GIMA
Geographical Information Management and Applications

Third wave open data in the Netherlands: Identifying and overcoming the barriers towards open geographical data of public undertakings

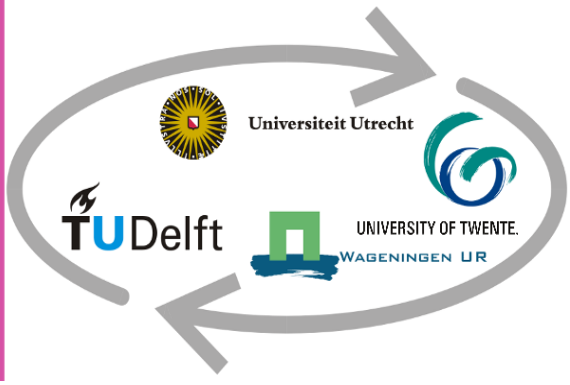
Frida Boone

February, 2021

Supervisors:

Bastiaan van Loenen - TU Delft

Peter van Oosterom – TU Delft



Third wave open data in the Netherlands: Identifying and overcoming the barriers towards open geographical data of public undertakings

Master's thesis

In Fulfilment

of the Requirements for the Degree of

the master Geographical Information Management and Application (GIMA)

By

Frida Boone

Under supervision of Bastiaan van Loenen and Peter van Oosterom

26 February, 2021

ABSTRACT

The creation of open data has seen a series of waves in which every growing resources of data are becoming accessible to a growing number of users from a diversifying number of public entities. The European Commission anticipates this movement by setting a new scope to the re-use of Public Sector Information Directive. Instead of exclusively focussing on Public Sector Information (PSI), the new scope of its successor, the Open Data Directive (ODD), includes data from public undertakings as well. In order for public undertakings to comply with this future legislation research into the current openness of public undertakings and the barriers to open data is key. This research presents three different levels of open data for public undertakings and it shows which barriers they face to achieve more open data. In this case the public undertakings are Port of Rotterdam (PoR) and Schiphol Airport. The results showed that the data policy of PoR matches the level where data is considered not to be open, level 1. The data policy of Schiphol Airport matches the level where data is considered partly open, level 2. For both public undertakings, the barriers that are associated with achieving a higher level of open data are related to institutional, financial, legal, and quality and technical aspects. The Dutch grid operator Liander, which has provided open data since 2014, presented possibilities to overcome these barriers based on own experience. The open data policy of Liander corresponds with the third level of open data as every user can re-use the data of Liander. It can be stated that neither PoR nor Schiphol Airport is ready to comply with the future rules when the ODD requirements become mandatory. Barriers still need to be overcome which start with the internal motivation to provide open data and by aggregating data to satisfy future legal requirements.

TABLE OF CONTEXT

List of abbreviations

ADP	Aéroports de Paris.
AIVD	Dutch General Intelligence and Security Service.
AM	Asset Management.
API	Application Programming Interface.
AVG	Algemene verordening gegevensbescherming.
BRO	Basis Registratie Ondergrond (Dutch base registration underground).
CAD	Computer Aided Design.
CC	Creative Commons.
CC BY	Creative Commons Attribution license.
CC BY SA	Creative commons Attribution and under the same terms.
CC NC	Creative Commons Non-Commercial use.
CC ND	Creative Commons no derivative works.
CC SA	Creative Commons ShareAlike.
CC0	Creative Commons Zero.
CDLA	Community Data License Agreement.
CSV	Comma-separated values file.
DEM	Digital Elevation Models.
DP	Port Development.
DWG	Drawing File.
EM	Environmental Management.
ENC	Electronic Navigation Charts.
FTE	Full- Time Equivalent.
GDPR	General Data Protection Regulation.
GIS	Geographical Information System.
HR	Human Resource.
HTML	Hypertext Markup Language.
ISO	International Organisation for Standardisation.
IT	Information Technologies.
MXD	Map Exchange Document.
NBNL	Netbeheer Nederland (Dutch electricity and gas network operators).
NWB	Nederlands Wegen Bestand (Dutch national road network dataset).
ODD	Open Data Directive.
OECD	Organisation for Economic Co-operation and Development.
OKF	Open Knowledge Foundation.
OSM	OpenStreetMap.
PDOK	Publieke Dienstverlening op de Kaart (Public Mapping Service).
POM	Port Objective Management.
PSI	Public Sector Information.
RDF	Resource Description Framework.
SDI	Spatial Data Infrastructure.
SPARQL	Semantic Query Language for databases.
USGS	United States Geological Survey.
XML	Extended Markup Languages.

Introduction

Chapter 1: Introduction

1 INTRODUCTION

Globally, open data has played an important role in creating social and economic opportunities, solving public problems and empowering citizens to make better decisions (Verhulst et al., 2020). An example of this is the United Kingdom, where heart surgeons of the National Health Service published comparable data on individual clinical outcomes in 2004. In 2011, improvements are reported; the survival rate increased by more than a third (Open Knowledge Foundation, 2011). Another example is Nepal where open data regarding aid flows – expressed in geographical information – have contributed to building a transparent and accountable public institution after the civil war (Open Data Institute, 2013). Likewise within the European Union, open data is considered important for socio-economic developments of the society (European Commission, 2020a). Recently, the lack of effective data use to address the COVID-19 virus shows that this important development still requires further work. In April 2020, 500 data practitioners and organisations over the world engaged in the 'Call for Action' by GovLab, a commercial big data think tank, to develop an open data infrastructure which is capable of challenge the pandemic and other dynamic threats (Verhulst et al., 2020).

Most of the data which is considered most valuable for tackling dynamic threats in the world is generated and held by the private sector – collected and controlled behind closed doors (Verhulst et al., 2020). Unlike such closed data, open data is often associated with government-held data, Public Sector Information (PSI). Re-use of PSI have fostered economic growth, especially in two sectors: legal and administrative, and geographical. The legal and administrative sector have changed their data policy by offering more PSI on the internet. This resulted in an increase of 40% in the market, hence a stable increase in income. In the geographical sector, the download of PSI data grew with 350% from 2002 to 2007 (Barbero et al., 2018). It is expected that the value of PSI in Europe will increase from €52 million in 2018 to €194 billion in 2030 (ibid.). Although the value of the PSI is likely to increase, it is acknowledged that it will not fulfil the demand for open data in the future. The demand is not only the data itself, it is the broader technical, social, political, and economic context within which the data is used. In order to answer public questions on dynamic threats data publicly obtained data needs to be open, central and incorporated into both public and private sector (European Commission, 2020a; Verhulst et al., 2020).

The growing demand for open data is starting to have an influence on the open data policy of the European Union. The scope of open data is not limited to public entities, but are being extended to private entities. One key area is geospatial datasets, which are regarded as datasets that are of high value to society (European Directive, 2013). In addition, recent developments towards open Spatial Data Infrastructures (SDI) highlight that non-governmental parties can enhance the SDI decision-making processes and add value to EU-wide information products (Vancauwenberghe et al., 2018).

In 2019 open data and the re-use of PSI was enacted in a new EU Directive, the Open Data Directive (ODD). The ODD provides a common legal framework for a European market for government-held data (European Commission, 2020a). It builds on the Directives of 2003 and 2013, that focused on the re-use of records from public organisations, including national archives and libraries (Directive (EU) 2013/37, 2013). The ODD also applies to documents held by public undertakings, research performing organisations and research funding organisations. These are non-government parties that collect, produce, reproduce and disseminate documents to provide services in the general interest (Directive (EU) 2019/1024, 2019; European Parliament, 1997). Most often the data policies within public undertakings are restricted, not open data policies. The provisions of the new Directive are not yet mandatory for public undertakings. However, it is expected that new legislation will be more strict in the future (Dalla Corte & van Loenen, unpublished). Since the European legislation acts as a leading

framework for Dutch legislation, potential changes towards more open data will also apply to public undertakings in the Netherlands. For the Netherlands to comply successfully with future legislation the challenge is to identify the barriers and means of tackling them for public undertakings to achieve an open data policy in the future.

1.1 IDENTIFICATION OF THE PROBLEM

The decision of the European Commission to move towards more open data policies – by including the private sector in their future scope – raises the question of how best to achieve this. The new legislation does not supersede previous legislation of the ODD, hence it needs to be taken into account; the past can guide the present.

Starting in the 1980s, continuing over the last four decades, the PSI Directive in the EU, since 2019 known as the 'Open Data Directive', went through an evolution, which is presented below (Dalla Corte & van Loenen, unpublished).

1980's

Re-use of PSI was considered an important element of the European data economy. In 1989 the Commission of the European Communities set up the Synergy Guidelines that provided a first stepping stone for the development of public sector information sharing and re-use within the EU.

2001

In 2001 the Commission followed up this effort by developing an EU framework for PSI re-use. The focus here was the economic aspects of PSI re-use rather than on the link to freedom of information.

2003

In 2003 the PSI directive was the main instrument in stimulating PSI re-use at EU level. It mandated re-use of documents held by public sector bodies in Member States for commercial or non-commercial purposes. Public sector bodies were required to process requests for re-use of documents and make them available for this purpose.

2008

In 2008 the commission was tasked to review the application of the PSI directive of 2003, to communicate the results to the European parliament and the councils, and propose any modifications. After consulting with relevant stakeholders the Commission found it too early to amend the 2003 PSI Directive and postponed the review to 2012 at the latest.

2013

In 2013 the PSI Directive was amended to introduce changes in the EU PSI legislation relating to its strength, its scope and to align partially PSI re-use and open data. In terms of strength, EU Member States had to ensure that documents to which it applied, were reusable for commercial or non-commercial purposes. Moving on from the 2003 directive, the public sector was obliged to make all PSI they processed available. The scope was expanded to documents in which libraries, museums, and archives held intellectual property rights but only when re-use of such documents was permitted.

2019

The Commission revised the 2013 PSI Directive. The commission of the European Communities agreed that the Open Data Directive had met policy objectives by stimulating the re-use of data beyond borders in Europe; they also identified some areas that still needed improvement (Dalla Corte, 2020). There were four main drivers that were tackled in the review:

- Insufficient access and re-use of dynamic data through an application programming interface (API).
- Market entry barriers and fragmentation.
- Insufficient availability of public and publicly funded data for re-use.
- Distortion of competition in the internal market.

The Commission's evaluation of the 2013 PSI Directive, the impact assessment and the consultation process led to the pursuit of low-intensity incremental change. The PSI directive had extended its scope towards data held by public undertakings in the transport and utilities sector. It could be expected that the current voluntary provision of open data of public undertakings would over time change to a legally mandated one (Dalla Corte, 2020; European Commission, 2020b).

In the Directive, data and documents that had a guaranteed socio-economic and environmental value were defined as high valued datasets (Directive (EU) 2019/1024, 2019). Re-use of these datasets must be free of charge and must be made available by public bodies. These datasets were associated with important benefits for society and economy and categorised as:

- Geospatial
- Earth observation and environment
- Meteorological
- Statistical
- Companies and company ownership
- Mobility.

Half of the categories are associated with geographical data. This research will examine the barriers in and around public undertakings to making geographical data openly available and how to overcome these in order to comply with foreseen future legislation. Figure 1 shows where the problems arise (red).

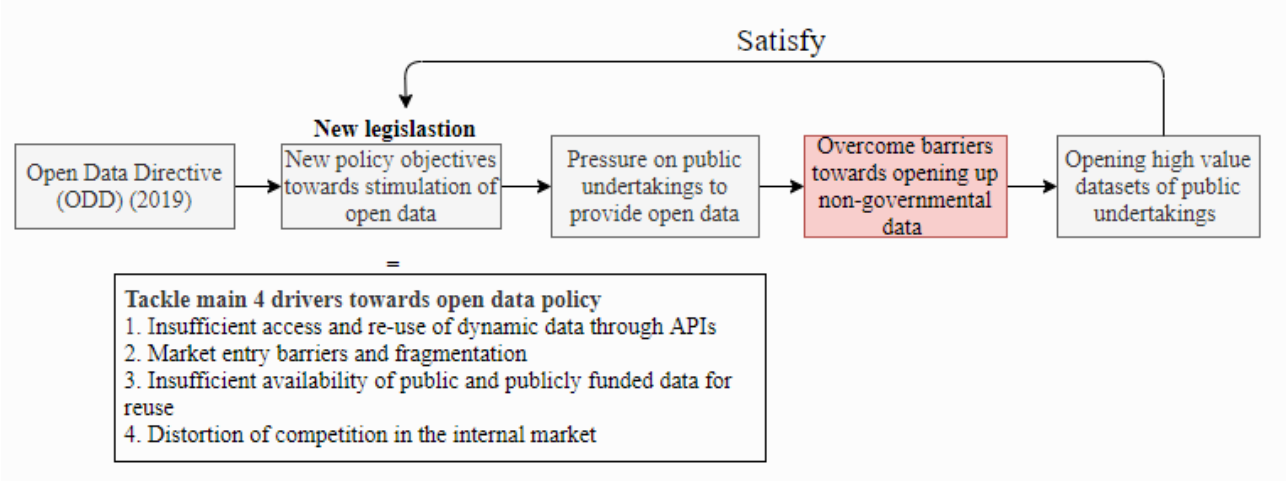


Figure 1. The situation explanation with the raised problem in red. It shows that overcoming barriers towards opening non-governmental data (high value datasets) will satisfy the new legislation proposed by the Open Data Directive (ODD) and the European Commission in 2019.

1.1.1 Relevance

Research has been done into to governmental bodies' overcoming barriers in order to achieve open data (Barry & Bannister, 2014; Janssen, Charalabidis, & Zuiderwijk, 2012; Martin et al., 2013); this has not been done from the perspective of public undertakings. The increased scope of the new Open Data Directive (Directive (EU) 2019/1024, 2019) to include public undertakings presents an opportunity for the similar research to be conducted with public undertakings as the focus.

This research seeks to identify such barriers and develop strategies to enable public undertakings to provide open data. Society should benefit as a result of companies adhering to the content of this research. By providing public undertakings guidelines to adapt to changing legislation, thus enabling The Netherlands to comply with EU directive (European Commission, 2020b).

1.1.2 Scope

In order to make this research specific and fit for purpose a scope has been set to maximise this research's utility. This scope is presented in Figure 2.

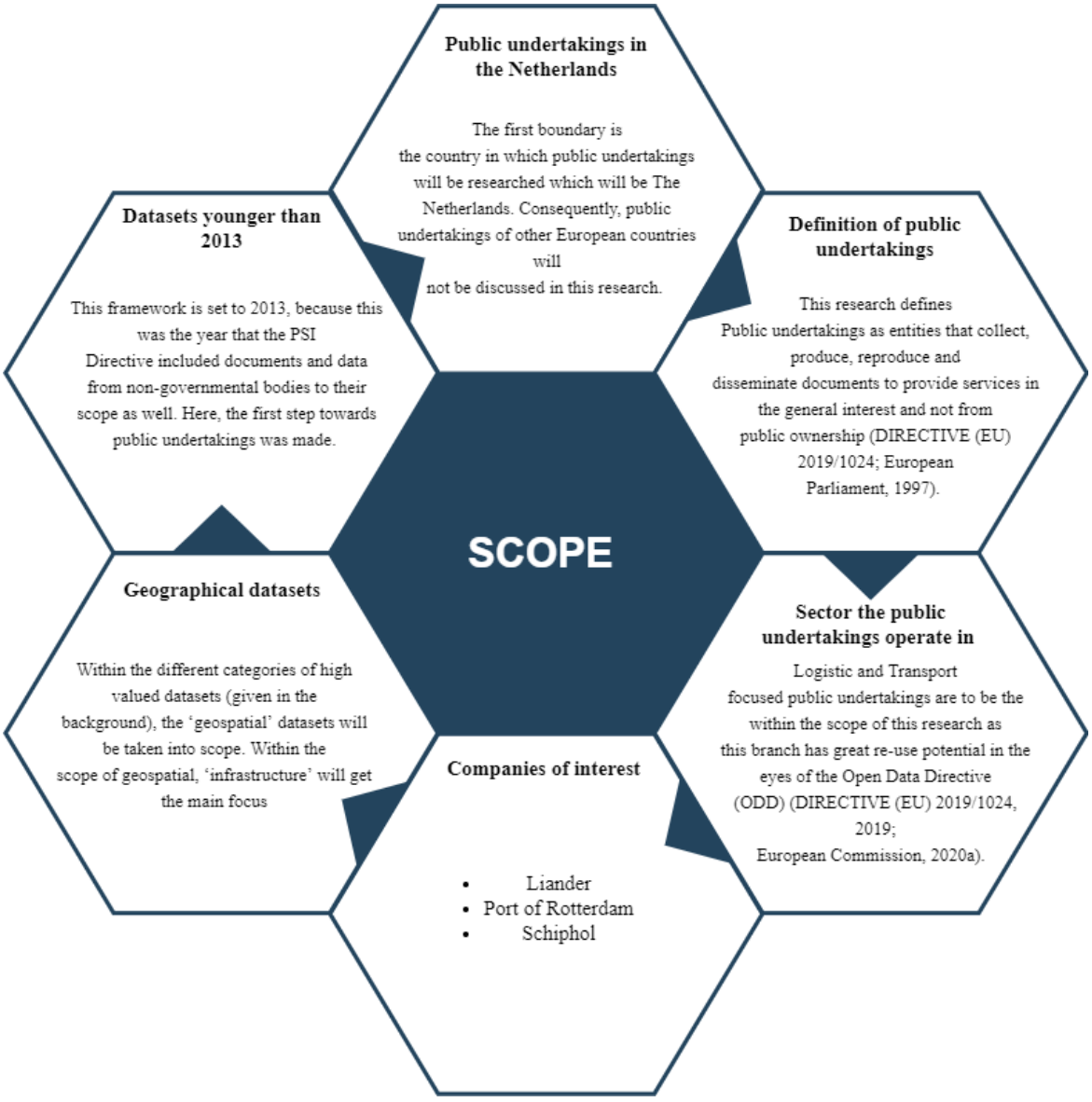


Figure 2. The scope set for this research.

In addition to the specific companies, some background information is provided. This research has been limited to Liander, Port of Rotterdam (PoR) and Schiphol Airport. Liander is a Dutch utility company that has been providing open data since 2014, hence was the leading example for this research. Both Port of Rotterdam and Schiphol operate in logistic and transport industry, which is considered as a potential branch for valuable data re-use by de ODD (Directive (EU) 2019/1024, 2019).

PoR can be defined as an unlisted public limited company. The shares are not listed on stock exchange but held by the Municipality of Rotterdam (70%,) and the Dutch government(30%). Port of Rotterdam is a data driven company with geospatial datasets of the Port. It is a public undertaking with high valued datasets. PoR provides open data on their Open Data Portal (data-portofrotterdam.opendata.arcgis.com), but only 17 datasets out of hundreds of datasets are openly available. These relate to infrastructure (1), weather (14) and harbour dues (2). Other valuable datasets on infrastructure in the water (wet infrastructure), logistic routes and many more are only accessible through a login with a PoR account and are not open publicly.

The third public undertaking that is included in the scope of this research is Schiphol Airport. Schiphol Airport is the biggest Dutch airport and is considered as a main port for logistics with high value for Europe. Schiphol is owned by the Royal Schiphol Group with the Dutch government, the municipalities of Amsterdam and Rotterdam and Groupe Aéroports de Paris (ADP), an investor of airports as main stakeholders. On their data portal (data.schiphol.nl), Schiphol holds datasets associated to Asset management, finance, HR, IT, operations, parking, passenger experience, real estate, retail, security and traffic and transport. However, not all datasets are openly available; to get access to all datasets a login with a secured Schiphol business account is needed.

1.1.3 Research objectives

The overall research objective of this thesis is to contribute to a solution dedicated to the perspective of public undertakings to meet expected future EU legislation towards open data. This is expressed as the research question:

"How can public undertakings in the Netherlands, similar to Liander, overcome the barriers to opening their geographical datasets in order to be prepared for expected future legislation towards open data?"

The next table presents the sub-research objectives and the sub-research questions.

Table 1. The research objectives and the linked research questions.

Research objective	Research question
1. Identify the different levels of openness for geographical data.	What are the different levels of openness for geographical data?
2. Identify the barriers towards opening geographical data for public undertakings.	Which barriers are faced by public undertakings to increase openness of their geospatial data?
3. Research how Liander succeeded in opening their data; which barriers did they meet and how were they overcome?	Which methods did Liander use to overcome barriers?
4. Discover which of the methods used by Liander can be successfully applied to two other public undertakings towards opening their data.	Which proposed methods used by Liander can help Port of Rotterdam and Schiphol to overcome the barriers towards open data?

1.2 METHODOLOGY

1.2.1 Identifying different levels of openness for geographical data

First the definition of open geographical data was researched by conducting a literature study and reviewing open geographical data from different perspectives. Definitions were given from the data provider perspective as well as from the data user perspective. In arriving at the definitions of open geographical data a holistic point of view, the data licenses, user types, data quality and data governance were considered. The main search terms used for the literature study were:

- Open (geographical) data.
- Availability of data.
- Levels of openness of data.

Documents retrieved from these research terms led to more literature for consideration, using the 'snowball method' for literature research. This literature study aided the creation of a multidimensional open data model as used in this research. This model presented three different levels of open data, all underpinned from the different perspectives. Extra input for this model was obtained in an online open data users event with the 'Ministry of Binnenlandse Zaken PSI users group', seeking findings on the openness of data in the Netherlands. Using the extra input the multi-dimensional model is used for the three remaining research objectives below.

1.2.2 Identifying the barriers towards opening geospatial data – for public undertakings

A literature review was conducted to identify the barriers between the different levels of open data of the multidimensional model. The main research terms used for the literature study were:

- Barriers of open data.
- Achieving open data.
- Open data for public undertakings.

As for the first literature review, documents were retrieved from these research terms; then more literature was found and used (the snowball method). This literature study helped identify the different barriers between the open data levels identified in the first research objective. Three semi-structured interviews were conducted with Port of Rotterdam personnel and one interview with two interviewees from Schiphol Airport. These interviews were approximately 45 minutes each. The interviews were focussed on the current data governance of the public undertakings, their main data users, the datasets they provided and the issues they expected to encounter when (more) open data was achieved. These interviews explored the (open) data governance of both public undertakings enabling their open data level to be placed in the multi-dimensional model. Discussion then focussed on any barriers they could face when striving for a higher level of open data. For both the Port of Rotterdam and Schiphol Airport the main barriers faced were summarised and were discussed in the third research objective (how to overcome the barriers, presented by the public undertakings) and used again in the fourth research objective (discover whether the methods used by Liander to achieve open data). For all the five

interviewees, only the job description is used as a reference in order to protect the privacy of the interviewees.

1.2.3 Research how one public undertaking, Liander, succeeded in opening their data, which barriers did they encounter and how were they overcome?

Liander NV, the largest utility company in the Netherlands distributing electricity and natural gas, has been providing open data since 2014. This third research objective discovered why and how Liander, as a public undertaking, provides open data. To achieve this objective an unstructured interview of 45 minutes was conducted with a product developer at Liander. This interview focussed on the current data governance, the data users, the datasets, the drive to provide open data and the barriers they encountered when open data was provided in 2014. The barriers mentioned by Port of Rotterdam and Schiphol Airport before were briefly mentioned to the interviewee of Liander. The retrieved information from Port of Rotterdam and Schiphol Airport, combined with the new information of Liander, was used as an input for an online data gathering between the three organisations. The online data gathering was part of the final research objective of this research and organised by myself.

1.2.4 Discover if any of the methods used by Liander can be successfully applied to two other public undertakings towards opening their data.

Information for the last research objective of this research was sought by organising an online meeting for the three subject organisations. This meeting was organised by myself. This meeting was organised on the third of February 2021 using Microsoft Teams software to this agenda:

- 11:30 – 11:40 Opening and elaboration on the subject
- 11:40 – 11:55 Discuss barriers of Port of Rotterdam
- 11:55 – 12:10 Discuss barriers Schiphol
- 12:10 – 12:50 Discuss the methods used by Liander
- 12:50 – 13:00 Ending

The fourth agenda point, successful methods for adoption, was the key part of the gathering. Open discussion considered methods used by Liander and whether they could be adopted by PoR and Schiphol Airport to overcome any barriers towards open data. Based on the four research objectives the overall research question could be answered: “How can public undertakings in the Netherlands, similar to Liander, overcome the barriers towards opening their geospatial datasets in order to be prepared for expected future legislation towards open data?”.

The conducted interviews and the organised data gathering for this research, including the people present, dates, and times are presented in the next table, Table 2. The overall methodology is presented in Figure 3.

Table 2. Conducted interviews and organised online data gathering presented, including the people present, date and time.

	Present	Date	Time
Interview Port of Rotterdam	1 interviewee	23-11-2020	10:00
Interview Port of Rotterdam	1 interviewee	23-11-2020	11:00
Interview Port of Rotterdam	1 interviewee	23-11-2020	13:00
Interview Schiphol	2 interviewees	27-11-2020	13:00
Interview Liander	1 interviewee	18-12-2020	13:00
Online data gathering	Contact person of PoR, Schiphol Airport and Liander.	03-02-2021	11:30

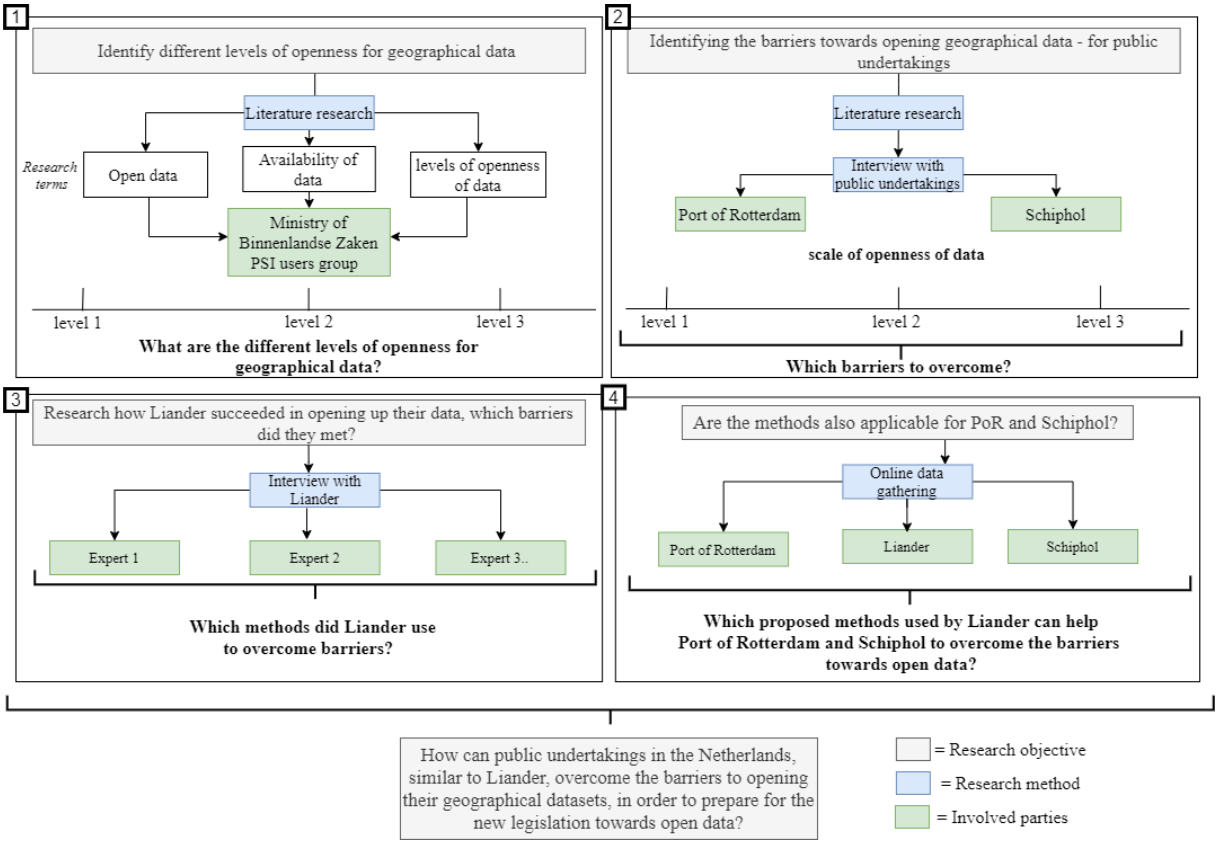


Figure 3. A schematic overview of the methodology of this research.

Literature review

Chapter 2 : Identifying open geographical data

Chapter 3 : Identifying the barriers towards open data

2 IDENTIFYING OPEN GEOGRAPHICAL DATA

Open data is a topic high on the European agenda (Directive (EU) 2019/1024, 2019). Over recent years it took on a more dominant role in both European legislation as well as in national legislation because of the increasing appreciation of its value towards society (Directive (EU) 2008/48, 2008; Directive (EU) 2013/37, 2013; Directive (EU) 2019/1024, 2019). Open data can play a significant role in the stimulation of socio-economic development as it stimulates decision making processes on both European and national level (Directive (EU), 2013; van Loenen, 2012; Vancauwenberghe et al., 2018). Not only the governmental data was considered valuable for re-use, but also non-governmental data such as documents of public organisations, national archives and libraries (Directive (EU) 2013/37, 2013). Although open data is high on the European agenda, it is still not delivered in an adequate and sufficient manner (European Commission, 2020b).

In line with the growth of open data, several efforts emerged to measure various aspects of open data readiness, implementation, outcomes and impacts (Caplyn et al., 2014). This resulted in open data assessment frameworks which were able to describe and monitor the state of open data in countries and organisations (Welle Donker & van Loenen, 2016). Open data frameworks are developed from different perspectives; it may be from a technical perspective, an organisational perspective or an holistic perspective (ibid.). Since the goal of this chapter is to identify the state of open data within organisations, from different perspectives, open data assessment frameworks are introduced to this research. Through this, different levels of open data can be identified for public undertakings.

2.1 OPEN DATA ASSESSMENT FRAMEWORKS

There are four open data assessment frameworks of specific interest to this research, each interpreting open data from a different point of view, presenting the majority of the requirements of open data. The four assessment frameworks are:

1. The three laws of governmental data of Eaves (2009)
2. Five-star model of Berners-Lee (2009)
3. Open knowledge Foundation
4. Open data from a holistic point of view (Welle Donker & van Loenen, 2016)

2.1.1 Three laws of governmental data of Eaves

When it comes to data, David Eaves (2009) describes three laws which make data 'open data'. The first one is straightforward and states that whenever the data cannot be found, the data does not exist. Therefore, this first law is called '*Find'*. This simply means that one needs to be able to find the data. The second Eaves law is called '*Play*' and asks the question whether someone can play/experiment with the data. In order to comply, the data needs to be delivered in an open machine readable format. The standard for this is compliance with the framework of the Open Data Barometer and The Organisation for Economic Co-operation and Development OECD OURdata index (e.g. an Application Programming Interface (API), subscription feed or a document file) (OECD Open and Innovative Government Division, 2019; World Wide Web Foundation, 2017). If the data is not open and available in a machine-readable format it is not possible to use the data. The third Eaves law is called '*Share*': after the data is found and played with (used) it needs to be possible to share the data with others. If the data cannot be shared for juridical reasons the data cannot be empowered and shared with others. The three laws of governmental data are written from a user perspective requiring stating that the user needs to be able to

find, play and share the data in order to be classified as open data (Eaves, 2009). The user perspective is discussed from the holistic point of view later in 2.1.4 (Welle Donker & van Loenen, 2016).

2.1.2 Five star model of Berners-Lee

The method of Berners-Lee describes the openness of the dataset itself and its interconnection with other datasets in order to classify its openness. Interconnected datasets have the opportunity to be re-used in unexpected ways, adding value to the information especially across the Web (internet) (Berners-Lee, 2009). According to this method, in order to be open, datasets need to interconnected across the internet.

The five star model of Berners-Lee (2009) offers the opportunity to rate datasets with one to five stars. One star is the lowest score a datasets can get and five the highest (Berners-Lee, 2009). It gives a rather straightforward description of the different ranks. The build-up of the five star model is presented in Figure 4.

In this model, datasets score one star when they are only available on the web in a fixed format but with an open licence. An example of such a dataset could be an image of a graph. Datasets ranked with two stars need to provide data in a machine-readable structure, such as an excel instead of an image of the table such as an Excel spreadsheet, a limited proprietary format. A dataset is given three stars when it comes in a machine-readable format, hence in an open format (CSV, instead of excel). Four starts are given to datasets that use open standards from World Wide Web Consortium[1], on top of the three stars datasets. Making use of the open standard from W3C means using RDF with the query language SPARQL[2]. RDF is a directed, labelled graph data format for presenting information on the internet. SPARQL is used to express queries across diverse data sources, where the data is stored natively as RDF or viewed as RDF (W3C Recommendation, 2008). Thereby, it gives people the opportunity to talk about the same information in the same language. The last and highest rank from the five star model of Berners-Lee (2009) needs all the aforementioned abilities plus the ability to link the dataset to other people's dataset in order to provide context. Thus, it needs to be able to use the dataset in combination with other datasets. T**Fout! Bladwijzer niet gedefinieerd.**he Berners-Lee method (2009), shows that the interconnection of datasets offer data the opportunity to be re-used in unexpected ways that can lead to valued added information by the Web.

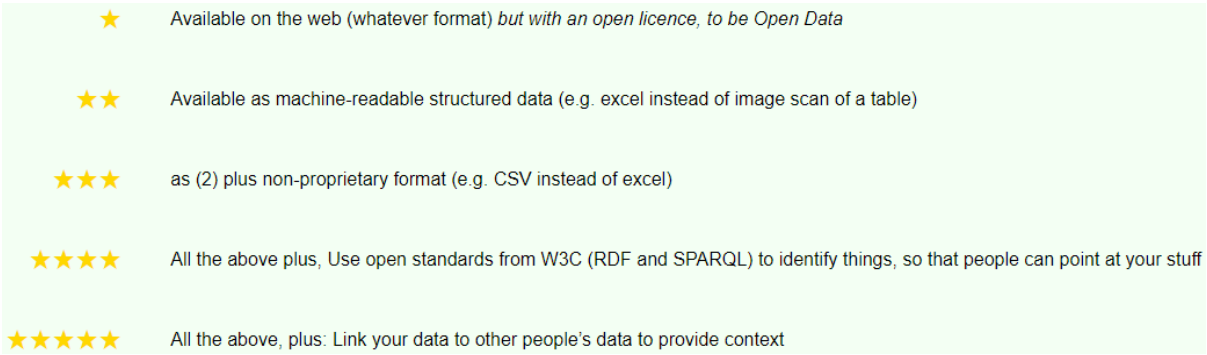


Figure 4. Five-star model of Berners-Lee (2009).

[1] The World Wide Web Consortium (W3C) is the main international standards organisation for the World Wide Web. Founded in 1994 and currently led by Tim Berners-Lee, the consortium is made up of member organisations that maintain full-time staff working together in the development of standards for the World Wide Web.

[2] a semantic query language for databases—able to retrieve and manipulate data stored in Resource Description Framework (RDF) format.

2.1.3 Open Knowledge Foundation

The definition of open data given by the Open Knowledge Foundation (OKF) focusses on open data for everyone: 'Open data is data that can be freely used, re-used and shared by everyone' (Open Knowledge Foundation, n.d.-b). There are three key features of data openness:

1. **Availability and access**. First, the data needs to be available, preferably downloadable on the internet and for no more than a reasonable production cost. The data also needs to be available in a convenient and modifiable form.
2. **Re-use and redistribution**. Secondly, the data needs to be provided under terms that permit re-use and redistribution, including the intermixing with other datasets. This can be put as an open licence that is compatible with other open licenses. According to the OKF the share alike licence is also included; this is further explained in a later section. Additionally, the data needs to be machine readable.
3. **Universal participation**. The last key feature is that everyone must be able to use, re-use and redistribute the data. There should be no discrimination against fields of endeavour or against persons or groups. This means, for example, that 'commercial' restrictions on data availability would prevent 'non-commercial' use is not allowed.

This definition is written with respect to knowledge. It is interoperability of data that is maximised. This is in line with the five star model of Berners-Lee (2009) that holds that linking data is key for data to be open (Berners-Lee, 2009). However, this definition does not satisfy the re-use of data completely; it contains share-alike conditions. This means that the re-used version of the dataset 'B' needs to be in the exact same condition as the original dataset 'A' (van Loenen, 2012). Only when the original dataset A is given with free references, the re-used dataset B can also be presented like that. When dataset A is not given with free references, this is also not possible for the re-used dataset B. Adding these conditions to the re-use of datasets can result in a loss of utility in the re-use of the data (ibid.).

2.1.4 Holistic point of view

The holistic framework goes beyond the assessment frameworks listed above (Berners-Lee, 2009; David Eaves, 2009; Open Knowledge Foundation, n.d.) by taking into account the quality of the data from a user perspective and is described by Welle Donker and van Loenen (2016) (Welle Donker & van Loenen, 2016). This view holds that the quality of open data will improve in line with the increase of positive impact of open data as acknowledged by society. This starts with the user, the user's perspective on open data, which is missing in some of the aforementioned open data assessment frameworks (Welle Donker & van Loenen, 2016).

To include the user's perspective, a multi-dimensional assessment framework was set up by Welle Donker and Van Loenen (2016). This was first established by providing a holistic, comprehensive approach through the re-use of elements of existing frameworks. The openness aspect of a dataset comes from, for example, the Open Data Index and the access through a portal (as part of the indicator 'recognisable') comes from CapGemini's framework. These parts of existing frameworks are then combined into an assessment of open data and considered from a user's perspective (Welle Donker & van Loenen, 2016).

The three output indicators that are set as conditions for a successful open data ecosystem are:

1. Data supply - the way data is provided as open data
2. Data governance - the way in which governance aspects are organised
3. User characteristics - the way in which the user characteristics enable the user to innovate with open data

The aspect of data supply is defined by the open data approach provided by Backxs (2003). This is a relevant approach for this research as it focuses on the question ‘how is data provided as open data?’. Moreover, it concentrates on the perspective of the user towards open data, which is considered vital for the quality of open data according to Welle Donker and van Loenen (Backx 2003, referred to in Welle Donker and van Loenen, 2016). The concentric shell model of Backx is shown in Figure 5.

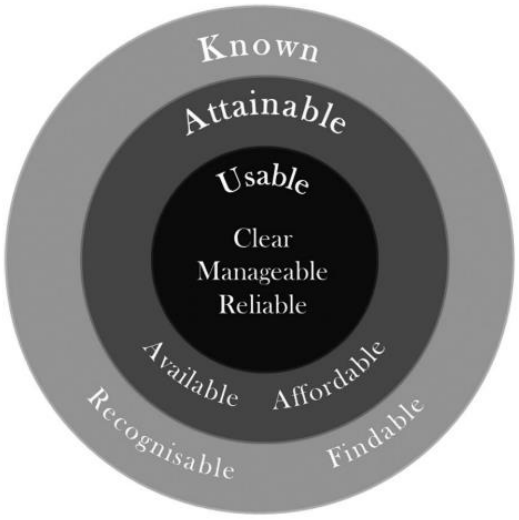


Figure 5. The concentric shell model of Backx on the definition of open data from a user’s perspective. Source: Welle Donker and van Loenen (2016).

For data to be considered of good quality from the user’s perspective, the data needs to be:

1. “Known to the user” - Is the data identifiable and where can data be obtained?
2. “Obtainable by the user” - Can the data be obtained by the user, and under what conditions?
3. “Useable for the intended purpose of the user” - Can the user assess the quality of the data?

First, the user needs to know that a certain dataset exists and where to find this dataset. This is similar to the law ’Find’, mentioned by Eaves (2009). The existence of a datasets can be discovered through resource metadata, data about the data, and can come in the form of titles, abstracts or textual keywords. For linked data, this can be done from resource description framework similar to the five stars model of Berners-Lee (Berners-Lee, 2009). A user’s first step is to find the data through a search engine (e.g. Google) before finding it in a data portal. Once the data is discoverable to a user, this is indicative of the accessibility of the data portal that provides this data.

Once the data is found it needs to be possible for the user (1) to access the dataset (through viewing and/ or downloading) (similar law of ‘Play’ (Eaves, 2009)), (2) to re-use the data through an allowance of a licence, and (3) reasonably to afford the data. This refers to the second circle of Backx (2003) model, attainability.

Lastly, the user needs to be able to assess the data on its suitability for its intended use. This relates to the last point of Backx (2003): the data needs to be useable for the intended purpose of the user. A user needs to be able to assess this based on the data quality, available documentation/metadata, level of coverage, timeliness and the update frequency. This is where ‘data quality’ makes an entrance, which is influenced by the type of user (Welle Donker & van Loenen, 2016). When the assessment shows that the data is clear, manageable and reliable it is considered usable according to shell model of Backx

(2003). The definitions of Welle Donker & van Loenen (2016), and Backx (2003) each add a relevant layer to the definitions already given on open data since the user is considered as well as the data quality.

2.1.5 Data licence

The requirement of open data to have an 'open licence' is mentioned by several research papers (Berners-Lee, 2009; OECD, 2017; Open Knowledge Foundation, n.d.; Welle Donker & van Loenen, 2016; World Wide Web Foundation, 2017). A data licence decides whether and by whom the data can be used and under which conditions it can be shared (Open Data Reader, 2016; Miller et al., 2008). The licence determines the openness of data. When there is no licence specified, nobody can use, share, distribute, re-post or transform the data (Data.world, 2019). Different licenses are provided by different data frameworks, enabling sharing and use of data openly. Examples of different frameworks are the Open Data Commons (Open Knowledge Foundation), the Community Data License Agreement (CDLA) and the Creative Commons (CC) (ibid.). All these frameworks offer their own licence allowing the data to be open, open under conditions, or not open at all. The most common licence types, used by these frameworks, are (ranging from open to more restrictive): Public domain, attribution, share-alike, non-commercial, database only, no derivatives (Data.world, 2019).

The Creative Commons framework (Figure 6) gives a clear overview on the variety of licence types and suits the levels of openness defined in this research. The © on the left side of the scale leaves the rights to share the data in the hands of the data creator (copyright), whereas the crossed © on the other side implies that the rights of sharing the data is in hands of the public domain. The variation of Creative Commons (CC) licence in the middle propose that the data creator preserves the rights as an owner, however, allow others to use the data to a certain extent (Open Data Reader, 2016). Choosing a CC licence offers the data creators the opportunity to share their data under conditions which suit their (open) data policy (ibid.).

Figure 6. Openness in terms of the Creative Commons licence: left not open (*Auteursrecht)*, right fully open (*Publiek Domein)*. Source: Open Data Reader, 2016.

Furthermore, according to the open data definition of the Open Knowledge Foundation, there are three licenses which satisfy the requirements of open data: CC0, CC BY, CC BY-SA (Figure 7). The description of the three licence is given below. The licence that is considered most open in terms of sharing is the Creative Commons zero licence (CC0). (Creative Commons, 2019; Open Knowledge Foundation, n.d.-b).

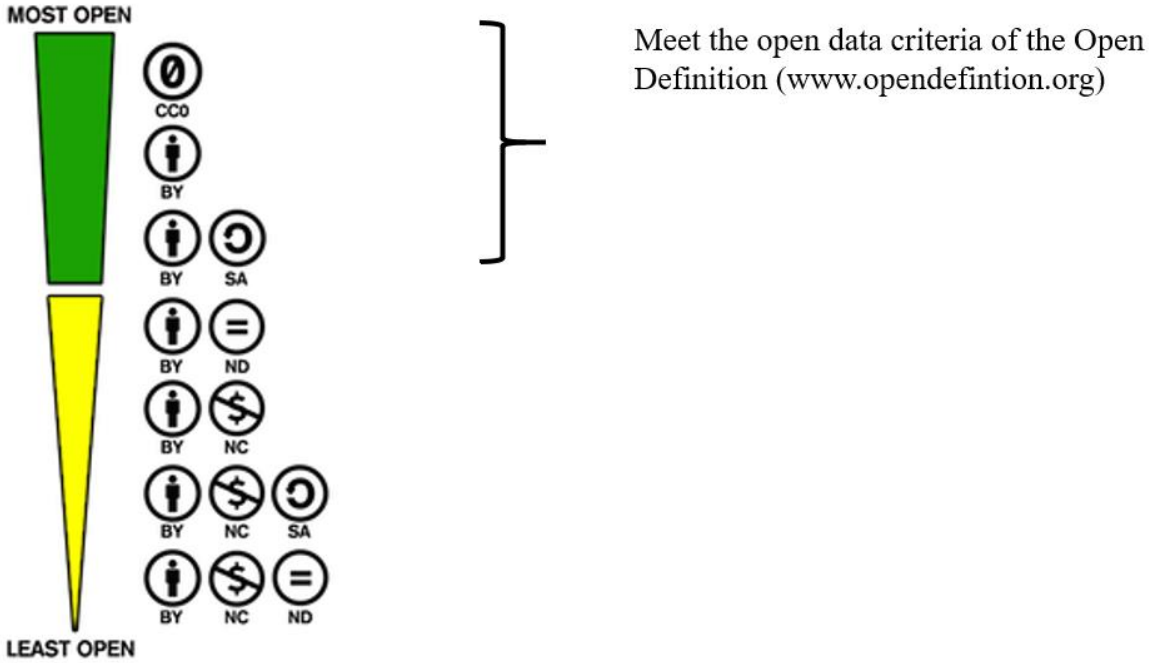


Figure 7. Creative Commons Licenses which do and do not satisfy the open data definition of the Open Knowledge Foundation. Source: Creative Commons, 2015; Open Data Reader, 2016.

- Creative Commons Zero (CC0) - the author declares all rights to be out of his/her hands as far as legally possible.
- Attribution (CC BY) - this allows re-users to distribute, adapt and build upon the data if attribution is given to the creator. Credits to the creator must be given.
- Attribution and under the same terms (CC BY-SA) - this allows re-users to distribute, adapt and build upon the data as long as attribution is given to the creator. Any modified data needs to be licensed under the same terms as the original data.

2.1.6 User Types

When considering non-governmental data sources the literature study suggests more factors that influence the openness of data. 'Users' is mentioned by several literature studies as being part of the open data approach (Berners-Lee, 2009; David Eaves, 2009; OECD, 2017; Welle Donker & van Loenen, 2016). Eaves (2009) requires that the user can play with the data in order for the data to be open (Eaves, 2009). Welle Donker and van Loenen (2016) refer to Backx (2003) stating that users need to assess the data on its suitability in order to decide whether they can use it for their purpose (Welle Donker & van Loenen, 2016). The data user uses the products and the services which are developed by the collectors, enablers, developers and enrichers of the data (Deloitte Analytics, 2012). However, there are different types of users with different purposes (Welle Donker et al., 2019). Certain data is kept close to the source

and used with the purpose of internal and commercial use in order to achieve business functions. The users are internal users embedded in the organisation and could contribute to the innovative process by using the data for business (Deloitte Analytics, 2012; Schweisfurth & Herstatt, 2014). Data may also be used for non-commercial purpose such as a research at an university. These external users on the other side of the spectrum are identified as consumers who use the data for societal or non-commercial and social purpose, contributing to the scientific body of knowledge (Deloitte Analytics, 2012). Yet, open data can be used for both commercial and non-commercial purposes simultaneously (Welle Donker et al., 2019).

The type of user gives an extra dimension to the identification of open data (Deloitte Analytics, 2012; Schweisfurth & Herstatt, 2014; Welle Donker et al., 2019) so the type of data user needs to be considered for further identification of open data in this research.

2.1.7 Data quality

Different types of data users (Degbelo, 2020; Safarov et al., 2017), will assess the quality of the data from different perspectives. This makes the requirement 'good quality data', mentioned in the holistic point of view, an arbitrary term (Corsar & Edwards, 2017; Welle Donker & van Loenen, 2016). For this research, the interpretation of 'data quality' is defined by the International Organisation for Standardisation (ISO) (ISO - ISO 19157:2013; ISO - ISO 9001:2015). ISO defines quality as "the totality of characteristics of an entity that bear upon its ability to satisfy stated and implied needs." This interpretation is used since it follows the idea of 'fitness for use' which implies that the concept of data quality is relatively, similar to the data user's role (Tayi & Ballou, 1998; Welle Donker et al., 2019). Quality that is considered appropriate for one user may not be of sufficient quality for another user (ibid.). The internal data user mentioned in 2.1.6 for example, draws on knowledge sources (use knowledge, solution knowledge and organisation knowledge) and social resources (relational- and cognitive capital) relevant for business innovation. Other data consumers, such as a researchers, may be more interested in the contextual qualitative data and precise knowledge for academic purposes (Safarov et al., 2017). The ability of an organisation to provide 'good quality data' depends on its ability of providing data that meet different needs and purposes (ISO - ISO 19157:2013; ISO - ISO 9001:2015).

2.1.8 Business Model Regime

As well as the different types of users and purpose of use for each open data level, different types of business regimes are applied to facilitate the user's needs (Welle Donker & van Loenen, 2016; Zuiderwijk & Janssen, 2014). These regimes are described as business models which can be associated with revenue and profit (for the internal user) or aim to generate public value (Janssen & Kuk, 2007; Zuiderwijk & Janssen, 2014). A business model can also be defined as the method by which a firm builds and uses its resources to offer customers better value (Afuah & Tucci, 2001). Data can be designed specifically around the customer, the data user (Welle Donker & van Loenen, 2016). When data is used for internal purposes, such as internal performance, often little attention is paid to possible external users and the openness of the data is limited to the internal user (Zuiderwijk & Janssen, 2014). As the next step towards opening data includes involving external (trusted) users as well, a suitable business model that generates both internal and external value is needed.

The next step would be to generate public value for everyone (citizens, researchers, etc.). This business model introduces an information intermediary ('infomediary') business model, positioned between the private sector and the external users (Figure 8) (Zuiderwijk & Janssen, 2014). The openness of the data is still not fully guaranteed because the consumers cannot access the data without the effort of an infomediary party – which is avoided when the public sector is considered as the data provider (Figure 8). No difference is made here between internal and external users since no other purpose than generating public value is applied (Zuiderwijk & Janssen, 2014). The openness of the data is fully guaranteed by a data regime that adopts the consumer as the leading user type. This regime seeks to focus on everyone – including all the aforementioned user types (Deloitte Analytics, 2012; Welle Donker & van Loenen, 2016).

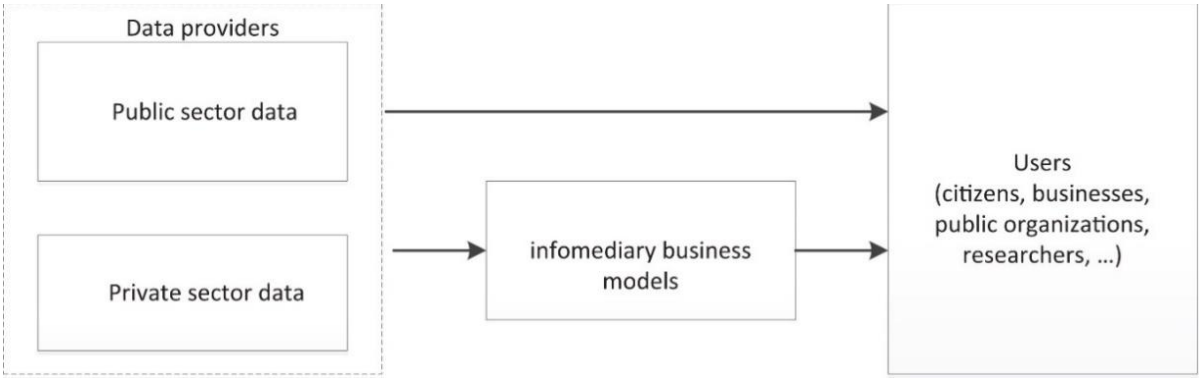


Figure 8. Position of the infomediary business model between the private sector data and the external users. Source: Zuiderwijk & Janssen, 2014.

2.1.9 Open GIS data portals

Currently, there are four GIS data sources that are considered free and most open (Table 3) (GISGeography, 2020). In the Netherlands, an open data portal which is considered of high value for the society is the Dutch Public Mapping Service (*Publieke Dienstverlening op de kaart (PDOK)*, which is also presented in Table 3.

Table 3. The four open data platforms, currently considered free and most open. Including a description, the features and the website link. Source: GISGeography, 2020.

	Description	Features
Esri Open Data hub hub.arcgis.com/search	Esri Open data hub provides over 250.000 open datasets from more than 5.000 organisations worldwide.	- Search on any topic and location - Downloadable in several GIS format (vector/raster/gdb/RDBMS) - Downloadable in bulk*
Natural Earth Data Naturalearthdata.com	The geospatial data provided by Natural Earth provides key cultural and physical vector and raster data. It is best suited for the purpose of cartographers.	- Downloadable in Map Exchange Document (MXD) and QGIS documents - Ability for the user to use, modify and disseminate (share) the data in any form
USGS Earth Explorer earthexplorer.usgs.gov/	The Earth Explorer of the United States Geological Survey (USGS), provides satellite and aerial imagery around the globe (USGS, n.d.). People from other places in the world can also download this data as well.	- Offers up-to-date remote sensing data - Downloadable Digital Elevation Models (DEM) - Downloadable in bulk*
OpenStreetMap Openstreetmap.org	OpenStreetMap (OSM) provides crowdsources based, free GIS data at a street level. It is the biggest inventory of buildings in the world. It is for the public and created by the public.	- Offers the ability to the user to give feedback to the system, e.g. participate - Downloadable in an Extended Markup Languages (XML) - Downloadable in bulk* - Share under a share-alike licence
Dutch Public Mapping Service (Publieke Dienstverlening op de kaart (PDOK) Pdok.nl	The PDOK provides geographical datasets and is developed by governmental bodies. This data is reliable and comes from both the public sector and the private sector. PDOK has 65.573.770 data calls per day and 213 high value datasets.	- offer the ability to download the data in bulk * - Share under attribution and under the same terms (CC BY-SA). - Downloaded data can be used (offline) In a GIS environment.

* No limit to the number of times a dataset can be downloaded

The requirements of open geospatial data as described above can be considered as the framework of the geographical open data 'must haves'. It complements the requirements of open data, with three requirements:

- Data can be modified
- The data is recent (up-to-date)
- The data is downloadable in bulk

2.1.10 Sub-conclusion

From this point, requirements that guarantee the openness of the data which are not influenced by the type of user (Deloitte Analytics, 2012; Safarov et al., 2017;Welle Donker et al., 2019), data quality (ISO - ISO 19157:2013; ISO - ISO 9001:2015), subsequently data regime (Welle Donker et al., 2019; Zuiderwijk & Janssen, 2014), are:

- Open data standards
- Provided free of charge
- A machine-readable format of the data
- Can be found through a search engine and/or data portal

Thus, the type of user, the quality of the data and the data regime are three additional dimensions that influence the openness of the data through the findability, the usability and in which way data is shared (Degbelo, 2020; ISO - ISO 19157:2013; ISO - ISO 9001:2015; Safarov et al., 2017).

The definitions of open data from the literature review were used as an input for the creation of a multi-dimensional model on distinct levels of open data (Figure 9). Three levels of open data were identified and used in this research: Not open, partly open and open. To specify the requirements of the three levels, they were categorised by *find*, *play* and *share* from Eaves (2009). *Find* and *play* are associated with how the data can be found and used, whereas *share* is associated with the person using the data and how the data can be shared based the data licence (ISO - ISO 9001:2015). In the first two levels the term *share* was used by Eaves (2009), claiming that once data is found and used, it needs to be possible to share it with others (David Eaves, 2009). When a licence does not allow sharing, the data cannot be empowered and shared with others. However, when considering level 3, 'sharing' was taken out of perspective as no licence means that sharing with others is no longer needed since everyone can access and re-use the data (Directive (EU) 2019/1024, 2019). The term *share* was then replaced by *re-use* in level 3.

The type of user, data quality and type of regime approach the openness of the levels from different dimensions, creating a separate influence on the openness of the data (Deloitte Analytics, 2012; Safarov et al., 2017; Welle Donker et al., 2019; Zuiderwijk & Janssen, 2014). At the first level data is considered not to be open at all and only accessible for the internal user. Here, the data cannot be found through a general search engine (Welle Donker & van Loenen, 2016). This makes the data invisible to everyone but the internal user. The absence of an open licence makes it impossible to share the data with external users (Eaves, 2009; OECD, 2017; Welle Donker & van Loenen, 2016; World Wide Web Foundation, 2017). This suits an internal regime that is focussed on using the data for internal purposes, limiting the data quality to the purpose of the internal user (ISO - ISO 9001:2015). At the second level data openness is improved by making it findable and accessible through a general search engine or data portal, presented in a machine-readable format (Welle Donker & van Loenen, 2016). However, fees may be charged and the data can only be shared under certain conditions and terms. This data policy generates both internal and external value. At the third level data can be considered most open. The data is findable through a general search engine and data portal, free of charge, comes in a machine-readable format and

with an open licence. Meaning that everyone can re-use the data (Berners-Lee, 2009; Directive (EU) 2019/1024, 2019; ISO - ISO 9001:2015). This can be applied through an infomediary business model, creating the space for the private sector to share the data through a third party either by keeping the data within external trusted parties or by sharing the data openly with everyone (Zuiderwijk & Janssen, 2014). An infomediary party can improve the quality of the data, making it to suit everyone's purpose (ISO - ISO 9001:2015). The regime applied to this level gives the space to generate public value from the published data. Nonetheless, it is impossible to make the data fit for everyone's purpose making 'data quality' a relative concept (ISO - ISO 9001:2015). The quality of data and its openness is still relative for each specific case and user (Safarov et al., 2017). Although this last level is preferred from a user's point of view it might lead to issues with the business model of a private company (Welle Donker & van Loenen, 2016). The three levels of open data, presented in Figure 9, were used to identify the barriers between the levels (chapter 3), and to identify the open data levels of Port of Rotterdam and Schiphol Airport (chapter 4.1 and 4.2).

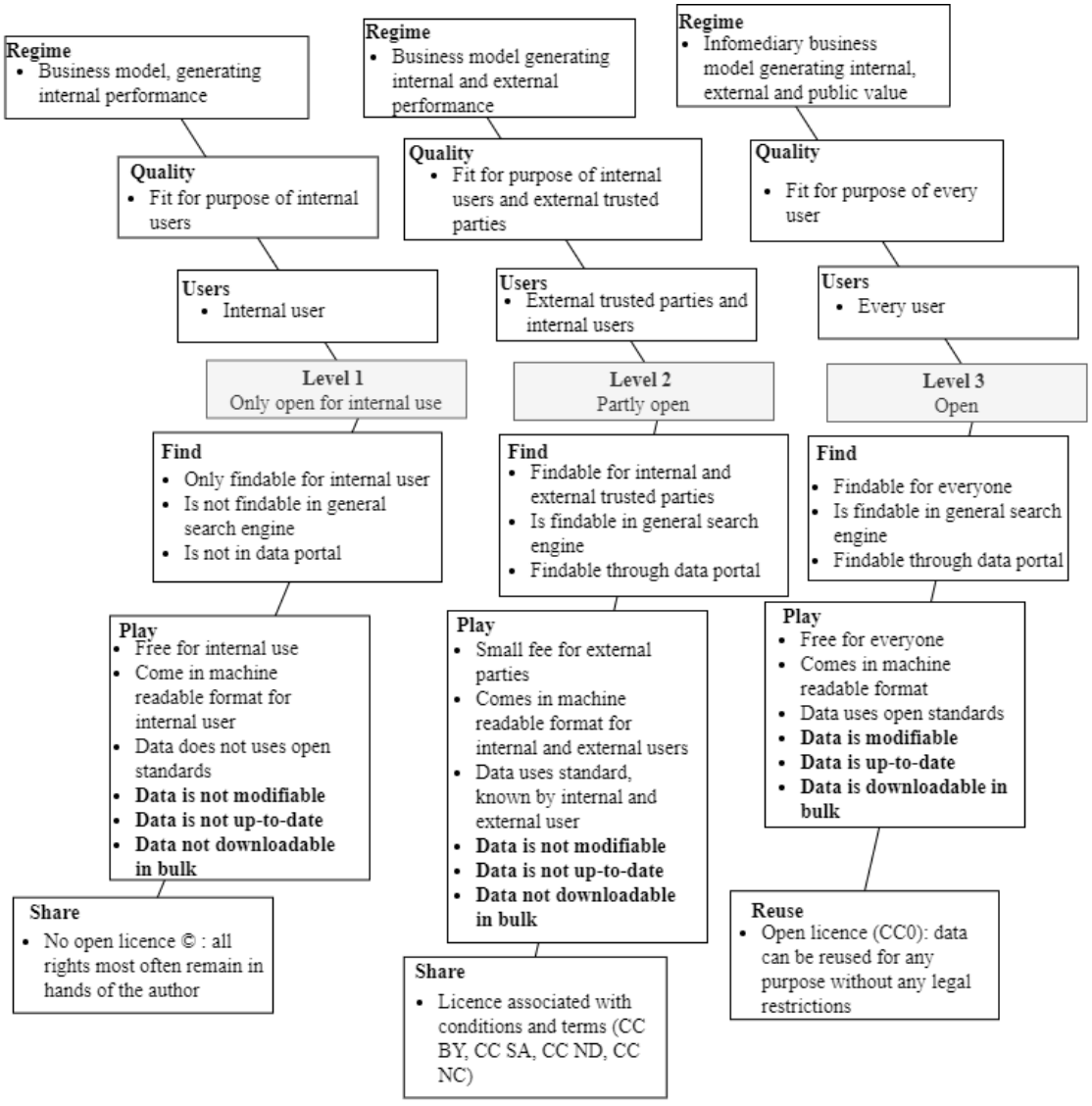


Figure 9. The multi-dimensional model of the three identified levels of open data, based on the literature study.

3 IDENTIFYING THE BARRIERS TOWARDS OPEN DATA

While open data can contribute to social and economic benefits, the adoption of open data also encounters numerous barriers (Janssen et al., 2012; Verhulst et al., 2020). To achieve open data these barriers need to be identified and overcome. According to Janssen et al. (2012) these barriers can be perceived from either the provider's perspective or the user's perspective.

Whereas each perspective is valid and with merit, data cannot be considered open if a potential user cannot access the data (Welle Donker & van Loenen, 2016). There are also barriers that arise from the relationship between the data provider and the data user (Martin et al., 2013). A lack of dialogue between the user and the provider about new updates of an already opened dataset or information can be considered as risks (ibid.). Additionally, visibility in both the data origin and data re-use is required for providers as well as users to identify optimal intervention points for mitigating data risk (Martin et al., 2013; Verhulst et al., 2020). Therefore, barriers for all data stakeholders need to be considered.

3.1 FINANCIAL BARRIERS

An open data initiative incurs costs that may not be recoverable within an existing business model (Martin et al., 2013). It takes human and financial resources both to collect, to maintain, to process the data and finally to distribute it as open data (Welle Donker, 2018). There are several types of costs that needs to be taken into account for realising open data: adaptation costs, infrastructural costs and structural maintenance / operational costs (Martin et al., 2013; Welle Donker, 2018). Adaptation costs include the effort of data improvements needed before it can be published as open data. Then, an infrastructure needs to be available to publish the data: Infrastructural costs. Even when the infrastructure for the data already exists, additional costs may arise to satisfy the requirements of data openness. Operational costs are incurred to make the data available through tools and apps, that might need to be developed first. Lastly, costs arise from the need to maintain and update the data. Potentially, a facility needs to be set up to deal with questions from users that could result in structural maintenance costs (Welle Donker, 2018). All together these costs can be substantial (Martin et al., 2013).

Another question arises when the data is openly published by the data provider, about the return on investment once data is open and freely available. Welle Donker (2009) offered a model that covers the production costs (adoption/adaptation costs and infrastructural costs) of open data. Organisations which rely on income of sales of their data can face budgetary problems when data is provided for free. Therefore, the choice to provide data against a fee through a cost recovery model is preferred over providing open data for free (Welle Donker, 2009). A cost recovery model means that equal amounts of revenue and expense are established as collections are made. Thus in the case of open data, fees are charged to equally recover the expenses made (ibid.). The transition from a cost recovery model to an open data model puts the data provider in a less secure, higher risk position since open data is provided as free (Martin et al., 2013; Welle Donker, 2009). Moreover, a common standard for assessing the costs as well as the benefits does not exist yet (Martin et al., 2013). In summary, the costs and uncertainty concerning the return value of providing open data creates a potentially financial barrier for the data provider. Open data means no recovery costs, which can result in financial problems for organisation (Martin et al., 2013; Welle Donker, 2018).

3.2 INSTITUTIONAL BARRIERS

Any unwillingness from data providers in terms of financial and legal risk to make data open available, is known as an institutional barrier (Janssen et al., 2012). A general lack of awareness of open data amongst public and private organisations in the EU often results in only the financial and legal risks being considered (Dalla Corte, 2020). A data provider could be held liable if publishing incorrect and incomplete data results in damage (Asser/Hartkamp & Sieburgh, 2008). For instance, when an organisation publishes a new road network of which two street names are mixed up, the municipality faces reputational damage since this incorrect data is processed and published in a city map, of which 10,000 examples are made. Moreover, this could cause financial and even physical damage as the ambulance, police and fire trucks are navigated to an incorrect location, and time and money are lost. When incorrect data is published by the data provider, it's liability can extend directly to the data users. Moreover, it can extend to the consequences from third parties' acting if it is found to be the provider's error (van Loenen et al., 2011). In such cases the liability for damages lies with the data provider as the source of the data is incorrect (Asser/Hartkamp & Sieburgh, 2008). Institutional risks like this make organisations cautious when providing data (Van Loenen et al., 2011). Such a risk-averse culture results in organisations preferring not take any risk to change (Barry & Bannister, 2014; Janssen et al., 2012).

Furthermore, without prior knowledge about the purpose of the provided data, false conclusions can be drawn by the data users. For example, demographic profile information of a neighbourhood might appear as unnuanced to a user without knowledge about the statistical method used. The data provider has a responsibility to rectify this false conclusion, which requires time and effort (Conradie & Choenni, 2014). On the other hand, data users can also intentionally distort and misuse the provided data in order to gain advantages or cause harm to the provider's reputation. Both the risk for false conclusions drawn by the user and intentional misuse of the user can cause a institutional barrier for the data provider (Geiger & Von Lucke, 2012).

3.3 TASK COMPLEXITY BARRIERS

Finding and using data tends to be challenging and often complex for the data user, due to high complexities. These complexities are worsened when there is no explanation of the context of the data or when the data formats and datasets are too complex to handle (Barry & Bannister, 2014; Janssen et al., 2012; Martin et al., 2013). For example, complexity becomes a barrier in geographical datasets for an unexperienced user when attempts are made to open an AutoCAD drawing (a detailed 2D or 3D illustration) in a geographical information system (GIS, ArcGIS pro for example). Matching of data formats with information systems can become more challenging and require more user knowledge/experience to manipulate the data. This is presented in Figure 10. Here, it is not clear what the actual purpose of the data is since architectural data is presented in the sea.

Therefore, use of data is considered only for those with domain knowledge which allow for opening, using and interpreting the data (Janssen et al., 2012). So for the data presented in Figure 10, this data can only be accessed and used by a user who has the technical skills to download the data, open the data in a GIS and analyse the data through tools. The format and complexity of data may contribute to a digital divide, a barrier, as the use of data might be limited to certain groups; only those with domain knowledge (Janssen et al., 2012). Furthermore, blocked use and lack of participation, possibly due to the complexity, might arise when there are no motives or there is no added value for the user to make use of the data and deliver feedback to the quality of the data (Janssen et al., 2012). User skills is a potential barrier that can be tackled by improved data format, structure and utility.

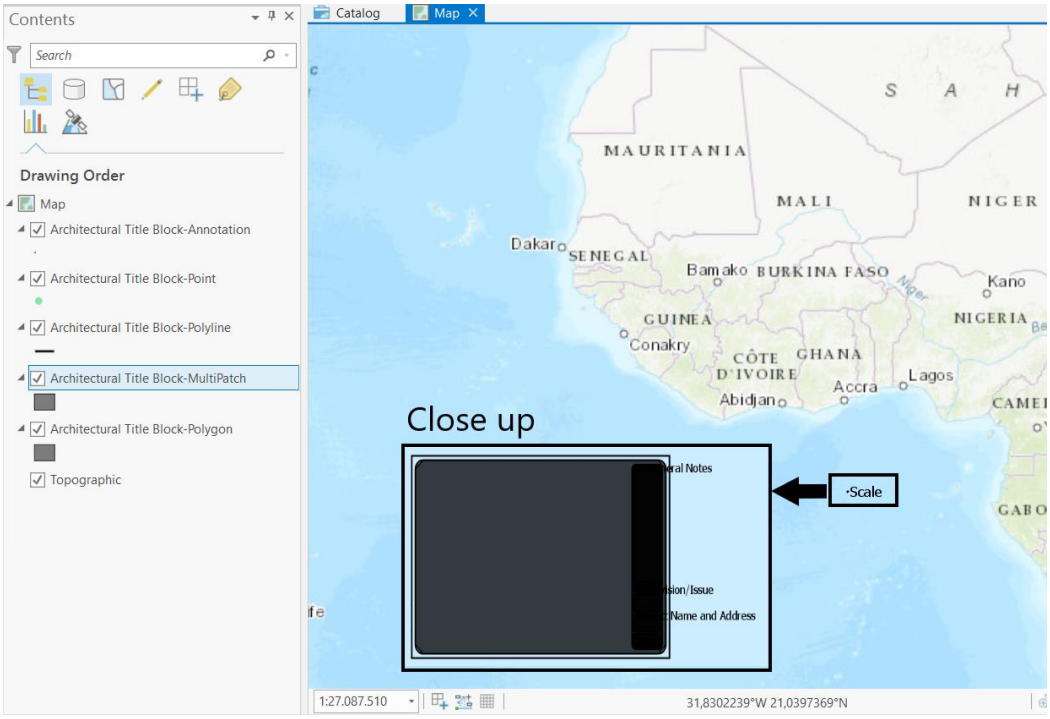


Figure 10. The .DWG dataset 'Architectural Title Block' (.dwg – 50.3 Kb) opened in ArcGIS Pro, making an appearance in the Golf of Guinee. The purpose of the dataset is not clear through this DWG format. Source: https://knowledge.autodesk.com/.

3.4 LEGAL BARRIERS

Legislation can be considered a barrier that can prevent organisations from opening their data for end users. Legal constrains are associated with the re-use conditions of the data, preventing the risk of data fragmentation at both international and national level (Martin et al., 2013). Legal interventions may arise from personal data protection, privacy violation (GDPR), security issues, confidential contracts or agreements with third parties (Janssen et al., 2012). This is most often done by implementing restrictive licenses or permitting (controlled) access. Such imposed limitations associated with the use of data protects the data provider against liability issues (Creative Commons, 2019; Open Data Reader, 2016 ; Miller et al., 2008; Open Knowledge Foundation, n.d.-a). Listed before in section 2.1.5 are examples of restrictive licenses: CC BY, CC SA, CC ND, CC NC (Creative Commons, 2019).

Legislation is also applied to prevent the market from unfair competition. Unfair competition takes place when consumers or businesses are engaged in deceptive business practices (Ginsburg et al., 2019). Examples of unfair competition are poor trading practice, trademark infringements and misappropriation of business trade secrets (ibid.). A fear of unfair competition was experienced by cartographer company Falkplan-Andes who makes cartographic maps for a living. When the Dutch national road network dataset (NWB) became openly available for re-use by the Dutch government, Falkland-Andes was afraid to lose their market share now that the similar data was published open and for free (Sanders, 2011). Publishing the NWB was considered as unfair competition since open governmental data could decrease the market share of Falkplan-Andes. Although organisations need to consider legislation to prevent the market from unfair competition, it hinders the possibility to provide open data (Ginsburg et al., 2019). So, legislation reasonably applied to prevent unfair competition can come at the expense of data openness as shown in the example of the Dutch national road network (Sanders, 2011).

Although legislation is not in favour for the concept of open data, data providers use legislation in order to protect their liability when publishing data (Asser/Hartkamp & Sieburgh, 2008; Van Loenen et al.,

2011). For instance, Pacific Gas & Electricity (PG&E), an American utility company, published their data without any restrictions toward the use of the data. After a spatial analysis, done with open data from the company on the electricity poles, PG&E were held liable for the cause of the largest and most destructive wildfires in state history. The study showed that the locations of the fires were often in the proximity of the electricity poles from PG&E (energy data request from public datasets from PG&E) (Figure 11). Their equipment of electric powerlines across the state evoked sparks that caused wild-fires which took the life of 84 people in 2018 (Kasler, 2018). In 2020, PG&E pleaded guilty and agreed to pay a maximum fine of 25.5 billion dollar for losses from the 2018 wild fire, blamed on the crumbling equipment of PG&E (CNBC, 2020). On the one hand it can be stated that open data is used correctly in this case by directing to the cause of the wildfires in California in 2018. On the other hand, this example highlights that there are risks associated with open data from a data providers point of view that may be mitigated by legal restrictions. Therefore, legislation could restrain data providers from publishing open data and can therefore be considered a barrier.

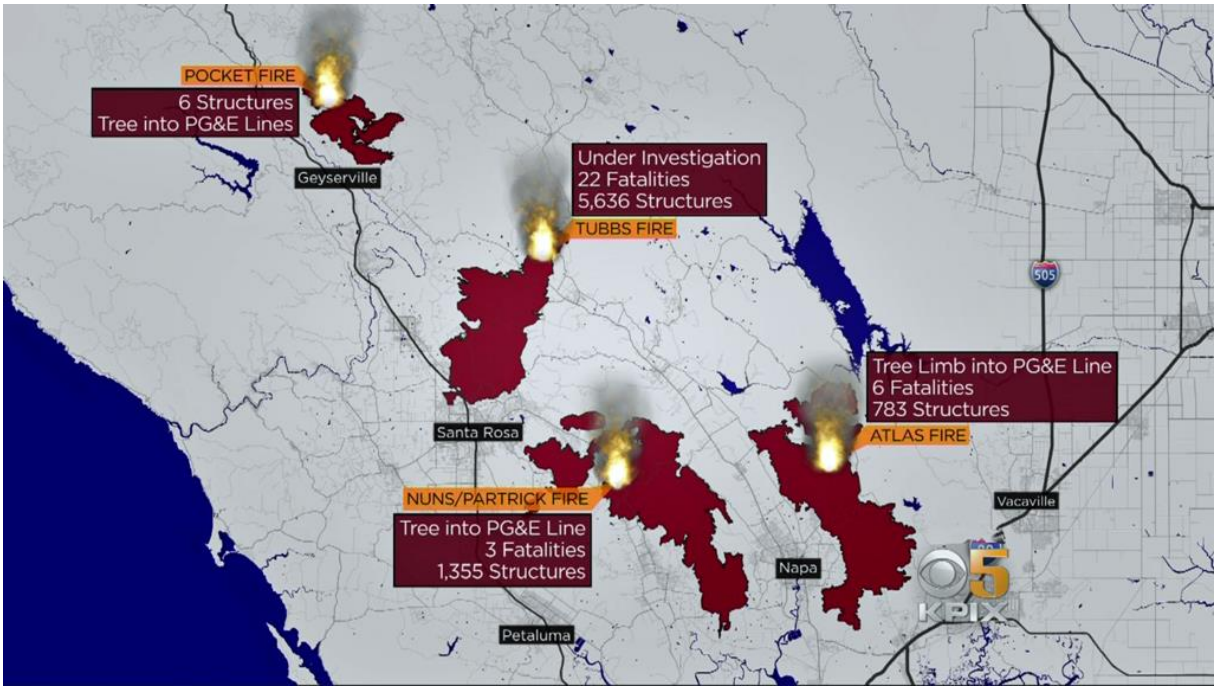


Figure 11. A map of California, presenting the wildfires that were caused due to the equipment of PG&E. Source: (KPIX 5 CBS San Francisco, 2018).

3.4.1 Privacy and security

Privacy means the right to keep someone's personal matters and relationships secret (Cambridge Academic Content Dictionary, n.d.-a). Security is defined as protection of a person, building, organisation, or country against threats such as crime or attacks (ibid.). Since personal data is information which can identify a person, direct or indirect from the data, the right of privacy is concerned (European Data Portal, 2018). In Europe, The General Data Protection Regulation (GDPR) concerns the protection of personal regarding free exchange of information (Directive (EU), 2016). It can be argued that the GDPR entangles the concept of open data since data cannot be provided as open data when it includes personal information. This regulation forces organisation to act more carefully and responsible when dealing with personal data of customers, personnel or others (Netherlands Enterprise Agency RVO, n.d.). This can be considered a barrier for organisations to provide open data. However, it can also be argued that the aim of GDPR is to encourage sharing and re-use of data, instead of penalising data use (European Data Portal, 2018). It aims to simplify the regulatory environment and highlight the

benefits of data re-use in accordance with data privacy regulations. Therefore, the GDPR sets a clear framework in which data sharing and re-use is safe and is not overshadowed by insecurity and anxiety of misuse (ibid.). This highlights the fact that the GDPR in fact supports the concept of open data through increasing transparency and knowledge on how to provide open data in a safe and legal way (European Data Portal, 2018).

Thus, the only situation when GDPR directly affects open data is when the data contains personal data (ibid). Before publishing open data, providers need to ensure that no personal sensitive information is disclosed (Beno, 2016). An example where disclosure of personal data can violate the privacy regulation, according to the GDPR, is given by Beno (2016): a school that decides to disclose students' personal performances, without concerning personal information harms the students privacy (Beno, 2016). The published data could release a student's address, surname and telephone number which could identify a person, therefore not in compliance with the GDPR (Directive (EU), 2016). Thus, legal precautions are made in order to ensure privacy and security (ibid.). Hence, to be able to provide open data without breaching the law (Geiger & von Lucke, 2012). This can be done for example by anonymisation of the data, which is the process of removing personal identifiable information (European Data Portal, 2018). As a result, the data can no longer be considered as personal data and is no longer subject to the GDPR.

Furthermore, disclosure of secured information, such as intellectual property, trade secrets and financial information can put an organisation at risk (Tankard, 2012). When such information is published an organisation might experience threats. Hence, it could provoke a terroristic attack (ibid.). Therefore, organisations are careful with the data they provide since sensitive data in terms of privacy and security might harm their overall security and liability (Conradie & Choenni, 2014) Although it is desirable, and often mandatory, to ensure data security, it takes a significant effort of the provider to do so (Beno, 2016). Maintaining a secure infrastructure to publish the data is the first effort to make. Furthermore, to avoid malicious data manipulation by third parties or other users, data integrity and authenticity need to be secured. The additional costs, effort and risks associated with the maintenance of a secure data infrastructure and those of securing the data integrity and authenticity present a barrier for providers (Beno, 2016).

3.5 TECHNICAL QUALITY BARRIERS

In order for the data to ensure a valuable return on both user and provider side, the data needs to be fit for use (ISO - ISO 19157:2013; ISO - ISO 9001:2015). Because every user may have a different purpose when using data, a guarantee of quality cannot be given (Barry & Bannister, 2014; Janssen et al., 2012)., An accuracy check on the data needs to be done before the data can be used for a certain purpose. Such a check can be accomplished through contact with the data creator and by enquiring about the correctness of the data in terms of the completeness of the metadata (Janssen et al., 2012). Often this is not possible as contact information, if present at all, does not trace back to the actual data creator (Janssen et al., 2012; Martin et al., 2013). Even when the metadata is present good data quality is not guaranteed as there is no single standard for metadata for all users (fit for use) (ISO - ISO 19157:2013; ISO - ISO 9001:2015) resulting in heterogeneity of metadata models and different vocabularies (Martin et al., 2013). At worst, this could limit or prevent the user from reusing the data (Janssen et al., 2012).

The absence of agreed quality standards, possible lack of a supporting infrastructure (data portal), as well as fragmentation of manipulation software and applications can present technical barriers to data openness.

3.6 UNKNOWN, UNATTAINABLE AND UNUSABLE

From the user perspective, barriers can be placed in three main categories: '*unknown*', '*unattainable*' and '*unusable*' (Welle Donker et al., 2019). Barriers associated with *unknown* appear when users are unaware of the data in general, e.g. the data does not exist for the user. Even when users are aware of the data, these barriers might still exists when the dataset cannot be found. Disintegration of data is also considered an unknown barrier if no complete version of the data set can be found in data portals (Welle Donker et al., 2019).

Applying organisational aspects to datasets such as a restrictive licenses, conditions of use, datasets without an open licence or against a fee makes them *unattainable*, the next type of barrier. *Unattainable* barriers for the data user are caused by the data provider determining which type of use/user they want to include in their data regime, and what the data quality should be (Welle Donker et al., 2019; Welle Donker & van Loenen, 2016; Zuiderwijk & Janssen, 2014). It may be that there is no specified licence with the dataset. Lack of clarity whether the user can re-use the data or how to use the data, together with technical aspects that keep users from re-using the data are considered as *unattainable* barriers. Furthermore, the prevention of the use of a dataset with a restrictive licence makes them *unattainable* (Creative Commons, 2019; Open Data Reader, 2016 ; Miller et al., 2008; Open Knowledge Foundation, n.d.-a). Legal barriers only arise when the licence is formulated in such a manner that use is forbidden in any form. Licence which do allow use under certain conditions and terms keep the data attainable for the user (Creative Commons, 2019).

Financial and legal barriers, formed by the data provider, have influence on the technical quality and use of the data. This is translated in *Unusable* barriers which occur when the re-use of the data is prevented due to the insufficient quality of the data (Welle Donker et al., 2019). When a certain domain knowledge is needed in order to open, use and interpret the data, the data is only attainable and usable for users with this knowledge and excludes those without. For the remaining users without the domain knowledge this data remains unattainable and unusable (Janssen et al., 2012). Technical quality barriers are the next type of barrier associated with usability issues from the user perspective. A lack of usability emerges when the data is unfit for the purpose of the user due to the quality of the data (ISO - ISO 19157:2013; ISO - ISO 9001:2015). Again, because not all users have the same purpose when using the data, a technical quality cannot be guaranteed (Barry & Bannister, 2014; Janssen et al., 2012). Because this barrier does apply to every user it is a difficult barrier to pin down especially when the data is incomplete, not up to date, not using standards or contains incomplete metadata. The lack of the facility for the user to give feedback or submit a request for desired data is seen as an *unusable* barrier (Welle Donker et al., 2019).

Barriers that cause the data to be *unknown*, *unattainable* and/or *unusable* for the data user are often interconnected (Janssen et al., 2012; Van Loenen et al., 2011). Institutional issues, expressed in unwillingness to change, may be associated with financial and legal barriers for the data provider (van Loenen et al., 2011). A financial barrier faced by the data provider as the result of costs that are not recoverable based on the current business model can result in a technical quality barrier faced by the data user as a result of poor quality data infrastructure. Hence, all the different barriers are a reflection of the willingness of the provider to make the data more open, the starting point to their removal (Bregt et al. 2012).

3.7 SUB-CONCLUSION

The barriers between the different levels of open data identified so far are presented at Figure 12. For this research the public undertaking is considered to be the data provider. Firstly, it shows the organisational barrier that affect the attainability of the data (Welle Donker et al., 2019). This is addressed by the data provider in terms of regime, quality of the data and the type of user (the upper part of the model). Open data starts with the institutional willingness of the data provider to adopt an open data regime, to include users beyond the internal user, and suits this purpose of 'every user'. This first dimension influences the openness of data (upper part of the model) (Bregt et al. 2012). The level of *play*, *find* and *share* (the lower part of the model, for the user) is determined by the data regime of an organisation that matches the quality to the type of data user (Deloitte Analytics, 2012; ISO - ISO 9001:2015; Safarov et al., 2017; Welle Donker et al., 2019; Zuiderwijk & Janssen, 2014). First of all, the regime faces institutional, financial and legislative barriers when steps towards an open data policy are made (Afuah & Tucci, 2001; Janssen & Kuk, 2007; Zuiderwijk & Janssen, 2014). Creating an open data regime requires willingness of the data provider to do so and this includes finding financial funds and applying licenses that allow the user to share and re-use the data (Afuah & Tucci, 2001; Janssen et al., 2012; Van Loenen et al., 2011). Again, the quality needs to fit the purpose of the user (ISO - ISO 9001:2015) and is influenced by the types of potential users. In order to create more openness through improved quality of the data, improving attainability and usability, financial and technical barriers need to be tackled. To modify the quality of the data for external and public users technical skills and money are required (Martin et al., 2013). Legal barriers may be faced when changes in licence are required enabling the sharing of data with external trusted parties whether or not under conditions. This is due to the fact that access to, and modification of the data is not only limited to the data provider. External parties also have the rights to access and modify the data through a the new licence. Therefore, new legal barriers are faced for the data provider to limit data misuse and data fragmentation, which might be caused by external parties as a result of more rights. When legislation prevents the *re-use* of the data for every user, as described by level 3, there are liability risks for the data provider when the step towards level 3 is taken. These risks can be expressed in financial, actual and/or reputational damage from false conclusions drawn from the data by the users, or from publishing private and secure data (Creative Commons, 2019). Financial barriers are encountered when making the data findable and accessible through search engines and/or data portals for external users. Subsequently, meet the users' purpose of the data (Welle Donker, 2018). Barriers associated with task complexity are faced when the users shift from being external trusted parties to public users. Contrary to internal and external users, identified in levels 1 and 2, the data user is unknown to the data provider in level 3. The domain knowledge of the user is difficult to assess which is expressed by the task complexity barrier (Barry & Bannister, 2014; Martin et al., 2013). So, it is difficult for the data provider to know whether the published data suits the knowledge domain of all the user (Janssen et al., 2012).

As a result of the *attainable* barriers imposed by the data provider, the ability for the data user to *find*, *play* with, and *share* or re-use the data can decrease (Van Loenen et al., 2011). In order to make the data more findable for users other than the internal users, financial and task complexity barriers are faced. To lower the task complexity for the user by making the data more findable in a search engine and/or data portal (known), a financial fund is needed (Martin et al., 2013; Welle Donker, 2018). The same barriers are faced when it comes down to *play.* Financial investment by the data provider is required to create the possibility for the user freely to use and modify the data (Janssen et al., 2012). The additional barrier of technical quality is faced by *play* since the published data need to be recent, in a machine-readable format and possible to be downloaded in bulk. This makes it more usable for the user (Welle Donker et al., 2019). The application of different types of licenses and conditions of use present barriers

to *share and re-use* as this decides whether and under which conditions. The barriers faced by the requirements of share/re-use are associated with the application of different types of licenses as this decides whether and under which conditions the data can be shared and re-used. The attainability of the data for the user is determined by the data provider (Creative Commons, 2019; Miller et al., 2008; Open Data Reader, 2016; Open Knowledge Foundation, n.d.-a; Welle Donker et al., 2019).

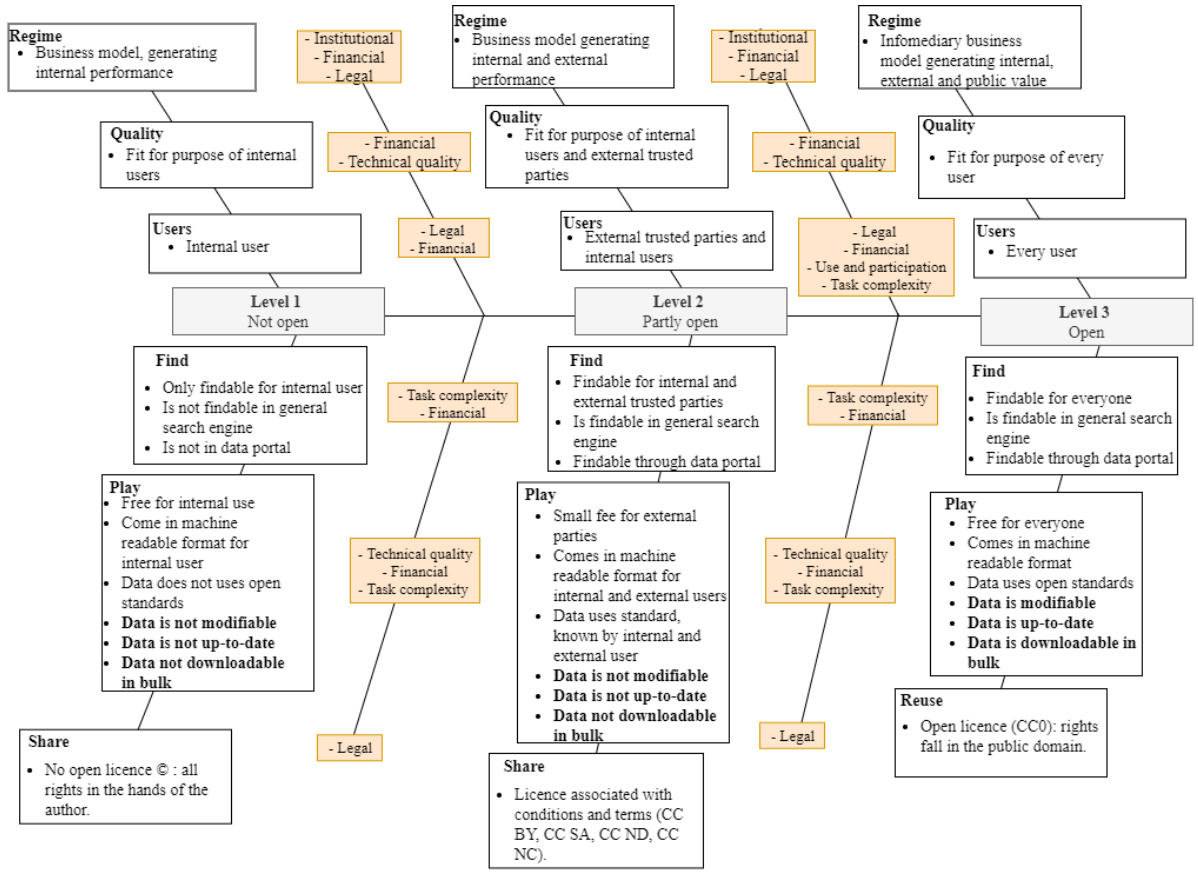


Figure 12. A multi-dimensional model of the three identified levels of open data (chapter 2), including the identified barriers between the levels which are faced when a transfer to a higher level is intended (chapter 3).

Figure 12 is a result of literature review on both open geographical data (resulting in the three open data levels) and barriers towards open data (the barriers between the open data levels). It provided the input for conducting the interviews with the two public undertakings of this research: Port of Rotterdam and Schiphol Airport. The model was presented to both the public undertakings to research at which level they can be placed and which barriers they face towards open data. The results are presented in the next chapter: Port of Rotterdam and Schiphol Airport.

Results

Chapter 4 : Port of Rotterdam & Schiphol Airport

Chapter 5 : Open data of Liander

Chapter 6 : Opportunities to overcome barriers

Chapter 7 : Discussion

Chapter 8 : Conclusion

Chapter 9 : Recommendations

4 PORT OF ROTTERDAM & SCHIPHOL AIRPORT

In common with many organisations, for both Port of Rotterdam (PoR) and Schiphol Airport providing open data is not the main focus of their data policy. An interest in data sharing is growing by both ventures and challenges are faced. From interviews that were conducted with the two public undertakings it is clear at which open data level they find themselves as are the challenges to be faced in the future.

4.1 PORT OF ROTTERDAM

Port of Rotterdam is the biggest sea harbour of Europe, situated in the Harbour of Rotterdam; it has a length of 42 kilometres and a land size of 12.600 ha. The harbour has deep-sea connections with more than a thousand harbours around the world. The Port Authority has an important role in developing, organising and managing the logistic activities in the Harbour. The companies' shares are partly held by the Municipality of Rotterdam (70%) and the Dutch government (30%). The shares are not listed on stock exchange which makes PoR an unlisted public limited company. PoR is a data driven company, specifically by geographical data. One such use of the data is the Harbour Master management Information System (HaMIS), an interactive system in which all shipping traffic is planned, monitored and administered (Port of Rotterdam, 2020a). Data plays a key role in achieving their vision: improving the port to the safest, most efficient and sustainable port in the world (Port of Rotterdam, 2020). The team responsible for managing the geographical data is captured in Port Development (DP), Environmental Management (EM) and Asset Management (AM). For this research two Analytics Specialists and one Advisor from AM were interviewed to assess the data openness of Port of Rotterdam. The next sub-sections are based on information gained during these interviews. The interviews were conducted on November 23, 2020.

4.1.1 Data Governance

According to the interviewee, when data is collected by Port of Rotterdam the first question is not 'how can we share this data with others outside the company?' (Analytics Specialists PoR, personal communication, 2020). PoR has around 1,200 employees operating in different departments, ranging from commercial to marine, and infrastructure areas. The organogram of PoR is shown in Figure 13. All the different departments hold data which is mainly used for internal operations within the department and are not shared with other departments. Hence, the different departments do not have access to datasets of other departments; there is no inter-department sharing within PoR. For instance, Asset Management, which primarily maintains geographical data, is not given access to contracts held in the department of finance. This is confidential information and only accessible to the financial department. The same applies vice versa; the department of finance does not have access the geographical information held by Asset Management (AM). An overview of the different data used by the departments is lacking (Analytics Specialists PoR, personal communication, 2020). Currently this results in insufficient data governance regarding internal data exchange. Permission to exchange data in an adequate manner is often withheld (ibid.).

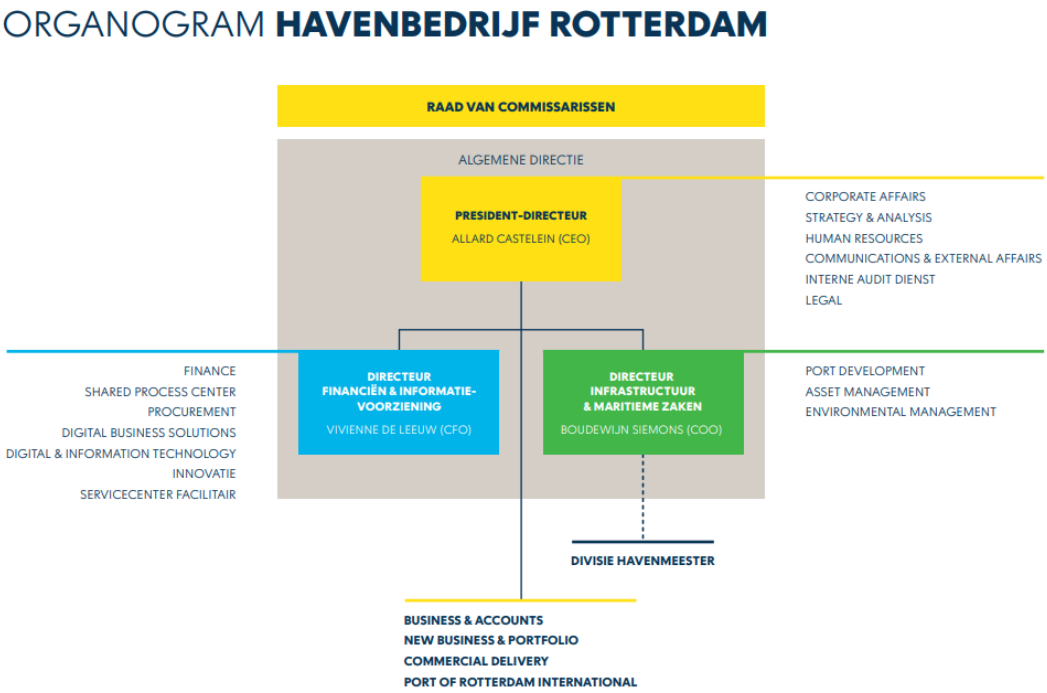


Figure 13. Organogram Port of Rotterdam, including the different departments. Source: (Port of Rotterdam, 2020).

To create an overview of the used data by the different departments, 12 different data domains were recently identified to get more insights in the data used by the different departments (Analytics Specialists PoR, personal communication, 2020). According to one interviewee, the data domains that contain valuable information are: commercial data, financial data (debtors, creditors), data related to client contact, seaport money, harbour money, contract revenues, trans-shipment of goods, data on sustainable energy and asset data (Asset Management). Asset management is taking the lead in identifying the data they use and try to encourage other data domains to do this as well. The goal of this development is to get a clear description of the data used for a project or data product. To achieve this goal Port of Rotterdam focusses on five key data questions (Analytics Specialists PoR, personal communication, 2020):

- Who is the data expert?
- What is the definition (context) of the data?
- In which processes (data domains) and products is the data used?
- Which data quality rules apply to the data?
- In which systems is the data being maintained / reported?

It is stated by one interviewee that the intended outcome would present an overview of the impact each data set has on other domains within the company (Analytics Specialists PoR, personal communication, 2020). Progress in this area is clear from some improved interconnected domain relationships, for instance between Asset Management and Port Development. Both departments (see Figure 13) started a project together regarding placement of sensors in the port area to see what data they could retrieve, such as quay information, and what could be done with this data internally. Once clarified, it might be possible to share this with third parties (clients) as well – if this lines up with the interest of PoR (Analytics Specialists PoR, personal communication, 2020). Thus, within PoR most of the collected data is still kept within specific departments and not shared with other departments.

4.1.2 Datasets

The geographical data within PoR is held primarily by AM and used to ensure a safe and smooth route for the arriving cargo. The data can be divided into datasets associated with infrastructure on land and nautical infrastructure (AM advisor PoR, personal communication, 2020). Thereby, the assets that can contribute to a cargo transitioning within the harbour are documented. Table 4 gives a selection of datasets per category that are used most to support a safe movement of cargo. Most of the information about nautical infrastructure is documented in the electronic fairway map (ENCs, Electronic Navigation Charts). ENCs can be divided into inland ENCs, representing the inland waters, and Port ENCs, representing approaching fairways to the sea port. The requirements for nautical information of a Dutch ENC are determined by the director-general of Rijkswaterstaat (ministry of infrastructure and waterways) and are based on international standards and contains waterway markings, bridges, locks, moorings and signs (Rijkswaterstaat, 2017).

Table 4. Most used geographical datasets of PoR. Divided in infrastructure on land and nautical infrastructure. Source: analytics specialist, personal communication, 2020.

Type of geographical data	
Infrastructure on land	**Nautical infrastructure**
Road networks	Constructing and dredging of waterways
Traffic signs	Bridges and locks
Harbour entrances	Waterways markings
Areas (not) within ownership of PoR	Moorings and signs

4.1.3 Users

Before sharing data with others outside PoR, it is important to get an understanding from the data used by internal departments (AM advisor PoR, personal communication, 2020). Now that data is becoming ‘the new oil’ in terms of company value, the awareness amongst internal data users is growing (Analytics Specialists PoR, personal communication, 2020). Meta data regarding the data held by 12 data domains are kept in a data catalogue called ‘Mavim’ and gives specific attribute information on all the different datasets. According to the interviewees, only internal users with a valid VPN connection can enter this data catalogue. Whether the data can be used for internal, external or only confidential purpose is documented per dataset to classify every dataset as internal, external (openbaar = public) or confidential data. The classification is determined by the level of confidential information the data contains (Figure 14). Data of which sharing is allowed is classified as external data (Figure 14a). In this case, only parts (attributes) of the data such as anchorage grounds, lampposts or trees can be shared outside the company. External sharing requires prior approval. ‘Others’ are mainly identified as third parties such as clients or building contractors that are of interest for the operations of PoR. The interviewees identified two main types of users: Internal users (employees) and external users (third parties, e.g. clients).

The decision to share data with clients is sometimes considered in terms of competition benefits (Analytics specialists PoR, personal communication, 2020):

"Data on, for example, inland cargo routes is shared with clients to encourage them to choose Port of Rotterdam for their cargo routes instead of the Port of Antwerp."

Internal data is only allowed for internal use and cannot be shared with others outside the company, for example data about gullies or sensors (Figure 14b). Sensor data is considered as valuable data for the internal performance. Misuse of this data by other users can harm the internal performance of PoR (Analytics specialists PoR, personal communication, 2020). Furthermore, confidential data is only available for specific commercial use as it contains personal data (data within the scope of the General Data Protection Regulation, GDPR) and business sensitive data (Figure 14c). Accordingly, this data cannot be shared with other data domains without the permission of the data owner and even then, only for internal purposes. One example is cost and budget data of assets, which is confidential to the department of AM (Analytics Specialists PoR, personal communication, 2020). Another example is information on details of cargo repair reports; sharing this data with other departments than AM has no additional purpose or benefit for PoR (AM advisor PoR, personal communication, 2020). The interviewees states that data is can also be classified as confidential data in terms of security; certificated data of cargo ships and routes can contribute to hijacking a shipment (ibid.). Finally, a payment is never charged for sharing and accessing the data. According to the interviewees this is not necessary since there is a data delivery agreement that is intended to avoid liability issues and potential cost issues for the company (Analytics specialists PoR, personal communication, 2020).

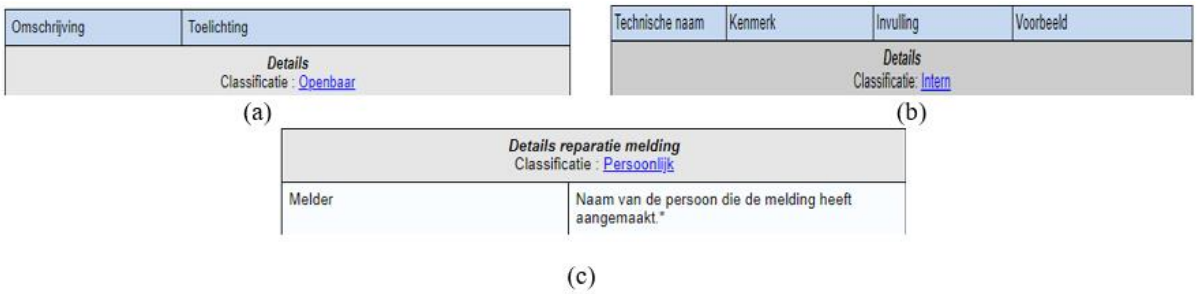


Figure 14. The three different classification categories of the datasets within Mavim. **(a)** *Openbaar* = public (external), which is available for external purposes. **(b)** *Intern* = internal, which only available for internal purpose or authorised clients. **(c)** *Persoonlijk* = personal (confidential), which is only available for specific commercial purposes as it contains personal data (i.e., data within the scope of the General data protection regulation), and can only be internally used after permission the department, responsible for the data. Source: Mavim, Port of Rotterdam, 2020.

Although the main data user is not identified as other users than employees and third parties, PoR does share data indirectly with others (AM advisor PoR, personal communication, 2020). This is done through the BRO (basisregistratie ondergrond = base registration underground). The BRO keeps track of changes in Dutch soils regarding the drilling and probing and must comply with legal requirements, set by the government. With the start of a new project, PoR requires information retrieved from new drilling and probing activties. As from 2 January 2021 on, PoR is obliged to share the new information from that moment on with the BRO (Analytics specialist PoR, personal communication, 2020). The information of the BRO is made available for everyone on the PDOK (Publieke Dienstverlening op de Kaart = Public Mapping Service), an open data portal that provides up-to-date governmental data. As a result of this government imposed obligation, this data is shared 'indirectly' with everyone on PDOK.

For users other than employees and third parties, PoR provides a free navigation tool for routes on container shipping on their website (https://rotterdam.navigate-connections.com/voyages). This data

shows the lead time, transfers, departure and arrival time of the cargo. This tool is provided with the next disclaimer: “The Navigate service may only be used by you as an end-user within the domains of PoR and may not be deep linked to and/or may not be embedded in end user’s and/or third party websites and/or services.” (rotterdam.navigate-connections.com). So this data can only be used by an end-user within the domain of PoR. Other data on infrastructure, weather, and tides and port charges is presented on an open data portal. Yet, the quantity of the data is minimal, possibly insufficient according to one interviewee (Analytics specialist PoR, personal communication, 2020). Moreover, the data is from the year 2014, thereby outdated. There is also data provided through ArcGIS online, however, this data is also outdated and therefore not suited for current research or other purposes. Hence, this data cannot be accessed without an ArcGIS online login, which is not free of charge.

Since most of the collected data is used by internal departments, generating internal performance, PoR can be placed in level 1 of the open data model. Regarding internal data exchange, the data policy can be placed prior to level 1 since collected data is still kept within specific departments and not shared with other departments. Most of the data is only findable and usable for the internal user and, therefore, unknown and unusable for everyone. Even for internal departments only internal-, and sometimes confidential data is findable and usable.

4.1.4 Legal barriers

A significant part of the datasets held by Port of Rotterdam contains confidential information which makes it not possible to share complete datasets with third parties (Analytics specialists PoR, personal communication, November 2020). The data of which sharing is allowed (classified as external) can be shared on request; if there is no reason for PoR to share the data in terms of performance increase, this is avoided (Analytics specialist PoR, personal communication, 2020). However, when the intention is to share the data as open data, legislation is not the most significant barrier, according to the interviewees. One interviewee quotes (Analytics specialist PoR, personal communication, 2020):

“Legal barriers are not the most difficult barriers as an agreement can state whether the data can be used, and for which purposes. It is not too challenging to determine what the best legal way is to ‘cover’ the data which is shared”

Here, ‘cover’ is interpreted as data protection from misuse that can lead to false conclusions or reputational damage (Beno, 2016). Currently, this is done with a *Gegevens Levering Overeenkomst* (data delivery agreement) between third parties and the data department within Port of Rotterdam (Analytics specialist PoR, personal communication, 2020). When the agreement does not suit the case or purpose of the data an adjustment can be made by the legal department of PoR to make it fit. For example, an agreement can state that the data can only be used by two employees with a secured account within the organisation of safety region Rotterdam Rijnmond. This is done to keep control over who uses the data, where it is used for, and when it is used. This is necessary to prevent the risk of data fragmentation outside Port of Rotterdam (Analytics specialist PoR, personal communication, 2020).

Most data is kept within the organisation as a result of fear for liability risks in terms of misuse by external parties or public users (Analytics specialist PoR, personal communication, 2020). It is stated that less than 50% of the data is shared with third parties and it is expected that this will not increase in the future (Advisor from AM, PoR, personal communication, 2020). This is due to the fact that more data sharing with third parties is not essential for achieving more internal performance (ibid.). Moreover, some of the data contains confidential information from clients, prohibiting sharing with others according to the data delivery agreement. An example is given by the interviewee who states that sharing data on maintenance reports of cargo ships with others than the clients is simply not allowed. Hence, it

will only lead to costs rather than benefits as the benefits of such action are not clear (Advisor from AM, PoR, personal communication, 2020).

In summary, PoR only shares their data with trusted external parties under legal conditions (Creative Commons, 2019), documented in a data delivery agreement (Analytics Specialist PoR, personal communication, 2020). This manner of sharing reflects the legal barrier between level 1 and 2 where the data delivery agreement decides who can use the data (internal or external trusted parties) and, if shared, with trusted external parties under conditions and terms (Figure 12).

4.1.5 Technical barriers

In contrast with legal issues, greater issues are faced with the development of a portal when not just one single department, such as Asset Management, needs access to the data but every department within the company as well as third parties (Analytics specialist PoR, personal communication, November 2020). Before developing an extra technical management department, devoted to setting up such a data portal, this needs to be discussed with the IT department. The interviewee explains that this requires work, time and, most importantly, costs. These costs include infrastructural costs as a new funding system needs to be defined that will facilitate potential revenue which are used to develop a data portal and publish the data (Welle Donker, 2018). Once the infrastructural costs are sunk, PoR will face costs to promote the use of the data portal, operational costs and to maintain the data quality and keep it up to date with the latest version: maintenance costs (Analytics specialist PoR, personal communication, November 2020). Thus, setting up a data portal which can be accessed by all the departments and third parties will lead to costs that are not recoverable by the current business model of Port of Rotterdam (Martin et al., 2013; Welle Donker, 2018). These technical barriers put PoR between level 1 and 2 of the open data model and are associated with *find* and *play* (Eaves, 2009). Overcoming these barriers will result in a data portal which improves the findability for internal departments as well as third parties. Moreover, the operational and maintenance costs will ensure a higher level of usability of the user. However, a new data portal will also be associated with a higher task complexity for the user (Barry & Bannister, 2014).

4.1.6 Institutional barriers

All the interviewees agreed that creating the internal awareness of the value of data is at the starting point for developing open data (Analytics Specialists PoR, personal communication, November 2020). Developing the 12 data domains is considered a first step towards more internal awareness. Yet, not all the data domains are willing to contribute to this new initiative. There is still a strong feeling of 'us and them' amongst departments. One of the interviewees explains this feeling as the fear of giving 'their data' away to other departments (Analytics Specialists PoR, personal communication, November 2020):

"By sharing the data with 'them', 'we' lose control over the data as it is not clear what the other department will do with it".

The other departments can (mis)use it in any manner but 'we' can be held liable for their actions. Hence, if 'we' lose control over the data to 'them', it is unknown under which terms and conditions the data can be used. Moreover, the initial purpose of the data may not be clear for 'them'. Consequently, control over own data is lost which can have consequences for the overall liability for PoR; it is unknown for what purpose the data is (mis)used (Analytics Specialists PoR, personal communication, November 2020). PoR could be said to be 'risk averse', fearing deviation from business as usual (Barry & Bannister, 2014).

The 'us and them' feeling also exists towards third parties such as container companies (Analytics Specialists PoR, personal communication, November 2020). When PoR shares data with container

companies, located in the port, the data is not solely owned by the PoR but also by the container company. Misuse by the container company, such as sharing or comparing the data with the port of Antwerp, may result in a reduction of the internal performance and decreased revenue (Analytics Specialists PoR, personal communication, November 2020). Therefore, sharing the data with external parties is still considered a great barrier to share the data with external parties (ibid.).

Another institutional barrier, as mentioned by one of the Analytics Specialists, is the significant effort needed to change to another data system; PoR already uses their data system for decades (Analytics Specialists PoR, personal communication, November 2020). The public undertaking is not yet familiar with the benefits of open data, only with the associated legal and financial risks (Analytics specialist PoR, personal communication, November 2020). This is a result of the general lack of awareness for open data amongst public and private organisations in the EU (Dalla Corte, 2020). Due to the general lack of awareness for open data amongst the employees of PoR, from a governance perspective fully adapting to an open data is not possible yet (Analytics Specialists PoR, personal communication, November 2020).

Furthermore, time and effort need to be spent by all departments to achieve the same level of skills and ability to work with the new data system (Analytics specialists PoR, personal communication, November 2020). A first step towards awareness is made through the development of the 12 data domains. It is now possible to understand the purpose of the data, where it comes from, and how it impacts other departments (Analytics specialists PoR, personal communication, November 2020).

The institutional barriers faced by PoR can be placed between level 1 and 2 of the multidimensional model (Figure 12). Together with financial and legal barriers, institutional barriers prevent PoR from generating more external and public value with their data as this is not their main vision.

4.1.7 Technical quality barriers

When data sharing is addressed, data which is classified ‘confidential’ needs to be filtered by attributes of which sharing is not allowed (Analytics specialists PoR, personal communication, November 2020). It is mentioned by an interviewee that a double check on the data quality is necessary when the data is shared in terms of liability (ibid.). Since every user has different purposes for the data, the utility/quality of the data changes per user. One user for example needs the road network in the port for calculating the total squared kilometres covered by the road network. Another user might need the road networks in order to revise the maintenance terms. Together with the data owner, a new system needs to be set up to assess the quality of the data per case (Analytics specialists PoR, personal communication, November 2020). This complies with quality check mentioned by Janssen et al. (2012) to make the data fit for the users’ purpose (ISO - ISO 9001:2015, 2015; Janssen et al., 2012). This will result in operating and maintenance costs (Barry & Bannister, 2014).It is recommended by an interviewee that questions need to be answered when checking the quality of the data (Analytics specialists PoR, personal communication, November 2020). Firstly, the data provided needs to fit the area, the location, of interest. For example, the nautical depth of the Port’s water will not fulfil the purpose of revising the maintenance terms of the road networks. Secondly, the data needs to be ‘complete’ in terms of the quality attributes needed by the user. For calculating the squared kilometres of the road network for example, the data should involve the total squared kilometres and not the layer construction of the roads. Lastly, data needs to be documented in a standardised manner so that other departments can understand the data equally. When the quality of the data is found not sufficient, a warning needs to be given (ibid.). This applies to sharing data both with other departments as well as third parties. Getting a clear image of the data requirements through these questions will benefit the data quality. One interviewee reported that the data is normally not sufficient because the information is not complete to begin with. Sharing old and

incomplete data could lead to liability issues and costs that are currently avoided where possible. To tackle this quality barrier, maintenance and operational costs need to be incurred to make the data fit for the purpose of other departments and external parties (Analytics specialists PoR, personal communication, November 2020).

The technical quality barriers faced by PoR can be placed between level 1 and 2 in Figure 12 since the next step is to make the quality of the data fit for internal departments and external users but not yet for everyone.

4.1.8 Financial barriers

It is stated by all interviewees that legal, technical and quality barriers are all grounded in one overarching barrier: the financial barrier. Extra costs as a result of legal and liability issues, e.g. misuse of the data by the user, are avoided with a data delivery agreement (Analytics Specialists PoR, personal communication, November 2020). Financial barriers in terms of technical facilitation of a data portal are faced due to the fact that setting up a new open data platform costs time and money. Costs are incurred when developing the data portal, when publishing the data, and when keeping the data complete and up-to-date. In terms of legal issues, the fear exists that liability issues, as a result of incorrect legal restrictions, leads to financial risks. When data is misused by an internal department, or even an external user, the costs are for PoR as the company is responsible for the source data (Analytics Specialists PoR, personal communication, November 2020). Moreover, financial benefits of open data are not clear yet for PoR. So, invested time and money spent in preparing the quality of the data for users outside the company is seen as a risk instead of a valuable investment (Analytics Specialists PoR, personal communication, November 2020). Beside the institutional barrier, unwillingness to change, the fear for financial risk is present especially because any financial benefits are yet unknown to the company (ibid.).

4.1.9 Sub-conclusion

Regarding the different open data levels, identified in chapter 2, Port of Rotterdam can be placed in the situation prior to level 1. Although data is shared with internal users, it is not yet shared with all internal users. Data is collected within departments and not shared with others. PoR can be placed in level 1 regarding the data regime as the collected data is used by the internal user, generating internal performance. Data is shared with third parties when this is in the interest of business activities. Until now, data has never been shared exclusively to generate public value. Awareness of sharing data is growing within the company. This has resulted in 12 data domains creating an overview of the data that is used by the departments and the impact it has. When this is complete, a next step will be to create more openness towards third parties to generate both internal and external performance. Public value for its own sake is not yet on the horizon. This next step will be towards level 2 of the different open data levels, dealing with, in order of significance to PoR, financial, technical, quality, institutional, and legal confidential barriers (Figure 15). Legal barriers are not considered to be the biggest issue since the conditions and terms can be determined in the data delivery agreement on sharing data. Technical and quality issues, however, are considered difficult barriers to deal with since a new technical department needs to be developed to make the data more findable through a portal for third parties (between '*find*' and '*play*' in Figure 15). The quality needs to be fit for the purpose of third parties costing the company time and money. The (un)willingness to share data with external parties is growing within the company but is still a significant issue, placing the institutional barrier not on the top of the list. The drive to share data is there but the next step is to find the most suitable technical and financial solution for it. As yet, level three, where data sharing is replaced by data re-use and the user is identified as everyone, is, according to the interviewees, a step too far away for Port of Rotterdam. The barriers currently faced are listed in Table 5, including the main reason why the barriers are faced.

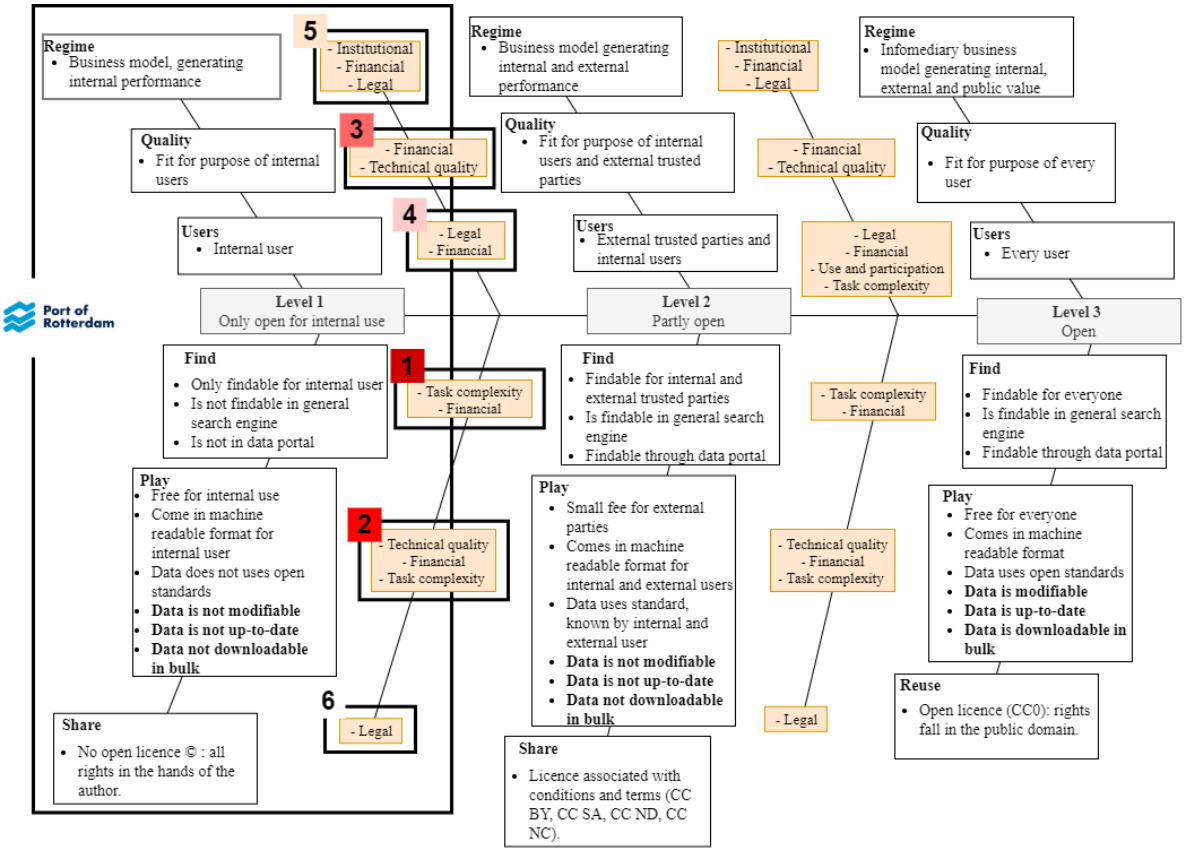


Figure 15. The level at which Port of Rotterdam can be placed in (level 1) and the barriers which are faced, numbered from most significant (1 between find) to least significant (6 between share).

Table 5. The barriers currently faced by PoR when making their data more open: ranging from most significant (financial) to least significant (legal).

	Barrier	***Reason***
1	Financial	Technical, quality, institutional and legal barriers are all grounded in the extra costs, such as adaption costs, infrastructural costs and structural maintenance costs. Costs of which the benefits are not guaranteed.
2	Technical	Data needs to be findable through a portal for third parties and other – external, individual – users. Therefore, technical issues are faced such as developing such a data portal.
3	Quality	Since every user has different purposes for the data, the utility/quality of the data changes per user. Therefore, the quality needs to satisfy the data requirements which requires time and money spent in still unknown internal benefits.
4	Institutional	Some data domains are yet not willing to contribute to the new initiative of sharing data amongst other departments. Let alone providing open data. There is still a strong feeling of 'us and them'. This is for example the case between the finance department and Asset Management; financial contracts are confidential to finance.
5	Legal	Misuse by the users can be minimised by a data delivery agreement with the data user. This is done to limit the liability of the company. Yet, a data delivery agreement is different for every case, therefore needs to be revised with a new intention of data sharing with third parties.

4.2 SCHIPHOL AIRPORT

Amsterdam Schiphol Airport is the largest airport in the Netherlands and plays an important economic and social role in Europe. It is considered one of the most connected airports in the world and facilitates 332 international connections. Regional airports, international alliances and cooperation enhance this international connection. Schiphol is held by the Royal Schiphol Group, with the Dutch government, the municipality of Amsterdam and Rotterdam and Groupe ADP (an airport operator) as stakeholders. Schiphol's annual contribution to the Dutch economy is estimated at 10,4 billion euro. Moreover, 114,000 jobs are created through the operations of Schiphol. The vision of Schiphol is to create value for their clients, partners and the environment and the society as a whole. Achieving this vision starts with establishing a quality of their network, managing passengers- and cargo movement. The airport holds datasets associated to asset management, finance, HR, IT operations, parking, passenger experience, real estate, retail, security and traffic and transport (Royal Schiphol Group, 2020). For Schiphol data offers potential for innovative solutions when it is processed and applied correctly. Because other consumers and businesses can also benefit from Schiphol's data Schiphol, data sharing with third parties and public users is encouraged. For this research, one interview with two interviewees were conducted. One interviewee is a GEO-IT solution architect and one is an Enterprise data architect. Instead of two separate interviews with the interviewees, one interview was conducted since both interviewees preferred the same date and time. The interview was conducted on November 27, 2020. The following sub-sections are based on information gained during this interview.

4.2.1 Users

For Schiphol, the main data user types are: internal users (employees), partners (third parties, such as airlines and main contractors), and public users (passengers at Schiphol and people who can access the data portal). Similarly, data have three purposes at Schiphol (GEO-IT solution architect Schiphol, personal communication, 2020). The first one is for the private user who is using the data for internal purposes. The second one is for partners who are using the data of Schiphol for their own interests, under an agreement with Schiphol. The third one is used by or for the public user such as passengers or clients who are provided with data to choose the best possible route from A to B at Schiphol. The private user is enabled to access most information of the dataset (attributes) with more rights while the public user only has access rights to certain attributes of the dataset (GEO-IT solution architect Schiphol, personal communication, 2020). The choice to provide different numbers of attributes with the data to the users is discussed later on in this chapter. Thus, Schiphol provides the internal user, external and public user with their data (Enterprise data architect, personal communication, 2020). Therefore, include the user of level 1, 2 and 3 of the multidimensional model.

4.2.2 Data Governance

For Schiphol, data sharing is the main goal of their data governance (GEO-IT solution architect Schiphol, personal communication, 2020). Through data sharing Schiphol wants to create extra value together with users and for users. Rather than making the data open, data needs to be accessible; if at all possible, the user can access the data if the user wants to. This is achieved by the working method of their data portal. Within the data governance, 'the user' was initially interpreted as the passengers starting their journey at Schiphol. The users were not passengers arriving at Schiphol for a transfer or end destination. From the data governance at Schiphol, the intention is to achieve the most positive effect with their collected data, for the included users. All the data provided to the passenger is supplied to ensure a smooth process for the passenger to get from point A to B (through customs, security, etc.) (GEO-IT solution architect Schiphol, personal communication, 2020).

According to the interviewees, collected data needs to play a leading role in ensuring a smooth route for the passenger from A to B. Schiphol intends to involve the highest number of users possible with the collected data (GEO-IT solution architect Schiphol, personal communication, 2020). An example of a successful use of data resulted from the revision of the Wayfinding service in the Schiphol app. By including a digital map and route navigator in the Schiphol app (Way Finding), data on the walking routes at Schiphol was collected in 2014 and used in 2015. This data includes the estimated time of arrival at the destination, shows the passenger if and where construction takes place and what the best alternative route is, for example (GEO-IT solution architect Schiphol, personal communication, 2020). The data could be used by everyone who had downloaded the (free) Schiphol app. For Schiphol, this was the first public app where you could navigate within Schiphol. After a year, Schiphol reviewed the feedback on the tool to see what the airport had achieved with this app for public users and partners (airlines). Positive feedback was given on the fact that the data was accessible for everyone with the app. Simultaneously, this was the negative feedback: the data was only available for passengers with the Schiphol app. Some passengers who arrived at Schiphol without the app did not know that Way Finding existed. The interviewee mentioned that the app was also quite 'heavy' with data initially: "You cannot imagine that now, but a couple of years ago an app could not function in an optimal manner due to too much data" (GEO-IT solution architect Schiphol, personal communication, 2020). A year later, the whole Way Finding part was dropped from the Schiphol app. The more valuable parts of the Way Finding tool were reviewed and found to be the wayfinding map data API (POI and Base map) part. Schiphol could offer this API part to other users such as airlines (DELTA, KLM) to include in their airline app. Hence, the technique to determine the passengers' location within the navigation tool is also provided as an API to other airlines (the beacon API). It turned out that most passengers at Schiphol use the airline's app instead of the Schiphol app. As a result, Schiphol could serve more passengers by providing the API to other airlines than only through the Schiphol app. Thus, by opening the Way Finding API to other airlines, the airlines could include that in their app, and so benefit more passengers (GEO-IT solution architect Schiphol, personal communication, 2020). By providing this API to other airlines as well, Schiphol included more data user in their data governance and created an added value for more than just passengers (Enterprise data architect, personal communication, 2020).

Supplying the data in a consistent manner is another goal of the data governance at Schiphol (GEO-IT solution architect Schiphol, personal communication, 2020). The time of departure on the monitors in the airport is the same time of departure in the app. Here, all information sources supply the same information resulting in a convenient manner of data sharing to the user. Although data sharing is the main goal for Schiphol, it cannot conflict with the responsibilities Schiphol has for their data. These responsibilities are:

- The aviation responsibility to the passengers: ensure a straightforward border transfer of passengers and make sure that they make their flight on time.
- The commercial responsibility (on land): ensure safe and good parking, safe Premium experience (access to lounge clubs, quick check-ins, and quick security pass for members), and safe shopping.

Data sharing is often encouraged by the aviation industry's responsibility towards the passengers, while it can interfere with commercial responsibility. Sharing more detailed data can help the passengers journey but also benefit the aviation industry. Contrary to the aviation industry's responsibility, data management is also a commercial responsibility to generate internal performance. It is important to find the balance between these two responsibilities when data sharing is considered. The recurring question regarding data sharing is formulated by the interviewee as (Enterprise data architect Schiphol, personal communication, 2020):

"Can the data be shared without harming the commercial (on land) responsibility?"

As for data governance of Schiphol it can be said that it intends to generate public value if it does not interfere with the internal performance. The intent of data governance is therefore in line with the data regime of level 3 of the multidimensional model.

4.2.3 Datasets

According to the interviewee, there is a significant amount of data collected by Schiphol:

"Almost everything you see at Schiphol is documented as geographical data in vector/point cloud data, aerial pictures, 360 degrees pictures, or 3d pictures".

This applies for data above and on the ground, and inside and outside buildings (GEO-IT solution architect Schiphol, personal communication, 2020). Currently Schiphol is working with data on low voltage (wall plugs and lights), luggage handling systems in 3D, flying aircraft, flying birds (which they detect with radar in the air) and driving cars on airfield sites. Everything that moves (or does not) is captured in geographical data (GEO-IT solution architect Schiphol, personal communication, 2020).

The data portal is divided into two segments, one for private, internal use (internal data portal) and one for public, external use (developer centre). The internal user has access to every dataset available. The data available for public use is provided through APIs on the data portal of Schiphol (developer.schiphol.nl). It presents those datasets that are available for public developers. These APIs are associated with information on scheduled flights, operational flights, waiting times, or boarding passes for example (developer.schiphol.nl, 2021). A registration on the website is, however, needed since Schiphol wants to keep track and control of the rights they give away to users and type of users. The type of users of which the API is considered valuable is given with the different APIs. For example, the Operational flight API (information on scheduled flights from and to the airport), adds value to the travellers' journey, the processes of airlines and ground handling agents. The Wayfinding API offers relevant information for travellers, airlines, floor managers and other staff. When an API is requested, sign-up information is required and agreement to the Legal Terms and Conditions of the Schiphol API platform, as presented in Figure 16. The Legal Terms and Condition states, amongst other subjects, that: "the users may not use the APIs to transmit information that is inaccurate, harmful, misleading, offensive or is perceived as unwanted mass communication (spam)." No further definition on the words inaccurate, harmful, misleading and office is given (developer.schiphol.nl).

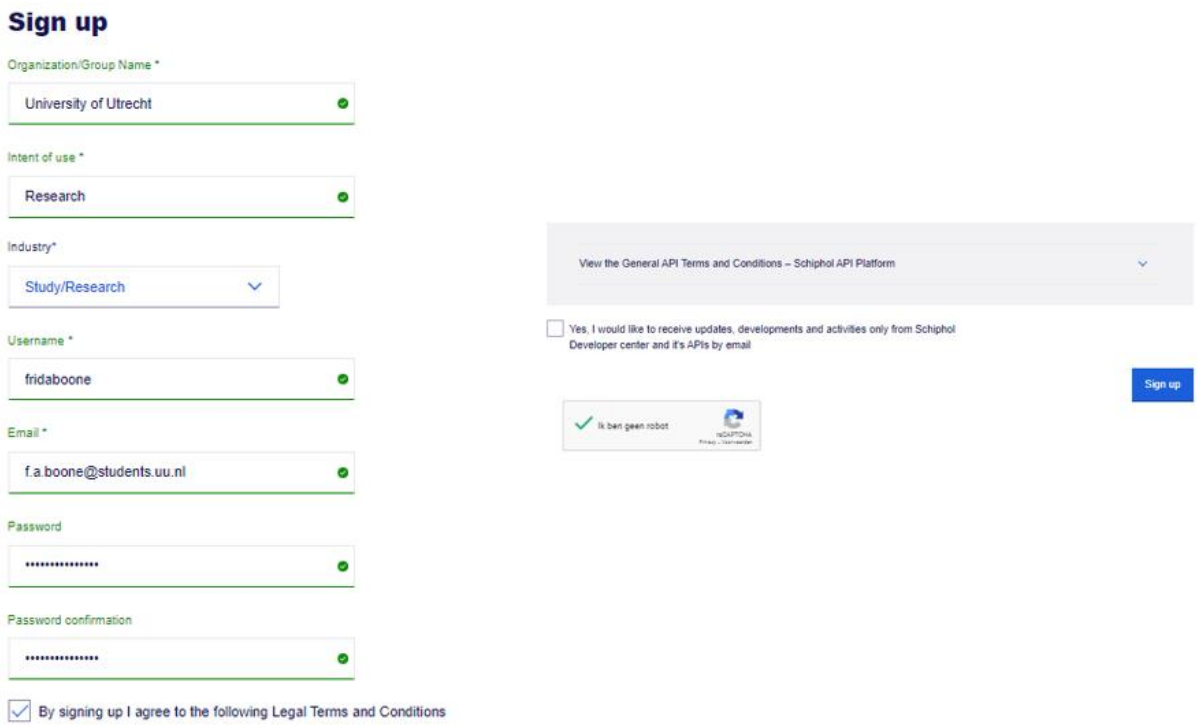


Figure 16. Screenshot of the sign up form for the API's on the Schiphol API platform, filled in with the authors' information. Source: developer.schiphol.nl.

The data portal is developed to ensure that users have access to the data that is available, not necessarily to provide open data as defined in this research (GEO-IT solution architect Schiphol, personal communication, 2020). They follow the concept of 'FAIR'[3] which seeks to improve the findability and accessibility of the data for the user (ibid.). If the user wants to access the data, a request can be made and the rights for either the internal user, trusted party, or external user can be given. Although there are currently no costs charged for this service, effort and time is involved providing this data; requests are considered and the source data is modified to the level of detail, related to the rights of the user. Because this service involves costs and time, not all datasets are added to this data portal. More time is spent on modifying datasets for third parties (airlines), who use the datasets for purposes that are also of interest for Schiphol. Subsequently, the right to access more attributes than external users is given (Enterprise data architect Schiphol, personal communication, 2020). The selection of the available datasets on the public data portal is based on three principles:

- The data has already been collected and modified for the purpose of internal projects and the data is already used for internal performance. In this case, Schiphol seizes this possibility to provide the collected data to the public user as well.
- The 'low hanging fruit principle'; if data can be collected for little costs and effort, the intention is always to provide this data to the public user as well.
- It is legal to share the data with the public user.

The data portal presents the available datasets but is also set up to gives some context to the data to show what the data represents and for what it used. Stated by the interviewee: "If data does not have any context, it can be misused. If the data states: there are 500 flights, it is not put into a context, which can result in wrong interpretations. Which flights are meant by this? Cargo flights? Military aircraft? In

[3] FAIR data: findable, accessible, interoperable, and reusable data.

order to put the data into context it needs to state which kind of flights are represented by the dataset, such as 500 passenger flights" (Enterprise data architect Schiphol, personal communication, 2020).

In sum, with the datasets provided on their data portal, Schiphol tries to improve the findability and accessibility of the dataset using an API. This is not done freely for all datasets and a user's data request is still needed (GEO-IT solution architect Schiphol, personal communication, 2020). This manner of working corresponds with level 3 in that everyone can find the Schiphol data. However, the fact that the data is not directly open for usage does not comply with the 're-use' of Level 3.

4.2.4 Legal barriers

According to the interviewees, published data by Schiphol needs to be protected from misuse that can result in liability issues (Enterprise data architect Schiphol, personal communication, 2020). Some data cannot be shared because it is retrieved from third parties and under an agreement that states that it cannot be shared with others outside Schiphol. This applies, for instance, for detailed flight information retrieved from KLM. Usage of this data is only allowed for specific operations of KLM and Schiphol, which is documented in an agreement between the two parties. This issue can be regarded as a legal barrier that prevents data sharing beyond the trusted parties. Hence, the conditions and terms are documented in a data processing agreement which contains:

- Subject matter and duration of the processing.
- Nature and purpose of the processing.
- Type of personal data to be processed.
- Categories of data subjects.
- Rights and obligations of the controller.

This agreement covers both the handling of personal data in accordance with the General Data Protection Regulation (GDPR) and the received confidential data from both parties (GEO-IT solution architect Schiphol, personal communication, 2020). This agreement between Schiphol and the trusted party states that this data can only be used by these parties so this data cannot be provided as open data to the public user. Should data be provided to the public user, Schiphol is faced with the same legal barrier to prevent misuse and distortion of the data but applied to a less detailed dataset (ibid.).

4.2.4.1 Privacy and security

By decreasing the level of detail of the public dataset Schiphol tries to protect secure and private data. This is for instance done to protect Schiphol from terrorism. According to the interviewees a floorplan of Schiphol that reveals too much detail can provoke terrorism (GEO-IT solution architect Schiphol, personal communication, 2020):

"Schiphol is after all a target for terrorism"

The floorplan that is presented to the public user only contains public areas, not all the areas from the original source data. The original source data, which comes from Asset Management, has information on a much more detailed level; it presents every asset at Schiphol ranging from the walls, to the light bulbs present in the arrival and departure halls at Schiphol. Furthermore, locations of, for example, the armoury and emergency routes are represented in the source dataset as well, however not in the public one, as this information can be misused by terrorists (Enterprise data architect Schiphol, personal communication, 2020). Moreover, the locations of the surveillance cameras are not present in the public dataset since these could show the blind spots in secured areas which can be used to plot the best route for an intruder or criminal (Wijk et al., 2020). The intention of Schiphol remains to supply a clear

floorplan to the public, without releasing sensitive data that could enable terrorist attacks. The floorplan provided for public use is presented in Figure 17. Besides security means, there are two other reasons why source data is not shared with the public according to the interviewees. First, it is assumed that the source data is not relevant for public user; they are not interested in the walls and light bulbs present at Schiphol (GEO-IT solution architect Schiphol, personal communication, 2020). Secondly, the source data has 'too much data' (attributes) which makes it impossible to supply a well-structured dataset to the public as presented in Figure 17 (ibid.).

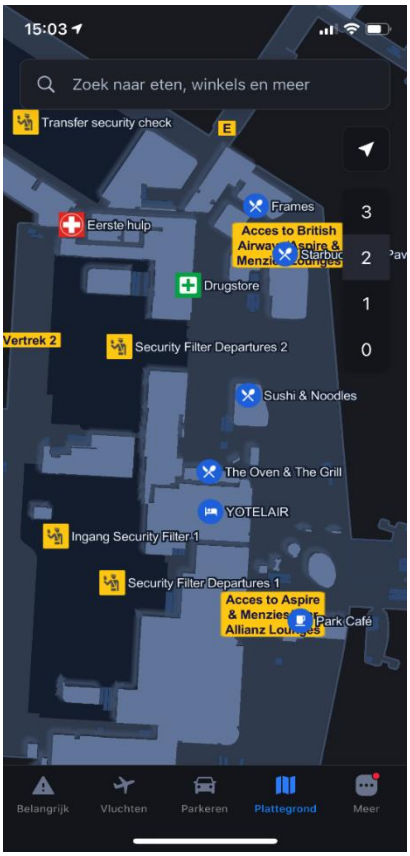


Figure 17. A screenshot of the floorplan of the second floor at Schiphol, presented in the Schiphol app. It shows shops, restaurants, toilets, heath care facilities and information on security points, departure halls and gates. This is the level of detail for the public. Source: Schiphol app.

Schiphol also faces privacy barriers when sharing their data (Enterprise data architect Schiphol, personal communication, 2020). A flight ticket with a flight number, an arrival time and date can practically identify a person. It presents information about a fixed number of people for one specific flight. One interviewee states that it is possible to pinpoint a person of interest when the information of the flight ticket is released with the slightest amount of extra information (Enterprise data architect Schiphol, personal communication, 2020). If Schiphol publishes this, it would conflict with the General Data Protection Regulation (GDPR) (ibid.). This is how privacy issues can stop Schiphol from publishing their data for every possible user (in level 3 of the model). Both privacy and security barriers conflict with the data quality for the user. Personal data or business confidential data can only be provided as open data if the data is aggregated to anonymous levels of detail which complies with the GDPR. This may significantly impact the usefulness of the data, i.e. the data quality for the end user (GEO-IT solution architect Schiphol, personal communication, 2020). This corresponds to the aforementioned conflict between data sharing and the responsibilities Schiphol has when it comes to data: although data sharing is the main goal for Schiphol, it cannot conflict with the commercial and aviation responsibilities Schiphol has for their data (GEO-IT solution architect Schiphol, personal communication, 2020).

4.2.5 Institutional barrier

According to both interviewees, people who disagree with the data-related activities of Schiphol could intentionally misuse the data: "These people misuse the interpretation of open data to damage the reputation of Schiphol" (GEO-IT solution architect Schiphol, personal communication, 2020). Through the open data of Schiphol, data users can intentionally distort and misuse the provided data to gain advantages or cause harm to the provider's reputation (ibid.). It is possible that the data is used for other purposes than the primary purpose intended by Schiphol. For example, through open data, information on building sites and nitrogen emissions can be retrieved from the open data of Schiphol. This allows the nitrogen footprint of Schiphol to be calculated, based on often assumed, inadequate calculation (GEO-IT solution architect Schiphol, personal communication, 2020). It is mentioned by the interviewee that this 'guessing' of the nitrogen footprint of Schiphol is not based on real numbers and facts. Schiphol is, nonetheless, faced with defamatory statements about their nitrogen footprint (ibid). This can be considered as false conclusions drawn by the user from the open data of Schiphol.

Furthermore, it is stated by both interviewees that every operation or action that takes place in the area op Schiphol is associated with the name 'Schiphol'. Misuse of open data can tarnish the reputation of Schiphol and this requires time and effort to repair (GEO-IT solution architect Schiphol & Enterprise data architect Schiphol, personal communication, 2020). The interviewees underpinned this statement with the case of wrong accusation as a result of traffic accidents. Recently, it was claimed that the road map of the public roads around Schiphol represented incorrect information and caused traffic accidents (ibid.). In reality, the data on road networks is kept by, and is the responsibility of the municipality. The news reported that the accidents were caused by incorrect data provided by Schiphol (Enterprise data architect Schiphol). Schiphol needed to prove that the traffic issues were not caused by their data. This was necessary as they could experience reputational damage because of incorrect accusations from publishing open data owned by others such as airlines (ibid.). This case was presented by both interviewees as an example where Schiphol's name was used in a wrong matter (GEO-IT solution architect Schiphol & Enterprise data architect Schiphol, personal communication, 2020).

Misuse of the data, as described in both cases above, resulted in the institutional barrier which prevents Schiphol from sharing the data with everyone in level three of the multidimensional model. Though, it could be argued whether such cases can be turned over and used in the favour of Schiphol. Publishment of the real nitrogen footprint of Schiphol could clear the air of false conclusions, drawn by users. In the case, misuse of open data by the user can be challenged by Schiphol through providing real facts and figures. One could even argue that Schiphol should publish the data it has on road networks within the property. By publishing this data as open, all can use the correct data, instead of making false accusations. As a result, Schiphol could provide more open data which encounters false conclusions and misuse, which could benefit the reputation of Schiphol.

4.2.6 Future data policy

Considering these barriers, Schiphol faces two options when they want to achieve open data as is described by the new open data directive (GEO-IT solution architect Schiphol & enterprise data architect, personal communication, 2020):

Firstly, a 'business as usual' approach can be adopted. Schiphol provides information on which data is readily available. Moreover, how this data can be requested by the use, yet only when Schiphol believes the user has good reasons to use the data. Schiphol decides whether to make the data available for the user; it is not directly accessible. This means that the data is not directly 'open' for the data user.

Alternatively, Schiphol could open all the data, including the source data, for everyone. For example, Schiphol could publish the source data of the floorplans or flight information for everyone without the requirement of a request. This would cause security, privacy and legal issues as sensitive data is released. Publishing source data on the floorplan for instance could lead to terrorist risk since it would expose security rooms, personnel rooms and the armoury.

Neither option 1 or 2 are a convenient option for open data. Option 1 tends towards open data, however can not be called open data since users still need to get the rights to access the data. Option 2, which tends more to open data, causes security, privacy, legal and liability risks for Schiphol. It is not possible to provide open data as it is described by the new open data directive according to the interviewees (GEO-IT solution architect Schiphol & enterprise data architect, personal communication, 2020). Moreover, Schiphol faces difficulties with its data responsibilities (aviation and on-land). From the aviation point of view (operations and process) more open data for the passengers means better performance as more data delivery to their passengers means better decision making regarding the processes from A to B. However, from the commercial point of view on land (Parking, Previum and shops), data is required to achieve commercial goals. Rather than opening all the data, data needs to be management for commercial purposes. This does not affect the continuous drive of Schiphol to provide open data since Schiphol knows the value of open data. The landscape of data is now captured to the internal purpose, the next step is to capture it to the external purpose. One interviewee postulated that 3 questions needed to be considered when open data were to be discussed in the future (Enterprise data architect Schiphol, personal communication, 2020):

1. 'Can open data be provided in terms of technical aspects?' For Schiphol, technical issues, such as developing an open data portal, will not be the main problem for Schiphol since they have the tools and techniques to make it happen. Thus, technical barriers are not faced by Schiphol when opening data.

2. 'Is providing open data the goal?' Schiphol wants to provide open data, however, a balance need to be found between the aviation data and the commercial data. Providing open data cannot come at the expense of the commercial business and position. This point refers to their data governance, their data regime.

3. 'Is it allowed to provide open data?' This can covers legal, security, and privacy barriers which can affect Schiphol's liability.

4.2.7 Sub-conclusion

The evidence suggests that Schiphol can be placed in level 2 of the multi-dimensional open data model. The goal of their data governance is to share data with internal, external and public users. Sharing data with the public user is however only executed when there is no interference with the commercial data responsibility of Schiphol. Sharing data with external trusted parties comes with an agreement that covers liability issues regarding misuse of the data. For the public user the available datasets are presented on the open data portal of Schiphol. Although the available data is presented here the data cannot be directly used by the public user. To control the data used by the public user a data request – subsequently a registration – is needed from the user. After a request, the data is provided to the user with the suitable rights and level of detail –which is less than the source data that is intended for the internal user. Even though the data is available for the external and public user through sharing, it is not directly accessible and so it cannot be 're-used' as it is described in level three. Barriers that will be associated with the next step towards open data are a legal security barrier, legal privacy barriers, legal confidential barrier and institutional barriers (Figure 18). The interviewees highlight that the main issue that causes the legal privacy and security barriers is the level of detail of the data. This applies first of all the legal, privacy barrier for Schiphol. The data reveals too much detail, such as the location of the

armoury, that could assist a terrorist attack. Schiphol is considered a target for terrorism and opening this data could harm the security of Schiphol. Secondly, too much detail can reveal private data about individuals at Schiphol and so breach the GDPR. These legal barriers are associated with '*share*', preventing re-use of the data. Confidential agreements with third parties cause the third legal barrier. Schiphol cannot share data that is retrieved from third parties if re-use is only allowed by internal users of Schiphol; this data cannot be shared with public users. This legal barrier is associated with the 'user' preventing the user to be 'everyone' rather a specific external user. Lastly, due to fear of false conclusions drawn from the open data of Schiphol, not all data is made openly available according to both interviewees. This is an institutional barrier. Both interviewees state that Schiphol has already experienced reputational damage as a result of false conclusions drawn by users and, as a result of that, they are not willing to adapt to a fully open data regime (GEO-IT solution architect Schiphol & enterprise data architect, personal communication, 2020). However, it could be argued that publishing open data could prevent reputational damage. By publishing open data, Schiphol creates the opportunity to provide good and correct data, which can prevent the risk of false conclusion drawn by the user. So instead of fearing open data, it could also be considered a solution. The barriers currently faced are listed in Table 6 and presented as barriers one, two and three in Figure 18.

In contrast to the Port of Rotterdam, financial, technical, and quality issues are not considered to be the main causes for the barriers faced by Schiphol. These barriers are listed as numbers four, five and six – associated with '*quality*', '*find*' and '*play*', in Figure 18. Financial issues due to development and maintenance costs of open data are not considered since costs for developing and distributing data for public use are already made and not considered a great issue. A technical barrier will also not be the main problem since Schiphol already succeeded in setting up a data portal for the users (developer.schiphol.nl). Quality barriers are not faced in that sense that modification of the data for public use is not possible; it is possible but does take some time and effort.

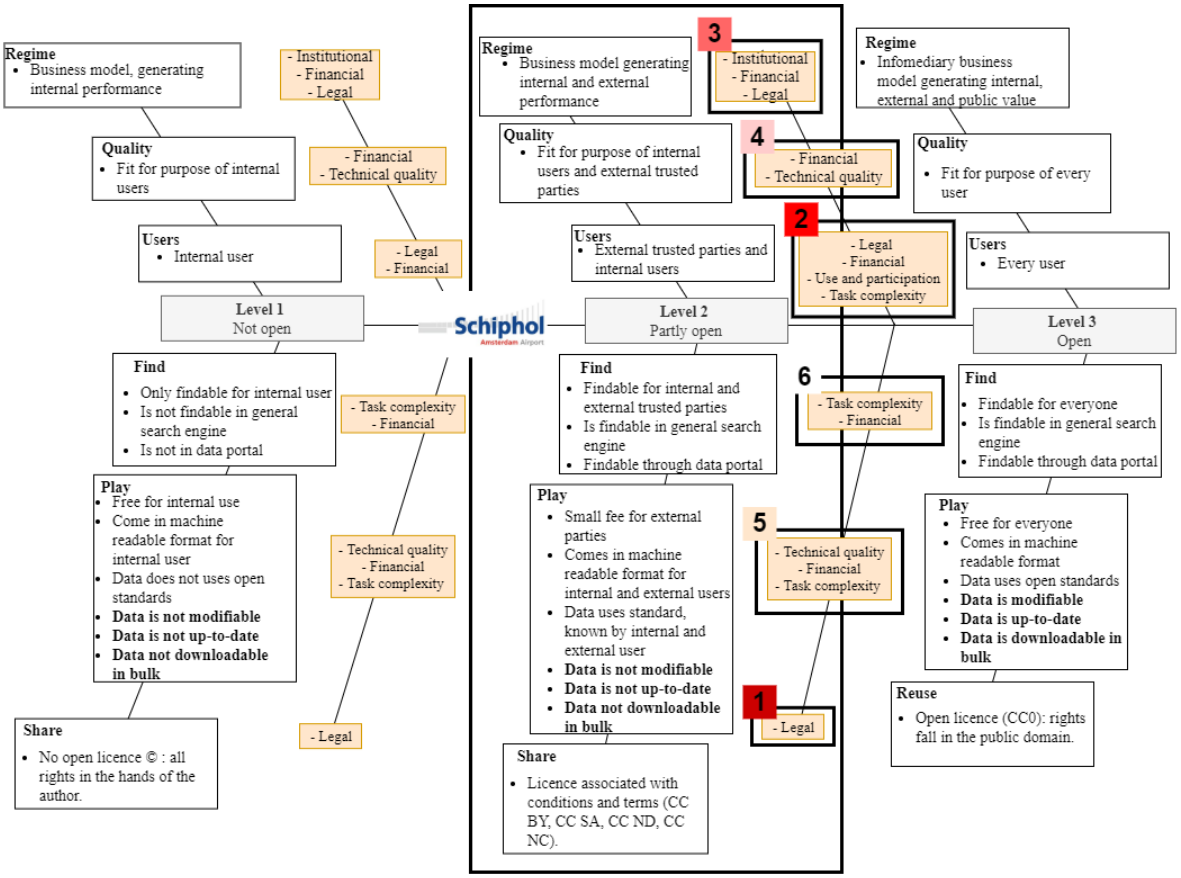


Figure 18. The level at which Schiphol can be placed in (level 2). The barriers which are faced are numbered from most significant (1 = share) to least significant (6 = find).

Table 6. The barriers currently faced by Schiphol when making their data more open: ranging from most significant (security) to least significant (legal).

	Barrier	***Reason***
1	Legal (security)	Data can reveal a level of detail which contains sensitive data that can provoke terroristic attacks.
2	Legal (privacy)	Data can contain personal information that cannot be shared in compliance with the GDPR.
3	Legal (confidential)	Data retrieved from third parties contains confidential data which need to be protected from sharing with others.
4	Institutional	False conclusions can be drawn by public users, from open data. Schiphol has experienced this already and is therefore not willing to open all data.
5	Quality	data needs to be aggregated to a level that is acceptable for the GDPR, however, needs to satisfy the data requirements of the end-user.

5 OPEN DATA OF LIANDER

At present, for most organisations it is clear what open data is, and which social and economic values it can offer (European Commission, 2020a). However, 'open data' is not a sort of package that organisations can buy in a shop that comes with instructions on how to apply it. This raises the question how organisations can best share their data or even provide it as open data. Since data sharing is relatively new, hard facts and figures are not yet available to indicate the best way of data sharing (Support Centre or Data Sharing, n.d.). It is often best to learn from the success of other organisations by learning how they overcame the barriers to open data. This was acknowledged by the European data organisations who focussed on data sharing for both governments as well as private organisations (Data.overheid.nl, user meeting, 2020).

One organisation that has successfully opened their data is Liander, which is a Dutch utility company. The company is providing open data since 2014 and define open data as digital data that is made available for everyone through the internet (Juffermans, 2015). They provide open data on:

- Energy and gas consumption per year, per type of house
- Energy day profiles
- Malfunctions;
- Smart meter data for planning and realisation
- Electricity and gas distribution networks, containing:
    - Number of gas connections
    - Length of gas pipes
    - Length of excavation sensitive gas pipes
    - Dominant small consumption electricity connections
    - Length electricity cables
    - Transformer capacity
- Small consumption gas use (up to 40 m^3/hour) and energy use (3x80 ampere)

This data is provided for their area of operation in the Netherlands (Figure 19) (Liander, n.d.). For the information of this chapter, an interview was conducted with a product developer of Liander on December 18, 2020.

Figure 19. Liander's area of operation in terms of gas and electricity. Source: Liander.nl/opendata.

5.1 INSTITUTIONAL MOTIVATION

According to the product developer at the 'change team' of Liander, providing open data starts with the question 'why open data?' (Product developer, personal communication, 2020). The motivation to provide open data needs to be clear because it costs money. Liander wanted to contribute to a better collaboration with the regions within their area of operation (van Loenen & Bregt, 2012). Furthermore, there were two motivation points for Liander to provide open data. The first one being the social benefits open data brings to the society. Liander often received individual questions regarding the utility usage of their network and the bottlenecks within their network. These requests needed to be answered one by one which resulted in a time consuming activity for Liander. By publishing their most sought-after data they wanted to create more opportunities for the users (such as municipalities) to work with their information without having to ask Liander for input every time. With data readily accessible, the user can work with Liander's information and substantiate their own plans with meaningful data. In turn, such plans could benefit to Liander's network. This potential societal benefit was the main motivation for Liander. Saving time on individual data questions was also considered a motivation to continue to provide open data (Product developer, personal communication, 2020). The other motivation arose from the fact that Liander wants to contribute to energy transition. Providing open data on for example consumption per year, per type of house can give an understanding on the best possible manner to lower the energy consumption. It did take some time for Liander to get fully behind this motivation because they were not legally obliged to provide open data. It was a returning question whether they were even allowed to do so in terms of data privacy (GDPR) (Product developer, personal communication, 2020).

5.2 USERS

The users were defined as key users of Liander's data (parties), and public users (everyone) who can access the open data portal (liander.nl/opendata). Within the key users, the parties are:

- Municipalities
- Regional energy strategics
- Research institutions
- Universities
- Housing associations

These parties request information on for example the location of the pipes and cables, the gas usages per household or the solar panel usage per household. No discrimination is made between the different parties, all data requests are handled equally. Hence, every party needs to have access to the same data in the same manner. The same applies to the public user, who can access the data through the open data portal and can use it for every purpose. Contrary to the work method of Schiphol, no registration is needed to access the data. Both the use of key user parties and the public users have proven to contribute to innovation. Open data of Liander is for example used for the visualisation of energy use (Municipality of Amsterdam), and the potency of solar energy through the Netherlands (zonatlas.nl).

5.3 CURRENT DATA GOVERNANCE

When a data request comes in that deviates from the data that is already open, the data is shared if it is allowed. However, against a fee as time is devoted to process the data (Product developer, personal communication, 2020). By doing this, Liander can keep track of who uses their data. However, to avoid time and costs Liander prefers to provide all the data online if sharing is allowed (in terms of the GDPR).

How does Liander determine when it is allowed to publish the data in terms of the general data protection regulation? The interviewee quotes: "Well, *the devil lies in the detail!*". For Liander 'aggregation is the answer to this question. When the data is aggregated to a level at which the GDPR does not apply, everything is possible' (Product developer, personal communication, 2020). As a result, information is mostly aggregated on the level of a postal code areas – avoiding data on one single house, which would go against the general data protection regulation. For housing corporations which need information on houses in different postal code areas on for example renovation status another aggregation is applied. In this case an aggregation is made based on (at least) 5 houses in the different postal codes. This information of interest is delivered by the housing corporation as an input for an information product. In this information product, the outcome is automatically provided in such a manner that it suits the corporation's interest. 'E-Atlas', which is the name of the information product is offered to housing corporations through an annual subscription. As a result, any specific request (for example the electricity usages) can be satisfied through a customised aggregation, presented in an automated information product (Figure 20) (Klep, 2018). This service has saved Liander time and costs compared to previous individual requests. It can be stated that providing open data starts with finding a motivation. When Liander reached this point the goal was to share all data within the terms of private or secret data controls (GDPR) (Product developer, personal communication, 2020).

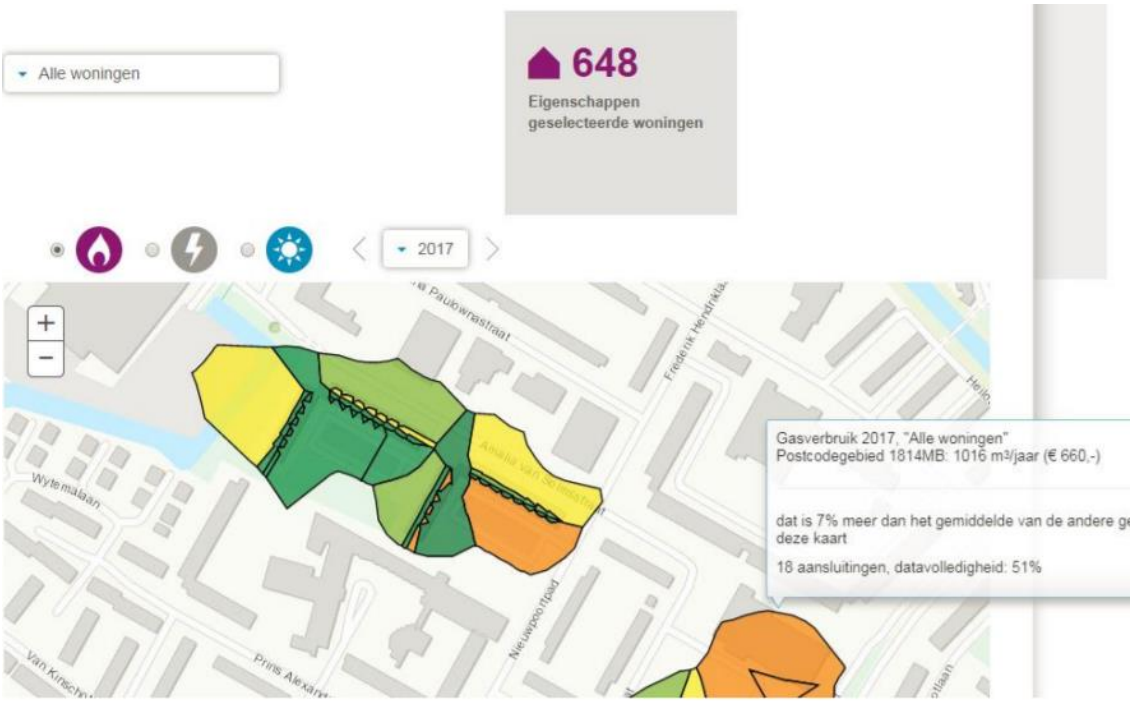


Figure 20. A screenshot of the E-Atlas for housing corporations. Here, the settings are set to the gas usages for all house types, in 2017, aggregated on at least 5 houses. Source: Klep, 2018.

5.4 Legal Issues

The interviewee indicated that the first question that needs to be answered is whether sharing of data is allowed under the law (Product developer, personal communication, 2020). Similar to Port of Rotterdam and Schiphol, the fear of liability issues is always present. Although Liander has experienced significant legal objection to data sharing from different Dutch legal authorities, liability issues were never experienced by Liander. This was prevented by setting a disclaimer on the data. The disclaimer used for this data is the Creative Commons BY (CC BY). The CC BY allows re-users to distribute, adapt, and build upon the data if attribution is given to the creator. Credits to the creator must be given (Creative Commons, 2019). According to the Open Knowledge Foundation, the CC BY satisfies the requirements of open data (Open Knowledge Foundation, n.d.-b). Although most of the liability risks are covered by this license, publishing open data is not risk free (Product developer, personal communication, 2020). If the data presents a wrong representation of the reality, Liander is responsible. For instance, when the dataset of the pipes and cables shows the wrong location compared to reality the organisation can be held liable for this issue and face extra costs or reputational damage. Nevertheless, this has never kept Liander from publishing open data. Indeed, feedback on the quality of the data is actively encouraged by Liander as it gives information on the quality requirements of the user (Product developer, personal communication, 2020). Moreover, the feedback given by the users can be used to improve the quality of the data so that other users will not face the same issue (Product developer, personal communication, 2020). Cases where the cruciality of feedback for improvement of quality is proven in several studies (Phillips et al., 2017; Russ et al., 2020; Samuel-Rosa et al., 2017).

Since Liander mainly provides their own collected data and not data from external parties, they do not face the same issue of sharing confidential data from clients as Port of Rotterdam and Schiphol. The interviewee mentioned that should this have been the case for Liander as well, they would have handled it the same way as Port of Rotterdam by adopting a data delivery agreement, an agreement that documents any attributes included (Product developer, personal communication, 2020).

5.4.1 Data security

Another issue mentioned by Liander for opening data, is the fear for terrorism and even robbery after publishing the location of the electricity cables in the Netherlands (Product developer, personal communication, 2020). This was already experienced by the train branch in 2007 when the connection between Nijmegen and Den Bosch was disturbed as a result of 'copper pipe thieves'. The interviewee quotes: "You ask yourself the question: "how bad can it be to publish the location of the electricity cables in the Netherlands? Well, nothing is further away from the truth" (Product developer, personal communication, 2020). Years of discussion were associated with getting the level of detail of the data right in terms of terrorism proof data (ibid.). Similar to Schiphol Airport, the fear of terrorism also affects Liander's data. The interviewee states that one bomb on the 'right' spot could lead to great damage to parts of the Netherlands and its inhabitants (ibib.). In order to get approval for such issues was discussed with the AIVD which is the Dutch General Intelligence and Security Service. Dutch Intelligence deals with the national security and falls under the responsibility of the Ministry of internal affairs and Kingdom relations. After years of discussion, the AIVD decided which of Liander's data could be published and which data could not be published in terms of security reasons. For national security reasons the AIVD imposed modification of the level of detail that was publicly available. This level of detail was achieved through aggregation of the data. This came down to modifying the data in such a way that the level of detail did not harm the guidelines of the AIVD (Product developer, personal communication, 2020). The AIVD decided that the electricity cable network could be published, whereas the location of the gas pipes was considered too sensitive to publish in terms of explosion risk (ibid.). Although it took years for Liander to satisfy the requirements of AIVD, the willingness to provide open data never gave way to fear for legal or terrorism issues according to the interviewee.

5.4.2 Privacy

Another issue associated with the level of detail is privacy. The data that Liander publishes as open data cannot concern personal data since the GDPR does not allow personal data to be published as such. Since the data collected by Liander contains personal data of Liander's clients (households) they would need a signature of each individual to release that data as open data when aggregation was not allowed (Product developer, personal communication, 2020). Thus, to satisfy the requirements of the GDPR, the level of detail needed to be aggregated to a level that prevented disclosure of personal data. Overall, the aggregated level of detail in the published data resulted in both protection and privacy warranty, enabling Liander to achieve open data (Product developer, personal communication, 2020). So in this case, compliance with the GDPR resulted in an opportunity for Liander to provide data instead of a barrier.

5.4.3 Unfair competition

In the Netherlands, the ACM (consumer association authority) ensures a fair balance between companies and protects consumer interest (Autoriteit Consument & Markt, 2021). Initially, the ACM considered the E-Atlas of Liander as a distortion of competition: unfair competition (see also Ginsburg et al., 2019). The fear was that other companies would be disadvantaged in their business if Liander put a similar business to the market, financed by public funds. In practice, this was not the case as other businesses were not allowed to access this source data on electricity and gas usage due to privacy rules. Due to market barriers, not related to data, other companies could not start a similar business. This is a legal barrier which was addressed and challenged by discussing the issue and proving the fact that unfair competition could not result from data issues (Product developer, personal communication, 2020).

5.5 TECHNICAL ISSUES

Technical barriers were not an issue for Liander when setting up an open data portal according to the interviewee at Liander. “Setting up the open data portal is done by internal employees so no extra, external costs are made” (Product developer, personal communication, 2020). Moreover, the data portal was developed by the internal employees to reduce the time spent on previous data requests. Therefore, the internal time spent on the development is an investment to would win time in the future. The internal expenses were estimated at 0,5 Full- Time Equivalent of an employee (FTE). This is equal to 0,5*40 (hours of a full workweek) = 20 hours. Ignoring opportunity costs and since these costs were made already, no additional financial issues were faced by Liander.

5.6 ACTION PLAN: UNIFORM DATA PUBLICATION

Figure 21 was provided by the interviewee of Liander and shows the action plan for realising open data. It shows all the barriers that Liander faced on this journey. It starts with an incoming request from a user, mentioned in section 5.2, which is input in Liander’s backlog. The request can seek the data already held by Liander, presented in the cloud in the bottom of the scheme (operational data, asset data, client data and measurement data). The framework then determines whether this data can be published. The security team is the first team that get a decision in this, based on the safety of the data in terms of terrorism. This assessment of the data is based on the WBNI (*Wet Beveiliging Netwerk- en Informatiesystemen)*. Next, advice is sought from a legal advisor (JZ team) not only whether the data complies with the gas and electricity law but also whether it complies with the law – enforced in 2018 – that forbids gas connections to new building construction ‘*WET VET’* (*De Wet Voortgang Energietransitie* = The Energy Transition Progress Act). Then, the ‘BEPS’ in the left corner of the diagram give advice on what can be shared in terms of the GDPR. After assessing the legal aspects of the data, the technical aspects are discussed in terms of quality and the standards suitable for the data (middle upper cloud). Finally, the data can be modified to fit the advice given by legal framework. This is done by setting the level of aggregation: does the data needs to be more anonymised to pass the legal advice? This last step is the key action in realising open data. The interviewee quotes (Product developer, personal communication, 2020):

“Almost all data can be published as long as it can be aggregated to a suitable level which satisfies the legal advice”

After the data is aggregated to a suitable level, a last overall advice is given by the NBNL, the trade association for electricity and gas network operators in the Netherlands. Then, the open data proposal is put to the other grid operators in the Netherlands who can decide to follow this advice and join in with their data. This decision can be different for every grid operator; one may join in, whereas the other could reject the proposal. This, for example, is the case for information on cables and pipes which is made available by every grid operator expect grid operator Rendo.

Publishing the location of the cables and pipes network was achieved through the action plan presented in Figure 21. It took Liander several years to get an internal approval to publish this data as open data since the data owner (Asset Management) did not comply with the societal benefits. “They did not see the benefits of publishing the location of cables and pipes network, only the difficulties” (Product developer, personal communication, 2020). Asset Management emphasised the difficulties in terms of terrorism and liability risks when publishing the data. Based on the retrieved advice from the framework of Figure 21, it was proven that it was possible to publish the data without harming the risk for terrorism or liability issues (ibid.).

The scheme in Figure 21 is made in order to create a greater picture of the actions needed for realising open data and to improve them. It is showed to other grid operators to make them aware of the process as well. Liander hopes that in the future more grid operators will see the benefits of open data and collaborate in starting up a shared open data portal.

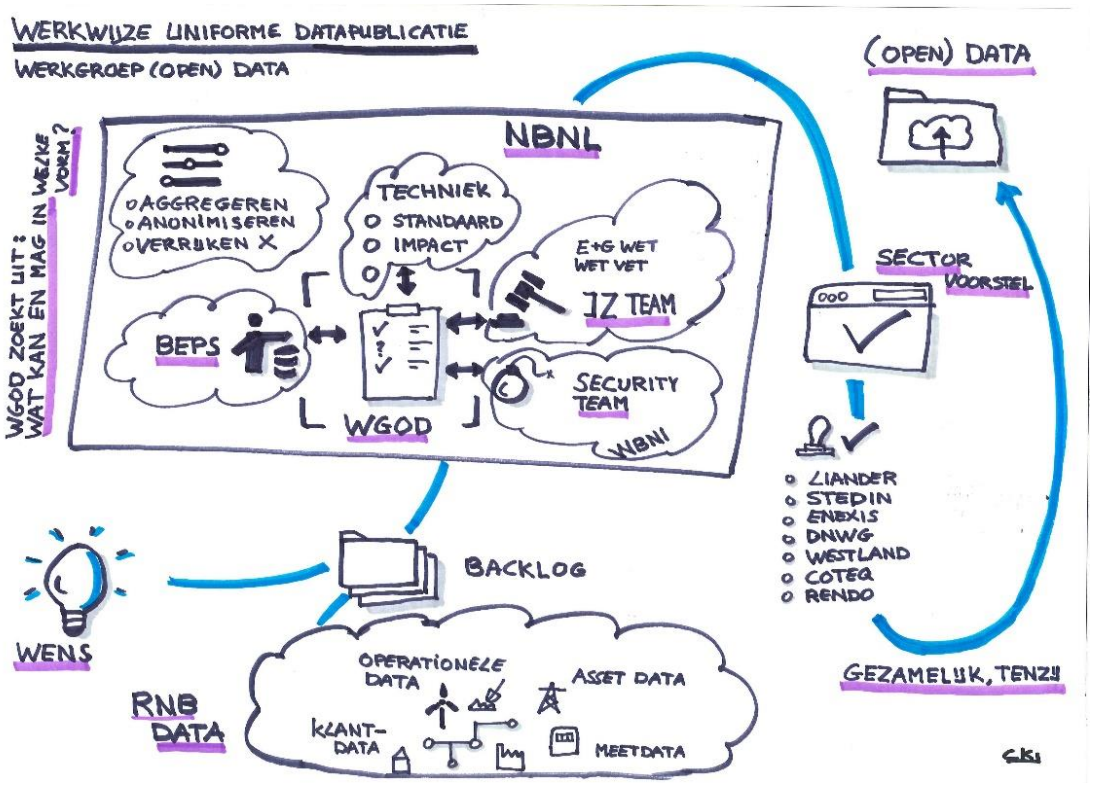


Figure 21. The action plan of uniform data publication of Liander. Presenting the different steps towards open data. The lower part (RNB data, *wens* + backlog) represent the start; the middle framework gives the legal, technical and quality advice which need to be satisfied by the data; and the right part (*Sector voorstel, gezamelijk, tenzij + (open) data)* propose the open data initiative to other gid operators, who can decide to join in or not. Source: (Product developer, personal communication, 2020).

5.7 FUTURE

In the future, Liander wants to start an open data collaboration with all the grid operators in the Netherlands by setting up one open data portal that is accessible for everyone and applies a fixed standard (Product developer, personal communication, 2020). However, there are some difficulties connected since every grid operator has a different set of rules regarding open data. For example, not every grid operator would aggregate on the level of 5 houses; some might prefer 10. This is a legal barrier which will be faced when all the grid operators are joining in. For now, Liander is taking the lead in this initiative.

5.8 SUB-CONCLUSION

The case of Liander proves that open data is not readily achieved and does not come with instructions how to attain it. However, their consistent determination to provide open data was key to achieving it. Liander faced mainly legal barriers associated with the level of detail of the data they could provide. The initial level of detail of the data interfered with both the guidelines of the AIVD and the GDPR. Aggregation of the data was key for the organisation to ensure open data without breaking the legal protection and privacy guidelines. Another legal objection was received from the consumer association

authority which accused Liander of stimulating unfair competition. Liander needed to prove to the authority that the data they provided was source data that was not available from other companies and organisations. In their action plan towards open data they dealt with legal, technical and quality issues that were challenged with data aggregation and legal discussions. By opening up Liander's data, the company experienced benefits in time and money saved on individual data request. Providing open data contributed to the energy transition because Liander's data informed on possible manners to lower energy consumption. Liander considered providing of open data a social benefit and which was a key motivating point. The next chapter assesses whether Port of Rotterdam and Schiphol could apply this working method as well in order to achieve open data.

6 OPPORTUNITIES TO OVERCOME BARRIERS

Having understood the barriers of Port of Rotterdam (PoR) and Schiphol Airport towards an open data policy, the next step is to discuss which methods can be used to overcome these barriers. This is based on the methods used by Liander to achieve an open data policy. The information for this last question is derived from an open data gathering between the three companies, organised by myself on February the 3th, 2021. During this open data gathering barriers from PoR and Schiphol Airport were presented and ideas and advice were shared. According to Liander, there is one overarching driver for overcoming barriers: internal motivation to achieve open data. If there is no internal motivation to achieve open data, it is difficult to find the willingness to overcome the barriers towards it (Product developer, personal communication, 2021). Table 7 places the barriers from both PoR and Schiphol Airport next to each other. It shows the similarity between the two companies regarding the type of barrier.

Table 7. Barriers faced by both PoR and Schiphol Airport, categorised on the type of barrier.

PoR		Schiphol Airport	
Barrier	**Reason**	**Barrier**	**Reason**
Institutional	'Us and them' feeling.	Institutional	False conclusion can be drawn from open data. Hence, data can be intentionally misused.
Legal (confidential)	Confidential can not be shared with everyone.	Legal (confidential)	Confidential data cannot be shared with everyone.
		Legal (privacy)	Personal data cannot be shared with everyone of it does not comply with de GDPR.
		Legal (security)	Sensitive data that can harm security cannot be shared.
Quality	Data needs to be fit for third parties' purposes.	Quality	Data needs to be fit for third parties and public user.
Technical	Technical development and skills are required to set up an open data portal.		
Financial	To develop and maintain open data extra costs are needed Adaptation costs, infrastructural costs and structural maintenance / operational costs). These cost are extra as they not paid now.		

6.1 INSTITUTIONAL

For both PoR and Schiphol it is difficult to find the internal motivation to provide open data, let alone getting an internal agreement. Within the companies, there are different interests which do not align to achieve open data. The internal lack of motivation to provide open data was also present for Liander in the beginning. According to Liander, a higher commitment from an overarching corporate body is needed to achieve an open data policy. The value and benefits of open data need to be clear for the higher overarching corporate body in order to get the institutional support. Appreciation of data sharing

and open data is growing within the PoR and Schiphol. For PoR through the development of the 12 data domains, for Schiphol through providing data in a FAIR manner to passengers. Yet, providing open data is not the main goal of the data policies of the companies. In order to create more internal appreciation of open data, experience of the benefits is needed. For Liander, the internal appreciation grew when it was discovered that the majority of the data requests could be covered with open data. Therefore, saving time and money for Liander. For PoR, developing the 12 data domains could show the internal body how data sharing between the departments could increase the internal performance. This could be underpinned by the argument that creating clarity in the different data, used by the departments, can result in more efficient data input and placement for projects. For Schiphol, sharing data in a FAIR manner to their client could show the internal body how it contributes to the aviation responsibility towards the passenger. So, for both PoR and Schiphol, it is possible to increase the institutional support of an internal corporate body.

6.2 LEGAL

Both companies deal with legal barriers, which were also main issues for Liander when opening data. To overcome legal barriers, the main advice from Liander is to use the available legal advice teams within the company. After receiving a legal advice for a particular dataset, aggregation can be used to satisfy the legal advice. According to Liander, private and confidential information can be published quiet flexible in terms of aggregation. It ensures a lower level of personal and confidential data. “From own experience, data sharing is actually always allowed when it is anonymised enough, in terms of personal data” (Product developer, personal communication, 2021). Both PoR and Schiphol own legal departments which can be utilised for advise on legal issues related to the data. These departments can give advice on the legal requirements for a dataset, after which aggregation can satisfy these requirements. Regarding own collected confidential data, it could be advised to remove attributes in terms of confidential and secured data. For PoR, attributes which classify the dataset to be confidential can be removed from the dataset. As a result, the dataset can be published without publishing confidential data. For Schiphol, decreasing the number of attributes which are considered confidential is done already, e.g. the floorplan, and can be done to more datasets. Regarding data retrieved from third parties, attributes that are considered confidential are documented in the data delivery agreement. This also applies for PoR. So for both companies the confidential attributes of the data is already documented. However, it needs to be considered whether the provided data still fits the purpose of the users after aggregation is applied since it comes at the expense of the data quality. Regarding personal and sensitive data in terms of security, aggregation is already used by Schiphol. Aggregation of the level of detail is mainly used by Schiphol to comply with the responsibility to be a transfer and transport hub. Though, the aggregated level of detail cannot conflict with the commercial responsibility of Schiphol. This is a recurring conflict that needs to be reconsidered by Schiphol when aggregating a new dataset. Thus, for both companies, utilising the legal advice teams within the companies combined with aggregation to satisfy this legal advice could aid the legal barriers. However, to get an approval from the legal advice teams could take years, likewise for Liander. According to Liander, internal motivation is required to get through this legal process.

6.3 QUALITY

Both PoR and Schiphol deal with clients with diffuse requests and purposes related to the data quality. This makes it difficult to map the data quality requirements of a dataset. Liander encountered scattered data requests from clients as well. In order to deal with this, Liander set up a ‘request window’ where external parties could requests the data they desired. As a result, the type of data requests were monitored

and the client's data requirements categorised. It was experienced that the majority of these requests could be provided through open data. Thus, by collecting and monitoring the different data requests, it was possible for Liander to decrease the difference and increase the similarities of the data requirements. Subsequently, to satisfy the majority of the requests with open data.

This method could also be applied by PoR and Schiphol. PoR already tries to get more insights in the type of data used by the internal user through the development of the 12 data domains. Categorising the type of data requests from the internal client could therefore be included as well. Mapping the majority of the data requests contributes to the initiative of the 12 data domains as it gives information on the data purposes of the internal user. For Schiphol, monitoring the different data requests can also be included in their current data policy. Currently, a registration and data purpose of the public user is requested from Schiphol in order for the user to access the data. Since the data request template shows these insights already, monitoring an categorising the type of data requests can be performed as well. The data requirements of third parties and clients, for both PoR and Schiphol, can also be mapped through contracts and data delivery agreements. Both contracts and data delivery agreements state which data the other party uses for which purpose. When requirements of the data quality are mapped, aggregation and anonymisation could be used again to satisfy these requirements.

6.4 TECHNICAL

Overcoming technical barriers requires insights in the technical skills within the company. An advice from Liander is to use the technical skills and resources which are already owned by the company to develop and maintain an open data portal. Liander did the same by deploying 0,5 fulltime-equivalent of an employee (FTE). This is 20 hours of a full work week. Although it came at the expense of available time for other work, Liander believed it was a good investment as it offered the possibility to save time and money in the future on other work. Providing open data through an open data portal saved Liander time and money on previous individual requests. Thus, it is advised by Liander to limit the technical barriers and the associated costs by using internal employment. Currently, Schiphol is sharing data through their data portal, which is developed and maintained by the internal IT department, the API support team, containing amongst others data analysts. For PoR, the technical resources are present as well. According to the Analytics specialist, developing an open data portal is mainly possible through the IT department, Port Objective Management (POM) and Team PortMaps. POM falls under Asset Management and is of great importance as they are the encyclopaedia behind the data. Hence, POM ensures that the meta-data is up-to-date and the data can be accessed and used. Besides POM and the IT department, Team PortMaps develops and maintains the data that is currently used for internal performance. When publishing the data on an open data portal, their input is needed as well. So, the possibility to tackle the technical issues is present for Port of Rotterdam. However, to put this in motion and to be able to use 0,5 FTE of the employees within these departments, allowance from the higher organisational body is needed. Again, this derives from the internal motivation to provide open data.

6.5 FINANCIAL

The internal drive to provide open data must contain a financial component as well. Equally to other business initiatives, it needs to be clear for the company what the financial returns will be in order to fund the initiatives to begin with (Product developer Liander, personal communication, February 2021). Setting up a new open data policy initially costs money and time. Hence, internal and societal benefits are not certain yet. For Liander, it took years before they could practice an open data policy, let alone the experience financial benefits through the time and money saved on individual data requests. The

invested time and money did not pay off from the beginning. Therefore, the only advice Liander gives in terms of financial barriers is to orientate on the foreseen financial benefits open data has to offer. For Port of Rotterdam, the financial benefits towards opening data to everyone is still one step too far. First, the financial benefits from internal data sharing needs to be clear in order to get that in motion. So: How can internal performance increase as a result of data sharing between the departments? Hence: how much time (thus money) can be saved when data is shared between the departments? After this is clear, it will be easier to foresee the financial benefits of sharing the data outside the company (Product developer, personal communication, 2021).

6.6 SUB-CONCLUSION

Thus, different types of barriers ask for different types of methods. Table 8 presents the advice given by Liander for the barriers faced by Port of Rotterdam and Schiphol in achieving an open data policy.

Table 8. The advice given per barrier by Liander for both Port of Rotterdam and Schiphol Airport.

PoR		Schiphol Airport	
Barrier	**Reason**	**Barrier**	**Reason**
Institutional	'Us and them' feeling.	Institutional	False conclusion can be drawn from open data.
Advice from Liander: Get appreciation from higher corporate body for open data, through indicating the internal benefits.			
Legal (confidential)	Confidential can not be shared with everyone.	Legal (confidential)	Confidential cannot be shared with everyone.
		Legal (privacy)	Personal data cannot be shared with everyone.
		Legal (security)	Sensitive data that can harm security cannot be shared.
Advice from Liander: Make use of the legal advice teams within the company to get an approval to open data. Although getting a legal approval may take years, it is possible through internal motivation.			
Quality	Data needs to be fit for third parties.	Quality	Data needs to be fit for third parties and public user.
Advice from Liander: Collect and monitor data requests from key data users to map the data quality requirements of the users.			
Technical	Technical development of open data portal is required.		
Advice from Liander: Use technical skills and departments which are already available within the company.			
Financial	New financial funding flow is needed.		
Advice from Liander: Orientate on foreseen financial, internal benefits and convince the internal management body through internal motivation.			

7 DISCUSSION

The Liander experience indicates that aggregation is an important tool to achieve open data that complies with legal requirements in terms of privacy, security and confidentiality of data. A discussion point could be whether open data can still be achieved without the option of aggregation. For PoR this could apply to the less sensitive datasets of information classified as 'public' instead of 'internal' or 'confidential'. This could, for instance, be the case for the datasets on traffic signs as often used by PoR for internal performance. Within this dataset, all attributes such as model number, location, and year of placement are classified as public; no aggregation is needed to provide this data as open data. Other datasets which seemingly do not hold confidential data, such as road networks, may prove otherwise. The dataset on the road network holds several attributes which are classified as public such as the road type, function, length, and hardening layer but this dataset also holds confidential information, such as inspection results and level of ambition (the desired maintenance level of the asset). For Schiphol, the same could be considered. Is it possible to open up datasets which presumably do not hold sensitive data? Of all the datasets on which Schiphol is currently working it may be possible for the dataset on flying birds. The approach could be that this information has no potential provoke terrorism or breach the requirements of the GDPR. This approach could be applied to more datasets than the flying birds. Moreover, without the possibility of aggregation it might be possible for both companies to provide open data through the CC-BY licence, as done by Liander. Since this licence only allows re-users to distribute adapt, and build upon the data if attribution is given to the creator, it could remain a sense of control for the companies. This way, it becomes clear for what purposes the data is used and by whom.

The lessons from Liander show that at the heart of open data sits an open data mindset; a fundamental belief in the concept that openness of data is desirable and a service to the common good. PoR and Schiphol, as mostly publicly owned organisations, could reasonable be expected to adopt a more socially responsible approach to their data. Liander's experience showed a more holistic approach to data by considering the cost of responding to questions in combination with the cost of data openness. They showed that the added expense was limited and outweighed by commercial as well as social benefits.

Liander also showed that careful management of accessibility of the data, for instance by aggregating, mitigates the risk to reputation and could be off-set by the benefit of being regarded as a transparent, accountable and socially responsible organisation. As a grid operator, Liander is a ready target for potential criticism, for instance on climate impact. PoR and Schiphol, being less open with their data but still significant potential targets for criticism, could benefit a more pro-active approach to open data. A pro-active approach to open data could anticipate such accusations and potential reputational damage. Hence, they could point at the readily available data; this is often enough to deflect more detailed investigation. For example, for Schiphol this could be done by providing correct data on road networks to avoid false conclusions drawn by the users, subsequently the media. The cost effort in responding to media or legal challenges, both in direct financial terms but also in reputational terms, should be included in the equation when considering data openness.

Finally, as described early on in the paper, the pressure to open up data in public and semi-public organisations will continue to accumulate. Both PoR and Schiphol would recognise the unavoidability of moving toward open data. Early recognition of the inevitability investment requirement would still give them an opportunity to plan, schedule, implement and finance their open data programme at their own tempo. Once regulation overtakes their effort, the tempo will be set from outside and may be less optimal.

8 CONCLUSION

In this research the following research question was addressed: How can public undertakings in the Netherlands, similar to Liander, overcome the barriers to opening their geographical datasets in order to be prepared for expected future legislation towards open data?

It can be stated that public undertakings, such as Port of Rotterdam (PoR) and Schiphol Airport, can overcome barriers towards open data to be prepared for the expected future legislation of the Open Data Directive. However, to do so changes need to be enforced to overcome the barriers. The multi-dimensional model in this research identified three different levels of open data for a public undertaking to reference it's data policy: not open, partly open and open. In this model the requirements of open data are interpreted from the data provider's perspective in order to make the data more open for the end-user. At the first level data is considered not to be open at all and only accessible for the internal user, using the data for internal performance; such data cannot be found through a general search engine. At the second level data openness is improved as it is findable and accessible through a general search engine or data portal, for the external data user as well as to the internal data user. Data is used for generating internal and external performance. At the third level data can be considered most open. The data is findable for the internal, external and public user, through a general search engine and data portal, free of charge and with an open licence for everyone to re-use the data. In this case internal, external and public value is generated from the data.

At present, PoR can be placed in level 1 of the multi-dimensional model as the collected data is used by the internal user, generating internal performance. Although data is shared with some internal users, it is not yet shared with all internal users. Some data is shared with third parties when this is of interest for the internal performance of PoR. Schiphol can be placed in level 2 of the multi-dimensional model since data is shared with internal, external and public users. Sharing data with the public user is, as yet, only executed when there is no interference with the Schiphol's commercial data responsibility. The main goal of both companies is to generate internal performance with their collected data. In between the levels, barriers are identified which are faced when a higher level is pursued. The identified barriers are financial, institutional, task complexity, legal, technical and quality. To achieve level 2 of open data for PoR, financial, technical, quality, institutional and legal barriers are faced. To achieve level 3 of open data for Schiphol, legal security barrier, legal privacy barriers, legal confidential barrier and institutional barriers are faced. For Liander, who provides open data since 2014, similar barriers were encountered and defeated on their path to open data. According to Liander, achieving open data starts with the institutional motivation to do so. A commitment to open data must stem from high in the corporate body to gain sufficient traction. To overcome the numerous legal barriers, Liander asked the internal legal teams to give an advice on the datasets of which open data was preferred. The legal advice covered the confidential, security and privacy information of the data. Most of the advice given could be mitigated through aggregation of the data. This is possible for both PoR and Schiphol as well. Both companies own legal departments which can give legal advice regarding the datasets. Decreasing the level of detail through aggregation could satisfy the legal advice and allow publishing the data. Despite the level of detail, the data still needs to satisfy the quality requirements of the end user. Through monitoring and classifying the incoming data requests, Liander tackled the diffuse data quality requirements of the end user. It proved that the majority of the data requests could be satisfied by opening up available data. In terms of the quality barrier, PoR and Schiphol both could apply the same method since both companies keep track of the data requests already. By categorising the different data requirements from the user it is possible to pin down the required quality to fit the purpose of the user. Regarding the technical barrier faced by PoR, the model offered by Liander is to utilise the technical departments already present in the

company to develop and maintain the open data portal. Within PoR, the IT department could facilitate the technical component of developing an open data portal; Port Objective Management (POM) and Team PortMaps could focus on developing and publishing the data itself. Again, to put this in motion and to be able to use the employees within these departments, anticipation from higher up in the organisational body is needed. Lastly, to overcome the financial barrier for PoR, it is Liander's experience to highlight foreseen financial benefits of open data for the company and so convince the internal managing body of the foreseen financial benefits.

Neither PoR nor Schiphol are ready to comply with the future rules when the Open Data Directive requirements become mandatory instead of the current voluntary requirements for public undertakings. Barriers still need to be overcome, but Liander has shown how this can be achieved. Starting with the internal motivation to provide open data and using aggregation to satisfy legal requirements. It is discussed whether aggregation is necessary to achieve open data; datasets which do not hold sensitive data could be published without aggregation. This is input for further research.

9 RECOMMENDATIONS

For future research it is recommended to take this research as a motive and reset the scope to the outcomes of this research. An interesting feature presented in the results was the use of aggregation by Liander. Aggregation was considered to be the key method to use for the achievement of open data in terms of legal requirements concerning security privacy and confidentiality. One proposed action would be to focus on the level of aggregation, suitable for the current data policy of both PoR and Schiphol. The question to consider would be: to what extent can the level of the datasets be aggregated and still contribute to the internal performance of the companies? This question interprets the level of detail from the data provider. However, the same question could be asked from the perspective of the users: how valuable is aggregated data for users? For this last question, a reference could be made to the case of Liander where external users, such as housing corporations, successfully used data from the E-Atlas. Thus, in the case of Liander, it is clear already that external users value the aggregated data, provided by Liander.

Another recommendation derives from the action plan used by Liander to achieve open data. The different legal and technical steps taken in this action plan could also be taken by PoR and Schiphol. Liander's action plan helped the company to map the different steps and actions needed to achieve open data; it is recommended to set up a similar action plan for PoR and Schiphol. Future research could develop a similar and suitable action plan for PoR and Schiphol that gives insights in the detailed actions needed to achieve open data for these public undertakings.

10 REFERENCES

Afuah, A., & Tucci, C. L. (2001). *Internet Business Models and Strategies: Text and Cases SEE PROFILE École Polytechnique Fédérale d…*. Retrieved from https://www.researchgate.net/publication/215915163

Asser/Hartkamp, & Sieburgh. (2008). Asser 6-I De verbintenis in het algemeen, eerste gedeelte, 358 | Navigator. Retrieved November 27, 2020, from https://www.navigator.nl/document/inodefde19b84da4b1a04d98245c166992c2/asser-6-i-de-verbintenis-in-het-algemeen-eerste-gedeelte-358

Autoriteit Consument & Markt. (2021). Autoriteit Consument & Markt | ACM.nl. Retrieved January 11, 2021, from https://www.acm.nl/nl

Barbero, M., Bartz, K., Linz, F., Mauritz, S., Wauters, P., Chrzanowski, P., … Osimo, D. (2018). *Study to support the review of Directive 2003/98/EC on the re-use of public sector information*. https://doi.org/10.2759/373622

Barry, E., & Bannister, F. (2014). Barriers to open data release: A view from the top. *Information Polity*, *19*(1–2), 129–152. https://doi.org/10.3233/IP-140327

Beno, M. (2016). *Open Data Hopes and Fears, Determining the barriers of Open Data*.

Berners-Lee, T. (2009, June 18). Linked Data - Design Issues. Retrieved October 16, 2020, from https://www.w3.org/DesignIssues/LinkedData.html

Bregt, A. K., Castelein, W., & Dignum, W. (2012). Deelrapport organisatorische aspecten voor een succesvolle open data strategie Deelrapport organisatorische aspecten voor een succesvolle open data strategie. *Delft: Onderzoeksinstituut OTB: 14*.

Cambridge Academic Content Dictionary. (n.d.-a). PRIVACY | meaning in the Cambridge English Dictionary. Retrieved February 17, 2021, from Cambridge University Press website: https://dictionary.cambridge.org/dictionary/english/privacy

Cambridge Academic Content Dictionary. (n.d.-b). SECURITY | meaning in the Cambridge English Dictionary. Retrieved February 17, 2021, from Cambridge University Press website: https://dictionary.cambridge.org/dictionary/english/security

Caplyn, R., Davies, T., Wadud, A., Verhulst, S., Alonso, J. M., & Farhan, H. (2014). Towards common methods for assessing open data: workshop report & draft framework. In *World Wide Web Foundation* . Retrieved from http://opendataresearch.org/sites/default/files/posts/Common Assessment Workshop Report.pdf

CNBC. (2020). *PG&E pleads guilty to 84 deaths in 2018 California wildfire*. Retrieved from https://www.cnbc.com/2020/06/16/pge-to-plead-guilty-to-deaths-from-california-wildfire.html

Conradie, P., & Choenni, S. (2014). On the barriers for local government releasing open data. *Government Information Quarterly*, *31*(SUPPL.1). https://doi.org/10.1016/j.giq.2014.01.003

Corsar, D., & Edwards, P. (2017). Challenges of Open Data Quality: More Than Just License, Format, and Customer Support. *J. Data and Information Quality*, *9*(3). https://doi.org/10.1145/3110291

Creative Commons. (2015, March 25). A Masterwork in Simplicity: The Story of the CC Logo - Creative Commons. Retrieved February 24, 2021, from https://creativecommons.org/2015/03/25/a-masterwork-in-simplicity-the-story-of-the-cc-logo/

Creative Commons. (2019). About CC Licenses - Creative Commons. Retrieved November 11, 2020, from https://creativecommons.org/about/cclicenses/

Dalla Corte, L. (2020). *Safeguarding Data Protection in an Open Data World : On the idea of balancing open data and data protection in the development door*.

Dalla Corte, L., & van Loenen, B. (n.d.). *The Recast of the PSI Directive*. 1–29.

Data.overheid.nl. (2020). *Gebruikersbijeenkomst Data Delen op 5 november 2020 | Data overheid*. Retrieved from https://data.overheid.nl/actueel/bijeenkomsten/gebruikersbijeenkomst-data-delen-op-5-november-2020

Data.world. (2019). Common license types for datasets – data.world help center. Retrieved November 23, 2020, from https://help.data.world/hc/en-us/articles/115006114287-Common-license-types-for-datasets

David Eaves. (2009, November 29). Three Laws of Open Data (International Edition) | eaves.ca. Retrieved October 16, 2020, from https://eaves.ca/2009/11/29/three-laws-of-open-data-international-edition/

Degbelo, A. (2020). Open Data User Needs: A Preliminary Synthesis. *The Web Conference 2020 - Companion of the World Wide Web Conference, WWW 2020*, 834–839. https://doi.org/10.1145/3366424.3386586

Deloitte Analytics. (2012). *Open growth Stimulating demand for open data in the UK*.

Directive (EU). (2016). *REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL*. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02016R0679-20160504&from=NL#tocId2

DIRECTIVE (EU) 2008/48. (2008). *DIRECTIVE 2008/48/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 23 April 2008 on credit agreements for consumers and repealing Council Directive 87/102/EEC*.

DIRECTIVE (EU) 2013/37. (2013). RICHTLIJN 2013/37/EU VAN HET EUROPEES PARLEMENT EN DE RAAD van 26 juni 2013 tot wijziging van Richtlijn 2003/98/EG inzake het hergebruik van overheidsinformatie (Voor de EER relevante tekst). *Publicatieblad van de Europese Unie*, *2013*(3), 1–8.

DIRECTIVE (EU) 2019/1024. (2019, June 20). DIRECTIVE (EU) 2019/1024 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL, on open data and the re-use of public sector information. Retrieved September 25, 2020, from https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32019L1024&from=EN

European Commission. (2020a, January 30). From the Public Sector Information (PSI) Directive to the open data Directive | Shaping Europe's digital future. Retrieved September 16, 2020, from https://ec.europa.eu/digital-single-market/en/public-sector-information-psi-directive-open-data-directive

European Commission. (2020b, March 20). A European Strategy for Data | Shaping Europe's digital future. Retrieved July 13, 2020, from https://ec.europa.eu/digital-single-market/en/policies/building-european-data-economy

European Data Portal. (2018). *Protecting data and opening data | Europees Data Portaal*. Retrieved from https://www.europeandataportal.eu/nl/highlights/protecting-data-and-opening-data

European Parliament. (1997). Public Undertakings/Services in the EU: Summary Part 1 (DG4 Study

ECON W21) - Public Undertakings and Public Services in the European Union: Summary - Part 1. Retrieved September 21, 2020, from https://www.europarl.europa.eu/workingpapers/econ/w21/sum-1_en.htm

Geiger, C. P., & Von Lucke, J. (2012). *Open Government and (Linked) (Open) (Government) (Data) 1*. *4*(2), 265–278.

Ginsburg, J. C., Litman, J., & Kevlin, M. (2019). *Trademark and Unfair Competition Law Cases and Materials, Sixth Edition 2019 Supplement*. Retrieved from www.cap-press.com

GISGeography. (2020). 10 Free GIS Data Sources: Best Global Raster and Vector Datasets [2020] - GIS Geography. Retrieved October 28, 2020, from https://gisgeography.com/best-free-gis-data-sources-raster-vector/

Government, & Taskforce. (2009). *Engage Getting on with Government 2.0 Report of the Government 2.0 Taskforce government 2.0 taskforce*. Retrieved from http://www.catchmedia.com.au/.

ISO - ISO 19157:2013. (2013). ISO - ISO 19157:2013 - Geographic information — Data quality. Retrieved November 20, 2020, from https://www.iso.org/standard/32575.html

ISO - ISO 9001:2015. (n.d.). ISO - ISO 9001:2015 - Quality management systems — Requirements. Retrieved November 20, 2020, from https://www.iso.org/standard/62085.html

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. https://doi.org/10.1080/10580530.2012.716740

Janssen, M., & Kuk, G. (2007). E-Government business Models for Public service networks. Retrieved November 25, 2020, from IGI Publishing website: https://pdfs.semanticscholar.org/876c/90c8c698373621c0b296d0d908b26d27eb42.pdf

Juffermans, P. (2015, May 20). Open data Liander. Retrieved December 18, 2020, from Presentations & Public Speaking website: https://www.slideshare.net/lisettevanbeusekom/open-data-liander-paul-juffermans

Kasler, D. (2018, March 22). PG&E says it will begin shutting down power lines when fire risk is extreme | The Sacramento Bee. *The Sacromentoo Bee* . Retrieved from https://www.sacbee.com/article206369044.html

Klep, C., & Productontwikkelaar Liander, sr. (2018). *Direct inzicht in verbruiksdata huurwoningen*.

KPIX 5 CBS San Francisco. (2018). *California Wildfires Blamed On Utility Power Lines, Equipment Failures – CBS San Francisco*. Retrieved from https://sanfrancisco.cbslocal.com/2018/06/08/12-northern-california-wildfires-caused-by-pge-power-lines-equipment-failures/

Kumar Tayi, G., & Ballou, D. P. (1998). Examining Data Quality . Retrieved November 20, 2020, from https://dl.acm.org/doi/pdf/10.1145/269012.269021?casa_token=1z1HGJBsrUwAAAAA:qXX0yv22vFqXRstnCE0ZUMXVoxIQmb39__1sP-xaFilNaPQCwTJ3Z7jbhHAAVtbr9RBXhOlDOxSWOg

Liander. (n.d.). Liander | Partners | Open Data. Retrieved December 21, 2020, from https://www.liander.nl/partners/datadiensten/open-data

Martin, S., Foulonneau, M., Turki, S., Ihadjadene, M., Paris, U., & Tudor, P. (2013). Risk analysis to overcome barriers to open data. *Electronic Journal of E-Government*, *11*(1), 348–359.

Miller, P., Styles, R., & Heath, T. (2008). *O P E N D ATA C O M M O N S , A L I C E N S E F O R O P E N D ATA.*

Netherlands Enterprise Agency RVO. (n.d.). How to make your business GDPR compliant. Retrieved February 25, 2021, from https://business.gov.nl/running-your-business/business-management/administration/how-to-make-your-business-gdpr-compliant/

OECD. (2017). Open Government Data. Retrieved October 21, 2020, from https://www.oecd.org/gov/digital-government/open-government-data.htm

OECD Open and Innovative Government Division. (2019). *OECD Open. Useful and Re-usable data (OURdata) Index: 2019*. https://doi.org/10.1787/9789264305847-en

Open Data Institute. (2013). Making aid more effective in Nepal. Retrieved November 13, 2020, from Open Data handbook website: http://opendatahandbook.org/value-stories/en/effective-aid-in-nepal/

Open Data Reader. (2016). Open Data Reader 2016. Retrieved November 18, 2020, from https://www.kl.nl/wp-content/uploads/2016/04/Open-Data-Reader-2016-1.pdf

Open Knowledge Foundation. (n.d.-a). Home — Open Data Commons: legal tools for open data. Retrieved November 23, 2020, from https://opendatacommons.org/

Open Knowledge Foundation. (n.d.-b). What is open? Retrieved October 19, 2020, from https://okfn.org/opendata/

Open Knowledge Foundation. (2011). Open data reduces mortality rate in UK hospitals. Retrieved November 13, 2020, from http://opendatahandbook.org/value-stories/en/uk-mortality/

Phillips, C., Kenny, A., & Esterman, A. (2017). Supporting graduate nurse transition to practice through a quality assurance feedback loop. *Nurse Education in Practice*, *27*, 121–127. https://doi.org/10.1016/j.nepr.2017.09.003

Port of Rotterdam. (2020a). Control & management | Port of Rotterdam. Retrieved December 30, 2020, from https://www.portofrotterdam.com/en/doing-business/port-of-the-future/digitisation/control-management

Port of Rotterdam. (2020b, October). Organisatiestructuur | Haven van Rotterdam. Retrieved January 4, 2021, from https://www.portofrotterdam.com/nl/havenbedrijf/over-het-havenbedrijf/organisatie/organisatiestructuur

Rijkswaterstaat. (2017, June 1). Elektronische vaarwegkaarten (ENC's) | Rijkswaterstaat. Retrieved January 8, 2021, from https://www.rijkswaterstaat.nl/zakelijk/zakendoen-met-rijkswaterstaat/werkwijzen/werkwijze-in-gww/data-eisen-rijkswaterstaatcontracten/elektronische-vaarwegkaarten.aspx

Russ, C. M., Stone, S., Treseler, J., Vincuilla, J., Partin, L., Jones, E., … Kelly, D. P. (2020). Quality Improvement Incorporating a Feedback Loop for Accurate Medication Reconciliation. *Pediatrics*, *146*(6).

Safarov, I., Meijer, A., & Grimmelikhuijsen, S. (2017). Utilization of open government data: A systematic literature review of types, conditions, effects and users. *Information Polity*, *22*(1), 1–24. https://doi.org/10.3233/IP-160012

Samuel-Rosa, A., Simão Diniz Dalmolin, R., Moura-Bueno, J. M., Teixeira, W. G., & Alba, M. F. (2017). Open legacy soil survey data in Brazil: geospatial data quality and how to improve it. *Sci. Agric. V*, *77*(1), 2020. https://doi.org/10.1590/1678-992X-2017-0430

Schweisfurth, T. G., & Herstatt, C. (2014). How internal users contribute to corporate product innovation: the case of embedded users. *R&D Management*, *46*(S1), 107–126. https://doi.org/10.1111/radm.12103

Support Centre or Data Sharing. (n.d.). What is data sharing? | Support Centre for Data Sharing. Retrieved November 7, 2020, from https://eudatasharing.eu/what-data-sharing

Tankard, C. (2012). Big data security. *Network Security*, *2012*(7), 5–8. https://doi.org/10.1016/S1353-4858(12)70063-6

USGS. (n.d.). EarthExplorer. Retrieved October 28, 2020, from https://earthexplorer.usgs.gov/

van Loenen, B. (2012). Open data and beyond: exploring existing open data projects to prepare a successful open data strategy. Deelrapport: overige gronden die zich tegen open data kunnen verzetten. *Delft: Onderzoeksinstituut OTB.*

Van Loenen, B., Welle Donker, F. ., Kulk, S., Groetelaers, D. A., De Jong, J., & Ploeger, H. D. (2011). *Databeleid Rijkswaterstaat: Deel: Hoofdwatersysteem; een overzicht van de juridische kaders omtrent het omgaan met data*. (november). Retrieved from http://repository.tudelft.nl/view/ir/uuid:540a04e6-d4ee-4d78-a526-7d6a8300009b/

van Loenen, Bastiaan. (2012). Open data and beyond: exploring existing open data projects to prepare a successful open data strategy. Deelrapport inleiding open data. *Delft: Onderzoeksinstituut OTB.*

Van Loenen, Bastiaan, & Bregt, A. K. (2012). Open data and beyond : Deelrapport : Instrumenten voor de monitoring van de impact van open data (bij Alliander) Deelrapport : Instrumenten voor de monitoring van de impact van open data (bij Alliander). *Delft: Onderzoeksinstituut OTB.*

Vancauwenberghe, G., Valečkaitė, K., & van Loenen, B. (2018). Assessing the Openness of Spatial Data Infrastructures (SDI): Towards a Map of Open SDI. *International Journal of Spatial Data Infrastructures Research*, *13*(706999), 88–100. https://doi.org/10.2902/ijsdir.v13i0.468

Verhulst, S. G., Young, A., Zahuranec, A. J., Aaronson, S. A., Calderon, A., & Gee, M. (2020). *The Emergence of a Third Wave of Open Data How To Accelerate the Re-Use of Data for Public Interest Purposes While Ensuring Data Rights and Community Flourishing*.

W3C Recommendation. (2008, January 15). SPARQL Query Language for RDF. Retrieved October 19, 2020, from https://www.w3.org/TR/rdf-sparql-query/

Welle Donker, F, Braggaar, R. C., & Loenen, B. Van. (2019). *Hergebruikers van open data in beeld*.

Welle Donker, Frederika. (2009). The Socioeconomic Effects of Public Sector Information on Digital Networks. In *The Socioeconomic Effects of Public Sector Information on Digital Networks*. https://doi.org/10.17226/12687

Welle Donker, Frederika. (2018). *Funding Open Data*. (April), 55–78. https://doi.org/10.1007/978-94-6265-261-3_4

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

Wijk, A. Van, Torre, E. Van Der, Barneveld, G. Van, & Wolsink, J. (2020). *Ondermijning op en rond luchthaven Schiphol*.

World Wide Web Foundation. (2017). The Open Data Barometer | Open Data Barometer. Retrieved October 21, 2020, from https://opendatabarometer.org/barometer/

Zuiderwijk, A., & Janssen, M. (2014). Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly*, *31*(1), 17–29. https://doi.org/10.1016/j.giq.2013.04.003

Appendix

Interview Questions

11 APPENDIX A INTERVIEW PORT OF ROTTERDAM

Aim of the interview

Identify the level of openness in which Port of Rotterdam is situated in achieving open data. Subsequently, identifying which barriers they face in achieving open data.

Introduction

1. What kind of geographical dataset does the company use?
    a. Can you come with an example of dataset?
    b. Are these datasets open to use?

Current situation

2. For who are these datasets available?
    a. Is this intern or extern?
    b. Who is identified as the user in your (open) data policy?
3. How is the data made available for intern users?
    a. Which dataset is made available for internal users?
    b. Through a portal?
    c. Is it available in a machine readable format?
    d. Is it possible to download the data in bulk?
4. (How) is the data made available for external users?
    a. Which datasets are made available for external users (up-to-date)?
    b. Is this through a portal?
    c. Is this against a fee?
    d. Is it available in a machine readable format?
    e. Is it possible to download the data in bulk?

Future situation

5. Is achieving (more) open data included in the vision of the company?
    a. Yes? What kind of barriers will there be for the company to overcome?
    b. No? Why not, what is the (main) reason(s) for not achieving open data?
    c. If there will be a legal obligation to provide the data as open data, what will be the consequences for your organisation?
        i. Can you give an example?
    d. What will need to be done to implement an open data policy?
        i. Can you give an example?

12 APPENDIX B INTERVIEW QUESTIONS SCHIPHOL

Aim of the interview

Identify the level of openness in which Schiphol is situated in achieving open data. Subsequently, identifying which barriers they face in achieving open data.

Introduction

1. Which geographical datasets does Schiphol have?
    a. Which are the most used ones?
2. From the open data portal of Schiphol it became clear that 11 datasets are openly available for every user. How many datasets does Schiphol has in total?
    a. Is this more when a someone with a Schiphol account logs in?
3. What was the underlying idea when making only these 11 datasets openly available?
    a. Was there a certain purpose/aim behind this decision?

Current situation

4. Who does Schiphol identify as the key user?
    a. Is that the internal user or the external user?
    b. Is the data published in a different way for internal and external user?
    c. Is that reflected in the downloadable format, costs or something else?
    d. Is this done with a purpose?
5. On which user does the current data policy of Schiphol focus the most?
    a. Is this the internal or the external user?

Future situation

6. Currently, there are 11 datasets openly available on the data portal of Schiphol. Imagine when all the datasets, owned by Schiphol, have to be open. Would this fit into the current data policy of Schiphol?
    a. Why yes or no? Can you give an example?
7. Which barriers would you face when Schiphol needs to make every dataset openly available?
    a. Is this associated with financial, legal, privacy or technical barriers? Or other barriers?
    b. Did Schiphol already face barriers when they made these 11 datasets openly available?
8. What adjustment is needed in order to make a suitable data policy which is focussed on open data?
    a. Is this an organisational, financial or technical barrier? Or another barrier? Or more barriers?

13 APPENDIX C INTERVIEW QUESTION LIANDER

Aim of the interview

Identify which barriers Liander faced toward to road to open data and which methods they used to overcome these barriers.

Introduction

1. Which geographical datasets does Liander have?
2. Which datasets are available as open data?
    a. What was the underlying idea to provide the datasets as open data which are currently provided?
3. Who can be identified as the key user of Liander's data?
    a. Who can be identified as the key user of Liander's open data?

Current situation

1. What was the main motivation for Liander to provide open data?
    a. When did this arise?
    b. From who did this initiative to provide open data arise?
    c. Was the drive to include more users in the current data policy?
    d. Was it desired that your data was also used for academic research?

Barriers

A small introduction of the barriers mentioned by Port of Rotterdam and Schiphol Airport was given (legal, technical, quality, institutional and financial).

1. After providing open data, did Liander experienced liability issues as a result of publishing poor quality data?
    a. If yes, how was dealt with this?
    b. If no, why did Liander not faced this barrier?
2. After provider open data, did Liander experienced liability issues as a result of publishing confidential data?
    a. If yes, how was dealt with this?
    b. If no, why did Liander not faced this barrier?
3. After provider open data, did Liander experienced liability issues as a result of publishing sensitive data (in terms of terrorism or personal data)?
    a. If yes, how was dealt with this?
    b. If no, why did Liander not faced this barrier?
4. Did Liander faced internal conflict as a result of providing open data?
    a. How was this dealt with?
5. How were the costs of open data funded?
    a. Were external costs made?
6. What was the most difficult barrier/issue to deal with in order to achieve open data?
7. What is the most successful method used in order to achieve open data?