities: Utrecht, The Hague and Haarlem agreed to participate as subjects in this research project.

The municipalities of Rotterdam and Amsterdam declined to participate. In 2018, open data publishing activities in Rotterdam have been suspended. The open data portal listed a program manager, he was contacted with a request for participation in this research project. He replied that he had worked on the open data project for at least two years. He forwarded the request for participation to a colleague (Gemeente Rotterdam a, personal communication, November 14, 2017). Further contacts with his colleague made clear that the municipality of Rotterdam had suspended open data publishing activities and there were no plans to resume open data publishing activities in the near future (Gemeente Rotterdam b, personal communications, November 27, 2017). Due to the suspended open data publishing activities it was determined that the municipality of Rotterdam would not make a relevant case in this research project. Had Rotterdam participated, the research would focus on procedures that have not been used for two years and might not be used in the future. This does not fit with the criteria of a noteworthy municipality regarding the active publication of open data.

The municipality of Amsterdam responded but declined to participate. The documents regarding open data publication were qualified as drafts at the time of the request. These documents would not become definitive documents until mid-2018 (Gemeente Amsterdam. December 12, 2017). From a research standpoint it would not be problematic to use documentation qualified as drafts. Amsterdam would still be an interesting case due to the large number of published datasets. This was communicated to the municipality. However, the municipality would be unable to share documents that were classified as drafts and declined to participate. They did indicate that they used data classification norms by the Information security service (IBD) documents in combination with the draft documents (Gemeente Amsterdam. December 12, 2017).

4.2 Sub-question 2: How are comprehensive open data policies and procedures applied?

The application of open data policies and procedures are discussed per risk management element. These elements match the elements of the analysis framework used to analyse the documents. The results will be discussed per municipality.

4.2.1 Set-up: Municipality Utrecht

4.2.1.1 General open data policy

A total of seven documents were analysed and two interviews where held with the municipality Utrecht. The data protection officer (DPO) and the open data coordinator where present. The general open data policy is embedded two documents: commission letter 'approach data driven steering and open data', hereafter: commission letter and action plan indexation municipal datasets, hereafter action plan (Gemeente Utrecht, 2014a; Gemeente Utrecht, 2014b). Utrecht has been working on open data since 2014. The open data policies and procedures are part of a broader data policy. The data policy seeks to anticipate upon demand of organizations in the public and private sector that have an interest in using municipal data (Gemeente Utrecht, 2014a). On the other hand, the "*municipality itself can work smarter by actively using data*" (Gemeente Utrecht, 2014a, p.2). In the commission letter the following general open data policy is established: "*Open, unless…*" (Gemeente Utrecht, 2014a, p.7).

The commission letter dedicated a paragraph to the balance between privacy and as phrased in the document: “to dare”. This paragraph acknowledges that privacy risks can occur as a result of the new data strategy. However, the exact risks are not fully described (Gemeente Utrecht, 2014a). One of the identified privacy risks, the unintended disclosure of personal data through combining datasets, can be deducted from an assurance regarding a question from the council from 2013. This assurance lays down the foundation for a decision-flowchart that will be used for publishing open datasets (Gemeente Utrecht, 2014a, p.9).

The last part of the general open data policy concerns the other risks of open data, excluding privacy. These risks are not explicitly mentioned in the general data policy documents. However, there is awareness of the possible risks in practice. The base for this claim comes from the first interview, concerning a dataset that was reviewed for possible publication as open data (Gemeente Utrecht, personal communication, December 13, 2017). The dataset included the number of persons per address. In this case, it didn’t include personal data, however there were serious safety risks. And to quote the open data coordinator: “There *can be situations where we […] where we are perfectly within the legal boundaries. But where we think […] it does not feel right. That is actually ethics.”* (Gemeente Utrecht, personal communication, December 13, 2017).

4.2.1.2 Legal framework

Utrecht has adopted a comprehensive privacy municipal ordinance. This policy is meant to further specify the Wbp. It refers back to this law for definitions and norms. However, this policy is also very much based on the Gdpr. It includes the use of Privacy Impact Assessments and appointing a DPO (Privacyverordening gemeente Utrecht, 2016). This ordinance includes one article specifically on open data. This article, article 6, further refers to Regulation re-use of government data (Wet hergebruik van overheidsinformatie, 2016). This article also explicitly states that open data will not include any data that is traceable to a person. The privacy policy is an official publication of the municipality an available online (Privacyverordening gemeente Utrecht, 2016).

4.2.1.3 Selection of open datasets

The document ‘processes data catalogue’ includes all the main processes concerning the selection, publication and updating of open datasets. This document lays out the exact steps and roles in the selection process. This document includes three different processes for how datasets are selected. The first two processes are reactive to requests or initiatives from within the organization (Gemeente Utrecht, 2017a). The first process in based on demand from the organization. employees can request for data to be released. The second process is based on an initiative from the employees or departments to publish certain data. In both these processes the request or demand lands at the desk of an administrative employee and will be handled by the so-called data squad. The data squad is collaboration between the different officials working the open data project (Gemeente Utrecht, 2017a). The third process is an active search for data that can be published. The data coordinator will initiate action, these actions can be based on contacts with the organizations or stakeholders in the city of Utrecht (Gemeente Utrecht, 2017a).

4.2.1.4 Procedure for publishing data

The procedure for publication is based on a flowchart made by the municipality of Rotterdam (Gemeente Utrecht, personal communication, December 13, 2017; Gemeente Rotterdam, 2017). This process is described with individual steps and based on roles. Certain parts of publishing process depend on the way the dataset has been selected for publication. If a dataset is selected in an active way, there are steps added before coinciding with the same steps after reactive selection. These extra steps concern communicating with the city and preparing data to be published (Gemeente Utrecht, 2017a). After these steps, all processes arrive at the same step: the proposed dataset is discussed in the data squad. The data squad discusses possible edits or adaptions to the dataset. Then the dataset is reviewed for privacy sensitive data (Gemeente Utrecht, 2017a). If data is found to include privacy sensitive data, the data squad will make adjustments to remove the privacy sensitive data. After a dataset is determined not to contain privacy sensitive data or the data squad has removed privacy sensitive data from the dataset, the procedure for connection to data catalogue starts (Gemeente Utrecht, 2017a). The last phase concerns connecting the data in the data catalogue to the open data platform. This is mainly a technical process of adding the proper meta data and making the right connections between the source and the data layers (Gemeente Utrecht, 2017a).

Determining if the data contains privacy sensitive data has become easier over time. The data owner is responsible and the data coordinator will determine if a datasets contains personal data (Gemeente Utrecht, personal communication, December 13, 2017). In the beginning there was more insecurity on the ability of making the proper judgements. The DPO would regularly be involved in judging the contents of data (Gemeente Utrecht, personal communication, December 13, 2017). The procedures have not changed over time but the involved actors have gained more confidence to make the right judgements on the privacy sensitivity of data (Gemeente Utrecht, personal communication, January 15, 2018).

Assessment of risks other than privacy sensitivity are not part of the ensured processes. The municipality has an Ethical Data Assistant (DEDA) (Dataschool, 2018). The DEDA is a questionnaire that includes the assessment of risks other than privacy; although it also includes privacy issues. The DEDA includes some straight forward questions about the use and access to the data but also more ethical questions that have the potential to trigger ethical debates (DEDA, 2017). However, the use of this questionnaire is not part of the regular process and is not used on a regular basis. There is an intent to expand the use of the DEDA because there is uncertainty if the questionnaire is used enough (Gemeente Utrecht, personal communication, January 15, 2018).

4.2.1.5 Removal of personal data

The editing of datasets is part of the publication process. Utrecht has adopted several privacy-protecting measures in the privacy regulation and created a framework for using these measures in preparation for the Gdpr. These techniques are developed in the 'framework privacy by design'. This framework used during a privacy impact assessment (Gemeente Utrecht, 2017b).

After personal data is removed from a dataset there is a check if all sensitive data has properly been removed. This step is not documented in the 'process data catalogue'. This part of the policy is embedded in the work methods of another department. However, these activities are part of normal work methods and appear to be an ensured procedure: *"One person blacks it out and the other checks it. […] there are four eyes that look through it*." (Gemeente Utrecht, personal communication, January 15, 2018). Coordinators of employees removing data are responsible for ensuring this procedure (Gemeente Utrecht, personal communication, January 15, 2018).

4.2.1.6 Management of published datasets

The roles and procedures for updating datasets are set out in the document 'processes data catalogue'. The actions are divided per role. These actions do not include specific periods for updating. There are no pre-set periods for updates because this very much depends on the specific dataset (Gemeente Utrecht, 2017a). Some data is static and does not require updates, for example reports (Gemeente Utrecht, personal communication, January 15, 2018). The terms for updating are set in agreement with the data owner in the organization. The data owner is responsible for preforming the necessary updates. These update agreements are put in writing. In practise the made agreements are not always honoured (Gemeente Utrecht, personal communication, January 15, 2018).

The municipality does not preform checks for traceability after publication. The current procedures are implemented to make sure that re-identification is not possible after publication (Gemeente Utrecht, personal communication, January 15, 2018).

The community manager is responsible for the collaboration between open data users and the municipality (Gemeente Utrecht, 2017a). The communication with the end-users of the data is an important aspect of the open data project. The data coordinator collects input from the end-users through meet-ups with users and maintains contacts with users (Gemeente Utrecht, personal communication, December 13, 2017). Aside from these actions there is an email address for users of the data catalogue. An administrative employee maintains this email-box (Gemeente Utrecht, 2017a).

4.2.1.7 Risk management

Municipal risk management is set out in 'Note risk management and buffer 2015-2018' (Gemeente Utrecht, 2015). This document is a product of the concern-management Finances and Control. The development of the risk management within the organization is laid out in this document. The methodology is described and provides a broad overview of the implemented risk management framework. The risk management responsibilities are decentralized, departments are responsible for their own risks. (Gemeente Utrecht, 2015). The document identifies nine categories of risks: economic/market, political/society, nature, organizational, political/governance, implementation, resources, personnel and legal (Gemeente Utrecht, 2015).

The municipality has a risk provision for the event of receiving a fine for a personal data breach. This provision was an initiative of the DPO. The risk of actually getting fined for causing a personal data breach is small, but the fines can be severe (Gemeente Utrecht, personal communication, December 13, 2017). Aside from this risk provision the influence of risk management on the policies and procedures concerning open data has been limited. However, there is a growing awareness of risks and risk management. Information security officers had a workshop organized by the department responsible for risk management on the theme of risk management. The value of risk management methodology is acknowledged as the DPO sees possibilities to incorporate risk management activities in the PIA process (Gemeente Utrecht, personal communication, January 15, 2018). The interview itself, on the fifteenth of January and the discussion of the topic of risk management, brought about more awareness and possibly actions to further ensure procedures. Specifically, further ensuring the compliance with the agreements made on updating of datasets (Gemeente Utrecht, personal communication, January 15, 2018).

4.2.1.8 Personal data breaches policies

The procedure and the relevant roles in the case of personal data breaches are set-up in the privacy policy. Municipality Utrecht maintains a public register of its own data breaches in addition to the required notifications to the Autoriteit Persoonsgegevens and the data subject (Gemeente Utrecht, personal communication, December 13, 2017).

4.2.1.9 Availability and awareness of policies and procedures

Many of the documents used for the analysis are available online. The privacy policy, the commission letter, the approach for adding onto the data catalogue and the note on risk management are all publically available (Gemeente Utrecht, 2014a; Gemeente Utrecht, 2014b; Gemeente Utrecht, 2015). Availability is important however, awareness is crucial. The DPO and data coordinator refer to it as “missionary work”, sending the message of the benefits of open data (Gemeente Utrecht, personal communication, December 13, 2017). The publication of open data mainly depends on the willingness of the business to invest time and money into open data. It is a change of culture and it takes time. There is willingness but also caution as the final responsibility lies with the data owners. While there is still some way to go to build awareness of the value of (open) data, significant progress has been made (Gemeente Utrecht, personal communication, December 13, 2017; Gemeente Utrecht, personal communication, January 15, 2018).

4.2.2 Set-up: Municipality The Hague

4.2.2.1 General open data policy

A total of six documents were analysed and one interview was conducted with the coordinator administrative unit Open Data. The general open data policy is laid out in several letters from the Mayor and Aldermen to the City council. These policy-indicating documents mainly point to the benefits of open data and open government. The municipality is of the opinion that open data will contribute to transparency, co-creation with the city, contribute to innovation, stimulate education and will stimulate the economy (Municipality The Hague, 2011). The Hague has practically started over with the open data publication activities. The organization is in the early stages, working on the easier datasets, focussing on easy to publish data, the low hanging fruit, and working to increase awareness and visibility of open data (Gemeente Den Haag, personal communication, January 31, 2018).

The municipality has adopted a policy of 'Open, unless….'. Municipal data should be 'open' unless the data contains privacy sensitive, confidential material, cannot be opened due to technical reasons or due to relatively high financial costs. The municipality is working on adopting the degree of openness in the information architecture, which will in time eliminate the ad hoc selection of open datasets. The municipality seeks cooperation with different partners in the development in the open data policy and open data activities. These partners include: universities, communities of open data users and other governmental organization, for example the four largest municipalities (Gemeente Den Haag, 2013).

The privacy risks are acknowledged in different policies and the alderman had to answer questions from the council members on the issue of privacy affirm the need for protecting privacy and set out the procedures for conducting a privacy check before data is opened up (Gemeente Den Haag, 2013; Gemeente Den Haag, 2017a).

General risks are discussed in less detail in the analysed documentation. These risks are not directly discussed in the policy documents however, these risks are included in an online tool that is used to determine if data is appropriate to publish as open data. The municipality developed a checklist, made available online, that includes among others: (copy-) rights of third parties, security risks and possible damages to individuals or society. The tool indicates that procedures for privacy protection and prevention of general risks are in place (Gemeente Den Haag, 2016b).

4.2.2.2 Legal framework

The municipality does not have an overview of all applicable laws concerning open data. The legal information is scattered throughout the organization. If that information is needed, it can be found (Gemeente Den Haag, personal communication, January 31, 2018). One important part of the legal framework is the concept of personal data. A description of the concept of personal data is available in the online tool.

4.2.2.3 Selection of open data

The selection process is specified with steps and roles in the document ‘description process open data’. The selection of potential open datasets is mainly driven by demand from users. Requests for data come from inside and outside the organization. The first step, after a request is received, is to find the requested data. The requested data is not always available in the manner that is expected by the person requesting the data (Gemeente Den Haag, personal communication, January 31, 2018). The open data project is still in the early stages. This causes the selection process to be mainly demand-driven and focused on data from the so-called ‘High value data’- list (Gemeente Den Haag, personal communication, January 31, 2018). This list of high value datasets is created in cooperation with the Ministry of Internal Affairs and includes among others: city council information, WOB-requests, events and the locations or catering industry and retail business (Ministerie van Binnenlandse Zaken, 2018).

Depending on the data, it might be placed in a source. Data can be static or dynamic, this requires different approaches regarding updating the data. After this process is completed, the data will be entered in the data catalogue (Gemeente Den Haag, personal communication, January 31, 2018). This process is executed by an employee of the administrative unit.

4.2.2.4 Publication of open data

The publication process is specified with steps and roles in the ‘description process open data’ document. After the data has been entered in the data catalogue it is determined to what extent the data can be opened. The online flowchart can be filled in by an employee of the ‘administrative unit open data’ or by the requesting party. The results of the flowchart are collected and saved. The online tool includes privacy-related issues, safety issues, competition issues, rights of third parties, supervisory tasks of administrative authorities, international relations, confidentiality, costs and data format (Gemeente Den Haag, 2016a).

After this check, the type of data is assessed; it is either structured or unstructured data. Based on the type of the data the next steps differ. The steps for unstructured data and for structured data are aimed at preparing the data for publication on the data platform. The responsible role is identified for every step in the publication process (Gemeente Den Haag, 2016a).

These processes are executed by members of the administrative unit open data. The processes are ensured because they are all essential for publishing open data that meets all the standards for usable open data. If a step would be missed, it would be noticed. In cases where procedures where not followed it is known why this happened (Gemeente Den Haag, personal communication, January 31, 2018).

4.2.2.5 Removal of personal data

The document ‘description process open data’ also includes the role and process on removing data that cannot be released. This includes but is not limited to personal data. Removing information from a dataset is part of the analysis of a dataset. This task is performed by a member of the administrative unit open data (Gemeente Den Haag, 2016a). The procedure described in the document does not include a check if all data has been properly removed. However up till the time of the interview, there have not been datasets that required the removal of personal data (Gemeente Den Haag, personal communication, January 31, 2018).

4.2.2.6 Management of published datasets

The procedures for updates in published data is partially set out in the document 'description process open data'. For every dataset the update requirements can be different. Most of the geo-data is connected to a source. This data is updated daily according to changes made in the source (Gemeente Den Haag, personal communication, January 31, 2018). Preparing a connecting between geo-data and a source is part of the process document. Some data does not require updates. Reports for example, are finished and uploaded and do not require updates (Gemeente Den Haag, personal communication, January 31, 2018). For all the other data that does require updates but is not connected to a source, the source owner is responsible. The source owner needs to determine how often updates are required. The administrative unit has a supportive role. They maintain their own register that indicates when data should be updated and they should contact source owners about preforming updates (Gemeente Den Haag, personal communication, January 31, 2018).

The municipality does not do checks for traceability. The municipality is very careful with what data is opened up because the open data project is still in early stages (Gemeente Den Haag, personal communication, January 31, 2018).

The administrative unit open data has an email address where users can contact the municipality about open data. Users can request data and the unit will then assess that request and possibly release open data (Gemeente Den Haag, personal communication, January 31, 2018).

4.2.2.7 Risk management

The municipality does have its own accountancy department. This department is also responsible to conduct IT-audits that might include compliance to privacy regulation (Gemeente Den Haag, 2017b). The tasks and roles of the Auditing Committee Municipality The Hague is determined in an order by the Mayor and Aldermen in 2004 (Gemeente Den Haag, 2004).

Based on the interview it can be concluded that up till now there has been little connection between risk management and open data policy. There is awareness of the risks, especially of the risk of deanonymization. This awareness is accompanied with a lot of uncertainties. The legal exposure concerning these risks is not clear. These uncertainties and the relatively new open data activities lead the administrative unit to be careful (Gemeente Den Haag, personal communication, January 31, 2018).

4.2.2.8 Personal data breaches

The procedures in the event of personal data breaches are not included in the documents of the administrative unit open data. However, the municipality has a data protection officer and she has set up a protocol in the case of personal data breaches. There are contact points throughout the organization where employees need to report incidents (Gemeente Den Haag, personal communication, January 31, 2018).

4.2.2.9 Availability and awareness of policies and procedures

Several documents are publicly available as they are communications from the political spectrum. This includes questions from city council members, the answers to those questions from the mayor and aldermen and information on the open data project provided by the mayor and aldermen. These documents are accessible through the online council information system. The online tool is also available, a link is provided by the administrative unit open data. The 'description processes open data' is not directly available. However, this includes many, very technical procedures that are not necessarily of interest for employees that are not working for the administrative unit. More information can be requested through the email address of the administrative unit.

General information on the audit committee, that works on risk management, can be found online. Most information is not publicly available online (Gemeente Den Haag, 2017b). Because this research project is focussed on open data, the audit committee was not further investigated. The interview indicated that a connection between risk management and open data had not yet been established (Gemeente Den Haag, personal communication, January 31, 2018).

The awareness on the activities of the administrative unit open data is limited. The unit used a poll to test the awareness in the organization. That poll indicated that hardly anyone in the organization knew about the unit. Working with data is a challenge for some employees. Open data is still a very new concept. The mindset that open data can contribute to the city has not been established in the organization. Changing the mindset is part of building the open data activities (Gemeente Den Haag, personal communication, January 31, 2018).

4.2.3 Set-up: Municipality Haarlem

4.2.3.1 General open data policy

Haarlem has been working with open data for three years and adopted a new comprehensive data management policy in 2017(Gemeente Haarlem, personal communication, February 7, 2018). The open data policy of the municipality Haarlem is part of a general data policy. Open data is essentially a by-product of properly managed data (Gemeente Haarlem, personal communication, February 7, 2018). The newly developed data is based on the arguments that data management is crucial for efficient and effective operations of the organization, providing efficient and effective services, compliance to legal requirements regarding confidentiality and privacy, accessibility and usability of data and protect privacy and security risks (Gemeente Haarlem, 2017a). The design of current open data policies was partly modelled after procedures of other municipalities. Haarlem used open data procedures developed in Rotterdam and The Hague (Gemeente Haarlem, personal communication, February 7, 2018).

The municipality aims to structurally offer open data. The only pre-condition is that open data does not include personal data or harm the public interest. This pre-condition includes that open data may not contain other data that might lead to identification of personal data through combination of datasets (Gemeente Haarlem, 2017a). The data policy mainly focusses on the general benefits of data management. The publication of open data is mainly based on the legal framework. The privacy risks are acknowledged. Possible other risks are also acknowledged, however, in a lesser extent, in the document (Gemeente Haarlem, 2017a). The awareness of risks other than privacy can be distinguished in the interview: can the data be properly understood by the user? (Gemeente Haarlem, personal communication, February 7, 2018).

4.2.3.2 Legal framework

The data policy includes an extensive legal framework regarding open data publication. This framework functions both as a ground for publishing data as well as for setting the limits for publishing open data. This legal framework is a summary of different articles from several different laws. This legal framework is publicly available because it is part of a public records (Gemeente Haarlem, 2017a).

The concept of personal data is part of this legal framework and is also described in other parts of the document. This description also includes data that indirectly can identify individuals as personal data (Gemeente Haarlem, 2017a).

4.2.3.3 Selection of open data

The data policy of Haarlem is very new, it was established in early 2017. The policy makes clear that all the data that should eventually be published as open data, cannot be published all at once. Therefore, the following priorities have been formulated for what data to publish first:

1. data must already be adequate to publish without further adjustments,
2. connected to tangible societal issues even though adjustments need to be made to the available data,
3. Without demand and data that requires adjustments to the available data (Gemeente Haarlem, 2017b).

This part of the policy is generally worded and only includes a few distinct procedures. The roles in the open data policy are divided over different departments. One of the distinct actions in the selection process is creating an overview of all the municipal datasets This overview will also indicate what data has been published. The municipality is mainly focussing on publishing the easier datasets that do not require adjustments first (Gemeente Haarlem, 2017b).

4.2.3.4 Publication of open data

The publication process will start after data is available in the data register and might be published for open data. The first step is to use a decision-flowchart. The decision-flowchart includes privacy-related issues, safety issues, competition issues, rights of third parties, supervisory tasks of administrative authorities, international relations, confidentiality and costs. The decision-flowchart includes the usability after problematic data has been removed, traceability of masked personal data, if the municipality may require permission from third parties to use the data, weighing the public interest versus the interest of parties mentioned in the data (Gemeente Haarlem, 2017b). The data is published if the result of the completed decision- flowchart is positive. The decision-flowchart is filled-out by the department Data, Information and Analysis (DIA). This department is the data-owner and is responsible for this step of the publication process (Gemeente Haarlem, personal communication, February 7, 2018; Gemeente Haarlem, 2017b). If the decision- flowchart results lead to a negative result there is an option to write a motivation on why the data should be opened up. This might result in the data being released (Gemeente Haarlem, 2017b).

The data owner is responsible to register the relevant documents in a document management system. The department Base-registrations is the registration-owner and facilitates the data owner for this process, if necessary. The department base registrations reviews if all the necessary steps have been followed. The status of openness is changed on the data register. This is done by the Legal department. The Legal department also registers this in the data register if data cannot be opened up (Gemeente Haarlem, 2017b).

4.2.3.5 Removal of personal data

The municipality uses different techniques to remove personal data from datasets, for example aggregation or removal particular parts from the dataset. The rules for traceability are based on the law and on statistical rules. The customization of the datasets can vary per dataset (Gemeente Haarlem, personal communication, February 7, 2018). The exact measures for customizing a dataset are not clearly described. However, the complete procedure includes a four-eyes method that ensures that the data owner follows the necessary steps prior to publication (Gemeente Haarlem, 2017b; Gemeente Haarlem, personal communication, February 7, 2018). The data owner is responsible for making adjustments to the dataset. The registration holder then checks through the data register if all necessary steps have been taken. This process ensures a four-eyes method (Gemeente Haarlem, personal communication, February 7, 2018).

4.2.3.6 Management of published datasets

The municipality does not conduct checks for traceability after data has been published. This should be ensured in the processes prior to publication (Gemeente Haarlem, personal communication, February 7, 2018). The open data platform is updated every night. In the future the updates will be real-time. This is an automated process (Gemeente Haarlem, personal communication, February 7, 2018).

The municipality has made a conscious choice to use a proclaimer for open data. The proclaimer includes the following part: "*Do you come across something that is incorrect, outdated or incomplete? Report it on the tab Report. Do you have other questions, suggestions for improvements or proposals for new dataset? Respond via: opendata@haarlem.nl*" (Haarlem open data, 2018). The municipality invites users to provide feedback. Using a proclaimer is expected to increase the creation of value from open data (Gemeente Haarlem, personal communication, February 7, 2018).

4.2.3.7 Risk management

The municipality has internal audit. Details on how the processes and policies are designed where not available to review. However, an external accountant, PricewaterhouseCoopers, has reviewed the audit and internal controls. The report that was provided does provide sufficient evidence on the implementation of internal controls. Cyber risks are part of this report. The overall conclusions are that the information security is properly dealt with. The necessary roles are filled, there is an effort to increase awareness of information security risks and the technical measures are taken (Gemeente Haarlem, 2017c).

The municipality has a list of identified risks available. Personal data breaches are identified as a risk that can result in a fine, investigation costs, repair costs and legal costs (Gemeente Haarlem, 2016).

There is awareness that open data can have negative effects. However, there is uncertainty about the size of the risks. This risk is another reason for using a proclaimer, it stimulates use of open data while also requiring the user to act in a fair and lawful manner (Gemeente Haarlem, personal communication, February 7, 2018; TU Delft, 2013).

4.2.3.8 Personal data breaches

The municipality has adopted measures to mitigate the risk of a personal data breach. These measures are in accordance with the national guidelines (Gemeente Haarlem, personal communication, February 7, 2018). These measures are part of information security efforts. Internal measures include demands on connections between sources, encryption and contracts with other processors of personal data. The municipality also started to collaborate with an independent partner on information security in order to further ensure information security measures (Gemeente Haarlem, 2016).

4.2.3.9 Availability and awareness of policies and procedures

All of the newly introduced data policies and procedures are available online. These are public documents. The report by the external accountant is not yet available online. The risk management documentation is not available.

There is awareness of the data management policies. The mindset within the organization on the importance of data management and open data has been established over the past three years, according to the coordinator Base registrations (Gemeente Haarlem, personal communication, February 7, 2018). The policies that have been analysed are less than a year old. However, the open data efforts have started three years ago. The development that is sought right now is focussed on making more data available in more ways, other formats, and meet the demands of users (Gemeente Haarlem, personal communication, February 7, 2018).

4.3 Sub question 3: To what extent do comprehensive open data policies and procedures implement prevalent risk management methodology?

The last part of the research examines to what extent open data policies and procedures integrate components of a comprehensive risk management framework. Risk management methodology is meant to deal with the uncertainty and risks that face organizations. A fully integrated risk management framework will allow an organization to deal with uncertainty and risk more effectively than organizations that do not implement all components of the framework (COSO, 2004).

4.3.1 Achievements of objectives

The top side of the COSO framework concerns the achievement of objectives. The COSO framework identifies four categories of objectives: strategic, operations, reporting and compliance. These categories overlap on certain areas but also have distinct differences. These objectives are placed at the top of the framework because they descend through the organization (COSO, 2004). The objectives regarding open data are set out and discussed as part of the internal environment and objective setting components of the ERM framework in the following paragraphs.

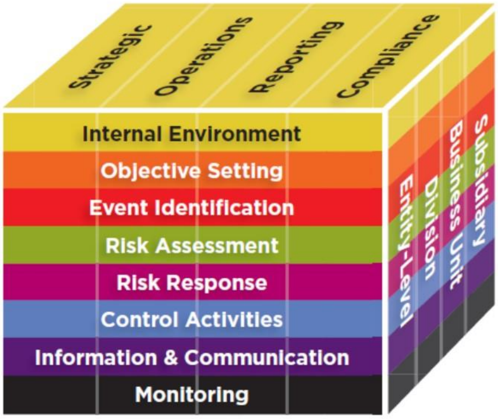


Figure 2,COSO ERM Framework ©, 2004

4.3.2 Entity Units

Entity units make up the right side of the COSO framework. This side of the framework indicates the levels on which the framework needs to be implemented. Specific risk management measures can differ depending on the entity level that is responsible for ensuring the measures (COSO, 2004). In all three municipalities data ownership lies with departments or divisions. Open data coordinators or data administrators do not have ownership of the data that is published. The responsibility to update or make mutations to data, lays with the data owner, usually data ownership lies with the department that is responsible on the data topic. (Gemeente Utrecht, personal communication, January 15, 2018; Gemeente Den Haag, personal communication, January 31, 2018; Gemeente Haarlem, personal communication, February 7, 2018).

4.3.3 Components of ERM Framework

	General observations	Utrecht	The Hague	Haarlem
Internal environment: *The internal environment includes the culture of an organization, the way risk is viewed and the risks are addressed. Internal environment includes the values of the organization, the ethics, the willingness to accept certain risk and how they operate (COSO, 2004).*	In all cases the municipalities have determined that open data has important societal benefits. The efforts to expand open data activities follows a national effort to increase transparency in government (Ministerie van Binnenlandse Zaken, 2017).	Transparency, allowing society to use data that the government collected with from tax money and to stimulate the economy are the main arguments to publish open data[1].	Transparency, allowing society to use data that the government collected with from tax money and to stimulate the economy are the main arguments to publish open data[2].	Transparency, allowing society to use data that the government collected with from tax money and to stimulate the economy are the main arguments to publish open data[3].
Objective setting: *In order to identify risks, it must be clear what objectives an organization wants to achieve with its activities. It is important for municipalities to clearly define the objectives regarding open data publication (COSO, 2004).*	The objectives in the cases of The Hague and Utrecht are broadly formulated. Broadly formulated objectives can make it more difficult to identify specific risks. For example, the objective of more transparency, it is a broad objective that without further specific measures is difficult to determine when it is achieved or what risks could hinder the achievement of that objective (COSO, 2004).	One current objective is increasing the number of available datasets, from 300 datasets to 500 in 2018. More general objectives are using data to tackle societal challenges and stimulating innovation using municipal[1] data[1].	Current objectives are mainly oriented towards transparency and expanding the open data activities within the organization[2].	Creating "self-service" for users to retrieve municipal data without having to request data every time[3].
Event identification: *Event identification depends on the objectives*	Wrongful publishing privacy sensitive data or otherwise inappropriate data is an event	The expanding open data activities are causing Utrecht to identify	Actions have been described under general observations[2].	Based on the objective to provide self-service, risks and opportunities can be identified.

[1] Gemeente Utrecht, personal communication, December 13, 2017; Gemeente Utrecht, personal communication, January 15

[2] Gemeente Den Haag, personal communication, January 31, 2018

[3] Gemeente Haarlem, personal communication, February 7, 2018

that have been set. An organization can identify events that pose either risks or opportunities based on these objectives (COSO, 2004).	that is identified by **all three** municipalities as a risk [1;2;3]	controllability as a possible risk. The open data coordinator was initially responsible for controls. Now the open data activities have expanded the control task has become large to be placed with one employee[1].		These risks are mutations that are not timely or accurately carried out thereby hindering users to obtain accurate data or possible misinterpretations of the data that can be caused by the use of labels that users do not properly understand[3].
Risk assessment: *Risk assessment concerns the methodological analysis of risks, this includes determining the likelihood of an identified event (COSO, 2004).*	Risks exclusive to open data are not identified and assessed methodologically in any of the participating municipalities. One important risk of open data publication is re-identification of individuals. It is a difficult risk to assess [2];[3]. First, the supply of data is growing, the available datasets per municipality and of course all the other data that is available online. It is difficult to determine how far people can go when they combine multiple datasets [2]. Second, the implications of incidents where combinations of open datasets, published either by municipalities or other organizations, are unclear. There is a lack of	Personal data breaches are identified as a risk and there is a risk provision. This financial provision would be used in the event that the municipality was fined for a personal data breach. However, a personal data breach is a risk that is not exclusive to open data publication. Personal data breaches can occur in ways that are unrelated to open data. Therefore, the risk provision needs to be considered as an organisation-wide risk management effort and cannot only be attributed to measures that mitigate open data risks[1].	The Hague is still in an early stage of developing open data publication. The rest of the organization is not very aware of the activities of the administrative unit open data. Risk management is currently not part of the work methods of the administrative unit[2].	Risk management methodology to identify risks and the possible impact of these risks. The municipality publishes the full list of identified risk including the financial impact. This list includes three identified risk regarding data management activities: delays in phasing out the paper archives, personal data breaches and failure of IT systems. These risks include a likelihood and estimate of the impact. These risks are related to open data but are not exclusively related to open data. The identified risk of delayed mutations or misinterpretation of data are

[1] Gemeente Utrecht, personal communication, December 13, 2017; Gemeente Utrecht, personal communication, January 15

[2] Gemeente Den Haag, personal communication, January 31, 2018

[3] Gemeente Haarlem, personal communication, February 7, 2018

[1] Gemeente Utrecht, personal communication, December 13, 2017; Gemeente Utrecht, personal communication, January 15

	jurisprudence on this specific topic[3]. Even without a formal identification of the risk, municipalities apply take measures to anonymize datasets. For example: changing dates of birth to an age or not including aggregates of 10 items or lower [1];[3] .			not formally assessed based on likelihood or impact[3].
Risk response: *Risk response concerns the reaction to an identified risk. An organization can choose to accept a risk, avoid a risk, reduce the risk or share the risk. Based on this determination actions are developed to match the risk to the risk response (COSO, 2004).*	Municipalities have not formally assessed all privacy risks. However, the implemented procedures and policies are aimed at reducing the informally identified risks. All three municipalities designed open data policies and procedures with the aim to reduce the risk of unlawfully disclosing personal data (Gemeente Utrecht, 2014; Gemeente Den Haag, 2017a; Gemeente Haarlem, 2017b	Inconsistently uses a tool to check for sensitive data, other than privacy related data[1].	Consistently uses publication decision-flowcharts that include non-privacy risks[2].	Consistently uses publication decision-flowcharts that include non-privacy risks[3].
Control activities: *Control activities are measures to ensure the risk response policies and procedures are carried out effectively (COSO, 2004).*	No general observations.	The open data coordinator has the main responsibility for controls. There is a transition planned to transfer some responsibilities for controls from the open data	Does not have no explicit controls aside from the stand procedure for open data publication. The publishing process is designed in such a way that missing steps would hinder	has Implemented internal controls conducted by the municipal accountant to monitor the collection, processing and opening of data from the base registrations. These controls are mainly

[3] Gemeente Haarlem, personal communication, February 7, 2018

[2] Gemeente Den Haag, personal communication, January 31, 2018

		coordinator to the department of information facilities. Before this transition the controls on the following of procedures are the responsibility of one employee. The expansion of open data activities created risks concerning the controllability. This is the reason for the transfer of ownership towards the department of information facilities[1].	the publication process. The administrative unit, amongst themselves are aware when diversions from the normal process occur. Ensuring the normal publication process lies solely with the unit. The use of the online tool allows the data owner to decide if data is appropriate to publish and by sending in the results the administrative unit has access to the results from the online tool[2].	focussed on proper maintenance of the data in the base registration. This is a control activity on open data because some data from base registrations can be published as open data. Haarlem also implemented a data management system where the data owner reports activities and the registration owner reviews if all necessary activities have been completed[3].
Information & communication: *Relevant information that enables employees to carry out their responsibilities, is identified and communication through the organization. The information is communicated timely and in a manner that contributes to people carrying out their responsibilities (COSO, 2004).*	In order for the data owner to invest time and possible resources to do fulfil their tasks they need to understand the importance of (open) data management. All three municipalities have gone through or are still going through a change in mindset among employees to see the value of data and open data. Employees need to embrace the value of open data in order for them to invest their time	Progress has been made over the last few years. There is still room for improvement when it comes to creating an organization-wide data-oriented mindset[1].	Open data is a generally unknown topic. There is little awareness regarding the existence of the administrative unit open data. The administrative unit is generally working independently from other departments. A flow of identifying and sharing relevant information has not been established[2].	Progress has been made over the last few years. There is confidence that the mindset and culture have sufficiently changed[3].

[1] Gemeente Utrecht, personal communication, December 13, 2017; Gemeente Utrecht, personal communication, January 15

[2] Gemeente Den Haag, personal communication, January 31, 2018

[3] Gemeente Haarlem, personal communication, February 7, 2018

	and resources into open data activities [1];[2];[3]			
Monitoring: *After implementing procedures to mitigate the risks an organisation implements controls to make sure the procedures are actually carried out (COSO, 2004).*	Municipalities are legally required to have an accountant review their annual report (Gemeentewet, 1992, art. 213). This legal requirement was introduced to review how the municipalities manage financial risks.	The monitoring of policies and procedures does not take place through an embedded method. Based on the developments and the progress that is made regarding open data policies and procedures are modified[1].	The open data publication procedure in needs to be followed to ensure proper preparation and publication of open data. Deviations from this procedure and the motivations to deviate from the procedure are monitored among the members of the administrative unit open data. The Hague also set-up an online questionnaire to evaluate if a dataset can be published as open data. The results of this questionnaire are saved and available for review by the administrative unit[2].	The annual report includes some other controls in their review for the year 2017. This review included the implementation of the Gdpr and controls on cybersecurity (Gemeente Haarlem, 2017c). Haarlem is working on adopting the ISO27002 standards into a municipal compliance tool used for base registrations and norms like the Gdpr. This is a specific internationally acknowledged control standard for information management (Forum standaardisatie, 2018a). Haarlem also introduced a new policy for open data publication in 2017. The policy will not be modified in the near future to preserve some stability in the work methods. Haarlem also reports the achievement concerning the data warehouse in the annual report of the municipality[3].

[1] Gemeente Utrecht, personal communication, December 13, 2017; Gemeente Utrecht, personal communication, January 15

[2] Gemeente Den Haag, personal communication, January 31, 2018

[3] Gemeente Haarlem, personal communication, February 7, 2018

4.4 Conclusion

This chapter presented the results of the research. The results were discussed per sub question. The publishing activities of Dutch municipalities and the noteworthy cases were presented first. Followed by the implemented open data policies and procedures. And lastly, the policies and procedures were compared to prevalent risk management methodology.

5 Conclusions

This chapter discusses the final conclusions based on the results that have been presented in the previous chapter. First, the research question will be answered. The second part of this chapter concerns the discussion and the third part discusses recommendations.

5.1 Conclusion

Using the data that has been presented in the previous chapter, the following research question can be answered:

How have Dutch Municipalities, that are noteworthy regarding the publication of open data, designed and implemented comprehensive open data policies and procedures to protect citizens' privacy when they publish open data? And to what extent does this design and implantation integrate prevalent risk management framework methodology?

5.1.1 Noteworthy municipalities

Several noteworthy municipalities have been identified. These municipalities are: Amsterdam, Utrecht, Rotterdam, The Hague, Haarlem, Eindhoven and Leeuwarden. After further inquiry three municipalities agreed to participate. Participating municipalities were: Utrecht, The Hague and Haarlem.

5.1.2 Application of open data policies

Open data policies are part of a broader data management strategy or policy in all three municipalities. Open data publication is a component of general data management policies (Gemeente Utrecht, 2014a; Gemeente Haarlem, 2017a; Gemeente Den Haag, 2011). Preventing unlawful publication of personal data is one of the most important aspects of the publication process. Procedures include a description of the necessary steps prior to publication. Responsibilities for carrying out of the necessary actions are assigned based on roles or function (Gemeente Utrecht, 2017; Gemeente Den Haag, 2016a; Gemeente Haarlem, 2017b). Procedures used in The Hague and Haarlem also include a check for unlawful or otherwise sensitive data unrelated to privacy, for example sensitive data concerning security or market competition. Utrecht uses the Ethical Data Assistant; a tool covers also covers data unrelated to privacy and ethical considerations. This tool is more extensive than the flowcharts used by Haarlem and The Hague. However, this tool is not structurally used and it is expresses that there are intentions to increase the use of the tool (Gemeente Utrecht, personal communication, January 15, 2018; DEDA, 2017).

5.1.3 Extent of implementation of risk management methods in open data policies

The policies and procedures regarding open data publication do not explicitly integrate prevalent risk management methodology. There is little collaboration between risk management or audit activities and development of open data policies (Gemeente Utrecht, personal communication, January 15, 2018; Gemeente Den Haag, personal communication, January 31, 2018; Gemeente Haarlem, personal communication, February 7, 2018). The policies and procedures do share elements similar to the components of the COSO framework. Municipalities adopted general data policies with general objectives. Open data policies are one part of this general data policy. The open data policies acknowledge the privacy risks and are mainly focussed on mitigating the privacy risks. These policies include procedures and roles for the selection, publication and updates for open data. Some of the elements are documented in policies and other aspects are not documented but are implemented work methods.

The privacy risks connected to open data are informally acknowledged in policies and procedures of all municipalities (Gemeente Utrecht, 2014a; Gemeente Haarlem, 2017a; Gemeente Den Haag, 2011). However, these risks are not formally identified and assessed and listed with other formally identified risks. The risks, both privacy related and not privacy related, are not formally identified and qualified. One exception to this is the risk of personal data breaches. This risk is formally identified and qualified by the municipalities Utrecht and Haarlem. However, this is an organization-wide risk not exclusively related to open data publication (Gemeente Utrecht, personal communication, December 13, 2017; Gemeente Haarlem, 2016). Another disparity between a component of the COSO framework and the implemented polices is monitoring. Monitoring and controls are not all described in policies and procedures. Based on the interview, it can be concluded that several monitoring and controls measures are implemented in work methods (Gemeente Haarlem, personal communication, February 7, 2018; Gemeente Haarlem, personal communication, February 7, 2018).

5.1.4 General conclusions

All three municipalities take different approaches to open data publication. Utrecht puts more focus on societal challenges whereas Haarlem focusses on creating self-service to use public records like register of addresses and buildings. Both municipalities have achieved a change in culture of the organization to be more aware of the value of (open) data (Gemeente Utrecht, personal communication, December 13, 2017; Gemeente Haarlem, personal communication, February 7, 2018). The Hague has rebooted open data activities and is working to rebuild the open data publication activities. There is still little awareness within the organization about the open data activities (Gemeente Den Haag, personal communication, January 31, 2018). Open data policies and procedures are based on the methods of other municipalities and participation in regional and national pilots, projects and forums. One example of methods of other municipalities, is the procedure for publishing open data that was developed by the municipality of Rotterdam. This procedures was used as a base for the procedures of Haarlem and Utrecht (Gemeente Utrecht, personal communication, December 13, 2017; Gemeente Den Haag, personal communication, January 31, 2018; Gemeente Haarlem, personal communication, February 7, 2018).

One important risk is re-identification of individuals through combining several datasets. Municipalities do take measures to minimize the risk of re-identification but have not formally assessed this risk (Gemeente Utrecht, personal communication, December 13, 2017; Gemeente Haarlem, personal communication, February 7, 2018; Gemeente Den Haag, personal communication, January 31, 2018).

The risk of re-identification is difficult to assess. Municipalities take measures to make sure that published datasets cannot be used to unmask personal data. However, it is almost impossible to completely eliminate the possibility unmasking personal data (Meijer, Conradie & Choenni, 2014). And the legal consequences and responsibilities, if such unmasking would occur, are difficult for individual municipalities to assess. The Gdpr sets a several conditions for processors of personal data. However, there is uncertainty if there are limits to the legal responsibilities of processors in the event of unmasking and what these limits are (Gemeente Haarlem, personal communication, February 7, 2018; Gemeente Den Haag, personal communication, January 31, 2018).

When the implemented policies and procedures are compared to the COSO framework it appears that these policies have not been designed from a risk management perspective. However, these policies do share several elements of the COSO framework. The policies set out objectives, work methods that are aimed at preventing personal data breaches and establish monitoring measures. The COSO framework is a prevalent methodology that serves as a helpful comparative tool to analyse municipal open data policies. It addresses the different elements of risk mitigating policies such as: objective setting, risk assessment, risk response and monitoring. The framework also includes also the implementation of policies on the different entity levels and identifies several different types of objectives. It provides guidelines on implementing risk management activities. The framework does not necessarily needs to be implemented fully in order to assure privacy protection concerning open data publication. Every organization can integrate (parts of) the framework in a manner that matches best with the activities of the organization (COSO, 2004). Therefore, unintegrated COSO components do not necessarily indicate insufficient risk mitigating measures.

5.2 Limitations and further research

Initially this thesis included a survey among employees that worked with open data policies and procedures in practice. This survey was meant to collect supporting evidence of the application of formulated policies. This would help to establish existence of the policies and procedures. This survey collected too little response to provide reliable data. The survey was published online and the link to the survey was distributed among employees by the official that was interviewed. After three weeks a reminder was sent. However, the response did not increase and no reliable data could be collected. Due to the removal of the survey from of the research it is difficult to determine to what extent the implemented policies and procedures are ensured. The interviews and procedures could provide some indications on the extent of ensured procedures. However, this is too little evidence to based conclusions about existence of policies and procedures on.

This research focussed on the mitigation of privacy risks when municipalities publish open data. This is a limited view on all the possible risks not only regarding open data but data management in general. In all three cases open data policy is part of a broader data management policy. The scope of this thesis was limited and therefor might give too little attention to strengths and weaknesses of the comprehensive data management. The privacy risks can appear to be more urgent, especially with the Gdpr officially having taken effect. The non-privacy risks are just as important. These risks are discussed less in policy document and are not formally identified. The research was limited to risk management measures regarding open data policies and procedures. The full scope of the risk management activities of the participating municipalities were not included in the research. All three municipalities do have an audit service or other department with the task financial and internal control. These services or departments are responsible for financial and internal control. A review of the complete risk management framework was not part of the research. The results on application of risk management methodology do not have to be representative for other risk management efforts within the municipality. The documents that have been analysed were open data policies. Formal risk management was part of these policies in a very limited way. A review of the work methods and methodologies used by audit and internal control services might have provided more insight on how risk management is embedded in municipal policies.

5.3 Recommendations

The purpose of the research is to make recommendations to improve protection of citizens' privacy regarding open data. The recommendations are aimed at making the publication of open data safer regarding the privacy of citizens.

The first recommendation is to increase shared learning. Municipalities already share knowledge on different platforms and in several pilots. However, municipalities could increase shared learning on the level of specific procedures or policies. All three municipalities have implemented measures that could be recommended for other municipalities or organizations to implement:

- Haarlem uses a data management system where the data owner needs to report the steps that have been taken making it easy for the registration owner to monitor if all necessary steps have been taken (Gemeente Haarlem, personal communication, February 7, 2018).
- Utrecht has the most complete checklist to determine if data is appropriate to publish. This checklist was developed in collaboration with Utrecht Data School and University Utrecht but can be used by any municipality. The Ethical Data Assistant (DEDA) includes both privacy and non-privacy issues and includes ethical considerations (DEDA, 2017).
- The Hague uses an online tool as a decision flowchart. The online tool makes it prevents skipping steps or questions and all answers are saved. The administrative unit can review the results. This facilitates monitoring (Gemeente Den Haag, 2016b).

Increasing shared learning can help municipalities improve by learning from other organizations instead of every organization having to develop their own measures. Shared learning should not only be limited to municipalities, instead shared learning should cross over to other organizations working on open data.

The second recommendation is to formally identify and classify all the risks that are associated with open data publication. There is awareness of the risks of open data however these risks are not formally identified and classified. One important risk in particular is that of re-identification of individuals through combining datasets. All three municipalities are aware of the possibility of re-identification through a combination. However, there is a lot of uncertainty around this risk. The likelihood is unknown, the impact is unknown and the legal implications are unknown. It is recommended that more research to be done on the likelihood and impact of re-identification through a combination of datasets that includes open data published by either municipalities or other public authorities. Aside from this particular risk, there are possibly more risks either related to privacy or not related to privacy. The identified risks may vary between municipalities due to different objectives that have been set with their open data activities. It is recommended that municipalities identify the specific risks that could hinder the achievement of objectives.

The second recommendation ties into the first recommendation of shared learning. It appears that municipalities are not able to fully interpret the (legal-) liability if a re-identification, using municipal open data, would occur (Gemeente Haarlem, personal communication, February 7, 2018; Gemeente Haarlem, personal communication, February 7, 2018). The Gdpr sets standards for a processor of personal data, however, it is still unclear to what extent a processor remains responsible if a third party would misuse open data. It is expected that individual municipalities or organizations are not able to conduct the necessary (legal) analysis to interpret risks of re-identification. Arguably, it would not be desirable for every municipality or organization to have to conduct a such an extensive review of the risk of re-identification. Instead, it is recommended that the risk of re-identification is classified through collaboration between publishers of open data, legal experts, policy makers and other relevant parties on the topic of open data. A cooperation between these parties is expected to bring together different field of expertise, share expertise and partly relieve individual parties of having to gather information individually. Possibly, such a cooperative effort could be assisted by and overarching party.

The third recommendation concerns identifying specific and measurable objectives regarding open data publication. Clear objectives for open data publication will help with identifying and classifying risks. The approaches to open data publication can vary between municipalities and different approaches can be based on different objectives. Without clear objectives it can be difficult to determine when open data activities have been successful. Openness and transparency are generally set objectives; however, these objectives alone are difficult to measure and open to interpretation. Municipalities can take different approaches and publish different types of data. Specific and measurable objectives serve as guidelines alongside of broad objectives of creating more transparency and stimulating innovation. Measurable objectives cover an intended number of data sets over a specified period or concern contacts with end-users of the data. General strategies concerning open data can vary between municipalities and measurable objects depend on the strategies the municipalities have formulated.

The fourth recommendation is to review the strength of monitoring measures. Monitoring measures are meant to ensure that the designed procedures are followed and to be able to identify the instances when the procedures are not followed. The monitoring measures were not part of the analysed documentation. They were discussed in the interviews. However, due to the design of the research it is difficult to determine to what extent the implemented policies and procedures are ensured. Monitoring is an essential element to ensuring that privacy protecting policies are enforced properly. Monitoring can concern all the different stages of open data publication from selecting datasets to updating published data. The monitoring measures depend on how an organization has designed open data policies and procedures. Automatic updates of datasets require a different approach compared to updates that periodically done by employees.

6 References

Alamgir Hossain, M., Dwivedi, Y.K., & Rana, N.P. (2016). State-of-the-art in open data research: Insights from existing literature and a research agenda. *Journal of Organizational Computing and Electronic Commerce, 26(1),* 14-40. doi: 10.1080/10919392.2015.1124007

Alashwal, A.M., Abdul-Rahman, H., & Asef, A. (2017). Influence of Organizational Learning and Firm Size on Risk Management Maturity. *Journal of Management in Engineering, 33(6).* doi: 10.1061/(ASCE)ME.1943-5479.0000553

Algemene Rekenkamer (2016). *Trendrapport open data 2016.* Retrieved from: https://www.rekenkamer.nl/publicaties/rapporten/2016/03/24/trendrapport-open-data-2016

Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government information quarterly, 32,* 399-418. doi: http://dx.doi.org/10. 1016/j.giq.2015.07.006

Autoriteit Persoonsgegevens A. (2017a). *Algemene Verordening gegevensbescherming.* Retrieved from:https://autoriteitpersoonsgegevens.nl/nl/onderwerpen/europese privacywetgeving/algemene verordening-gegevensbescherming

Autoriteit Persoonsgegevens B. (2017b). *Meldplicht datalekken*. Retrieved from: https://autoriteitpersoonsgegevens.nl/nl/onderwerpen/beveiliging/meldplicht-datalekken

Autoriteit Persoonsgegevens C. (2017c). *Artikel 16 Wbp*. Retrieved from: https://autoriteitpersoonsgegevens.nl/nl/over-privacy/wetten/wbp-naslag/hoofdstuk-2 voorwaarden-voor-de-rechtmatigheid-van-de-verwerking-v-30

Badrul, N.A., Parslow, P., Lundqvist K. O., & Williams, S.A. (2016). Investigating Employees' Understanding of the Concept of Privacy for the Open Government Initiative. *2016 7th International Conference on Computer Science and Information Technology (CSIT)*. doi: 10.1109/CSIT.2016.7549446

Bargh, M.S., Choenni, S., & Meijer, R. (2017). On addressing privacy in disseminating judicial data: towards a methodology. *Transforming Government: people, process and policy, 11(1),* 9-41. doi: 10.1108/TG-12-2015-0051

BIS (Department for Business, Innovation & Skills) & DCMS (Department for Culture, Media & Sport). (2009). Digital Britain. *Building Britain's Future.* Retrieved from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/228844/7650. df

Brooks, s., Garcia, M., Lefkovitz, N., Lightman, S., & Nadeau, E. (2017). An Introduction to Privacy Engineering and Risk Management in Federal Systems. *National institute of Standards and Technology*. doi: https://doi.org/10.6028/NIST.IR.8062

CBS. (2017). *Informatie voor gemeenten.* Retrieved from: https://www.cbs.nl/nl-nl/dossier/nederland regionaal/informatie-voor-gemeenten

COSO. (2004). Enterprise Risk Management- Integrated Framework. *Committee of Sponsoring Organisations of the Treadway Commission (COSO)*. Retrieved from: *https://www.coso.org/Documents/COSO-ERM-Executive-Summary.pdf*

Creative Commons. (2017a). *Frequently asked questions.* Retrieved from: https://creativecommons.org/faq/#can-governments-and-intergovernmental-organizations igos-use-cc-licenses

Creative Commons. (2017b). *Creative commons licences.* Retrieved from: https://creativecommons.org.nz/licences/licences-explained/

Data.overheid. (2017). *Dataportaal van de Nederlandse Overheid.* Retrieved from: https://data.overheid.nl//

Dataplatform. (2018a). *Over Dataplatform.* Retrieved from: https://www.dataplatform.nl/over dataplatform

Dataschool. (2018). *Deda*. Retrieved from: https://dataschool.nl/deda/

DEDA. (2017). *De Etische Data Assistent.* Utrecht Data school, march 2017, Utrecht.

European Commission. (2010). Riding the wave, How Europe can gain from the rising tide of scientific data. Retrieved from: http://ec.europa.eu/information_society/newsroom/cf/itemlongdetail.cfm?item_id=6204

European Commission. (2011). Communication, Open data, an engine to innovation, growth and transparent governance. *COM(2011) 882*. Retrieved from: http://eur lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0882:FIN:EN:PDF

Forum Standaardisatie. (2018a). *NEN-ISO/IEC 27002*. Retrieved from: https://www.forumstandaardisatie.nl/standaard/nen-isoiec-27002

Forum Standaardisatie. (2018b). *Pas-toe-of-leg-uit.* Retrieved from: https://www.forumstandaardisatie.nl/lijst-open-standaarden/in_lijst/verplicht-pas-toe-leg-uit

Gemeente Den Haag. (2004). *Uitvoeringsbesluit Audit Comittee Gemeente Den Haag*. College van Burgemeester en Wethouders, BSD/2003.2806 – RIS 112318.

Gemeente Den Haag. (2011). *Beschikbaar stellen 'Open data' door de gemeente Den Haag.* Wethouder van Onderwijs en Dienstverlening Wethouder Volksgezondheid en Zorg, Duurzaamheid, Media en Organisatie, PBS/2011.255 – RIS 181617.

Gemeente Den Haag. (2013). *Voortgang open data.* Wethouder van Volksgezondheid, Duurzaamheid, Media en Organisatie, BSD/2013.1432 - RIS 268722

Gemeente Den Haag. (2016a). Beslisproces. Internal documentation.

Gemeente Den Haag. (2016b). Open data beslisboom Versie 1.0. Retrieved from: https://docs.google.com/forms/d/e/1FAIpQLScq9 EK4uovdimvvFjOlwIpzDUfF9r5ie1KGODuGVj7jeCI_w/viewform?c=0&w=1

Gemeente Den Haag. (2017a). *Antwoord van het college op de vragen van het raadslid mevrouw Zandstra, luidend "Open Data".* Beantwoording schriftelijke vragen, BSD/2017.177, RIS296495.

Gemeente Den Haag. (2017b). *Gemeentelijke Accountantsdienst (GAD).* Retrieved from: https://www.denhaag.nl/nl/algemeen/gemeentelijke-accountantsdienst-gad.htm

Gemeente Haarlem. (2016). *Jaarverslag en Jaarrekening 2016 Gemeente Haarlem.* Retrieved from: https://gemeentebestuur.haarlem.nl/bestuurlijke-stukken/2017155018-3-Jaarverslag-en Jaarrekening-2016-1.pdf

Gemeente Haarlem. (2017a). *Bijlage 1: Nota Haarlems gegevensmanagement.* Retrieved from: https://gemeentebestuur.haarlem.nl/bestuurlijke-stukken/2017183621-2-Bijlage-1-Nota Gegevensmanagement-3.pdf

Gemeente Haarlem. (2017b). *Bijlage 2: Handreiking publicatie open data.* Retrieved from: https://gemeentebestuur.haarlem.nl/bestuurlijke-stukken/2017183621-3-Bijlage-2 Handreiking-Publicatie-open-data-3.pdf

Gemeente Haarlem. (2017c). *Rapportage interim-bevindingen 2017, Samen het verschil maken*. Internal document

Gemeente Rotterdam. (2017). *Beslisboom Rotterdam open data.* Retrieved from: http://beslisboom.rotterdamopendata.nl/beslisboom_rotterdam_open_data.pdf

Gemeente Utrecht. (2014a). Aanpak Data gedreven sturing & Open Data, *College van Burgemeester en Wethouders.* 19-12-2014. 14.503919.

Gemeente Utrecht. (2014b). *Plan van aanpak indexeren gemeentelijke datasets.* Retrieved from: https://online.ibabs.eu/ibabsapi/publicdownload.aspx?site=utrecht&id=14130

Gemeente Utrecht. (2015). Nota Risicomanagement en Weerstandsvermogen 2015-2018. *Bestuurs- en Concernstaf: Concernmanagement Financien en Control*.

Gemeente Utrecht. (2017a). *Proces Data catalogus.* Internal document gemeente Utrecht

Gemeente Utrecht. (2017b). Framework Privacy by Design. Internal document gemeente Utrecht

Green, B., Cunningham, G., Ekblaw, A., Kominers, P., Linzer, A., & Crawford, S. (2017). Open Data Privacy, A risk-benefit, process-oriented approach to sharing and protecting municipal data. *Berkman Klein Center Research Publication.* Retrieved from: https://dash.harvard.edu/handle/1/30340010

Haarlem open data. (2018). *Proclaimer.* Retrieved from: https://opendata.haarlem.nl/public/portal.html#collapseProclaimer

Hader, I., Hasson, T., Ayalon, O., Toch, E., Birnhack, M., Sherman, S., & Balissa, A. (2017). Privacy by designers: software developers' privacy mindset. *Empirical Software Engineering, 22,* 1 31. doi: 10.1007/s10664-017-9517-1

Hardy, K., & Maurushat, A. (2017). Opening up government data for Big Data analysis and public benefit. *Computer Law & Security Review, 33(1),* 30-37. doi: https://doi.org/10.1016/j.clsr.2016.11.003

Hillson, D.A., & P. Simon. (2007). *Practical Project Risk Management*. The ATOM Methodology. Vienna, VA: Management Concepts

ISO (ed.). (2009). *Risk Management - Principles and Guidelines, ISO 31000:2009*. International Standard, International Organization for Standardization, retrieved from: https://www.iso.org/standard/43170.html.

ISO. (2005). *International standard ISO/IEC 27002- Information technology, Security techniques, Code of practice for information security management*. International Standard, International Organization for Standardization, retrieved from: http://www.slinfo.una.ac.cr/documentos/EIF402/ISO27001.pdf

Janssen, M., & Van den Hoven, J. (2015). Big and Open Linked Data (BOLD) in government: A challenge to transparency and privacy? *Government Information Quarterly, 32,* 363-368. doi: http://dx.doi.org/10.1016/j.giq.2015.11.007

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012) Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management, 29(4)*, 258-268, doi: 10.1080/10580530.2012.716740

Keenan, T.P. (2012). Are They Making Our Privates Public? – Emerging Risks of Governmental Open Data Initiatives. *Privacy and Identity Management for Life, 375*, 1-13. doi: 10.1007/978-3 642-31668-5_1

Kuk, G., & Davis, T. (2011), "The roles of agency and artifacts in assembling open data complementarities", Proceedings from 32nd International Conference on Information Systems (ICIS), Shanghai, China.

Kusera, J., & Chlapek, D. (2014). Benefits and Risks of Open Government Data. *Journal of Systems Integration, 5(1):* 30-41. doi: 10.20470/jsi.v5i1.185

Lavrenovs, A., & Podins, K. (2016). Privacy Violations in Riga Open Data Public Transport System. *4th IEEE Workshop on Advances in Information, Electronic and Electrical Engineering.* doi: 10.1109/AIEEE.2016.7821808

Mantelero, A. (2017). Regulating big data. The guidelines of the Council of Europe in the context of the European data protection framework. *Computer Law & Security Review, 2017.* doi: http://dx.doi.org/10.1016/j.clsr.2017.05.011

Meijer, R., Conradie, P., & Choenni, S. (2014). Reconciling Contradictions of Open Data Regarding Transparency, Privacy, Security and Trust. *Journal of Theoretical and Applied Electronic Commerce Research, 9(3),* 32-44. doi: 10.4067/S0718-18762014000300004

Ministerie van Binnenlandse Zaken. (2015). *Actieplan Open Overheid 2016-2017.* Retrieved from: https://www.rijksoverheid.nl/documenten/rapporten/2015/12/01/actieplan-open-overheid 2016-2017

Ministerie van Binnenlandse Zaken. (2017). *Over open data.* Retrieved from: https://data.overheid.nl/over-open-data

Ministerie van Binnenlandse Zaken. (2018). *Gemeentelijk high value datalijst.* Retrieved from: https://data.overheid.nl/gemeentelijke-high-value-datalijst

Ministerie van Economische Zaken. (2016). *Digitale agenda, vernieuwen, vertrouwen, versnellen.* Retrieved from: https://www.rijksoverheid.nl/documenten/rapporten/2016/07/05/digitale agenda-vernieuwen-vertrouwen-versnellen

Narayanan, A., & Shmatikov, V. (2008) Robust De-anonymization of Large Sparse Datasets. *2008 IEEE Symposium on Security and Privacy*, 111-125. doi: 10.1109/SP.2008.33

Open Government Partnership. (2017). *About OGP.* Retrieved from: https://www.opengovpartnership.org/about/about-ogp

Open overheid. (2017). *Open data.* Retrieved from: http://www.open-overheid.nl/open-data/

Oulasvirta, L., & Anttiroiko, A.V. (2017). Adoption of comprehensive risk management in local government. *Local Government Studies, 43(3),* 451-474. doi:10.1080/03003930.2017.1294071

Plasterk, R.H.A., (2015). *Actieve beschikbaarstelling van overheidsinformatie*. Retrieved from: https://data.overheid.nl/sites/default/files/aanbiedingsbrief-inzake-actieve-beschikbaarstelling overheidsinformatie.pdf

Privacyverordening gemeente Utrecht. (2016). Retrieved from: https://zoek.officielebekendmakingen.nl/gmb-2016-135657.html

Project Management Institute. (2008). *A Guide to the Project Management Body of Knowledge (PMBOK® Guide)*. 4th ed. Newtown Square, PA: PMI.

Rijksoverheid. (2017). *Beleid open data*. Retrieved from: https://data.overheid.nl/beleid-open-data-1

Ruijer, H.J.M. (2017). Proactive Transparency in the United States and the Netherlands: The Role of Government Communication Officials. *American Review of Public Administration, 47(3),* 354-375. doi: 10.1177/0275074016628176

Simperl, E., O'Hara, K., & Gomer, R. (2016). Analytical Report 3: Open Data and Privacy. *European Data portal.* Retrieved from: https://www.europeandataportal.eu/sites/default/files/open_data_and_privacy_v1_final_clean pdf

Solove, D.J. (2002). Conceptualizing Privacy. *California Law Review, 90(4),* 1087-1155. doi: https://doi.org/10.15779/Z382H8Q

Sunlight Foundation. (2010). *Ten Principles for Opening Up Government Information*. Retrieved from: https://sunlightfoundation.com/policy/documents/ten-open-data-principles/

TU Delft. (2013). *De mogelijkheid van een open data beleid voor het Actueel Hoogtebestand Nederland nader onderzocht.* Retrieved from: uuid:e106773c-33e6-4ad0-ae42-064761e36346

Van Dijk, N., Gellert, R., & Rommetveit, K. (2016). A risk to a right? Beyond data protection risk assessments. *Computer Law & Security Review, 32(2),* 286-306. doi: 10.1016/j.clsr.2015.12.017

Viale Pereira, G., Macadar, M. A., Luciano, E. M., & Testa, M. G. (2016). Delivering public value through open government data initiatives in a Smart City context. *Information Systems Frontier, 19,* 213-229. doi:10.1007/s10796-016-9673-7

Warren, S. and Brandeis, L. (1890), "The right to privacy", *Harvard Law Review,4 (5),* 193-220.

Weerakkody, V., Irani, Z., Kapoor, K., Sivarajah, U., & Dwivedi, Y.K. (2016). Open data and its usability: an empirical view from the Citizen's perspective. *Information Systems Frontier, 19,* 285-300. doi: 10.1007/s10796-016-9679-1

Welle Donker, F., & Van Loenen, B. (2016). How to assess the success of the open data ecosystem? *International Journal of Digital Earth, 10(3),* 284-306. doi: 10.1080/17538947.2016.1224938

Wet bescherming persoonsgegevens (2016). *Meldplicht datalekken Wet bescherming persoonsgegevens.* Retrieved from: http://wetten.overheid.nl/BWBR0037346/2015-12-16

Wet hergebruik van overheidsinformatie. (2016). Retrieved from: http://wetten.overheid.nl/BWBR0036795/2016-10-01

Wieczorek-Kosmala, M. (2014). Risk management practices from risk maturity models perspective. *Journal of East European Management Studies, 19(2), 133-159.* doi: 10.1688/JEEMS-2014 02-Wieczorek-Kosmala

Zuiderwijk, A, & Janssen, M. (2014). The negative effects of open government data- investigating the dark side of open data. *Proceedings of the 15th Annual International Conference on Digital Government Research,* 147-152. doi: 10.1145/2612733.2612761

Zuiderwijk, A., Janssen, M., Choennie, S., & Meijer, R. (2014). Design principles for improving the process of publishing open data. *Transforming Government: People, Process and Policy, 8(2),* 185-204. doi: 10.1108/TG-07-2013-0024

7 Appendix

Analysis framework

1. Open data context missie/visie/strategie	**1.1 Societal gains of open data**
	1.2 General risks of open data
	1.3 Privacy risks of open data
2. Legal framework	2.1 Framework of Laws and regulatory policies regarding the publication of personal data
	2.2 Summery/easily accessible legal information available to employees
	2.3 Description of personal data
3. Selection of potential data sets	3.1 Procedure for selection of potential open data sets
	3.2 Roles in selection of potential open data sets
4. Publication of open data (after selection)	4.1 Description of the procedure for publishing open data
	4.2 Description of roles in the open data publication process
	4.3 Privacy as a ground for refusal of publication
	4.4 Risks excluding privacy as grounds for refusal
5. Removal of Personal data	5.1 Description of personal data removal procedure
	5.2 Description of different anonymization and pseudonymizing techniques
	5.3 Check after personal data removal
6. Updating procedures for published data/ management of published data sets	6.1 Monitoring for traceability
	6.2 Procedures for regular updates for published data sets (at least once a year)

	6.3 Roles in regular update activities for published data sets
	6.4 Responsive to user feedback
7.1 Risk assessment methodology	7.1.1 Description of risk assessment methodology
	7.1.2 Description of identified privacy risks (including re-identification)
	7.1.3 Description of other identified risks
7.2 Risk management governance	7.2.1 Description of risk management roles
	7.2.2 Description of risk management procedures
8. Personal data breach response	8.1 Description of roles in case of personal data breaches
	8.2 Description of personal data breach procedures
9. Availability of open data and risk management documentation	9.1 Are documents on open data policy available to all employees
	9.2 Are documents regarding risk management available to all employees
	9.3 Is there awareness about open data policies among employees
	9.4 Is there awareness about open data risk management activities related to open data among employees

Reseach data of this thesis can be requested by contacting the author: www.linkedin.com/in/eva-peeters-weem-051865102
The research data contains privacy sensitive data and therefore is not directly available