MSc thesis in Geomatics

# All in? Identifying and tackling private sector's barriers to data sharing: A Perspective on geospatial data in the Netherlands

Chrysanthi
Papadimitriou

2023

TU Delft Delft University of Technology

**MSc thesis in Geomatics**

# All in? Identifying and tackling private sector's barriers to data sharing: A Perspective on geospatial data in the Netherlands

Chrysanthi Papadimitriou

October 2023

A thesis submitted to the Delft University of Technology in partial fulfillment of the requirements for the degree of Master of Science in Geomatics

The work in this thesis was carried out in the:


Knowledge Centre Open Data
Delft University of Technology

| | |
|---|---|
| Supervisors: | Dr.ir. B. van Loenen |
| | Dr. S. Calzati |
| Co-reader: | Dr. A. Rafiee |

# Abstract

There is a drive to create an inclusive open data ecosystem, that includes public, private and academic open data [Loenen et al., 2021]. There is a lot of existing research in Open Government Data [Janssen et al., 2012; van Panhuis et al., 2014; Martin et al., 2013] and plethora of OGD in the Netherlands. Although the private sector produces a lot of data, those are not open, as they are not participating in data sharing. In order, to fill the existing data gap, that OGD create, the EU [Commission, 2018], specifically with the Open Data Directive, started to promote more openness in private sector data, especially geospatial data. So far this, it is only done for public undertakings, as a sector in between public and private Boone and van Loenen [2022]; van Veenstra and van den Broek [2013]. The private sector is not bounded by legislation to share their geospatial data.

This research aims to identify what are the challenges to arrive at an ecosystem with more open private sector data, through the identification of the barriers in the process, in relation to the level of openness they are and how to move forward. Five categories of barriers are identified, strategic, technical, legal, economic, cultural and a multi model with 4 levels of openness is used, to identify the current state of geospatial data sharing of 9 companies in the Netherlands. The results show that companies that are providers of data are mostly sharing internally, trying to share with external users, while companies that are intermediaries are mostly sharing with some external users but they are not sharing fully open data yet, and the companies as user of existing open data are difficult to identify. This research demonstrates the influence of the role of the company in data sharing and the level of dataset, project, department in the barriers that the private sector faces and prevent them from sharing geospatial data in the Netherlands.

**Keywords**: barriers, private sector, data sharing, geospatial data, level of openness, open data ecosystem

# Acknowledgements

I am blessed to have been surrounded by supportive colleagues, friends and family throughout my entire MSc Geomatics journey.

Thank you to my supervisor and advisor team including Dr. Bastiaan van Loenen, Dr. Stefano Calzati and Dr. Azarakhsh Rafiee for your constant support and guidance throughout this process.

Mom, Dad, Vaso I couldn't have done it without your encouragement! Niko, Thano and Apostoli thank you for believing in me.

Last, but certainly not least, a big thank you to the rest of my friends for your support through the ups and downs even from afar. Especially Marianna for your blind confidence that I could do this!

# Contents

# List of Figures

# List of Tables

# 1. Introduction

There is a trend to create an inclusive open data ecosystem, that involves public, private and academin open data [Loenen et al., 2021]. The rapid developments of the last 30 years in automated data collection and communication technologies created the basis for the vastly increase of the number of the data collected, analyzed and stored [Tenopir et al., 2011]. The use of data has started to become crucial in the process of decision making, while data sharing has been acknowledged as a means to transparency, reproducibility of research, cost-efficiency and acceleration of innovation [van Panhuis et al., 2014].

According to the OECD OURdata Index [OECD, 2019], the Netherlands is placed on the 13th place among 36 countries regarding data availability, data accessibility and government support for data reuse. In addition to this the Netherlands has at its disposal a specialised platform for high quality geodata called PDOK, with 239 open datasets (PDOK). According to [van Loenen, 2018], The Netherlands can be considered as one of the global best practice countries concerning open geographic data, with already good practice in providing open government data. The Netherlands could be a good indication for the state of the private sector open data. Moreover, data sharing barriers in academic and public sector context have already been researched extensively, while research on data sharing barriers for the private sector started to become significant the last few years [Laia and Jonathan, 2020].

According to [Fassnacht et al., 2023], a significant amount of the data are accumulated and managed by the private sector, which renders the private sector as an important actor in data sharing. Additionally, In 2019, the European Commission realized the importance of the private sector data and involved public undertakings in the Open Data Directive (Directive 2019/1024) [DIRECTIVE (EU), 2019], which are private companies that perform tasks for public service. This importance of private sector data could involve the private sector as a next step in the sharing of open data. The EU started to show their interest of involving the private sector in data sharing with the proposed Data Act and trying to find how to make possible B2B and B2G partnerships [Richter, 2021]. Additionally, geospatial data was on of the categories recognised by the [Commission, 2020] as high-value datasets, that is important to be research further to help in solving issues on the topics of safety of the environment and as to get information that can be used to achieve the goals of the 2030 goals of the EU ( [Commission, 2020]). Research in public undertakings data sharing exists as a sector in between public and private, although not as detailed as public and academic data [van Veenstra and van den Broek, 2013; Boone and van Loenen, 2022]. It important to identify what are the challenges to arrive at a open data ecosystem with the contribution of private sector data.

Research into private sector data sharing is scant. This research will focus on *"What are the barriers to private sector data sharing of geospatial data in the Netherlands?" and "How can those barriers be approached to reach (open) data sharing?"* Chapter 2 presents the related work to the research topic, about data sharing, open data, barriers of data sharing for the private sector and levels of openness, followed by Chapter 3 where the research question and sub-questions are presented, next to the scope of research. In Chapter 4 the methodology of

the research is described, followed by 3 case studies in Chapter 5 and the results of the interviews in Chapter 6. In Chapter 7 a discussion of both literature review findings and interview findings is presented, follow by the conclusion of the whole research in Chapter 8.

# 2. Related work

In this section, the theoretical background of the topic is presented, by explaining how the data sharing of the private sector evolves from the existing literature. While the levels of openness of geospatial data is researched through existing open data assessment frameworks.

## 2.1. Data Sharing

The [Commission, 2022b] recently acknowledged the importance of data sharing between sectors and presented two situations in which it can be addressed, business-to-business (B2B) and business-to-government/public sector (B2G). [Fassnacht et al., 2023], on their research for B2B they define data sharing as "a collection of practices, technologies, cultural elements and legal frameworks that are relevant to transactions in any kind of information digitally, between different kinds of organizations". [Waterman et al., 2021], on their research of a flood resilience system that consists of B2B and B2G data sharing they refer to it as a process of data and technologies that are used to gather, process and share data, based on a set of skills, tools, privacy and data sharing capacity.

The primary role of data sharing is to provide access for use and reuse of data, while it can bring more advantages that are connected to verification of data, different secondary uses according to purpose, maintaining data integrity for reuse, minimizing of resources used for recollection of data, and availability of data that promotes transparency and trust [Tenopir et al., 2011].

### 2.1.1. Private, Public, Public Undertakings

Although this research will have a focus on private sector data sharing, it is important to give a description of public sector and public undertakings. That is to highlight the use of existing research on public sector data sharing that may differ from private sector data sharing, as they goal is not only to maintain their business model, but also achieve public value.

When describing the stakeholders taking part in data sharing, the public sector is consisted of the government, public agencies, municipalities and other local authorities. The public sector, is a portion of the economy composed of all levels of government and government-controlled enterprises. It does not include private companies, voluntary organizations, and households [Britannica, nd].

The private sector can be described as organizations with a majority of private ownership seeking to generate profit and not being owned or controlled by the government [Cambridge Dictionary, nd]. For example in the case of a resilient flood management system the

private sector consists of water companies, insurance organizations, consultancies, technology, infrastructure, utilities, communications. catastrophe modelling consultancies, flood management technology industries, insurance companies, re-insurance companies and environmental consultancies [Waterman et al., 2021].

While the main goal of the private sector is to generate profit there are different types of partnerships created to achieve innovation, a public undertaking is one of them. As described by the [Parliament, 1997], a public undertaking is a non-government party that carriers out a service, which can be a public service that is economic activity of general interest and is defined by public authorities, or it can be a service in areas that are entirely market based where the public authorities play no part. [van Veenstra and van den Broek, 2013], refer to a semi-public organization which performs public services for the public authorities by providing their datasets to the national data portal to achieve the benefits of transparency and re-use, while developing a business model for its data to achieve their own commercial purposes. Some examples of semi-public organizations are cultural heritage foundations, public transport organizations and research institutes.

It is important to state the differences in the business model and the type of services that the three categories provide as they might have an impact on how restricted the policies they follow are. The organizational structure of each sector differentiates in order to achieve its individual goals that may have an effect on the barriers that are created in data sharing. In order to better understand the data sharing barriers it is necessary to connect them to each category in order to potentially use them to achieve the benefits of data sharing.

### 2.1.2. Open Data

The majority of the most valuable data are owned by the private sector, and are collected and managed only for internal use [Verhulst et al., 2020], making difficult or even impossible to be used externally or for other purposes other that achieving the internal business model. In the last decade open data has played an important role in generating social and economic opportunities, while contributing in solving public problems and allow citizens to make better decisions [Verhulst et al., 2020]. The open data sharing of the private sector data could be used in assisting those social and economic opportunities.

The importance of open data has been recognised from the European Union as a tool in the socio-economic development of the society [Commission, 2020] and its contribution to stimulating citizen participation and innovation, transparency [Zuiderwijk and Janssen, 2014]. The growing demand for open data has started to have an effect on the open data policy of the European Union. The scope of open data has started to expand from to public entities to also private entities.

To give an overview of the progress on the involvement of EU in data sharing policies, the Directive of 2013 focused on the re-use of information from the public sector (Directive (EU) 2013/37) [DIRECTIVE (EU), 2013], in continuation in 2019 the Open Data Directive (ODD) [DIRECTIVE (EU), 2019] was developed where open data and Public Sector Information (PSI Directive) was enacted (European Commission, 2020; Directive (EU) 2019/1024, 2019). Building on the public sector documents, the ODD applies also to public undertakings, a specific type of private entities that collect, produce, reproduce and disseminate documents to provide services in the general interest [Boone and van Loenen, 2022].

Non-government data were added with the ODD (2019) to fill the gaps of existing public data. Data Governance Act (2022) [Commission, 2022a] was entered into force in 2022 to set up to facilitate data sharing across sectors that would benefit citizens and businesses. The Data Act that was proposed in February of 2022 comes as a continuation [Commission, 2022b] to ensure fairness in the allocation of value from data among actors in the data economy and foster access to and use of data to achieve data-driven innovation. The direct involvement of the private sector in data sharing could be derived by the goal of the Data Act (2022) to develop mechanisms for public sector bodies to access and use data held by the private sector in cases of public emergencies such as floods and wildfires [Commission, 2022b].

## 2.2. Literature Review

To address the existing literature on the barriers for the private sector to share their data, a literature review was conducted. The search keywords that were used were: ”data sharing” + ”barriers”, ”data sharing” + ”private sector”, ”private sector” + ”data sharing” + ”barriers”. From the research through keywords, data sharing in the private sector is very limited, while most literature is about public sector data sharing. The search for ”public sector data sharing barriers” gives 140 results in Scopus [Scopus, nd], while private sector data sharing barriers” gives 67 results. Additionally, in IEEE Explore [IEEE Explore, nd], the first search gives 8 results, while the second one gives 2 results. For AIS Electronic Library [AIS, nd], we get the results 3275 and 1966 accordingly. Therefore the search term is not further limited to private sector data sharing.

At the time of the literature review, February of 2023, the results from Google Scholar of the search based on ”keywords” were selected, to have a first understanding of the barriers and then continue by relevant literature from those articles. A manual assessment of each article’s relevance was made based on first reviewing the abstract, which were then reviewed based on their full-text. The reduction of articles is mostly based on data sharing not being the main focus of them and articles with a strong technical focus of data exchange, which only describes the technical transmission of data, ignoring the organizational, legal and cultural aspects of data sharing [Awada and Kiringa, 2015]. At this stage of the research, the term ”open data” was completely neglected to have a general idea of data sharing barriers. After the review of the aforementioned articles, only 2 of them were referring to ”open data sharing”. A new search term was created to incorporate ”open data”, ”open data” + ”sharing”, without specifying the sector again, due to the limitation in the private sector data sharing literature. Only 3 articles were chosen based on their relevance after reviewing the abstract.

Although, the keyword ”private sector” was removed from the search term, 2 articles were specifically about ”private sector data sharing”. Those were used as a starting point to find relevant articles based on the references used, implementing the snowball effect. The analysis of the literature review led to the identification of 24 barriers to data sharing for the private sector, which were divided into 5 categories.

## 2.3. Barriers to Data Sharing of the Private sector

Through the literature review there were identified a number of categories of barriers of sharing data by the private sector (Table 2.1):

Strategic as a term was only found in Fassnacht et al. [2023], which is also the only paper in the literature review that refers to private sector data sharing. Fassnacht et al. [2023], definition of strategic barriers refers to barriers that take place on the the highest level of the organization and affect the strategic direction. Although the term "strategic" was not identified in the other papers, strategic barriers identified by [Fassnacht et al., 2023], were part of the Organizational and Managerial barriers [van Veenstra and van den Broek, 2013], Political barriers [van Panhuis et al., 2014; Edelstein and Sane, 2015], Institutional [Janssen et al., 2012] and Administrative [Barry and Bannister, 2014]. Barriers in those categories were identified to have be part of the strategic category.

The second category that was identified in all papers mentioned above is the technical barriers, which shows similarities with the technological barriers of Fassnacht et al. [2023], information technology of [van Veenstra and van den Broek, 2013], technical quality and task complexity of Boone and van Loenen [2022] and appears to involve data characteristics, metadata, access and skills of Martin et al. [2013].

The legal barriers category is a collection of barriers that have legal repercussion that are related to privacy, regulations, security, licenses. In the literature review that categories that were identified related to that were legal [Waterman et al., 2021; van Panhuis et al., 2014; Edelstein and Sane, 2015; Barry and Bannister, 2014], regulatory [Fassnacht et al., 2023], legal and regulatory [van Veenstra and van den Broek, 2013], institutional [Boone and van Loenen, 2022], legislation [Janssen et al., 2012], license and legal frameworks [Martin et al., 2013]. The category was populated with identified legal barriers that fit the description of the category that was created [Tenopir et al., 2011; Adair et al., 1998].

Economic barriers as a term was mentioned in van Panhuis et al. [2014], Edelstein and Sane [2015], Barry and Bannister [2014], Martin et al. [2013]. The barriers of this category were also present in other categories from the literature, Commercial [Waterman et al., 2021], Institutional [Boone and van Loenen, 2022], Institutional and Information Quality [Janssen et al., 2012].

Cultural barriers that were present in van Panhuis et al. [2014] and Edelstein and Sane [2015], were mostly connected to political will, as their focus health data, that are mostly government data. However, Waterman et al. [2021], that refers to data sharing between different sectors, recognizes cultural differences that could result to barriers. Additionally, Fassnacht et al. [2023], goes into more detail about barriers that can be created from cultural differences between different organizations.

Many organizations in the private sector have enormous resources to provide for whatever issue they care to focus on [Serena et al., 2021], but they seem to not participate in data sharing activities. This literature review led to 24 barriers that will be presented in the following sections, divided into five perspectives: strategic, technical, legal, economic, and cultural. This analysis focuses on the barriers for individual companies to share their data. The barriers can be derived from the Tables 2.2, 2.3, 2.4, 2.5, 2.6 and are described in more detail in the following subsections.

| | |
|---|---|
| Security | Waterman et al. (2021) |
| Legal | Waterman et al. (2021); van Panhuis (2014); Sane and Edelstein (2015); Barry and Bannister (2013) |
| Privacy | Waterman et al. (2021) |
| Technical | Waterman et al. (2021); van Panhuis (2014); Sane and Edelstein (2015); Barry and Bannister (2013); Janssen et al. (2012) |
| Commercial | Waterman et al. (2021) |
| Cultural | Waterman et al. (2021); Fassnacht et al. (2023); Barry and Bannister (2013) |
| Strategic | Fassnacht et al. (2023) |
| Operational | Fassnacht et al. (2023) |
| Technological | Fassnacht et al. (2023) |
| Regulatory | Fassnacht et al. (2023) |
| Motivational | van Panhuis (2014); Sane and Edelstein (2015) |
| Economic | van Panhuis (2014); Sane and Edelstein (2015); Barry and Bannister (2013); Martin et al. (2013) |
| Political | van Panhuis (2014); Sane and Edelstein (2015) |
| Ethical | van Panhuis (2014); Sane and Edelstein (2015) |
| Information Technology | van Veenstra and van den Broek (2013) |
| Organizational and Managerial | van Veenstra and van den Broek (2013) |
| Legal and Regulatory | van Veenstra and van den Broek (2013) |
| Institutional and Environmental | van Veenstra and van den Broek (2013) |
| Institutional | Boone and van Loenen (2022); Janssen et al. (2012) |
| Task Complexity | Boone and van Loenen (2022); Janssen et al. (2012) |
| Technical Quality | Boone and van Loenen (2022) |
| Administrative | Barry and Bannister (2013) |
| Risk related | Barry and Bannister (2013) |
| Use and Participation | Janssen et al. (2012) |
| Legislation | Janssen et al. (2012) |
| Information Quality | Janssen et al. (2012) |
| Governance | Martin et al. (2013) |
| Licenses and Legal Frameworks | Martin et al. (2013) |
| Data Characteristics | Martin et al. (2013) |
| Metadata | Martin et al. (2013) |
| Access | Martin et al. (2013) |
| Skills | Martin et al. (2013) |

Table 2.1.: Categories of barriers of sharing data by the private sector, found through literature review

### 2.3.1. Strategic barriers

Strategic barriers appear to take place in the highest level of the organizations as they affect the strategic direction, design and decision-making to participate in data-sharing activities [Fassnacht et al., 2023]. The strategic barriers that may exist are connected to lack of organizational motivation to enable data sharing, which can lead to management commitment and corporate strategy integration.

- *Lack of organizational motivation to enable data sharing:* Developing a data sharing strategy requires organizational transformation, with changes taking place on multiple levels [van Veenstra and van den Broek, 2013]. Organisational siloes may need deeper evaluation to enable data sharing [Waterman et al., 2021]. The motivation to enable data sharing should occur at a managerial level, but due to risk-averse culture data sharing is usually avoided, without it being transferred to other levels of the organization, while more focus is given to the barriers of data sharing and not the opportunities that might be created [Janssen et al., 2012; Barry and Bannister, 2014].
- *Lack of management commitment and corporate strategy:* In van Panhuis et al. [2014] and Edelstein and Sane [2015], the lack of management commitment is presented as lack of political will and commitment to promote data sharing. While in Fassnacht et al. [2023] the barrier is translated for the private sector. Specifically, data sharing in most companies is not part of the core business, for many private sector organizations data sharing is project-based and driven by individual departments, to benefit from the potential of data sharing, a top-down commitment is needed so the organization could have a long-term strategic direction [van Veenstra and van den Broek, 2013]. The lack of a corporate strategy could result in unclear responsibilities and decision-making processes [Fassnacht et al., 2023].
- *Lack of policy coherence:* After the organizations makes a management commitment for the sharing of their data, a strategy needs to be created. A policy should be created and implemented for the private sector, with the lack of a framework there is a risk of inconsistent policies across the private sector, because each organization developed its own based on their characteristics [Waterman et al., 2021]. Janssen et al. [2012], when they are referring to sharing of open data, mention even the lack of a uniform policy specifically for publicizing data, which can create technical barriers regarding standards of data for publications. When referring to sharing of open data, the choice of policy is connected to the core decision of the organization to become an open organization [Martin et al., 2013]. A lack of consistency and perseverance in political behaviors can also affect the sustainability of an open data policy, which could effect the update and availability of datasets [Janssen et al., 2012].
- *Lack of use case identification:* Identifying use cases for data sharing for the private sector is a time-consuming and complex process and incentives should be found to persuade the private sector in participating [van Veenstra and van den Broek, 2013]. No approaches exist to identify use cases systematically, as the complexity of the process is connected to challenges such as lack of transparency about existing data, creativity for novel value-added services, unclear benefits for each participant, and scalability of potential use cases [Fassnacht et al., 2023; Barry and Bannister, 2014].
- *Lack of revenue models and scalability:* According to Janssen et al. [2012], in the Netherlands some organizations' revenue models are based on creating income from the dissemination of data. The process of opening up their data could harm their existing

| Strategic Perspective | | | |
|---|---|---|---|
| **Barrier** | **Dataset** | **Sector** | **Literature** |
| Lack of organizational motivation to enable data sharing | geospatial data, flood risk data, transportation data, geological data | private sector, semi-public sector | Waterman et al. (2021); van Veenstra and van den Broek (2013); Janssen et al., (2012); Barry and Bannister (2013) |
| Lack of policy coherence | geospatial data, flood risk data | private sector, public sector | Waterman et al. (2021); Martin et al. (2013); Janssen et al., (2012); Barry and Bannister (2013) |
| Lack of management commitment and corporate strategy | public health data, geospatial data | private sector, public sector | Fassnacht et al. (2023); van Panhuis (2014); Sane and Edelstein (2015) |
| Lack of use case identification | geospatial data | private sector, public sector | Fassnacht et al. (2023); Barry and Bannister (2013) |
| Lack of revenue models and scalability | transportation data, geological data | private sector, semi-public sector | Fassnacht et al. (2023); van Veenstra and van den Broek (2013); Janssen et al. (2012); Martin et al. (2013) |
| Lack of feedback process | geospatial data | private sector, public sector | Janssen et al 2012; Martin et al. (2013); van Veenstra and van den Broek (2013) |

Table 2.2.: Strategic barriers for private sector data sharing, result of the literature review

business model. While semi-public organizations open up their data with the support of funding from the government, there is a lack of business cases for generating revenue from reuse that could be used for the development of new business models for the private sector [van Veenstra and van den Broek, 2013].

- *Lack of feedback process:* The feedback process, specifically the lack of it, appears only on the papers of open data barriers, and not in the general data sharing barriers. That is related to open systems requiring an understanding of the external world as they must consider the feedback and insights of users in order to continuously improve [van Veenstra and van den Broek, 2013]. The lack of a feedback process to realize the needs of the user and adapt the business model accordingly, could create barriers in understanding the value of some datasets, while the development of a feedback process would be helpful to avoid economic damage related to the publishing of datasets that are not reused, and as a result do not create value [Janssen et al., 2012].

### 2.3.2. Technical barriers

The second category constitutes of five technical barriers that are connected to the coordination and alignment of data models, the technical barriers could be connected to interoperability of data, inconsistent formats, data availability, data quality or a lack of metadata, all of them connected to IT infrastructure where data is a digital asset that can be shared [Fassnacht et al., 2023; Waterman et al., 2021].

- *Lack of data standards:* While data has been extensively collected by the private sector, there is not one framework for standardization of the data that can be used all the variations of data quality, that is based on the purpose of re-(use) of the data [Welle Donker and van Loenen, 2016]. For example for geospatial data the quality may need to be increased [Waterman et al., 2021; van Veenstra and van den Broek, 2013; Adair et al., 1998], or in the case of public health data it might need to be aggregated for privacy concerns [Edelstein and Sane, 2015; van Panhuis et al., 2014]. Additionally, there is a lack of metadata describing the content, data quality quality, date of collection, which limits secondary data usage [Fassnacht et al., 2023; van Panhuis et al., 2014; Boone and van Loenen, 2022; Martin et al., 2013]. Furthermore, different datasets often are collected by in a local language, such as public health data [van Panhuis et al., 2014; Edelstein and Sane, 2015], making difficult both the access and the re-use of the data (Martin et al., 2013). Standardization of methods and the development of robust metadata can increase data access (Janssen et al., 2012).

- *Lack of accessibility and management:* Even if the private sector is willing to share their data, the lack of standards has an effect on the technical infrastructure that is not often available connected to the accessibility and management. To access the data you come across a fragmented data landscape of historically developed databases, systems and applications that lacks transparency (Waterman et al., 2021; Fassnacht et al., 2023). Moreover, connected to the access of data, many datasets are collected for a specific purpose without being maintained, due to lack of data science skills for management of data and lack of the needed software and hardware for data management (Tenopir et al., 2011; Boone and van Loenen, 2022).

- *Lack of data processing and validation:* For the dissemination of the data in an open manner a prior processing and validation of the data is needed, which can create a lot of challenges to the private sector. Data processing requires data formatting and interoperable infrastructures, when there is not a universal framework to be followed, as well as data science skills that the employees may not be familiar with (Martin et al., 2013; Fassnacht et al., 2023). Additionally, the validation process may differ depending on the purpose of re-use of the data and there is a lack of a support system that endorses feedback from the user to help in the validation process, and better realize the needs of the user, especially for open data (Martin et al., 2013).

- *Lack of data security mechanisms:* For the secure dissemination of data the security of databases and systems is important both for the private sector providers and the secondary users (Fassnacht et al., 2023). Moreover, when the data is shared in a open manner and the user is not known, there might be issues of unauthorised third-party access, cyber crime that requires data encryption, anonymization to promote trust (van Veenstra and van den Broek, 2013).

- *Lack of technical infrastructures and compatibility:* There is a lack of reference architectures and models to connect all technical infrastructure in an interoperable system that connects the different incompatible organizational systems (Waterman et al., 2021; Fassnacht et al., 2023). Additionally, for open data the value of the data is increased with the use of data, the creation of a data portal to find and access the data is important, this is connected to databases, systems, applications, software, hardware that may be needed to maintain the data portal (Janssen et al., 2012).

| Technical Perspective | | | |
|---|---|---|---|
| **Barrier** | **Dataset** | **Sector** | **Literature** |
| Lack of data standards | geospatial data, flood risk data, transportation data, geological data | private sector, semi-public sector, public sector, public undertaking | Waterman et al. (2021); Fassnacht et al. (2023); van Panhuis (2014); van Veenstra and van den Broek (2013); Adair et al. (1998); Sane and Edelstein (2015); Martin et al. (2013); Boone and van Loenen (2022); Barry and Bannister (2013) |
| Lack of technical infrastructure and data compatibility | geospatial data, flood risk data, utility data, ship data | private sector, public undertaking | Waterman et al. (2021); Tenopir et al. (2011); Fassnacht et al. (2023); van Panhuis (2014); Boone and van Loenen (2022); Martin et al. (2013); Janssen et al. (2012); Barry and Bannister (2013) |
| Lack of data accessibility and management | transportation data, geological data, utility data, ship data | private sector, public undertaking | Waterman et al. (2021); Fassnacht et al. (2023); van Veenstra and van den Broek (2013); Adair et al. (1998); Sane and Edelstein (2015); Boone and van Loenen (2022); Tenopir et al. (2011); van Panhuis (2014) |
| Lack of data processing and validation | utility data, ship data | private sector, public undertaking | Fassnacht et al. (2023); Sane and Edelstein (2015); Boone and van Loenen (2022); Martin et al. (2013); Janssen et al. (2012); Tenopir et al. (2011) |
| Lack of data security mechanisms | transportation data, geological data | private sector, semi-public sector | Fassnacht et al. (2023); van Veenstra and van den Broek (2013); Barry and Bannister (2013) |

Table 2.3.: Technical barriers for private sector data sharing, result of the literature review

### 2.3.3. Legal barriers

The second category consists of five legal barriers that are connected to the development of data sharing infrastructure to enable sharing of sensitive data and manage the risk of economic loss, loss of data, loss of reputation (Waterman et al., 2021).

- *Restrictions by law:* The restrictions by law can be on a regional, national and international level. The differences in restrictions and the lack of a harmonization of the different levels create additional barriers in creating an open system that is interoperable (Waterman et al., 2021; Sane and Edelstein, 2015). The ambiguous legal framework can lead to ownership and privacy constraints (Sane and Edelstein, 2015). For open data specifically, the opening of a dataset should abide by the existing competition laws in order not to disturb the business model of another private organisation with a similar dataset (Waterman et al., 2021).

- *Privacy constraints:* While there are attempts by the European Commission (2022b) with the European Data Act, to open the barriers for data sharing for the access and use of the data by the public, at the same time the General Data Protection Regulation (GDPR) imposes restrictions on data sharing (van Veenstra and van den Broek, 2013; Boone and van Loenen, 2022; Sane and Edelstein, 2015) and access to the data (Adair et al., 1998; van Panhuis, 2014). Moreover, for open data the cleansing and anonymisation of data can create delays that could affect the business model of the private sector, those delays could be connected to the process of obtaining consent from data owners and members of the community (Waterman et al., 2021, Martin et al, 2013).

- *Ownership constraints (Licensing and Copyright):* The quality of the data is relevant to the purpose of re/use of the data, if shared data are used in others ways than those intended that might create liability issues connected to the quality of data (Tenopir et al., 2011). Specifically for open data, there is a higher liability risk due to errors in the data, after the data is shared as open it is often not clear who to hold liable for them (Boone and van Loenen, 2022; Janssen et al., 2012). This can lead to conflicts of data ownership and usage rights (Fassnacht et al., 2023). In order to open data without liability risks, the framework of the existing intellectual property rights and restrictive licenses should be taken into account for the adoption of open licenses (Waterman et al., 2021). For the adoption of open data licenses, there are already different frameworks such as the Open Data Commons (Open Knowledge Foundation), the Community Data License Agreement (CDLA) and the Creative Commons (CC) that allow the data to be open, open under conditions, or not open at all (Boone, 2021). A framework of all the open licenses should be created to take into account all the variations that can affect data sharing.

- *Contractual boundaries (standards, frameworks):* The lack of standards and a framework for contract design can create barriers related to a secure and trustworthy data sharing, although there is an attempt to create those standards with the project GAIA-X (Choi and Kröschel, 2015). This lack of a framework, leads to data sharing through existing contracts from the private sector, where especially in the process of sharing data as open it might lead to breaches of existing contracts (Boone and van Loenen, 2022). Moreover, this lack of contract design framework for open data sharing, and the reuse of previous restrictive contracts could limit the re/uses of geospatial data (Waterman et al., 2021; Janssen et al., 2012).

| Legal Perspective | | | |
|---|---|---|---|
| **Barrier** | **Dataset** | **Sector** | **Literature** |
| Restrictions by law | non-personal data, transportation data, geological data, public health data | private sector, semi-public sector, public sector | Waterman et al. (2021); Fassnacht et al. (2023); van Veenstra and den Broek (2013); Boone and van Loenen (2022); Sane and Edelstein (2015) |
| Contractual boundaries | non-personal data, public health data | private sector | Waterman et al. (2021); Fassnacht et al. (2023); Sane and Edelstein (2015); Boone and van Loenen (2022); Choi and Kröschel (2015); Janssen et al. (2012) |
| Privacy constraints | personal data, public health data, utility data, ship data | public sector, private sector, semi-public sector, public undertaking | Waterman et al. (2021); Fassnacht et al. (2023); van Panhuis (2014); van Veenstra and van den Broek (2013); Adair et al. (1998); Sane and Edelstein (2015); Boone and van Loenen (2022); Martin et al. (2013) |
| Ownership contraints (Licensing and Copyright) liability | personal data, public health data | public sector, private sector | Waterman et al. (2021); Fassnacht et al. (2023); Boone and van Loenen (2022); van Panhuis (2014); Adairs et al. (1998); Martin et al. (2013); Tenopir et al. (2011); Janssen et al., (2012); Barry and Bannister (2013) |

Table 2.4.: Legal barriers for private sector data sharing, result of the literature review

### 2.3.4. Economic barriers

Economic barriers are usually connected to commercial losses that might occur. It is a result of companies treating data as confidential in order to protect their commercial model or competitiveness in the market (Waterman et al., 2021).

- *Fear of economic damage:* Data sharing for the private sector could cause economic damage to the data providers that could result to loss of brand reputation , financial losses, and legal penalties due to misuse of data (Fassnacht et al., 2023; Barry and Bannister, 2013). Waterman et al. (2021) specifically refers to insurance claims that might be affected by the sharing of flood related data. Although the private sector might decide to share their data as open data to enhance their value with reuse, this should not hinder their commercial purposes (van Veenstra and van den Broek, 2013). The economic damage could be related to errors in the data and liability issues related to the quality of the data, if they are not clearly stated in the license of the data, that is especially important for open data that the end-user is not known (Martin et al., 2013).
- *Loss of income from change in business model:* For the private sector their revenue system is based on creating income from data (Janssen et al., 2012). Their income is related to a cost recovery policy, in order to open up their data, they have to change to an open data policy that may cause loss of income (Boone and van Loenen, 2022). To change to an open data policy it means to provide data for free or charge at the basic cost of making the data available, which can cause the organization's business model to fail, especially when referring to high value datasets, with high demand, such as geospatial datasets (topographic maps, addresses, utility) (Barry and Bannister, 2013).
- *Uncertainty about the value of data:* Many organizations are in the beginning of recognizing data as an asset and have a difficulty in determining their value as intangible assets (Fassnacht et al., 2023). The lack of successful business cases where the value of open data is showcased, prevents private organizations from sharing their data (Barry and Bannister, 2013). As the value of open data is strongly connected to the reuse of them, it is more difficult to be determined (Janssen et al., 2012).
- *Lack of resources:* For most private organizations, data sharing is not part of their business core, which translates to additional human and technical resources that are needed to realize data sharing (Fassnacht et al, 2023; van Panhuis, 2014; Sane and Edelstein 2015). This resources are connected to the collection, maintenance, processing, validation and distribution of data (Boone and van Loenen, 2022). These could be entirely new for open data, as there are standards (metadata, quality, language), licences and security processes (anomyzation) that need to be followed for the data to be open (Janssen et al., 2012; Barry and Bannister, 2013). For Small and Medium companies (SME) the resources with which to publicize data might be nonexistent (Janssen et al., 2012).

### 2.3.5. Cultural barriers

Socio-cultural aspects of an organization and its employees can hamper the willingness to share data, risk-averse attitudes and siloed thinking towards sharing data across different sectors can amplify their unwillingness to share data (Waterman et al., 2021).

| Economic Perspective | | | |
|---|---|---|---|
| **Barrier** | **Dataset** | **Sector** | **Literature** |
| Fear of economic damage | non-personal data, transportation data, geological data | private sector, public undertaking | Fassnacht et al. (2023); van Veenstra and van den Broek (2013); Adairs et al. (1998); Boone and van Loenen (2022); Martin et al. (2013); Barry and Bannister (2013) |
| Loss of income from change in business model | utility data, ship data | public undertaking | Boone and van Loenen (2022); Martin et al. (2013); Barry and Bannister (2013); Janssen et al. (2012) |
| Uncertainty about the value of data | geospatial data | private sector | Fassnacht et al. (2023); Barry and Bannister (2013); Janssen et al. (2012) |
| Lack of resources | non-personal data, public health data, utility data, ship data | private sector, public sector, public undertaking | Fassnacht et al. (2023); van Panhuis (2014); Sane and Edelstein (2015); Boone and van Loenen (2022); Janssen et al. (2012); Barry and Bannister (2013) |

Table 2.5.: Economic barriers for private sector data sharing, result of the literature review

- *Cultural differences:* Cultural differences might exist between different regions or countries and even different organizations, that led to restrictive data access policies and bureaucratic hurdles, which prevent the private sector from data sharing (Sane and Edelstein, 2015; Waterman et al., 2021). The organizational cultural differences between historic openness versus restrictions to sharing may magnify the barriers to data sharing between different organizations (Fassnacht et al., 2023; Barry and Bannister, 2013).

- *Unwillingness to share data:* The unwillingness could be connected to the lack of political will and commitments to promote data sharing in the country where the private sector acts (Sane and Edelstein, 2015). The lack of consistency in political behaviours can even hinder open data sharing, when the established policy is not considered sustainable, as it can affect the maintenance and availability of open data (Martin et al., 2013). The culture of risk aversion to data sharing for competitive reasons, has been set in the past and it remains as part of the culture of the company on a managerial level (Fassnacht et al., 2023).

- *Lack of trust:* The lack of trust is mainly focused on the usage of the shared data, as the further processing and analysis of the data is not known to the data providers, which is linked to fear of inappropriate use, such as misinterpretation of data (Sane and Edelstein, 2015), inference with core business (Fassnacht et al., 2023), or economic, reputational, or social harm (Waterman et al., 2021; van Veenstra and van den Broek, 2013). Especially for open data, the misuse of data imposes an important risk in data sharing, if the terms of use are not explicitly stated in the license of the data (Boone and van Loenen, 2022).

- *Fear of loss of control:* Private organizations tend to think that loss of control of their data through sharing, might result to loss of power that comes with information, to prevent that, they tend to guard their data (Barry and Bannister, 2013). Fassnacht et al. (2023) go more into detail on the loss of control being connected to external dependencies

| Cultural Perspective | | | |
|---|---|---|---|
| **Barrier** | **Dataset** | **Sector** | **Literature** |
| Cultural differences | transportation data, geological data | private sector, semi-public sector | Fassnacht et al. (2023); van Panhuis (2014); van Veenstra and van den Broek (2013); Sane and Edelstein (2015); Barry and Bannister 2013 |
| Unwillingness to share data | geospatial data, flood risk data, public health data | private sector, public sector | Boone and van Loenen (2022); Martin et al. (2013); Barry and Bannister (2013); Janssen et al. (2012) |
| Lack of trust in data usage | transportation data, geological data, public health data, utility data, ship data | private sector, public undertaking | Waterman et al. (2021); Fassnacht et al. (2023); van Veenstra and van den Broek (2013); Sane and Edelstein (2015); Boone and van Loenen (2022); Barry and Bannister (2013) |
| Fear of loss of control | geospatial data | private sector, public sector | Fassnacht et al. (2023); Janssen et al. (2012); Barry and Bannister (2013) |
| Fear of transparency and disclosure of competitive knowledge | transportation data, geological data | private sector | Fassnacht et al. (2023); van Veenstra and van den Broek (2013); Barry and Bannister (2013) |

Table 2.6.: Cultural barriers for private sector data sharing, result of the literature review

on data and infrastructure, but also about lack of knowledge about data usage or even potential. Sharing their data as open data might create new connections in the network with advantages for the private sector which might come at the expense of less control (Janssen et al., 2012).

- *Fear of transparency and disclosure of competitive knowledge:* Many organizations are reluctant to share data due to fear of transparency related to confidential knowledge about their products and the disclosure of competitive knowledge (Fassnacht et al., 2023). The inference of knowledge about research and development of the private sector through the release of their data as open can even interfere with their business model (van Veenstra and van den Broek, 2013).

## 2.4. Levels of open data

The emergence of open data is evident in the research the last few years. Open data are data in a machine readable format that require a minimum of technical, economic and legal barriers for the re-use to achieve societal and commercial purposes [van Veenstra and van den Broek, 2013]. With the development of the ODD (Directive (EU) 2019/1024, 2019) that takes into account non-government data and the amount of valuable data that are held by the private sector [Verhulst et al., 2020], it is important to research the state of the open data for the private sector, through existing open data assessment frameworks.

### 2.4.1. Open data assessment frameworks

Open data frameworks are developed from different perspectives; it may be from a technical perspective [Berners-Lee, 2009], an organisational perspective [Open Data Institute, nd], or a holistic perspective [Boone and van Loenen, 2022]. 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 the private sector.

There are three open data assessment frameworks that are relevant to this research, each presenting a different point of view of open data. The three assessment frameworks are:

- Data Spectrum of the ODI
- Five-Star model of Berners-Lee
- Multi-dimensional model of Boone and van Loenen

#### The Data Spectrum of ODI

The Open Data Institute, created the Data Spectrum (Figure 2.1) as a common graph that gives an understanding of small, medium or big data, whether they are personal, commercial or government data and their place in a spectrum from fully closed, only for internal access data to fully open, accessed by anyone, and the main connection to them all is how they are licensed [Open Data Institute, nd]. It is a general open data model that identifies five levels of open data that is focused only on the legal part of data sharing. The levels are described as follows:

- Internal access, for an organization, between different departments where the access is defined by the employment contract and the internal policies.
- Named access, can refer to a data sharing by request, and the conditions of the process are explicitly defined in the contract.
- Group-based access, can refer to a group of experts related to the shared data, for geographical data these could be topographers, data scientists. The access to those is guaranteed through a process of authentication in relation to the data shared.
- Public access, on that level the data are available through a geodata portal only for view, as they are limited by licenses that limit reuse. The users in this level do not have to be experts on the field.
- Anyone, this is the fifth and final level, the open licence of the data allows for access and reuse without limitations.

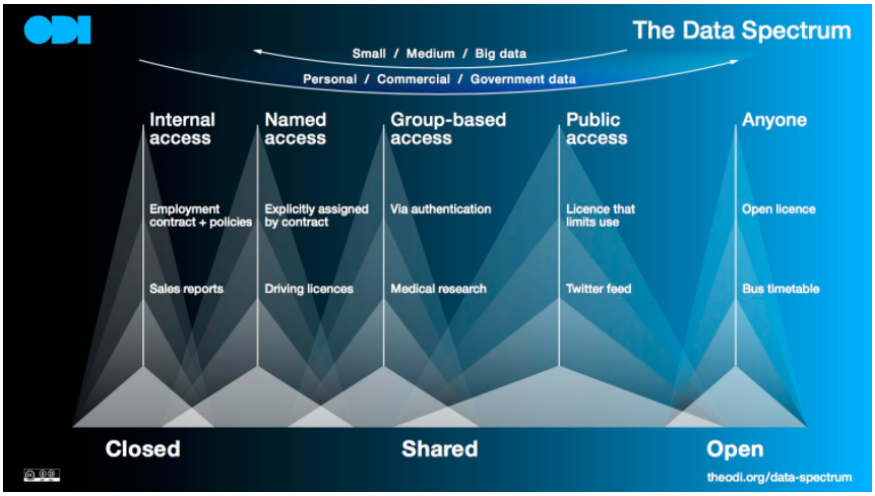


Figure 2.1.: The Data Spectrum [Open Data Institute, nd]

#### Five-Star model of Berners-Lee

The main focus of the Berners-Lee model is the interconnection with other datasets that is used as a mean to classify its openness. In addition, interconnected datasets can be reused in a variety of usages, that can contribute to the value that is added on the information on the internet (Berners-Lee, 2009). The description of the model is straightforward and it is presented in Figure 2.2. The datasets are rated from one to five stars based on the model by Berners-Lee (2009), where one star is the lowest and five stars is the highest score a dataset could get.

Starting from one star, the dataset should be available on the web, in any format, but with an open licence. For the dataset to get two stars, it should be in a machine-readable structure, such as an excel spreadsheet, which is a limited propriety format. For three stars, the dataset should be in a machine-readable format that is non-proprietary, which is an open format, such as CSV. The fourth star is given when all of the above are true and the dataset uses open standards from the World Wide Web Consortium (W3C). The open standards that need to be followed are referring to using Resource Description Framework (RDF) with the query language SPARQL. The data is stored as RDF and SPARQL is used to perform queries across different data sources [W3C Recommendation, 2008]. The fifth star of the model is possible for datasets that follow all the aforementioned criteria, while having the ability to link the dataset to other datasets.

#### Multi-dimensional model of Boone and van Loenen

The Multi-dimensional model of Boone and van Loenen (2022) (Fig. 2.3) [Boone and van Loenen, 2022] is built upon six dimensions, the type of user, data quality, data regime, findability, usability and in which way data is shared. Three levels of open data were identified: not open, partly open and open.

At the first level data is considered not to be open and only accessible to the internal user, and it is not findable through a general engine. The absence of an open licence makes

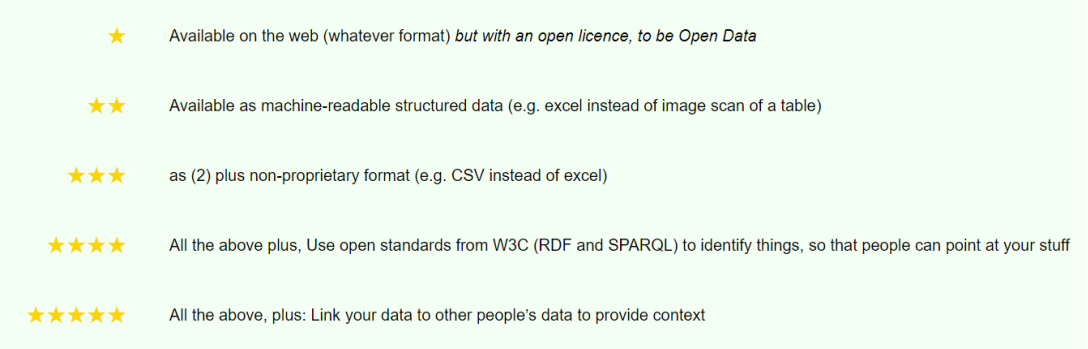


Figure 2.2.: Five-Star model of Berners-Lee [Berners-Lee, 2009]

it impossible to share the data with external users [Welle Donker and van Loenen, 2016; World Wide Web Foundation, 2017]. This is on par with the internal regime that focuses on the internal purposes of data, limiting data quality to the purpose of the internal user.

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 and van Loenen, 2016]. However, fees may be charged and the data can only be shared under certain conditions and terms. This data policy generates internal and external value.

At the third level data can be considered most open. This 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 reuse the data [DIRECTIVE (EU), 2019]. This can be applied through an informediary 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 and Janssen, 2014]. The quality of the data and its openness is still relative for each specific case and user [Safarov et al., 2017]. Although the last level is preferred from the user's point of view it might lead to issues with the business model of the private sector [Welle Donker and van Loenen, 2016].

### 2.4.2. Addressed barriers of open data

Going through the open data assessment frameworks in relation to the data sharing barriers that were identified in Section 2.3, there are three categories that are addressed through them, those are legal, technical, and economic. This will be further described in this subsection.

Starting from the Data Spectrum of ODI, mainly legal barriers are considered in the process of moving from one level of openness to the next on. The different terms in licences determine also the type of user, from an internal user of the organization, moving to a group-based user, that can be a group of experts in the field, to any user not taking into account the level of skills that may be needed for the reuse of data [Martin et al., 2013]. In the fourth level, public access, the characteristics of the user that is limited are not presented, which means there is no indication for the design of the appropriate licence to be used [Welle Donker and van Loenen, 2016].

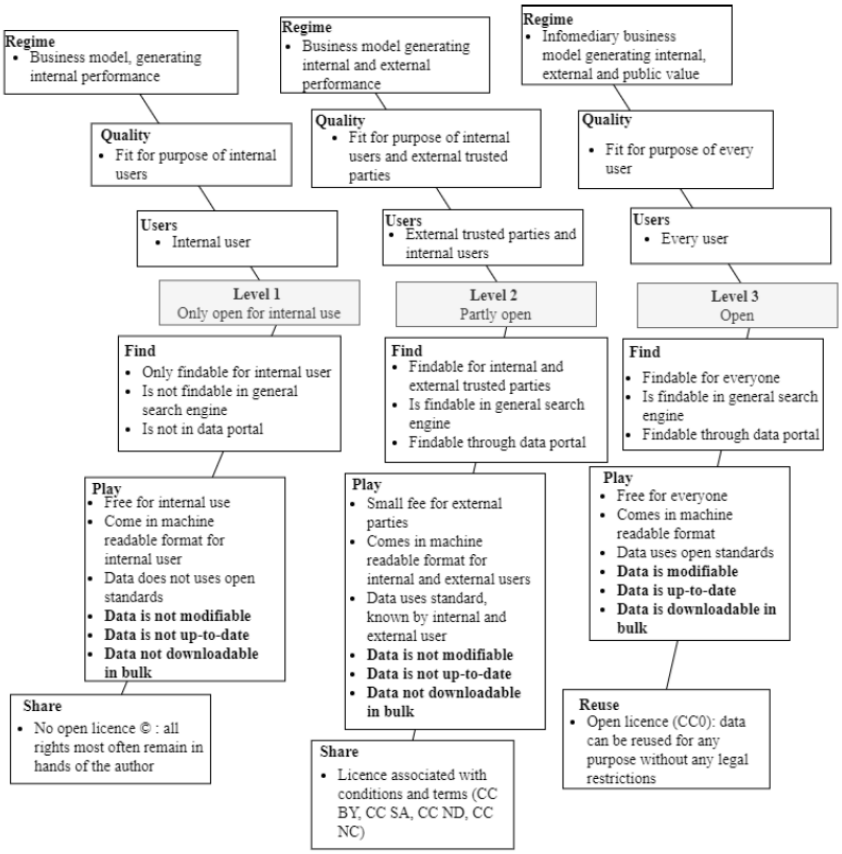


Figure 2.3.: The multi-dimensional model of the three identified levels of open data [Boone and van Loenen, 2022]

The framework of Berners-Lee (2009) [Berners-Lee, 2009], focuses on the interconnection of the dataset itself, creating the basis of the legal framework, with an open licence and builds upon it [Janssen et al., 2012]. In the next level, the focus is on the technical aspects of data standards, starting from a limited propriety format to an open format, such as CSV. While maintaining these criteria it focuses on the importance of metadata as a way to connect data with similar content. At the highest level of five stars the interconnection of the dataset with other datasets is the drive to the creation of value from the dataset. This framework takes into account the legal and technical barriers, while also proposing a solution for the development of business models, connected to services that can be created, through the added value of linked data.

The Open Knowledge Foundation, defined openness based on technical characteristics of the existence of data standards (format, metadata), data portal and its connection to a data portal through the use of metadata. The reuse and redistribution of open data is connected to the legal aspects of open licences. Part of the licences in the framework, are also share alike licences which could hinder the reuse from everyone, but could be an opportunity for the private sector to maintain their business model through commercial licences.

The multi-dimensional model of Boone and van Loenen [2022], takes into account the holistic framework of Welle Donker and van Loenen [2016], to consider the user perspective in the levels of openness, which is in agreement with three out of five of the levels of users presented in the Data Spectrum [Open Data Institute, nd], internal access - internal use, named access - external trusted parties and internal users, and anyone - every user. Through out the six elements of the framework the legal aspects that are addressed are licences in the share element, the technical aspects are addressed in find and play elements, through the data standards (format, metadata, quality), data portal, and general search engine, and the economic aspect through the development of a regime connected to the characteristics of the user for every level.

The common characteristics of all four frameworks are their focus on the technical, legal and economic barriers, while there is no mention of strategic or cultural barriers and ways to address them. The importance of a regime is highlighted in the multi-dimensional model, without mentioning the cultural differences that may exist on a national and an organisational level for the private sector connected to historic openness or restrictions of sharing, that prevent the private sector from even developing a strategy for open data [Fassnacht et al., 2023; Barry and Bannister, 2014]. Establishing an open data culture for the organization is important to enable open data sharing to the other levels of the organization [Janssen et al., 2012; Waterman et al., 2021].

The enabling of data sharing, could lead to management commitment and the creation of a corporate strategy, that involves all levels of employees and departments of the organization, while making data sharing part of the core business, which translates to a long term internal strategic direction, that takes into account technical (data standards,data portals, management systems), legal (licences, liability) and economic barriers (use case identification, development of open data revenue models), by stating clear responsibilities and decision-making processes [van Veenstra and van den Broek, 2013; Fassnacht et al., 2023].

Although, the establishment of an open data culture is important, the lack of trust was identified as a barrier that could prevent it. Lack of trust for data providers was linked to fear of misuse and misinterpretation of data, which comes as a result of the user not being known in the open data system [Edelstein and Sane, 2015; Boone and van Loenen, 2022].

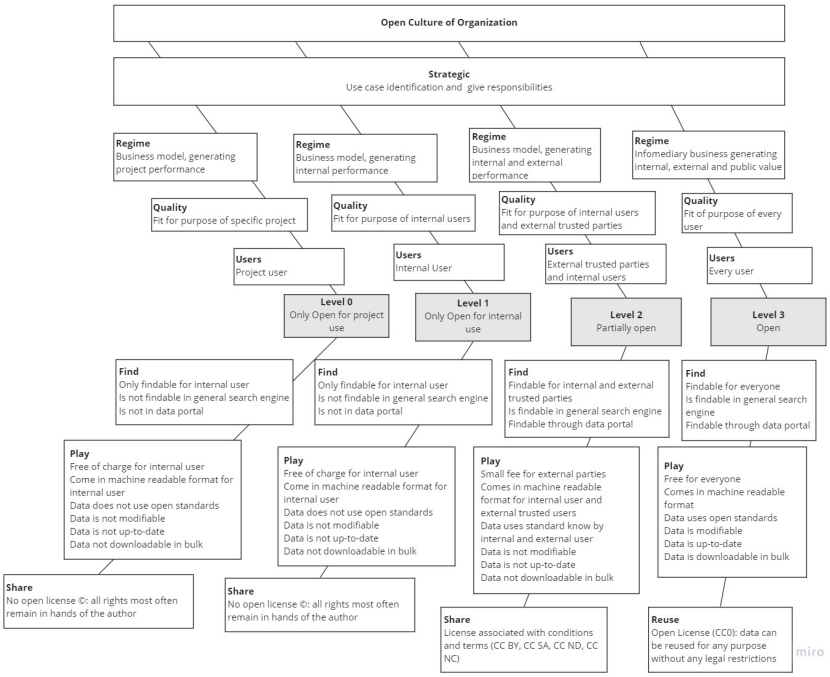


Figure 2.4.: The updated multi-dimensional model of the four identified levels of open data, created from the literature review

Based on the literature review and the multidimensional model of Boone and van Loenen (2022) at its core, the updated multi-dimensional model of the four identified levels of open data is presented in Figure 2.4. To give an explanation of all the components of the previously mentioned assessment models that were used in the creation of the new multimodel. The holistic approach of [Boone and van Loenen, 2022] created an solid base to built on it with the addition of elements that were missing using the other two models.The five stages of the Data Spectrum and the Five Star model give an indication of the need of intermediate levels for a more thorough analysis of the existing open data situation. That is why the Level 0 of the new model was created, giving an indication that even for internal access the sharing of data can be limited to even specific department excluding the rest of the company, bounded by the contacts of people that are part of this department.

The Five star model was integrated on the find section of the new model were it is integrated for Level 2 partially as level two involves dataset that are available on the web but not with an open licence and Level 3 that can vary from datasets with 1 star to 5 stars. The fifth star is an indication of the link between datasets that can be possible though the existence of metadata that is also mentioned in the multimodel [Boone and van Loenen, 2022]. The multidimensional model was created with a public undertaking in mind that at some point because of its special businesss model to achieve and the public service that provides on the governments command, has as an end goal open data. The difference with the private sector is that this might never be the goal of their business, which is why the addition of the regime of a business model that generates project performance might be more realistic.

### 2.4.3. Research Approach

Based on the literature review that was conducted the private sector is not sharing their data with other public or private organizations. Most of the existing literature was focused on public sector, public-undertaking data, where the intention to open their data was clear due to the objectives of the ODD (2019) [DIRECTIVE (EU), 2019]. With the clear intention, the focus moved to how to tackle the barriers to open data, those barriers were technical, legal and economic.

For the private sector data sharing, there is a different starting point, because there is a lack of research regarding the willingness of the private sector to share their data, as they are not bounded by the ODD or other legislation. This different starting point, is the base to identify different categories of barriers that may hinder private sector data sharing. Through the literature review two additional categories were identified, those of strategic and cultural barriers. From the cultural barriers, cultural differences on a national and organizational level and the unwillingness to share data due to cultures of risk-aversion that were established for years, can be indications on why there is lack of organizational motivation to enable data sharing on a strategic level. Those are some barriers that indicate that unwillingness of the private sector to share data, is not rooted only on technical, legal and economic barriers.

The intention is to move forward from data sharing to the private sector opening up their data, to achieve that it was important to identify whether different levels of openness have an impact on these barriers. Three open data assessment frameworks, that were identified as relevant to this research, were further researched. The common characteristic of all three was their focus to technical, legal and economic barriers, neglecting the strategic and cultural barriers. The lack of an open culture in the private sector, might be preventing from even developing a strategy for open data, which is why it is important to be researched further. The focus of this research will be ”What are the barriers to private sector data sharing of geospatial data in the Netherlands?” and ”How can those barriers be approached to reach (open) data sharing?”.

# 3. Research Questions

## 3.1. Objectives

The main research question for this thesis is ”What are the barriers to private sector data sharing of geospatial data in the Netherlands?” and ”How can those barriers be approached to reach (open) data sharing?”

In order to answer the main research question the following sub-research questions should be answered:

- What are the different levels of openness for geospatial data?
- How do the different barriers relate to each level of openness of the developed model for private sector data sharing?

## 3.2. Scope of research

The scope of this research is presented in Figure 3.1 to better represent the aspects that affect the current research questions. This research will be focused only in private companies in the Netherlands that are involved in geospatial data sharing as producers, users or developers of services for geospatial data. For these companies the barriers that prevent them from data sharing will be researched in relation to their level of openness with a focus on strategic and cultural barriers.


Figure 3.1.: The scope set for this research

# 4. Methodology

This section elaborates on the methodology that will be followed to answer the research questions presented in the previous section. Although, there is existing research that focuses on data sharing for the private sector to private sector [Fassnacht et al., 2023] , and research for barriers for the public sector to open up their data [Janssen et al., 2012], this research will investigate the barriers for the private sector to share their data in a more open manner. The topic will be researched through an exploratory research, as there is not much research performed on this topic.

Merkus (2022) [J., 2022], describes a general methodology for an exploratory research, which can be divided to primary and secondary research based on the data collection and analysis. In primary research, the data collection is being done through surveys, focus groups or interviews, while in secondary it is being done by using results of a primary research, such as case studies, literature review or published results of surveys. Furthermore, Crinson (2006) [I. and M., 2006] gives an overview of the types of interviews, which are structured, unstructured, semi-structured (a combination of both, and focus group interviews.

The methodology for this research (Fig. 4.1) will be based on a literature review of the related work regarding the categories of barriers that effect private sector data sharing, as well as, the levels of openness of data that can affect the barriers to data sharing. The literature review will be used as the theoretical basis, but to have a better understanding of the reality of private sector data sharing semi-structured interviews will be conducted with selected companies. In preparation for the interviews, an interview structure is designed taking into consideration the proposed sub questions mentioned in Section 3.1.

The research will be conducted with semi-structured in-depth interviews. According to Crinson and Leontowitsch (2006) [I. and M., 2006] the qualitative interviewing could be used to go into more detail about the barriers of data sharing for the private sector and how to address them. Interviews will be conducted with the private sector as an expert on the field to get more meaningful insights on the barriers that may exist in data sharing in relation to the level of openness. In the following Section 4.2 the interview process is explained in detail.

## 4.1. Case studies

Next to the literature review, the three case studies of AND-OSM, Seabed 2030 and Microsoft Footprints were used as success stories and provided the benefits of open data sharing even for the private sector that are connected to the brand reputation and sustainability, marketing of the companies products and tools in the creation of the open datasets and the reuse of open data with multiple research papers as a result. The case studies are presented in Chapter 5.

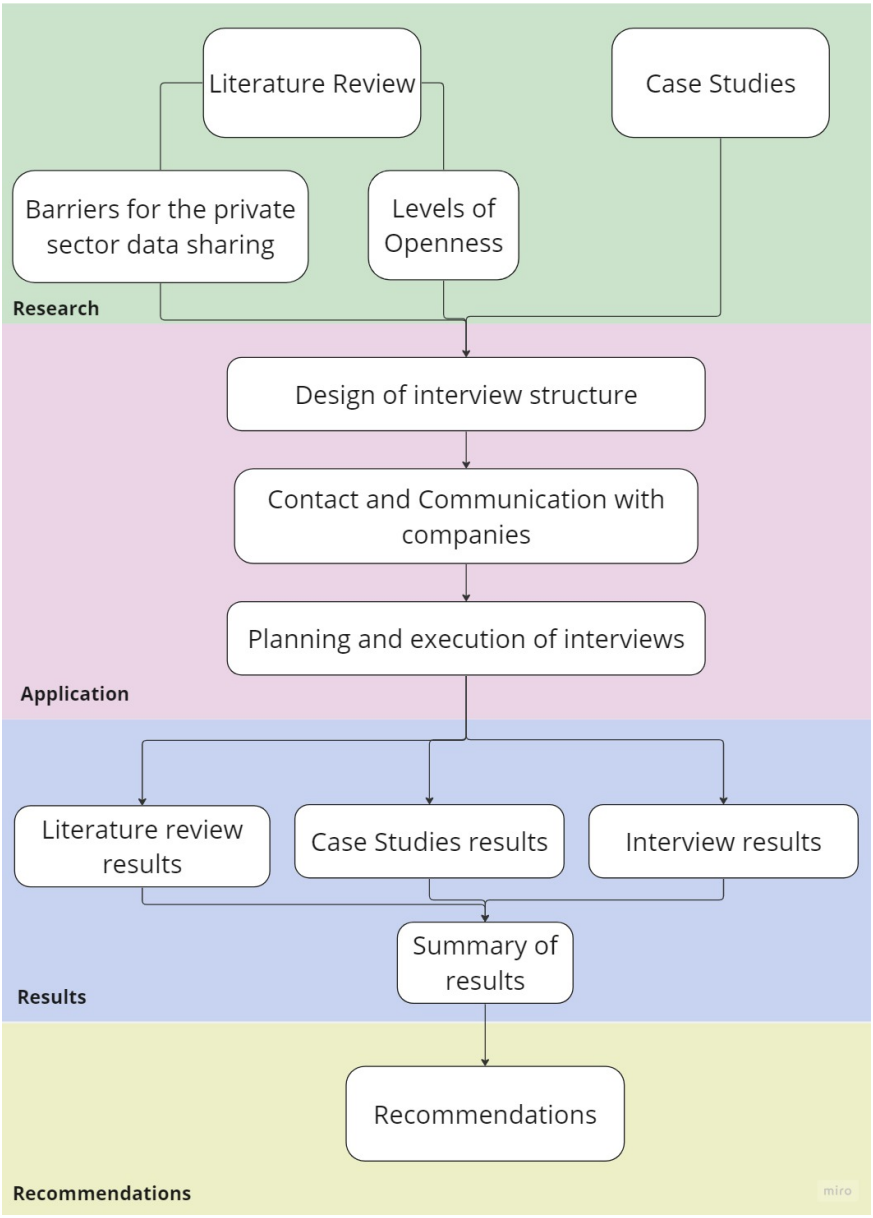


Figure 4.1.: Research Approach

## 4.2. Interview Process

Initially the focus of the research was to contact companies that produce geospatial data in the Netherlands and each one of them fits to one of the four levels of openness from the schema in Figure 2.4. The difficulty to identify the level of openness of a company upfront led to the identification of the level being part of the interview process. The research for the suitable companies for the interviews started from the members of the Geobusiness Nederland trade association for companies that work with geospatial data [Geobusiness.nl, nd], they were asked whether they could suggest companies from the clientele that fit the four levels of openness, based on the model (Fig. 2.4). The Geobusiness.nl were not able to identify and suggest companies, for that reason the research for companies turned from finding a specific company that fits one of the levels to contacting as many companies in the Netherlands as possible that are involved in geospatial data sharing to get an overview of the current situation. The final selected interviewees were contacted with one of the three ways, through a mutual contact, through the official website of the company, or through a research and connection on Linkedin.

In the process of contacting the interviewees, the topic of the research was mentioned on "What are the barriers that exist and prevent companies in the Netherlands from sharing their geospatial data with others, such as government, universities, or other companies. More importantly, how to tackle those barriers?". They were also informed that the interview will take from 45 minutes to 1 hour of the time and it could be arranged in person or online. After arranging the day and time of the interview, the interview questions were sent to inform them of the structure and the content of the interview in more detail. The interviewees were asked to study the schema in Figure 2.4 before the interview, and to try to identify in which level of the schema their company is at the moment. Subsequently, the identifications of the level of openness of the company in the schema was the starting point of the semi-structured interview.

At the beginning of the interview all interviewees were asked if they allow the recording of the interview in order to keep a record for the later analysis of the results, and whether it is allowed to mention their name in the final report. Additionally, the interview started with stating the aim of the interview to identify the level of openness of the company in achieving open data and to identify barriers they face in relation to sharing open data, followed by a brief description of the schema in Figure 2.4, to make sure that they have a good understanding of its content. When the identified level was not clear, but it contained parts of different levels the interviewees often presented an example dataset or a project to explain their identification of the level of openness, and that dataset or project was used as a reference point for the following questions.

Moving from the current level of openness that they identified, in order to identify the barriers that they faced in the process in relation to their level of openness, the interview is focused on their past situation. Starting from Question 2 "how did they arrive at this level", that could help in describing the process that they followed and if it was intentional, and whether there is a strategy behind it, giving the focus to the cultural and strategic layer of the schema. The interview continued with Question 3 "what are the barriers encountered", to go more into detail of the specific barriers for each company, or dataset that was used as an example. Question 4 "how did you resolve/overcome them" was used as an indicator on how they moved from one level to the other, and if that could be used to move to another level.

To cover the situations where the company identifies itself in the left side of the schema, and using a hypothetical situation to move towards level 3 of open data, the section of the future situation of the interview was asked. Question 5 and 6 focus on the barriers that they expect and if they have some suggestions on how to overcome them in the future to show that it might be possible. The last Question 7 is focused on what it might be needed to move to the next level, and it was used as a way to identify potential benefits and incentives that could make the company willing to participate in geodata sharing in the future.

While following the structure of the interview as stated in Section 4.3, additional follow up questions were added in the process of the interview, according to the topic of discussion. All 10 interviews took place during April and the first week of May 2023, with the companies SWECO, Asset Insight, Esri Nederland, Fugro NL, Cyclomedia, Tensing, GeoJunxion, Imagem NL, Geodan and Microsoft. The format of the interview is presented in Section 4.3 as it was described in detail in this Section.

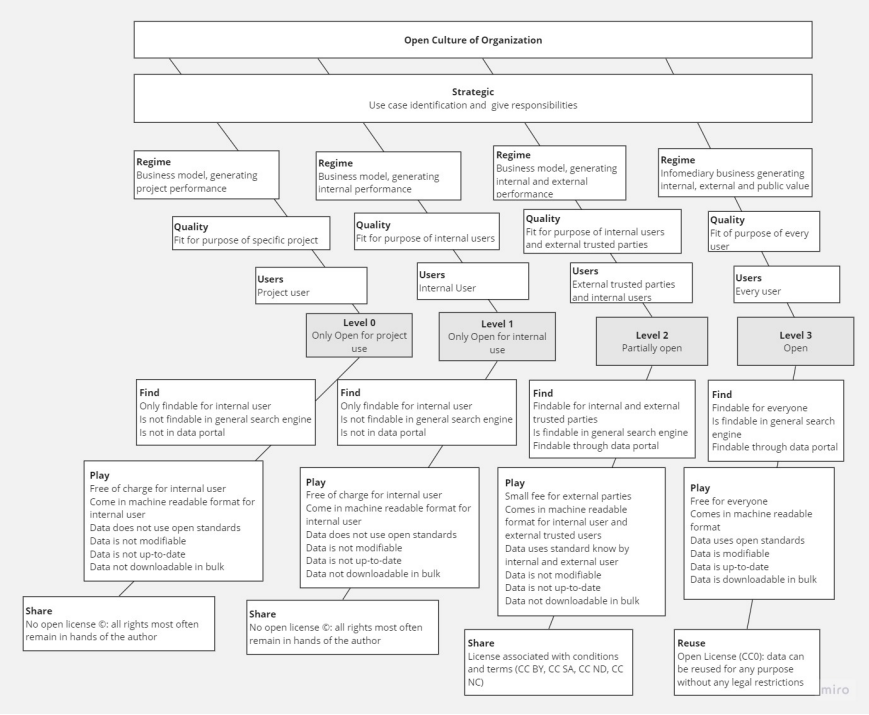


Figure 4.2.: The updated multi-dimensional model of the four identified levels of open data

## 4.3. Interview Questions

**Aim of the interview**: Identifying the level of openness of the company in achieving open data. Identifying barriers they face in relation to sharing open data.

**Current Situation**

1. Where do you think you are in this schema? (level of openness model)

   **Past Situation**

2. How did you arrive at this level in the schema?
3. What are the barriers encountered?
4. How did you resolve/overcome them?
   **Future Situation**
5. What are the barriers you expect to move to the next level?
6. How could you overcome those barriers to move to the next level?
7. What is needed to move to the next level? (benefits economic or other)

# 5. Private Sector Data Sharing Case Studies

During the research for companies in the Netherlands that would be suitable for the interview process, there were identified three case studies of private companies that shared their data as open data. Those could be used as a indication for the interviewed companies of the incentives that helped them share their data. Also, as those case studies are considered success stories it could help the other companies by showing them how they managed to overcome their barriers in the process of sharing their data with others to move to level 3 of the schema in Figure 2.4. The three case studies refer to two Dutch private companies and one American which is active in the Netherlands, involved in the geospatial data production. They will be presented in a chronological order starting with AND in 2007 with the sharing of the streetmap of the Netherlands to OSM, Fugro in 2017 with the sharing of their bathymetry data to the GEBCO Ocean Map and Microsoft in 2018 with their building footprints data.

## 5.1. AND Streetmap of the Netherlands OSM

"Automotive Navigation Data" (AND) is a Dutch start-up company founded in 1996 [GeoJunxion, 2020a]. In 2001, they focused in Digital Map Data. In 2007 they decided to donate the whole street network of the Netherlands to OpenStreetMap [OpenStreetMap, 2007]. OpenStreetMap (OSM) is a community based project that is "built by a community of mappers that contribute and maintain data about roads, trails, railway stations around the world" [OpenStreetMap, 2004]. In 2008, AND released the first navigable map of Western Europe, part of which was the dataset that they provided to OSM. In 2020, the company is renamed as GeoJunxion as their focus sifts to location intelligence services, on the clients request [GeoJunxion, 2020a]. These services are Research, Data Collection and Consolidation, Geo-Analytics and Normalisation, Global Sources Search, Outdoor Venue Plan [GeoJunxion, 2020b].

On July 4th 2007, it was announced by Steve Coast, the founder of OSM, that AND donated a street network of the entire Netherlands to OSM, and with this geodata Netherlands would be the first complete country in OSM [OpenStreetMap, 2007]. AND donated digital maps of the Netherlands, China and India with the aim to create the most up-to-date map in the market, recognizing the power of the community and the Web 2.0 developments. A small part of the data donated was not up-to-date and according to the CEO of AND, Maarten Oldenhof, at that time, it was their goal out of this project to close this gap with the help of the OSM community [Internet Archive Wayback Machine, nd]. At the time of the announcement a plethora of linked wikis were created to inform all OSM contributors of the AND codes that need to be integrated to the OSM existing codes, and information about the integration of the Roadmap dataset and the Borders dataset [OpenStreetMap, ndb]. Additionally,

two conversion tools from AND to OSM data were created by the OSM contributors Import Floris and Import Marc to help the whole community process [OpenStreetMap, nda].

To give a little more context to the case study that is not documented anywhere else, an interview was conducted with a researcher in open data at the time of the donation, that discussed informally with AND on the motivation for the donation to OSM. At the time of the donation, in 2007, there were discussion within the national government on more open geo-data. As mentioned by the interviewee "the Rijkswaterstaat as part of the ministry of Infrastructure and Environment (Infrastructuur en Milieu, IandM) was actively participating in this attempt. Rijkswaterstaat was sued for opening the NWB (national road database) by a company that had invested in developing a similar database and viewed publishing the NWB as unfair competition, but Rijkswaterstaat won the case as the company concerned could or would not provide an insight into their costs and lost income." As a result of the agreement on the INSPIRE directive in 2006 (enacted in 2007), there were discussions in opening more datasets including the topographic dataset (TOP1ONL). At the time AND's business model was based on a similar dataset and the freely available topographic dataset would be a competitor that would result in the end of AND's business model.

Through this case study AND as a new start-up company, found the OSM data contribution as a way to market their company as a company that cares for the geodata community and the involvement of everyone in them, while showcasing their expertise and attracting new customers. "Rather that fighting the upcoming ACT, AND donated the dataset to OSM under a CC-BY-SA and gained free advertisement for their products by the in the national press, while also competitors that used it would have to credit AND". In the process of doing that there were some barriers that venture difficult to achieve, those were in relation of who would be responsible for keeping the data up to date, the quality of the data gradually degrading, while also worrying for potential misuse of the data available on OSM, by end users that not always have the level of data science skills needed.

## 5.2. Fugro Bathymetry Data Seabed 2030

Fugro is a Dutch multinational company, with 9000 employs in 59 countries, active in four regions: Americas; Asia-Pacific and Australia; Europe and Africa; and Middle East and India [Fugro, nda]. The company was founded in 1962, as an Engineering Company for Foundation Technology and Soil Mechanics and went public in 1992 on the Euronext Amsterdam stock exchange and in 2021 they generated a total of 1.463 million EUR [Fugro, ndd]. The headcourters of the company are in Leidschendam, Netherlands. The services that they provide are divided into four categories Marine Site Characterization, Marine Asset Integrity , Land Site Characterization, and Land Asset Integrity [Fugro, ndb].

In 2017, they devised a method for collecting valuable high-resolution bathymetry datasets from their survey vessels as they transit between survey projects, this developement is called Fugro's Office Assisted Remote Services (OARS®), and it is a technology that "enables safe and efficient data acquisitions without the need for dedicated survey staff onboard" Fugro [ndd]. In the same year, Fugro started to support the project Seabed 2030, by sharing their collected in transit bathymetry data. Seabed 2030 is a collaboration between the Nippon Foundation and General Bathymetric Chart of the Ocean (GEBCO), that wants to inspire the complete mapping of the world's oceans by 2030 and to compile all bathymetric data into the freely available GEBCO Ocean Map [Sea, nd]. By 2018, the company had contributed

65,000km2 of high resolution crowdsourced bathymetry data to the project, according to GEBCO "crowdsourced bathymetry (CSB) is the collection of bathymetry data from vessels, using standard navigation instruments, while engaged in routine maritime operations" [GEBCO, nd]. while in 2020 Fugro's crowdsourced mapping programme has expanded to include seven survey vessels with data contributions now totalling more than 500,000km2 (Millar, 2020). The amount of data keeps growing and on December 2021, it reached a major milestone of more than 2 million km2 of high-resolution, in-transit bathymetry data collected by Fugro vessels [Fugro, ndc].

Along with all of Fugro's data contributions to the project, they are also in the process of investigating and advising their clients on how their existing of planned datasets can be incorporated to the programme, including datasets with sensitive information [Fugro, ndc]. As of 2021, the Seabed 2030 Project is formally endorsed as a Decade Action of the UN Decade of Ocean Science for Sustainable Development, as part of Goal 14 of the UN to "conserve and sustainably use the oceans, seas and marine resources for sustainable development", which could be of help in informing more private companies on how they could contribute to the project, apart from Fugro's clientele [Sea, nd].

The Seabed 2030, and the contribution of Fugro with their bathymetry data was a way for Fugro to market the development of OARS®, which allows efficient data acquisition of bathymetry data without the need for extra survey staff to do the acquisition. Showing to their customers the possibilities of this data collection process and instrument, at the expense of less resources. At the same time, as Fugro is a publicly traded company that also has to maintain a sustainable image, the Seabed 2030 project was a way to show that they are also contributing to their sustainability goals for a "better livable world". The motivation for this contribution is visible, but nonetheless they encountered some barriers in the process of making it happen, that relate to finding data that they could legally own, without using extra vessels and survey staff, but by using the least resources possible to achieve it. That is why the solution of data in between other projects with no need for extra staff was the solution. In the process, the data standards of bathymetry data, became an issue, that was solved by the existence of the GEBCO database, with already determined standards, and an already developed database for them to be stored.

## 5.3. Microsoft Bing Building Footprints

Microsoft is an American software company that was founded in 1975, the headquarters are in Redmond, Washington, USA. It has 221.000 employs and in 2021 it generated 168 billion USD. Microsoft develops, distributes, licenses and supports a wide variety of computer-related products and services CCN [2019]. Microsoft has its own web mapping service Bing Maps [Maps, nd]. On 28th June 2018, Microsoft Bing released 130 million Building Footprints in the USA as Open Data. Specifically, the Bing Maps team provided the dataset to OpenStreetMap, with a goal to increase the coverage of building footprints available. The full dataset is available on a GitHub repository (USBuildingFootprints, 2018), alongside a description of the whole model on how the building footprints where derived from their computer vision algorithms and Bing Maps satellite imagery [Blogs, 2018].

Although the reproducibility of the building footprints requires a level of data science skills, the dataset is available for download in a bulk through the GitHub repository. To avoid the need of a lot of storage, there is also the option of native calls to the Bing Maps API

(Application Programming Interface), through a REST service (RESTful API is an interface that two computer systems use to exchange information securely over the internet (Amazon, nd)). There is available an Imagery API documentation [Microsoft, nd] to make it possible for every user to retrieve the building footprints at a specific zoom level, for a given set of coordinates. The dataset is provided in GeoJson format, which is an open format that is used to encode a range of different geographic data.

The model to generated high-quality building footprints from open satellite data and it relies on Open Source CNTK Unified Toolkit which was developed by Microsoft. Using CNTK we apply our Deep Neural Networks and the ResNet34 with RefineNet up-sampling layers to detect building footprints from Bing imagery. Both tools from Microsoft are used in the process, while also, to run this model you need a lot of computational power that is possible on the Microsoft Cloud.

As soon as the dataset was made available as open data, a lot of applications found use for it, with the most popular being the one from HOT (Humanitarian OpenStreetMap Team), who is using it to map places that are vulnerable to disaster and poverty. Once the mapped areas are verified the humanitarian organizations can act in times of need (HOT Tasking Manager, nd). In additions, with a Google Scholar search, there are at least 15 more applications where the US building footprints where used since 2018, to identify informal subdivisions and manufactured home communities [Durst et al., 2021], to study of the population exposed to flood hazards [Huang et al., 2019], for Automatic Damaged Building Detection from Post-Hurricane Remotely Sensed Imagery (Chen et. al, 2018) and to uncover the dominant social class in Neighborhoods through Building Footprints [Liang and Wang, 2019].

This case study was an opportunity for Microsoft to show their expertise to potential clients, with the advertisement of the storage capacity and computational power of their API and cloud where it is possible to run this model. Additionally, following their sustainability goals it was a way for them to show that they are actively contributing by providing insight on existing problems, like the one of providing clean water in Africa, and doing that by contributing to the data sharing community with the development of open source tools such as the CNTK Unified Toolkit. The satellite data that they used were already available as open data, which did not create extra costs for the purchase and use of the data, and the whole development required extra human resource, that is was provided by the Microsoft Research section of Microsoft, where researchers can apply for a grand for their research and Microsoft gains potential high skilled employs that work for other development projects of Microsoft.

# 6. The current state of the geospatial private sector in the Netherlands

The case studies presented in Chapter 5, can be used as an indication of the barriers that the three companies (AND, Fugro, Microsoft) faced in the process of sharing their geospatial data as open data. They can also be used as examples of the potential benefits that they gained through this process, and those can further be used to persuade other private companies to share their geospatial data. Having those case studies in mind going into the interviews with the companies in the Netherlands, which is the focus of this research can help in having a better understanding of the interview results. The importance of this section is not only the barriers that were mentioned during the interviews with the experts in the selected companies, but also how these barriers are connected to the level of openness identified by each interviewee.

## 6.1. Asset Insight

Asset Insight is a company that was founded in 2018, in the Netherlands and it is part of VolkerWessels, a Dutch construction group. The company was created to focus on all services related to measurements and inspection, it consists of the 'Inspectation', 'Data Science', 'Paving Control and Advice' and 'Tecson Inspections' departments, inspects, measures, monitors and predicts the quality of assets with 60 employs [Insight, ndc]. The services provided by the company focus on asset management for track, road, water and industry, by inspections, digital measurements, monitoring and creating predictive models using data analytics and artificial intelligence, all of their customers are infrastructure related [Insight, nda,n].

For this research the interview was conducted with the Project Coordinator of the Track Department, and the focus was on the data collection with the Measuring Train Sheloc, the geodata of geometric parameters of the track are measured with a laser scan system, and a video inspection system that takes a HD photo of the track every half meter, the raw data have a cm accuracy [Insight, ndb]. According to Insight [ndb] "the measuring train complies with the European NEN standard for measuring trains and carries a ProRail certificate". The measurements are meant for clients like ProRail, a public undertaking that owns the tracks, but also for other inspections companies such as VolkerRail that maintains the track.

When presented with the schema of the level of openness, the interviewee, identified Asset Insight in level 0 to level 2 depending on the project, while stating that level 3 would never happen for their department as, they are not the owner of the data, and are only hired to do the collection and processing of the data. When referring to the project with Prorail and VolkeRail, there are several contract restrictions that arise that do not allow them to share the data even within the company, between different departments. They are not allowed to

share their data with the other department, as Prorail uses them to check whether Volkerail is doing their part on maintaining the quality of the track, which leads to the collection of data twice for different projects. Out of the contract Asset Insight is allowed to use the data develop Image Recognition, to automate the visual inspection, but they cannot share the raw data to other clients. Even within the company, the analytics departments is not allowed to share datasets from competitors of the track department.

Other than the contract restrictions that exist, the accuracy of the track data is temporal, which it often does not allow the reuse of them after several months, because there could be changes that render the data out of date for the use case of inspection. Because of their constant contracts with ProRail and VolkeRail, the data standards and formats that they follow were developed through their needs, and because Prorail is a public undertaking most of them are open standards like .csv. For the video data there are specific requirement and accuracy of the data as .geotiff. While following open formats for theses project, it is not always the case with all their clients, which effects their processing process. Moreover, part of the data that they collect with the Train Sheloc are video data, that create privacy concerns, to connect an individual to a location, but those are covered by the GDPR. However, the GDPR restrictions requires that they blur the faces of people that are captured in them, but this has an effect on the quality of the data, for the measurements, and it slows down their processing, which then means a delay on the delivery of the data to the customer, an potential loss of income.

When asked if they are familiar with open data and whether they use them in the processes, they talked specifically for data provided by trainspotters they replied that they are not part of their general processes. The reason for the lack of use was that they are unreliable, their sources are not stated in the metadata, and they are not sure if they can use them. They stated that ”although we know that most of the times they are true, our clients do not believe that and we are held reliable for errors. Still, the datasets are used for validation, of the development process of a product.

In a hypothetical scenario of moving to level 3 of the schema, the interview identified potential barriers that may arise. As part of the data that are collected with the Train Sheloc, there are also sensor data that are calculated towards a model, and within that model they can do measurements. Sharing the raw sensor data would mean giving information to their competitors on how they do it, if the data were shared completely open that would not be valuable to them and it could affect their business model. Also, it would give a lot of information on their expertise, that would arise questions, and it would be time consuming to answer all those questions, that could affect them from focusing on the other project that are attached to a business model.

The barriers that were highlighted to move to level 3 were that the data produces is technically complicated to share, as they are not always using open formats, sharing the raw data that are really specific as open can rise national security issues. Moreover, or licensing, their software is really specific based on the type of measurements they do that could create issues in sharing raw data. Also, for some of the datasets it is a technical challenge to put it into the storage Esri because of the massive amount of data. It was specifically mentioned that from the commercial aspect it is not in their interest to want to share data because ”information is knowledge, knowledge is money” as stated by the interviewee.

## 6.2. Cyclomedia

Cyclomedia is a Dutch company, that was founded in 1980, it has offices in the US and Germany and has 250 employees. It is a private company that specialises in the large-scale and systematic visualisation of environments based on 360° panoramic photographs (Cycloramas). Additionally, it also provides services of data analytics and delivers insights to their customers [Cyclomedia, nda]. They have developed an interactive webviewer for all Cyclomedia content that is called Street Smart, to view, analyze and use their data [Cyclomedia, ndd]. Moreover, they collect three types of data: Cycloramas that are 360° panoramic photographs with an ultra-high resolution from street level [Cyclomedia, ndb], LiDAR point clouds, and Street Ortho [Cyclomedia, ndc], a product generated by LiDAR point clouds and Cycloramas. From the collected datasets they provide three categories of data analytics: Mobility pack, Utility pack, Geo Objects + [Cyclomedia, nda].

The interview was conducted with a Consultant on Positioning System that is part of the Research and Development department, that makes the systems for the processing of the data. Within the Research and Development department there are base products team (images and point clouds), data analytics team for the processing of the data, platform team to make sure everything runs smoothly, they have a database of 5 petabytes in ESRI, front-end API and portal team. The GIS knowledge is shared in all those teams, with a limited amount of employs with knowledge about high-accuracy GPS, geometry to achieve 10 cm accuracy.

When presented with the schema (Figure 2.4), they stated that they collect the data, so they are always the owner of the data, as they do not sell them, but they license access to them. Cyclomedia was identified at level 1 and 2 by them and that depends on the product. They have data that are collected by them, that are also used from their analytics department, and they have some external parties that they are working with like the company Centric and IT company. These partnerships puts them in level 2, but from the characteristics of level 2 in the schemas the ”small fee” does not apply and they do not use CC licenses, but everything is stated in a contract that both legal teams approve. All the other characteristics of level 2 are valid. Also, they use open data standards (OGC), with a clear API documentation, and the data is up to date every 14 months, those are part of level 3.

In the process of describing the barriers that they faced so far, most of them focused on the data standards and metadata specifications that were resolved after discussion and suggestions from the customers to provide a better product. When asked about the barriers that would come up if they tried to move to level 3, the main barrier was the financial support that they would need, possibly from the government, as funding. The business model of licensing access to their data works for now and it is pretty successful in the Netherlands, so there is no intention from them to share their data as open. The only occasion that they provide data for free are within their collaborations with TU Delft and TU Twente, for research data, and in those cases they still create a contract that limits the use of the data to the use case that is researched, but in exchange they have the potential to find highly skilled future employs through their collaboration.

## 6.3. Fugro NL

The interview regarding the Fugro Seabed Case Study, was conducted with an employ of Fugro in America, which could potentially have a different internal structure and barri-

ers than Fugro NL. As this research focuses on the companies in the Netherlands and the barriers they face in sharing their geospatial data, a second interview was conducted with the Production Manager Rail, about the Track department specifically and it's findings are presented in the following section.

The interviewee identified Fugro in the schema (Figure 2.4) at level 2, because the datasets for the track department are shared internally and sometimes with external trusted parties. When they presented the barriers that they face at the moment to share their data, they mentioned that they are not the owner of the data and they are usually very restricted on the use of the data, and who are the parties that are allowed to share it with. There are several datasets that they are not allowed to share with other departments for security reason, those are raw datasets of high accuracy. Although an internal portal for all datasets exists and for all their contract there is always stated a number of years that the dataset should be maintained by Fugro, not everyone in the company has access to all datasets. The same portal is used for external trusted parties, with specific restrictions on access and use.

In a future scenario where they would have to move to level 3, according to the interviewee, first they would have to be the owner of the data to make that decision. It could be used as a way to advertise their expertise through their portal, but because of the technical characteristics of the data that require high data science skills, they are in fear of misuse of the data from third parties. Specifically, they mentioned that "the data can end up on a search engine where everyone, with or without experience has access to, the company could face a lot of questions, not that the company is hiding something but questions about how it was created, when it was created, or maybe the dataset is not up to date, because the dataset is out of the control of the company", so there would be people out there using data that the company is liable for. Another barrier that could come up is, because of the use of the Amazon cloud to store their data, every time a dataset is downloaded from there Fugro has to pay the fee, which it is small, but if the data where open, this fee would multiply. Those extra costs add to the already high cost of collection and processing of the data, that sharing those datasets as open data is not viable.

## 6.4. SWECO

Sweco is a a European architecture and engineering consultancy firm that was founded in 1958, with a total of 20.000 employs in Sweden, Norway, Finland, Denmark, Germany, Netherlands, Great Britain, Belgium, Czech Republic, Poland, Estonia and Lithuania. The focus of this research is only on Sweco Nederlands, that has 1.900 employs. The total turnover for 2021 was 2.2 billion EUR [SWECO, nde]. They provide consulting services in four segments, Buildings and urban Areas, Water, energy and industry, Transportation infrastructure, Architecture [SWECO, ndc].

There are 85 GIS consultants within Sweco NL, and their customers are government agencies, municipalities and private undertakings, such as Port of Rotterdam, Schiphol, Rijkswaterstaat, Prorail, Provinces, Waterboards, Municipalities. The focus of their department is software development regarding Asset Management, GIS / BIM, Urban Planning, Surveying and Mapping, and from 2011 the are based on Dutch Open Data provided by PDOK (Open Geo Data – Dutch Cadastre) and CBS (Statistics Netherlands). Some of the software that they developed are Obsurv [SWECO, ndb], a Web-based asset management platform

for infrastructure, Staat van je Straat an asset management app [SWECO, ndd], dg DIALOG BGT , a web service for The Register of Large-Scale Topography (BGT) is a digital map of the Netherlands on which buildings, roads, waterways, sites and railways are unambiguously recorded [SWECO, nda].

The interview for SWECO was conducted with the Head of the GIS and ICT Department. SWECO was identified by the interviewee at level 2 or 3 of the schema in Figure 7, depending on the product or service. Most of the products that they create are in order to help others make their data open. This clients are usually, municipalities or government agencies and SWECO develops applications based on already open data provided by PDOK, to avoid having liability issues, as those are not their data. When asked about barriers that they faced to reach level 2, they mentioned that they started with the development of closed software that was used only within the company, and at the start of 2000s when the shapefile was a popular format to export your data to, they tried to avoid the extra step of having to export their data into a shapefile, by developing WFS services in collaboration with ESRI, to minimize the processing time. In the process of choosing the right datasets to use for their applications, issues with the accuracy of the data and the lack of metadata standards were mentioned, while also mentioning that not all employs have the geodata science skills to differentiate the best accurate and reliable datasets, which leads to them asking for help from the few people in the department with GIS knowledge.

SWECO as a user of open government data that are already available, when asked about what to do to move to level 3, they focused their opinion on the how to better prepare the open governement datasets for users with lower level of skills.

## 6.5. Tensing

Tensing helps organizations get the most out of geographic information. With one of the largest teams of certified geodata experts, multiple offices across Europe and over 30 years of experience at the intersection of GIS and data integration, they bring consulting across the board - from sound advice to delivering innovative geodata solutions and maintenance. In every step of the process, Tensing can form or support geodata teams with capacity and knowledge. They are providing services that involve creating a strategy for their clients location data, design of software that fits best to a case, software development that builds on top of ArcGIS and FME, system integration and 3D GIS consultancy that focuses on Digital Twins [Tensing, nda]. Their focus is on six industries: governments, utilities, contractors, public safety and security, mobility, water authorities [Tensing, ndb]. The main focus of the company is share their knowledge as a service on everything related to geodata. That is being done, by the Tensing consultants being hired by different organizations, municipalities, other companies to find solutions for their geodata. The interview to identify the level of openness of Tensing in achieving open data and to identify the barriers that they face in the process was conducted with a GIS Developer in the company.

From the moment the schema of level of openness was presented the interviewee stated that they could not identify Tensing in the schema as an owner of data. The role of Tensing is to help their clients with finding and creating solutions for their data, this could be by developing software, by managing, maintaining and analyzing their datasets or by creating new datasets for them the client might own or other datasets that exist. The interviewee felt more confident in answering where Tensing would be in the schema based on their level of using

existing open data, and they would identify Tensing in level 2, the schema does not apply to Tensing because of the kind of work they do. Tensing as a user of open data, would be at level 2, where there are aware of existing open datasets and they are actively using them in their processes. The information of the existence of the datasets, comes from their extended network within the Netherlands, that involves, government agencies, municipalities and other private companies that they worked with before.

When they tried to elaborate on the barriers that they faced in relation to sharing geodata the interviewee approached this question from their experience from a recent project at a Municipality, in trying to develop a dataset for them, based on their data and an existing aerial photos that was available in PDOK. For Tensing this was good, because they could help the Municipality. The aerial infrared photos are every year available, which was also good for the use case because the municipality wanted to measure the difference in green per year. The barriers that they encountered with the Municipality were focused on the level of skills of the current employs of the Municipality, which called for the hiring of Tensing with the level of data science skills to complete the task. It is possible that due to the low capacity of the team, shortage of team members with other set of skills the dataset will not be updated this year. Additionally, a barrier that was mentioned in the process was that the municipality did not have the resources to keep the Tensing expert longer to maintain the dataset, which resulted in allocating that task to an employ that already worked for the municipality, with no specialty in geospatial data science.

When faced with a future situation on how to potentially move to level 3, the interviewee pointed out that since they are not the owner of data it is difficult to answer that question. Although, during the interview it was mentioned that Tensing is in the process of developing new services for Digital Twins, and they own some 3D datasets that were created in the process. In the event of wanting to share those 3D datasets as open data, the interviewee could not pinpoint one specific barrier, but they mentioned that for Tensing there is no business value to be willing to share open data, so it is not a goal for the company at the moment. They could share their 3D dataset partially as open data if that meant that it would be an opportunity to attract potential clients.

It is important to mention that their extended network of partners allows them to have knowledge on the new trends in geospatial data, and to actively pursue them. Specifically, they have a clear organogramma of three teams ArcGIS, FME, data science, where the manager of each team is responsible for the task of research on existing datasets, new trends that could be developed into services by using some hours or some days per week on that with a clear goal. The are called ”Evangelists” according to the interviewee and their task is to inform, help and motivate the team, like a geodata ambassador. In addition to that, the company organizes a knowledge exchange session every month, that everyone can present and inform others about new developments. Tensing’s business model shows adaptability to the new trends, that does not inquire them to share data as open data.

## 6.6. Imagem NL

Imagem is a Dutch private company that specializes in solving complex spatial issues with location intelligence solutions, they automate, visualise, analyse, and translate data into actionable information [Imagem, nda]. To find those solutions through data automation they developed a pipeline that consists of four steps simplification with automate processes by

combining innovation with under-the-hood location data, production of datasets through collection, processing, analysis of geospatial data, management of data through organizations of geospatial and business data in one library, and finally delivery of information to the customers through web services [Imagem, ndb].

The interview for Imagem NL was conducted with the Commercial Director. They work a lot with geospatial data, but those are either provided by the customer, they are purchased from the customers or they purchased them for their customer or we use open data available from PDOK. They process data, take raw imagery, put an algorithm and information comes out of it. The data always goes back to the customer so it is not their to share,as they say everything that was explained in the schema is geared towards companies that provide data. The did not identify Imagem NL on a specific level. When describing if we created a new algorithm or model we want to share the working of that model, they have to ask the owner if we can use their data to demonstrate, governments they are usually fine, but with commercial companies it is different. It is not fair to strive for open data in all cases, because there are companies that have a lot of time a lot of energy and a lot of financial means to create specialized data, then they don't have no means to get a return on investment from their efforts, and those efforts should be linked to a business model.

As they are software development company, they have a very strong advocacy for open standards, because they believe that the value of their software solutions should come from what this software is doing and it should not come from the fact that it is build in to closed data and that people have no other choice to use the software, because they want to use the data, they support the Geodatabase format, to the extend that Esri has made it possible to support it through publishing the ability to support, but that is not always the case. Even if they decide to be open there are cases where it is not possible. For example they work with the Ministry of Defense and they are obliged to secure the data, with no appropriate level of clearance they are not even allowed to see those data, as a matter of national security.

The barriers that they usually face is not about open culture, but how to carefully choose which data standards to support. "If our resources were infinite we would support all standards", they need to choose which of the standards is important for our clientele, because it takes time, to create the standard, create support for it, test it, put it through approval and retest it, then they can support the standards. Pretty often organizations use different standards, and then the company has to choose, sometimes even in exchange for the investment in research that this organization would otherwise provide.

According to the interviewee, in a future situation to move to level three they would not expect new barriers, because their culture is open so they would not have to persuade their company about the importance of open data. According to the interviewee moving to level 3 it should not always be the solution as it should not always be the goal to the most open level. It should be distinct on which cases it should happen, for public data with public investments, open data should not move back to not open, if they want to make data open that was not created publicly they should put in place framework for anyone that has put in investments, to make sure that the investment has returns to get back the cost that they put into creating it.

## 6.7. Geodan

Geodan is a Dutch private company, located in Amsterdam, that specializes in location intelligence solutions, that employs 140 people in the Netherlands. Their role is to collect and combine data to provide smart insights to their clients based on the processing of their data. Their services focus on location intelligence, logistics and mobility, vital infrastructures, smart society and energy transition [Geodan, nd]. Some of their products are STTR builder to create easily applicable rules for the Environmental Act, Phoenix an app that provides the same overview of the situation to all stakeholders involved in the project. Decentralized Counter, you offer citizens and companies a counter for all permits, including those not covered by the Environmental Act.

The interview was conducted with a Business Consultant. When presented with the schema they had a difficulty in identifying Geodan in the schema as they do not own data. They are a consultancy company that helps other make their data open, and those might be the public sector or some companies, such as electricity companies that asked from them to inform them on the benefits that open data could have for their company and how to achieve that. So according to the interviewee, the company is on the higher levels about the culture of open data and their benefits, but they are a private company so their primary goal is money.

"Open data is not a business goal for us but it is an important trend", as they believe in open source, open standards, they are not working with vendor lock-in companies, and because of that interoperability is not usually easy with ESRI, because they do not follow the OGC standards completely. In an attempt to contribute to the development of open source they contribute to the OGC standards, with pilot programs, taking part in the OGC conferences and testing them new standards. One of the most common barriers is that if they use data from their clients they have to make agreements on the use and sharing of the data. Internal use is usually possible, but not sharing them with others, they need the data property to resale data, it all depends on the license it has and if it is not classified as open data they always have to ask the owners.

In the attempt to move to level 3, some things that were mentioned as important in the process are for Geodan to have data property, but also that even though it is not that hard to publish open data, it is difficult to maintain it, latest version, latest data model, and so far it is usually treated as a project that when it is over it is forgotten. Moreover, as privacy is an important issue in the Netherlands, if they have data on location of people the cannot share it, they have to anonymize it, and that could be time consuming and cause a delay to the deliverable. Also, it was stated that "if we don't have a driver, like a contract with a company or the government, we are not urged to share data".

## 6.8. Esri Nederland

ESRI (Environmental Systems Research Institute) is an American multinational geographic information system (GIS) software company, known for its ArcGIS products. It was founded in 1969, and it has 49 offices worldwide, while the total turnover for 2021 goes to 1.1 billion USD. This research is focused only in one of its offices, Esri Nederland, that has 160 employees, with a goal to help organizations make themselves smarter with geo-information

[Nederland, nd]. In the Netherlands, more than 1.000 Dutch organizations work with ArcGIS include major cities, governments and the business community.

The ArcGIS platform gives access to Esri clients to more than 200 geo datasets, the "Living Atlas of the World" data portal [ESRI, nda], from which almost all are open data according to their website. Esri Nedeland provides products in the following categories: Mapping, Field Operations, Spatial Analysis and Data Science, Imagery and Remote Sensing, Indoor GIS, Real-Time Visualization and Analytics, 3D Visualization and Analytics, Data Management and Developer APIs [ESRI, ndb].

The interview was conducted with the Manager in Product Management and Innovation, from ESRI. When presented with the schema, they identified ESRI at level 2 and 3 depending on the dataset. Specifically, two datasets at level 3, which are the postal code zones and the topographic map of the Netherlands. For the postal code zones, it is based on an open dataset of the postal codes in the Netherlands, but ESRI created a model for automatic processing of the post codes into postal code zones, based on topographic characteristics and natural barriers like water. For the topographic maps, they are maintaining a service of the topographic maps provided as open data by the Kadastre, where you can retrieve the data you need through WFS. Their level 3 datasets are not always up to date. A similar model of postal code zones is being sold still from other companies, when ESRI provides it for free, if they were a partner they would have to think to not disturb their business model.

When referring to their level 2 datasets, the Living Atlas of the World was mentioned, which is a geodata portal that it has the restrictions that it is free with licensing restrictions that only users of ESRI software can use them (Esri, 2023c). They also have some datasets offered for a fee for data their produce, such as orthophoto, aerial imagery that they are provided with a yearly subscription to get access to a dataset or a paper model. ESRI has three categories completely open, open for ESRI customers, for a fee. Their focus is not on producing geoinformation, but on the development of software to make it easy for their users to start using GIS.

The barriers that they encountered in the process are discussed through proposals to the director of ESRI, in relation to their strategy to make data accessible to their clients. The consideration is being done based on whether there is a market value or financial value to proceed and whether it would affect their partners. If it affects their competitions this might be even an incentive to make it happen. Another important aspect that can affect their decision is the man hours that are used in the development of the product, according to how technically complex the product may be, as they also have to think of the cost of maintenance. In the future situation of moving to level 3 they barriers that they are expecting are the benefit that they would get, their software has to be used to maintain their business model, the expectations of the users about the quality of data and keeping them up to date. "Customers care more for the high quality than paying the small fee" according to the interviewee.

## 6.9. GeoJunxion

AND rebranded in 2020 to GeoJunxion, and although they did an important contribution to OSM in 2007, it is important to follow their stance in data sharing in the present. That was why the interview was extending in trying to identify GeoJunxion in the schema in

the present. When talking about the role of GeonJunxion in the system of sharing data, the interviewee mentioned, that there were a lot of customers from the industry that approached them, to create datasets, without the knowledge that some of them already exist as open or commercial data, that a motivation to change their business model to location intelligence services, that showed a rise in the revenue.

The interview was conducted with the Sourcing Manager, their task involves, research of available datasets for their clients, whether they are open data or commercial data, and informing them of the license restrictions on use. When presented with the schema there was a confusion about the intend of the schema. Specifically, their concern was ”Is the goal for the company to become open or to use open data? Because those could be two completely different things”. In fact in the company they use open datasets that allow it based on their license, terms and conditions, to create products that will be at level 0 or 1. It was stated that the level depends on the dataset, and they had the occasion with AND OSM streetmap data, that was mentioned in section 5.2, which according to them is at level 3, and the basemap that they were licensing access to it to municipalities or private companies, as their previous business model, before 2020, at level 2, but most of the services that they offer right now are in level 0 or 1.

When describing the process of donating their streetmap data to OSM they mentioned that a barrier in the process was that there was not a clear agreement on who would maintain the quality of the dataset up to date, as it was no longer a responsibility of AND, also they were worried about the misuse of the data in OSM, as the skills of the user are not known, that could lead to more questions that could affect the brand of the company. On the other hand, with their current business model, and the task of the sourcing department to research for existing open or commercial datasets, the barriers as a user of open data focus on the licensing and copyright conditions of use, that prevent them from some existing useful datasets from being used, such as the license CC SA (Creative Commons Share Alike), that would force them to share their data as open data. Further limitations were mentioned related to contract restrictions, specifically the decision to share data changes drastically when referring to a data license, were the owner of the data can decide what to do with the data, and a service license where the data that were created for that service, as stated in the contract, it is not allowed to be used commercially for other reasons.

In a future scenario where they would like to move to level 3, from their experience with the OSM data, they expect to have barriers related to low quality, misuse of data by unknown end users. Although, as company they believe in the importance of open data, as they are a user themselves, they mentioned that ”we are a B2B company and a customer based company that requires us to be focused on commercial data, and not making data available publicly”. Moreover, even if they wanted to move to level 3, most of the times the data is not theirs to make that decision.

# 7. Discussion

## 7.1. RQs

The goal of this research was to find out "What are the barriers to private sector data sharing of geospatial data in the Netherlands?" and "How can those barriers be approached to reach (open) data sharing?"

This research was conducted in an attempt to identify the barriers that prevent private companies in the Netherlands from sharing geospatial data, and in that process, if it would be possible to find ways to address those barriers in order to reach open data sharing. In order to do that, a literature review was conducted that contains barriers that were identified in data sharing from the public sector, public undertakings and private sector, as the existing literature focuses on the public sector data sharing as it was presented in the use cases of Martin et.al. (2013), Barry and Bannister (2013), van Panhuis (2014), Janssen et al, (2012). With limited literature for public undertakings van Veenstra and van den Broek (2013) and Boone and van Loene (2022), and Fassnacht et al. (2023) that focused on data sharing between private companies.

To answer the main research questions the additional sub questions mentioned in Section 3.1 should be answered first. In the following section the sub questions will be answered one by one according to the existing research.

## 7.2. Barriers literature vs real life

Starting with identifying the barriers to private sector data sharing of geospatial data in the Netherlands, through the literature review there were five categories of barriers that were identified: strategic, technical, legal, economic, cultural. From the Tables 2.2,2.3,2.4,2.5,2.6 in Section 2.3.

Starting with the interviews we will try to identify the barriers that were mentioned from the different companies, and see if they fit in the identified categories of barriers that were found during the literature review. The barriers that were identified during the interview and fit with the existing literature review will be presented first based on the five categories of strategic, technical, legal, economic and cultural barriers, and in the next section the barriers that were not developed in the literature review but came up during the interview will be presented.

### 7.2.1. Strategic

The strategy of an organization is vital for its strategic direction and desicion making on whether and how the organization would participate in geospatial data sharing. Through the literature review there six strategic barriers that were identified, starting with the lack of organizational motivation to enable data sharing that should occur on a managerial level, and make changes in all levels within the company [van Veenstra and van den Broek, 2013; Waterman et al., 2021; Janssen et al., 2012; Barry and Bannister, 2014]. That can have as a result the lack of management commitment and corporate strategy that leads to data sharing being project-based or department based [van Veenstra and van den Broek, 2013; Fassnacht et al., 2023]. This lack of consistency is enforced by the lack of policy coherence that could lead to other technical and economic barriers connected to resources [Waterman et al., 2021; Martin et al., 2013; Janssen et al., 2012].

In the case of geospatial data sharing, it is usually difficult to realise the value of data, on a strategic level the lack of use case identification that could add value to the data lead to time-consuming and complex processes to identify value-added services [van Veenstra and van den Broek, 2013; Fassnacht et al., 2023; Barry and Bannister, 2014]. The lack of business cases for generating revenue from reuse of data leads companies to maintaining there current business models with no new revenue models [Janssen et al., 2012; van Veenstra and van den Broek, 2013]. In order to develop new revenue models you need a constant interaction with the external world, to better adapt the business model, but this is not possible with the lack of a feedback process [van Veenstra and van den Broek, 2013; Janssen et al., 2012].

The strategic barriers are difficult to give them a clear description as they are on a higher level and their effects are visible when translated into lower levels of technical, legal and economic barriers. Most interviewees, when asked if they have a strategy for data sharing or open data, their initial answer was no, while through the conversation some slivers of strategy for data sharing was identified, in this section this strategic barriers will be presented according to the interviews.

Asset Insight, as a company that collects and processed geospatial data based on specific contracts, a business model connected to data sharing and open data does not seem like an option. The did not have in mind revenue models for the reuse of data, that could become a better option than their current of full filling a five year contract with ProRail in the hopes that they would renew this contract for another five years. This led to lack of organizational motivation to enable data sharing, that translates into the managerial level, the person I interviewed, not committing to it and maintaining data sharing on a project based or department based goal, that creates into a siloed landscape that is visible through the technical, legal and economic barriers that are presented in the next sections.

Fugro presented a strategy for data sharing that it was more focused on the internal data sharing between departments and projects, as there is a clear communication between departments on a managerial level, with a clear organogramma, with the responsibilities of the employees connected to the internal data portal. Although, this is all geared towards maintaining their current business model on a department level, were there is potential communication with other departments and exchange of resources, but they is limited research on identifying alternative revenue models connected to geospatial data sharing. As stated, from the interviewee ”data sharing and open data not the current strategy of the company, so there are no steps taken in that direction”.

For Cyclomedia there is clear understanding of data sharing as being a new opportunity of licensing their data to maintain their business model. As stated by the interviewee, "this model works for now in the Netherlands, so there is not need for us to change it". The current management commitment in sync with that the internal reuse of data can be beneficial for them on a department basis, but they have no way of scaling that to an external data sharing revenue model, only in the case of their collaboration with universities in the exchange of high skilled future employees. This strategy does not appear sustainable or adaptable to future changes, and their interaction with the external world is through feedback for their new products, only on technical specific issues.

Geodan mentioned also a lack of motivation to enable data sharing for the company, as "if we don't have a driver, like a contract with a company or the government, we are not urged to share data, we will not spend time on things that do not benefit us", which shows a lack of research on new revenue models that do not involve a direct return on investment. Moreover, it was stated that "so far data sharing is usually treated as a project that when it is over it is forgotten", which show a lack of management commitment on a higher level.

SWECO maintains a business model of creating products in order to help others make their data open. This shows an organizations motivation to enable data sharing, and it is further enforced with the creation of the GIS department as knowledge exchange base. An example of their managerial commitment to an alternative business model was at the start of 2000s when shapefile was a popular format to export your data to, they tried to avoid the extra step of having to export their data into a shapefile, and developed software for WFS (Web Feature Service) which is an OGC standard. Their current business model of only using available open data from PDOK and their clients to develop their products the potential scalability of their revenue models, that the existing open data limits their legal and economic barriers. This is only possible, because of them not providing any data of their own, but contributing to data sharing with services.

GeoJunxion, although they stated in the beginning of the interview that "we are a B2B company and a customer based company that requires us to be focused on commercial data, and not making data available publicly", shows a motivation to enable data sharing with the creation of the Sourcing department that collects and reviews available open data and commercial data that could potentially be used to achieve their business model. It appears to be difficult for them to translate that into their main business model that is not only connected to return on investment, because there are not successful business cases of reuse that they could use as an example. Although, they also, contributed to open data, by donating their streetmap to OSM, it is not clear how this could be possible for their main data strategy.

The two companies that appear to have a more clear strategy for their data sharing are Tensing and ESRI Nederland and they will be presented in the following paragraphs. Tensing acts as a consultancy firm for geospatial data, which means maintaining a good network of partners that may need their accumulated knowledge in technical matters connected to geospatial processing software, existing datasets available and turning geodata into useful information. They are not actively sharing their own data, but their established network of partners allows them to have knowledge on new trends on geospatial data, and whether it is beneficial for them to pursue them.

To enforce this network they have a clear organogramma of three teams ArcGIS, FME, data science, and the manager of each team is responsible for the task of research on existing datasets, new trends that could be developed into services by using some hours or some

days per week on that with a clear goal. They are called ”Evangelists” according to the interviewee and their task is to inform the team, like a geodata ambassador. Additionally a knowledge exchange session happens every month were everyone can present and inform others about new developments. This business model shows adaptability to the new trends, that does not inquire them to share data as open data. As a result, they are currently developing new services for Digital Twins, with 3D datasets that were created in the process, and could potentially be shared as open after the end of the project to showcase their expertise. Although, all the information mentioned seems promising the interviewee stated that there is ”business value to be willing to share open data, so it is not a goal for the company at the moment”.

According to ESRI Nederland, their focus is not on producing geoinformation, but on the development of software to make it easy for their users to start using GIS. That resulted in the development of GIS software, processing, managing and storing tool, such as ArcGIS and the ESRI cloud that is used from companies that we already mentioned such as Cyclomedia, Geodan, GeoJunxion, Tensing. Although their main goal is not data sharing and open data, they contribute with their current business model. While they maintain services that are directly connected to return on investment, when selling specific high quality datasets and maps, they also developed the Living Atlas of the Earth as geoportal of existing open data from their partners, with the condition to use ESRI software, which it is not fully open, but under conditions. This allows them to have the human and technical resources to develop now business cases that are presented through proposals to the director of ESRI, in relation to their strategy to make data accessible to their clients, and there is a process on whether the new proposal could be beneficial to them. This process is slow and can prevent them from even reviewing potential new business cases, which shows a lack a standarized process. However, they have a specified department of Products and Services that works in researching potential value adding business cases from their existing projects .To enforce that they have a monthly knowledge exchange with presentations for all internal departments, to find potential innovative solutions that could be turned into a business model in the future.

### 7.2.2. Technical

The technical aspect was apparent in all the researched literature and it was connected to the interoperability of data, inconsistency of formats, data availability, data quality and lack of metadata ( [Fassnacht et al., 2023; Waterman et al., 2021]. According to [Janssen et al., 2012; Martin et al., 2013], standarization of methods and the development of robust metadata can increase data access and reuse of data. However, the lack of data standards was mentioned from almost all nine companies that were interviewed in different variations. A framework of data standards could be consisted of variations of data quality, based on the use case of the data [Welle Donker and van Loenen, 2016; van Veenstra and van den Broek, 2013], a clear description of the metadata [van Panhuis et al., 2014], a common language for the collection of data ( [Edelstein and Sane, 2015], and even standardized formats [Martin et al., 2013].

According to the interviewee of Asset Insight, the data standards and the formats that they follow were developed based on the conditions of two of their clients, ProRail and VolkeRail, which happen to be open standards, meaning that the format of their data is a .csv and for their video data in .geotiff. ”Following open formats for theses project, it is not always the

case with all their clients, which effects their processing process", in other projects. Specifically, for their video data, an extra barrier that is created and slows down the processing of the data is the effect of the GDPR. As the video data, could be used to connect a person to a location, it is required to blur the faces in the footage, which could effect thee quality of the data, that are used directly for measurements. According to [Martin et al., 2013; Fassnacht et al., 2023] data processing requires standard data formatting and interoperable infrastructure to be possible.

Additionally, in the case of the track data collected with the Train Sheloc, temporal changes in the data are important for the accuracy of the data and their use, which renders that existence of clear metadata important, or else the data could be rendered unreliable in the validation process and unusable. Even if the company is willing to share some of the data produced, it is technically complicated, as they are not always using open formats, which according to [Fassnacht et al., 2023] can lead to a difficulty of the access and management of from a fragmented data landscape of different databases and systems. This fragmented landscape makes it more difficult, because the software used from Asset Insight is specifically developed for the type of measurements they do with the train, that creates issues of technical compatibility when sharing the raw data [Waterman et al., 2021].

Another, technical challenge that was mentioned by Asset Insight was the massive amount of data, that is impossible to be stored, and maintained by the company, which leads into the storage of data in ESRI. The same issue was mentioned by Cyclomedia, with a database of five petabytes stored in Microsoft Azure cloud. Although, they also mentioned that the lack of data standards and metadata specification created problems in the dissemination of their data in the past, that was resolved after continuous discussion and suggestions from the customers, "to provide a better product". They are following the open standards by OGC, and the API documentation of the OGC because of the convinience that it already exists and can be used to explain to customers easily. The use of open standards, is not common in all projects, as the client usually, can also suggest a different format. The main technical barrier that was mentioned by Cyclomedia was the amount of employs with knowledge about high-accuracy GPS, geometry to achieve 10 cm accuracy, which leads to the rest of the department to relay only on a few people.

Fugro focused on the technical barriers that are connected to the maintenance and management of the data, because of the amount of data, the use of Amazon Cloud is necessary, when their contracts even state that some of their datasets should be maintained for a number of years after the end of the project. If the data were to be shared as open data, that would mean they would be retrieved from the Amazon cloud, which means additional costs for the company. Moreover, technical characteristics of the data that require high data science skills.

SWECO mentioned as one of their previous technical barriers the existance of a closed format that was the shapefile (.shp) in the start of 2000s, to avoid the extra step of exporting to a shapefile, that would slow down their processing, they decided to develop open software that would be compatible with WFS and WMS. They also had issues with the accuracy of the data, when they were using data from Overheid.nl, when the accuracy of a dataset was not always known from the metadata, when also the lack of metadata standards, creates additional delays in the validation of the datasets. The lack of data science skills was also mentioned,as an issue that make the process of validation of the accuracy and reliability of the dataset slower, especially for the Architecture or Marketing employs, that have to turn to the GIS department that is already busy.

The main concern for Imagem was ”how to carefully choose which data standards to support”, choose which of the standards is important for our clientele, because pretty often organizations use different standards, and then the company has to choose. They believe that the value of their software solutions should come from what this software is doing and it should not come from the fact that it is build in to closed data, that people have no other choice to use the software such as a geodatabase or a Revit format they do not support that.

Geodan mentioned that in order to avoid problems with interoperability in processing, they are working with open source software and open standards, and avoid working with vendor lock-in companies. Although, this problem can happen in projects were they have to interact with ESRI software, because they do not follow the OGC standards completely. It was stressed by the interviewee that ”it is not that hard to publish open data, it is difficult to maintain it, latest version, latest data model” which can create problems in technical infrastructures and compatibility.

ESRI maintains a service you can retrieve data you need through WFS about the topographic maps that are provided as open data by the Kadastre, these datasets are not always up to date, so they have problems with the quality of the data provided by the Kadastre. One of the most common barriers is the maintenance of the high quality that the users expect, and constantly keeping them up to date. ”Customers care more for the high quality than paying the small fee” this is always their argument for specific dataset or maps that they provide in exchange for a fee.

### 7.2.3. Legal

Regarding the legal barriers that were identified during the literature review, they were divided into four legal barriers, restrictions by law, which according to [Edelstein and Sane, 2015] the differences in restrictions can create additional barriers to the opening of data that can lead to privacy constraints and ownership constraints. The two barriers that were identified with a focus on the GDPR by [van Veenstra and van den Broek, 2013; Boone and van Loenen, 2022] and in connection with the liability that derives from licensing and copyrights [Janssen et al., 2012; Boone, 2021; Waterman et al., 2021; Open Knowledge Foundation, nd]. Th fourth legal barrier fit better the way that the way and reason the collection of data happened from private companies and that is contractual boundaries, because the collection of data from the private sector was to achieve the success of a project or business model, and the breach of existing contracts can be a barrier to sharing geospatial data [Janssen et al., 2012; Waterman et al., 2021].

The most common answer I got in the beginning of the interviews on what is the reason they are not sharing their data, it was ”we do not own the data, so it is not our decision to make”. Starting from Asset Insight and the departments that collects track geodata, it was stated that ”most times they are not the owner of the data, and are only hired to do the collection and processing of the data”, which does not allow them to make the decision on whether they should share them or not. A similar answer was given by Fugro’s Track Department, Geodan, Tensing and GeoJunxion, limiting their possibilities to share the collected geodata. Imagem went into detail, stating that even when they create ”a new algorithm or model and want to share the working of that model they have to ask the owner if they can use their data to demonstrate where governments are usually fine with, but it is different with commercial companies.

In continuation of issues that arise from the ownership of the data, who is held liable for the data was mentioned as much, from Asset Insight on the use of open datasets from trainspotters for validation, where the company would be help liable for errors. For Fugro liability was a potential barrier in the case that "the data can end up on a search engine where everyone, with or without experience has access to, the company could face a lot of questions". Moreover, GeoJunxion, presented the existence of licenses, terms and conditions as a barriers for them to access and reuse certain datasets "that prevent them from some existing useful datasets from being used, such as the license CC SA (Creative Commons Share Alike), that would force them to share the data that they created as open. Liability was also an issue for GeoJunxion, as it could affect the brand of the company.

Cyclomedia, as a producer of their own data that they license access to others through an API and do not resale, are capable to make the decision of sharing their data more openly, but choosing not to. While, SWECO, is capable of avoiding liability issues with the use of existing open datasets and a connection to PDOK for the development of all their products and services. ESRI appears as a different case as their legal barriers, are connected to the development of their software for the processing of geospatial data. Specifically, all datasets provided on the Living Atlas of the World, a geo portal created by ESRI, are free with the licensing restriction that only users of ESRI software can use them.

Although the existing licensing frameworks (CC, MIT) can be used to create a more interoperable landscape of data sharing they are not used by the private sector as part of their process. Specifically, Cyclomedia stated that they do not use CC licenses, but everything is stated in a contract that both legal teams approve for every project that happens in the company, even for partnerships with other companies. That is also the case with the contracts that they have in providing data to universities such as TU Delft and TU Twente, were the contracts limits the use of data provided, although the data is provided for free.

The boundaries that are created from existing contracts that bound the geospatial data can vary from barriers that are connected to having to make new agreements on the use and sharing of the data with the clients as stated by Geodan that can slow down the process and potentially cost them. To restrictions that specifically state who are the parties that are allowed to share it with or external trusted parties, with specific restrictions on access and use, as mentioned by Fugro, even when it could be beneficial internally. Asset Insight, even mentioned that "there are several contract restrictions that arise that do not allow them to share the data even within the company, between different departments", referring to the project with Prorail and VolkeRail, as Prorail uses the data provided by Asset Insight to check whether Volkerail is doing their part on maintaining the quality of the track, the same time that Volkerail is only a partner department of Asset Insight. The sharing of this data could cost them the renewal of the contract with Prorail.

Moving for the contract boundaries that limit their sharing of data, geospatial data have the characteristic that contain location information that can link a person to a location and that on its own created privacy concerns. Geodan mentioned that in order to share the data they have on location of people, "as privacy is an important issue in the Netherlands, they have to anonymize it", which creates delays in the proccessing and validation of the data. Furthermore, Asset Insight, on the data collected with the Train Sheloc, they cannot share the raw data, that could pottentially used in other use cases, because they contain video data, that create privacy concerns. For the EU the GDPR is in force to protect these privacy concerns, which translates to having to blur faces in the video data for Asset Insight, that can potentially compromise the high quality of the data needed for further measurements and processing.

### 7.2.4. Economic

According to [Waterman et al., 2021], economic barriers are usually connected to commercial losses. Through the literature review there were four economic barriers identified, fear of economic damage connected to brand reputation, and liability from data errors [Martin et al., 2013; Waterman et al., 2021; Barry and Bannister, 2014], loss of income from change in the business model because it is not connected to a cost recovery policy [Boone and van Loenen, 2022; Barry and Bannister, 2014; Janssen et al., 2012], uncertainty about the value of data as intangible assets and lack of successfully business cases [Fassnacht et al., 2023; Barry and Bannister, 2014], lack of resources human or technical connected to collection, maintenance, processing, validation and distribution of data [van Panhuis et al., 2014; Edelstein and Sane, 2015; Boone and van Loenen, 2022].

As they are private companies, maintaining their business model was top priority, that often seemed that there were no alternative options researched with future data sharing in mind. For Asset Insight, sharing data as open, means that they lose their commercial aspect, and it is not in the interest of the company to share data because "information is knowledge, knowledge is money", which shows that the value that could be generated through sharing of data in not clear to them. In the same manner, Cyclomedia believes that their current business model is "pretty successful in the Netherlands and they have no intention to share their data as open", because that could potentially cause loss of income. Whereas, Geodan perceives open data as an important trend for now, but not as a long term business goal that could create value for their data, because "they are a private company so their primary goal is money". For ESRI, whereas they understand the value of data through data sharing, there is a lack of standardized successful business cases to follow. That makes the process slower as a consideration is being done every time based on whether there is a market value or financial value to proceed and whether it would affect their partners.

Other than the lack of knowledge about the potential value from data sharing, the focus of the economic barriers was turned to maintaining their business model. Asset Insight, mentioned that sharing their data completely open would not be valuable to them as it could affect their current business model of a contract for the collection and processing of data with ProRail that has a direct return on investment. Cyclomedia, suggested that in order for them to change their current business model that is successful of licensing access to their collected data, to avoid income loss they would need financial support, from the government or another organization in the form of funding.

ESRI appears to have a more complex business model that allows them to share geodata as open data, but their software has to be used to maintain it, which creates other technical interoperability issues. In that way, they have the option to avoid loss of income, and develop new products considering the "man hours" used in the development, according to how technically complex the product may be, and considering the cost of maintenance, making data sharing possible that way, under conditions. In the meantime, ESRI maintains a service of selling some data they produce for a fee, such as orthophoto, aerial imagery that they are provided with a yearly subscription to get access to a dataset or a paper model, in order to have the income for the development of other products that may not be successful.

Sharing geospatial data is not always part of the strategy for a lot of the companies, when that becomes a possibility during the project it creates additional issues. These might be connected to their lack of human resources, such as support for the new open software, new open format needed and technical resources of the interoperability of the open software

and data standards with the existing technical resources. For Asset Insight and Geodan that means delays in the processing of the data, because of the lack of the needed software, that can lead to a delay on the delivery of the data to the costumed and potential loss of income. It also means, lack of human resources that would be needed in the case of on of their datasets becoming open, and needing extra personnel to keep up with the constant feedback and questions of the users, that would result in the delay of other projects that are connected to a business model. For Fugro, the maintenance of the datasets after the end of a project, imposes extra costs that are connected to the use of Amazon Cloud to store them, and in the event of the dataset shared as open, that means for them that every time a dataset is retrieved from the Cloud Fugro has to pay a fee to Amazon, which is small, but it would be multiplied by the users.

Imagem at the time of the interview focused that only data that are created with public investments should be public data, that is their understanding of open data. For that to happen there should be framework for anyone that put investment in collecting, processing and maintaining the data to make sure that there is a return on investment. According to the interviewee ”efforts should be linked to a business model” and ”it is not fair to strive for open data in all cases, because there are companies that have a lot of time a lot of energy and a lot of financial means to create specialized data. To get into specific economic issues that might arise, the interviewee mentioned the difficulty they have choosing the right data standards to follow, even when choosing to use open standards, there are a lot of different frameworks that take time to create the standard, create support for it, test it, put it through approval and retest it, so it is not viable for them to support all standards, even when a client has a specific request. At times, the change in data standards could cost them investment from a client with specific data standards needs. To quote them ”if our resources were infinite we would support all standards”.

### 7.2.5. Cultural

The cultural barriers identified in the literature review were focused five types. The cultural differences that are the result of historic restriction or openness of data within an organization [Edelstein and Sane, 2015; Barry and Bannister, 2014; Fassnacht et al., 2023], the unwillingness to share data that derives for a culture of risk aversion of a company for competitive reasons [Edelstein and Sane, 2015; Fassnacht et al., 2023], the lack of trust in the usage and further processing of the data, that is connected to fear of inappropriate use and misinterpretation of data [Waterman et al., 2021; van Veenstra and van den Broek, 2013; Boone and van Loenen, 2022], fear of loss of control connected to loss of power that information holds or lack of knowledge about data potential [Fassnacht et al., 2023; Barry and Bannister, 2014] and fear of transparency related to confidential knowledge [van Veenstra and van den Broek, 2013; Fassnacht et al., 2023].

The barriers that were mentioned during the interviews about the culture of the company were limited. As Imagem and Geodan both argued that because of the high level of open culture in the company there is no need to persuade the internal departments about the importance and the benefits of data sharing and open data. On the other hand, Fugro that if their datasets are shared as open data through a general search engine, then ”the dataset is out of the control of the company”, and that creates a fear of misuse of the data from third parties, when their geodata skills are not known. Additionally, GeoJunxion, referring to the streetmap of the Netherlands that was donated to OSM, that they were worried about its misuse, as the skills of the user are not known and the end user is not known, while

| Barriers per company interview identified | | | | | |
|---|---|---|---|---|---|
| **Company** | **Strategic** | **Technical** | **Legal** | **Economic** | **Cultural** |
| Asset Insight | ✓ | ✓ | ✓ | ✓ | |
| Cyclomedia | ✓ | ✓ | ✓ | ✓ | |
| Fugro NL | ✓ | ✓ | ✓ | ✓ | |
| SWECO | | ✓ | | | |
| Tensing | | | ✓ | | ✓ |
| Imagem NL | | ✓ | ✓ | ✓ | ✓ |
| Geodan | ✓ | ✓ | ✓ | ✓ | ✓ |
| ESRI Nederland | | | ✓ | | |
| GeoJunxion | ✓ | | ✓ | | ✓ |

Table 7.1.: Barriers identified during interviews per company

the company can be help liable for the dataset. Asset Insight, also refused to share some of its raw data, in fear the competitors will understand how their model works and this is confidential knowledge.

### 7.2.6. Barriers per company

Regardless, of the company as it is visible in Table 7.1 strategic barriers are evident as 6 out of the 9 interviewees mentioned barriers that could be described as strategic based on the literature review of Section 2. Specifically, SWECO, mentioned that they do not have a corporate strategy regarding data sharing, which according to van Veenstra and van den Broek (2013) leads to project-based data sharing from individual department. In the case of SWECO, it leads to the development of specific products, such as the Staat van je Straat, by the GIS department, without a strategy to use the benefits of this project to the rest of the company. Additionally, according to Fassnacht et al. (2023), this lack of corporate strategy, can result to unclear responsibilities for the manager of this specific department that if they lack the will to commit to this specific project and scale its benefits of data sharing to a long-term strategy, the value of the data stops at the end of the project.

In the process of the interview all 9 companies were asked if they have a strategy for data sharing, to have a better understanding on how the strategic barriers would be visible in the different levels of openness. From the 9 companies, Microsoft, Fugro, from the Fugro Americas interview and Imagem mentioned the achievement of their sustainability goals as the driver for the geodata sharing, and actions taken to get insights about biodiversity and climate change. ESRI Nederland, mentioned as their strategy to provide as much datasets possible for their customers, that are users of their software, that was the drive for the creation of the ”Living Atlas”. SWECO, mentioned that their strategy is connected to them being a user of open data and helping others make their data open, similar to Geodan that has a goal to help others with making their data open through consulting. Tensing and GeoJunxion, had a similar strategy of collection of information about open data and it trends and translating that into useful knowledge that could help in the development of new products, based of the open data trends at the moment, with the examples of Digital Twins and Environmental Zone alerts. Although, all those are not clear strategies with a goal to share geospatial data as open data, they are all involved in a way with open data, helping in parallel with their data sharing.

On the other hand, when the same question of the companies strategy about data sharing was asked to Asset Insight, Cyclomedia and Fugro NL (track department), they did not mentioned a strategy related to data sharing or open data, their goal was to maintain their business model for Cyclomedia, and to be successful in their current project to achieve contract renewal for Asset Insight and Fugro NL, as they are both working in collecting track data for Prorail. This lack of strategy led to lack of use case identification as mentioned by van Veenstra and van den Broerk (2013) and Fassnacht et al. (2023), were they are not able to systematically identify other use cases for their data, and even if they manage to do that, they do not have a framework for the whole company to scale those case studies to create an alternative business model for the future, after the end of the contracts that maintain their business model right now.

A common barrier for all 9 companies, although it is not directly identified, was the lack of a policy coherence within each company. This lack of policy coherence was implied by the repeating technical, legal and economic barriers mentioned from the companies when referring to different projects connected to geospatial data sharing. According to Janssen et al. (2012), this lack of policy could lead to technical barriers regarding standards, which was in reality on of the two most common indications next to the lack of a licensing and conditions framework. The data providers companies (Asset Insight, Cyclomedia, Fugro), has requirements both data standards and licences on a contract basis, which could change from a project to a different project, according to the client and they are not uniform even within the same department in some cases, such as Asset Insight. The lack of a licensing framework was also mentioned by Tensing, Geodan, SWECO, Esri Nederland, as they change on a contract basis.

Those technical and legal barriers that are a result of the lack of a policy coherence were mentioned in detail about following open standards, without specifying the standards that they are following, or in some cases not even sure if those exist specifically for data, or only software. Through the Microsoft case study, it was presented that there are a lot of layers of licences and conditions for data, code, product and not a framework for them to specify the whole process. Imagem NL, specified that it often an issue for them to decide what open data standards, to follow as their clients have different needs, sometimes really specified to their use case, NATO standards, marine standards. A solution to that was given in the Fugro bathymetry data case study with the already established GEBCO database. Although, Imagem NL, stated that this type of specific database can also lead to vendor-locked usage, like the software provided by ESRI. While, it was also, mentioned from Geodan, that in their process of contributing to the OGC standards, it not always possible to be interoperable with ESRI, as they are not following fully the OGC data standards.

This lack of a framework of data standards, is common in the private sector were data has been extensively collected without a framework in mind that connects to their purpose of use/reuse. This lack of standardization can lead to issues in access and reuse of data, as their data standards are only specified to the project at hand. As an example described by Imagem NL, there are specific marine data standards, or NATO standards that are important for the Ministry of Defense, but not for a local Municipality. Connected to the lack of accessibility and use of data that comes from the lack of data standards, that can also affect the technical infrastructure that is needed for the maintenance and management of those standards, that can be the needed software and hardware to support it (Tenopir et al., 2011; Boone and van Loenen, 2022). Tensing, in their example of Municipality, mentioned the lack of maintenance of the dataset after the end of the dataset that was connected to their data standards and

software used, and the lack of the needed software and data science skills in the Municipality to maintain the dataset.

Fugro mentioned another barrier in the management and maintenance of datasets after the end of a project as an economic barrier connected to the use of Amazon cloud, and the bills they have to pay for the maintenance, that is connected to the storage that is needed. This was described from van Panhuis (2014) and Sane and Edelstein (2015) as an economic barrier of lack of resources, whether they are human or technical. Additionally, Boone and van Loenen (2022), mentioned that resources could be connected to all stages of collection, maintenance, processing, validation and distribution of data. Imagem NL, mentioned the importance of resources to create and support specific data standards in all stages described by Boone and van Loenen (2022), while Microsoft through the development of the Planetary Computer, mentioned the importance of their collaboration with ESRI, for the development of Hundreds of Ready-to-Use Analysis Tools for geospatial data.

As it was described so far, most existing barriers are connected to strategic, technical, legal and economic barriers, in the process of the interview when asked about moving to level 3 in the future the most common barriers mentioned were cultural. The lack of trust in the usage of the shared data was mentioned by GeoJunxion, as a result of the donation of the streetmap of the Netherlands to OSM. This is a barrier that was mentioned in the literature review by Sane and Edelstein (2015) as a fear of inappropriate use that can lead to lack of trust. The main concern of GeoJunxion was that the donation of the streetmap could cause reputational and social harm for the company if any errors were detected (Waterman et al., 2021; van Veenstra and van den Broek, 2013). Asset Insight mentioned the lack of trust in the usage of data as a fear that would interfere with their core business, which is the collection, processing and dissemination of data and the economic harm that could result in the errors in the shared raw data (Fassnacht et al., 2023; Waterman et al., 2021).

Asset Insight elaborated on the lack of trust as a barrier in relation to the disclosure of competitive knowledge through data sharing, as it was described with their example of sensor data. This example refers to the creation of a model from their sensor data, that they can later do calculations with to have a more automated process with less resources needed. However, sharing the raw data of that model with the rest of the world shows how this model is working and another company could use it for their own business model (Fassnacht et al., 2023; van Veenstra and van den Broek, 2013). On the other hand, GeoJunxion connected the lack of trust in the usage of data to a loss of control over them that comes from the reuse of data, from people with not the right set of skills, that can create errors in the processing of the data, and GeoJunxion would not be able to control those errors after the reuse of the dataset (Barry and Bannister, 2013; Janseen et al., 2012).

All the cultural barriers mentioned led to an unwillingness to share data, that was connected to a culture of risk aversion from most companies, because the existing business model works and changes can affect it (Martin et al., 2013; Sane and Edelstein, 2015). When faces with the future scenario of moving to level 3, Imagem mentioned that they do not expect new barriers to arise, other than the existing technical and economic, because of the culture of the organization, they will not have to persuade any employ that is part of the company on the importance of open data and their benefits, in a way that was the most important barrier to address.

## 7.3. Levels

In order to address the barriers to reach open data sharing, it was important to create a schema that assesses the levels of openness of the companies based on characteristics that relate to the identified barriers through the literature review. The schema that was created was based on three assessment models from the technical perspective (Berners-Lee, 2009), the organizational perspective (ODI, nd) and the holistic perspective (Boone, 2021), with the addition of two layers for the culture and the strategy of the company for open data and the addition of level 0, because of the competition in the private sector data sharing may not be possible even outside the completion of a project.

The strategic layer as described in Section 2.4 it looks blank it the schema, but during the interview a specification was given about the four levels, level 0 being no knowledge of external open datasets available, level 1 being acknowledging the existence of available open datasets, but still not used in the process of the company, level 2 having an open data ambassador or department which role is to inform other departments about new opportunities, and level 3 being that the company has a clear open data strategy that they implementing with allocation to other departments. This specification was not part of the model in the beginning of the interviews, but it was developed in the process of conducting the interviews, using the input of the interviewees on the strategy of each company for data sharing.

During the interview, each interviewee was asked to identify the level of their company in the schema in Figure 2.4, and describe their reason for choosing that level. Although, all companies were approached for an interview with the same schema in Figure 2.4, it became clear during the interview that is more complicated to determine the level of openness in isolation and that there might be other factors that influence the result. For each company this level will be presented next to their reasoning, and then a new level is revealed based on the analysis of the interviews based on the schema and the barriers that were the result of the literature review.

Starting with Asset Insight, they identified the company to level 0-2 depending of the project they are referring to. Level 0 for projects that are only developed internally for the creation of a new service, and the dataset will not be used for other causes, to level 1 for data that were collected for the fulfillment of a previous contract and they were given permission to use internally to create extra services from them, to level 2 which is the collection of data with the train Sheloc for Prorail, or other public organizations, were the dataset have to be shared on the portal of the organization as open data at the end of the project, but that is the decision of the owner to share them. It was clear that the sharing of data even internally was bounded by different contracts, and Asset Insight that was not the owner of the data was not allowed to share them, in some cases even with other departments, as it was the case with the Prorail contract. The internal use of the data was only allowed with a specific agreement, for the development of extra services for the client, and rarely for the development of Asset Insight. According to their current state they are somewhere between level 1 and level 2 of the model, has the culture, strategy and share of level 1 and regime, quality, users, find and play of level 2 excluding that data are not found in a general search engine, there is a fee according to the contract and the data are downloadable in bulk. As for their culture and strategy they remain in level 1 because they are allowed to share data to get benefits as a department and are aware of existing open datasets, but in the case of Prorail and Volkerail the culture is at level 0 as they are competitors, but those are not part of their process, only used individually for verification (open dataset trainspotters).

Cyclomedia, identified themselves in level 0 and 1 depending on the product. The have a Research and Development department that is consisted of all the GIS departments and there is free data sharing internally between them, whereas the level 1 according to the interviewee, specifically data that are collected by them and they license access to external users, are free for their processing department to develop new products. Access to those data have some external partners such as Centric for free. The follow OGC standards for their API service, but also use ESRI cloud storage and software for processing, which is not considered entirely open standards. From their current situation they are mostly on level 1 for all categories except play which is in level 3 with the modification that the data are not free of charge and are not always open standards, additionally the data are findable in a data portal, in which internal and external trusted users have access to. As for their culture and strategy they remain in level 1 because they are allowed to share data to get benefits as a department and are aware of existing open datasets, but those are not part of their process, only used individually for verification of their quality, because they collect very similar data, the example was the AHN3 dataset.

Fugro described their level of openness as level 2, because their data are shared internally and sometimes with external trusted parties. There is an internal geodata portal with data of the company, that users have to ask for access, that puts Fugro at level 0, when there is a project that data can not be shared even internally with other departments for security reasons, as they are high quality track data. They follow open standards, when they are working with public organizations, but this may not always be the case as the client specifies the data standards for the project. According to the model, Fugro is at level 1 for culture, strategy, and share, everything legal is stated in the contract of the project. For regime, quality, users, find and play they are at level 2 excluding the findable in general search engine and that the data are not for free. Based on those characteristics Fugro is in between level 1 and 2.

SWECO was identified by the interviewee level 2 or level 3 depending on the product or service. As they are working with municipalities or government agencies, they are always working with open data, mainly from PDOK and other provided by their clients. Their goal is to create applications with those, meaning that they are following open standards, mainly with WFS, a format that they developed in collaboration with ESRI, but they license yearly access to the products they create. According to their current state the are following all the categories of level 2, except the play category that has the characteristics of level 3, without the free of charge data and although they use licences with conditions and terms (CC), this is not always the case, because for sometimes they create contracts for specific access. For those reasons they are closer to level 2 of openness. Additionally, the GIS department appears to be the open data ambassador of the company, with the role to inform and educate about open geodata availability and trends.

Tensing is a consultancy agency that provides services that involve creating a geodata strategy for another organisation private or public with the design of software and turning their location data into useful information. The interviewee had a difficulty identifying Tensing in the schema provided, because "they are not a data provider" and they do not fit in the model with their characteristics. Their role is to be hired by other organizations to find solutions, so they do not even have data that are their own to decide to share, or a specific application to share. The interviewee tried to identify Tensing on the schema as a user of open data, at level 2. Level 2 with the current situation is viable, because they encouraged to share data, but there is a small process to go through hierarchy, and they have a geodata ambassador for every team (ArcGIS, FME, data science), that has the role of researching new trends, existing

geospatial datasets and informing the team. As a user of data and developer of software solutions, they use open standards for the formats, metadata, and quality of the data, but their software is not always open source, both FME and ArcGIS need a subscription. Also, they are not using licenses associated with terms and conditions, and everything is stated in the contract of the project, regarding legal issues, which is a level 1 characteristic.

Imagem tries to solve complex geospatial problems, with the production of datasets through collection, processing, analysis and management of geospatial data. For that reason, Imagem is not the owner of the data in the projects and does not provide raw data, after the end of a project the processed data also go back to the owner and its their decision to share them or not. For the interviewee it was difficult to identify Imagem on the schema, because they do not provide data. They are mainly a software development company, and a strong advocate of open data, as they contribute in the OGC standards, but for them for data it is not always the goal to go fully open. From their current situation, they are a user of open data in their processes, so they are encouraged to share data if they have an agreement from the owner to showcase new open source software and the GIS department is responsible as an ambassador of open geodata. Both culture and strategy are at level 2, the play category is at level 3, except that their data are not for free, only if the owner decides to share them, and the rest of the categories are in level 2 except of the share, because they continue to draft contracts for every project, and the interviewee had limited knowledge on the legal matter. For these reasons, Imagem is mostly in level 2.

Geodan specializes in location intelligence by combining data and providing insights, so they do not collect data that they share. The interviewee had a difficulty identifying the company in the schema, because "they do not own data" and act as a consultancy inform other about the benefits of open data, and how to make their open data, these organization could be public like municipalities or private like an electricity company that wanted to learn about the benefits of sharing their data as open data. They are a strong advocate of open source and open standards, they are contributing to the development of the OGC standards, and avoid working with vendor lock-in companies like ESRI, because it creates issues with interoperability. From their current situation, they are mostly on level 2 of the schema with the exception that for play they are in level 3 and their data are not for free and for share they are still in level 1 as they are using contracts for all requirements of every project and not a licensing framework.

ESRI Nederland is a GIS software company, known for its ArcGIS products. The interviewee when presented with the schema identified ESRI Nederland in level 2 and 3 depending on the dataset. According to them, ESRI has three categories, completely open, open for ESRI customers and in exchange for a fee. They are developing their software based on OGC standards, but they are not following them entirely. They are at level 3 only for the postal code zones and the service to retrieve data from the topographic maps of Kadastre. For the Living Atlas of the Earth they are at level 2 as it is restricted only for ESRI clients. For the culture and the strategy of the company they are at level 2, because as described from the interviewee when the want to share data as open data they have to make a proposal to the ESRI director that will go through review. Their strategy is to make it possible for more people to use GIS software, so open data is not directly a goal, but they are maintaining a Product and Services department that acts as an open geodata ambassador for the company.

GeoJunxion's business model is connected to location intelligence services, were clients hire them to create specific datasets. The interviewee when presented with the schema, they were confused about its intention "Is the goal for the company to become open or to use

| Levels identified in interviews | | |
|---|---|---|
| **Company** | **Level by company** | **Level according to the model** |
| Asset Insight | 0-2 depends on project | 1 -2 in between |
| Cyclomedia | 1 and 2 depends on project | 1 |
| Fugro NL | 2 depends on project | 1-2 in between |
| SWECO | 2-3 depends on product/service | 2 |
| Tensing | no level as producer / 2 as user | 2 |
| Imagem NL | no level | 2 |
| Geodan | no level | 2 |
| ESRI Nederland | 2 and 3 | 2 |
| GeoJunxion | 3 OSM, 2 basemap, 0 and 1 | 2 |

Table 7.2.: Levels identified in interviews

open data?" They identified the company on different levels depending on the dataset. For the donated streetmap to OSM, they are at level 3, as it follows all categories of level 3 except that its license is not CC0, but CC BY. For a basemap that were maintaining and licensing access to municipalities and other companies until the rebranding of the company in 2020, they are at level 2, except that it was in exchange of a fee and all legal matters were stated in a contract. For their current businness model they use open datasets that allow it based on their license, terms and conditions, to create products that will be at level 0 or 1, as they do not know if they are shared after they give them to the clients. They have specifically a Sourcing Department with the role to research the available datasets for their clients, whether they are open data or commercial data, and informing them of the license restrictions on use. Based on that the culture of the company is at level 1 as they are allowed to use open data for the benefits of the department, but their strategy appears to be at level 2, with the role of the Sourcing Department to act as a geodata ambassador. All this, renders it really difficult to put in one level of openness, but they fit towards level 2 as they interact with external trusted parties, for the quality of the data, the existance of a geoportal where external parties have also access to, and the play characteristics of level 2.

Through the multiple interviews, and the analysis of them it is apparent that it is difficult to identify the level of a company as a whole. That was visible, through the confussion of the companies that do not specifically own geospatial data that they can provide through data sharing as open data. While the model was identified both from GeoJunxion and Tensing as data provider driven, not taking into account other roles that the companies may have in the data sharing process. Table 7.2 showcases, an initial attempt to categorize the companies based on their levels, with the identifications that the companies gave for themselves in one column and the result that was given after the analysis of the interviews and the characteristics of the model. Most of the companies as described in the previous paragraphs are in between levels, showcasing the need for in between levels in the model, and in the level by me column of table 7.2 it shows the level that had the most common characteristics.

This section was related to answering the subquestion "What are the different levels of openness for geospatial data?" and we see that for the most part companies like SWECO, Tensing, ESRI, GeoJunxion, Geodan, Imagem NL are at level 2 (Fig. 7.1) of the schema engaging in data sharing internally and with external trusted parties, while attempting to use open standards or at least common standards in exchange for a fee. It is difficult though

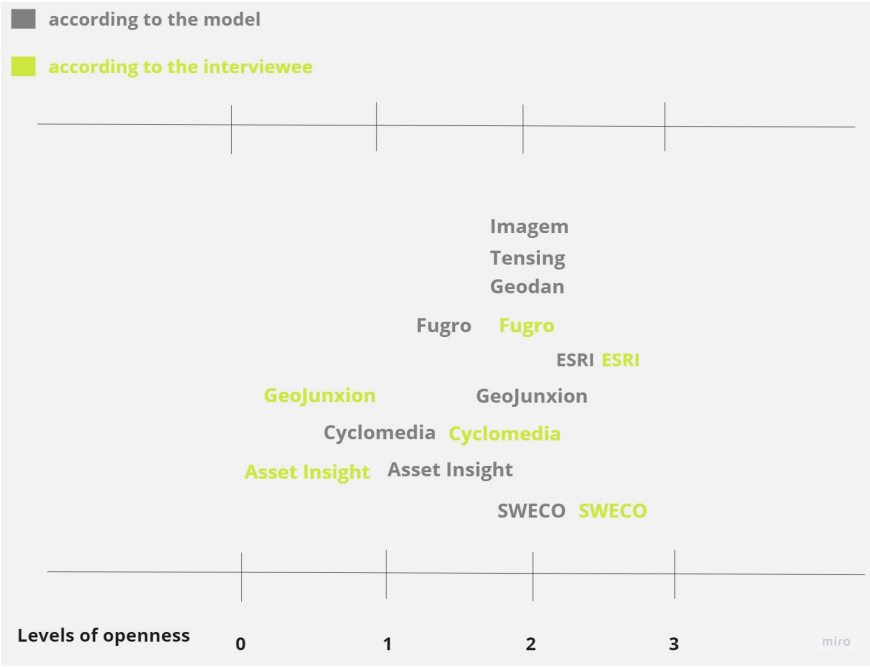


Figure 7.1.: Level of openness of each company according to the model (a) and the interviewee (b)

to really compare all the companies that are in level 2 between them because during the interview it became visible that the role of the company in data sharing and the level of a dataset, a project or a department can change the level of openness that is what will be discussed further in the next section.

## 7.4. Role of the company, Dataset Project Department

### 7.4.1. Role of the company

In the process of the interviews, although all interviewees were presented with the same model, there was a difficulty from some of them to identify the company in the schema. After a conducting a number of interviews this started to became a pattern, with the variety of the companies presented.

Specifically, it is visible in Figure 7.1 where 3 companies are missing from the graph when trying to identify the level of openness of the company. Tensing, Imagem and Geodan did not manage to identify the level of openness of the company. Their reasoning was that the existing model used in the interview (Fig. 2.4) was too provider driven, which meant it would be more suitable for companies that collect and disseminate their own data, which was not the case for these three companies. The interviewee from GeoJunxion, that also had difficulty identifying the company in the schema asked if the schema was meant to be used on ”How can a company provide open data or if they use open data”.

That contributed in questioning "if the role of the company in geospatial data could be a factor that affects the level of the company and the barriers that they face in that stage". There is a correlation between the role of the company in data sharing (provider, intermediary, user) and the level that they identified in the schema of the interview Figure 2.4. The three companies identified as providers, described datasets that are in level 0, 1 mostly and occasionally in level 2, with public agencies or the government as the external parties they are sharing their data with. On the other hand, companies that were identified in the interviews as intermediaries, mentioned that their datasets are in level 2 and 3, with the exception of Imagem NL.

While conducting the interviews there were three roles in geodata sharing that appeared for the private sector and those were data provider (Asset Insight, Cyclomedia, Fugro NL), data user (SWECO, Tensing, Microsoft, Imagem NL, Geodan, GeoJunxion, Esri Nederland) and intermediaries that help others make their data open, either by creating applications, services and software that supports it (ESRI Nederland, Imagem NL, SWECO, Geodan, Tensing, GeoJunxion, Microsoft).

The data providers are responsible for the collection and the dissemination of data to their clients, but it is up to their clients as owners to decide whether the data will be shared and under which conditions. The clients not being aware of the value of their own data, could refer to other companies with that knowledge to help them, by consulting them on the different barriers that they may encounter and how to proceed from there. This is where the connection with the other companies to the providers becomes meaningful, to lead to the access and use of data through enhanced solutions.

The intermediaries are companies that provide solutions to technical, legal or economic issues that may arise in the process of data sharing. This could be open source software, for processing, management, cloud storage (ESRI, Imagem), legal consultancy on licenses and conditions (GeoJunxion, Geodan), and creators of application and products (SWECO, Tensing). The last role that a company can have is that of a user of the available open geodata. In Figure 7.2 the levels based on if they are a user of open data are presented. Also, it is important to clarify that the categorization of the level as a user was mainly done with the culture and strategy layer of the schema (Fig. 2.4). In that way all companies are visible in the schema, but still there is a difference in the level that is identified as their role with Asset Insight being in level 0 and level 1. GeoJunxion in level 0 and level 2, Cyclomedia in level 1 and level 2, SWECO in level 2 and level 3, Fugro in level 1 and level 2.

That means that they also have different barriers that they have to overcome to move to the next level according to their role. As it was already discussed in Section 7.2.6 the three companies identified as providers appear to have more barriers to overcome to move to the next level, and they are mostly on level 1 going towards level 2. They have to solve technical barriers connected to data standards, interoperable software for collection, processing, maintenance, management and the educate their employees on the data skill needed. Also, solve their legal issues, with the use of a license framework and not with the creation of new contracts with different requirements.

To do all that they need a good allocation of their human and technical resources to avoid economic barriers and get value out of the data in a new business model that is not only grounded in return of investment. A common barrier that connects all the previously mentioned is the lack of a clear strategy about data sharing that creates those fragmented barriers, it is important to create on, even a department that would be the geodata ambassador of the company. It is often common with these companies that there is no will to contribute

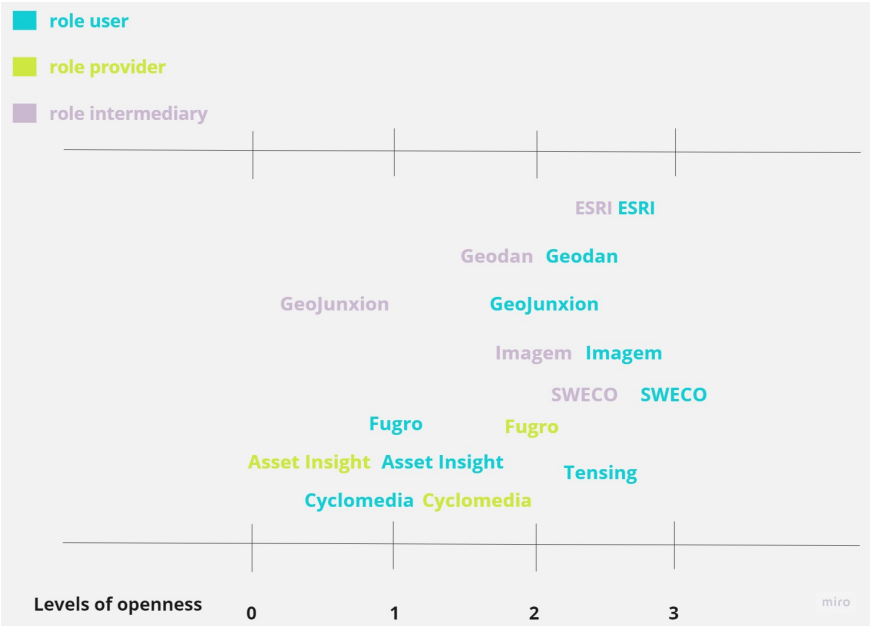


Figure 7.2.: Level of openness according to the model based on the role of the company in geospatial data sharing

to data sharing as they have all other barriers to solve first, but just to fulfill their contract and for that to happen, there is further education needed by the government an public institutions about the value of geospatial data with successful case studies.

The other 6 companies were identified as intermediaries during the interviews and they appear to be mostly on level 2 or in between level 2 and level 3. Although, these companies do not have a clear open data strategy, they have strategies that contribute in a way in data sharing. This allows them to better allocate their tasks and resources and as it is visible in Table 7.1 they identified less technical, legal and economic barriers. The existence of a strategy that contributes to data sharing already shows a high level of open culture within the company that drives those changes. Geodan is a little further back than the rest, because they are still more contract-project oriented and they do not have a geodata ambassador or a department that actively follows and informs the company about open data trends. The legal barriers were visible in all company interview except SWECO that uses only already existing open datasets, to avoid liability issues. All other companies are still working on a contract basis for most of their projects, without using the existing license frameworks of terms and conditions, and most of them are not even aware of their existence. It is important to follow existing licensing frameworks to create a more interoperable landscape internally and with external trusted parties, because most companies on the the share part of the schema are still in level 1.

Referring to the companies as users of existing open data can only be approached from the level of strategy and culture in the schema as the rest are provider or intermediary driven. The sections of Find, Play and Share of the schema in Figure 2.4 are criteria that connected to sharing the data of a provider as open data, or sharing products and services based on licenses and conditions, for a fee, that have specific data standards. The only interpretation

of the schema as a user is on whether your technical and legal interoperability is possible according to each level. This could mean, how prepared is your software to support open data standards, or how the contract restrictions within the company allow the use of open data for validation, processing or even to develop a new product or service. In that way the use of open data is connected mostly with open culture of the company and if they have a strategy for data sharing and using open standards, then this translates to the lower levels with high technical and legal interoperability preparation.

The different role of the company translates into different categories of barriers that the company is struggling with as explained above, but also means that the company can be categorized in a different level of openness according to their role, and a company can have a double role as a provider and user or as a intermediary and user and that complicates their categorization, because they can be in two different levels at the same time.

### 7.4.2. Is the Data, the Project or the Department open

Another element that affects the openness of data sharing of geo data is whether it is dataset-based, project-based or department based. According to Fugro NL and GeoJunxion, the level of data sharing could be project based for level 1 and level 2, because it is the decision of the owner of the data, the client of that project, and the specific conditions that are stated in the contract of the project. On the other hand it is possible to reach level 3 on a dataset basis, with the rest of the datasets remaining in other levels, a clear example of these are the three case studies of Chapter 5, in which each company for their own reason each time decided to share only this specific dataset as open data, without the company having a clear vision or a strategy towards sharing open data for the rest of their datasets. This stages are not linear, but are connected with the technical characteristics of a dataset, that are connected to technical difficulty that can be solved to move to even level 3 for that specific dataset, if it is not the core income of the company, as an example the bathymetry data of Fugro.

It can be project based as the project for a Municipality that Tensing was working on. Its initial goal was open the product to the municipality, not to the public, while using both existing open datasets and a dataset of the Municipality that was not open yet, to create a product for the identification of vegetation zones in between agricultural land. The open end product made the second dataset open by default, although this was not the intention when it was created. When referring to a department-based openness, the example refers Asset Insight, with the Track department responsible for the collection of data at level 1 and the Analytics department of the same being at level 2. Cyclomedia has also a similar case where the Data Acquisition department is at level 1 and is not allowed to share externally, while the Analytics department is at level 2, and shares data with external users. That shows that there could be a difference in the level of openness between different datasets, projects or departments, without it guaranteeing that the whole company is on the same level of openness.

## 7.5. Recommendations to approach barriers to reach open data sharing

As mentioned before because of the complexity of the current situation, there are different options on how to proceed, depending on the role in the data sharing system and the level of openness that you are currently in towards the level that you want to reach as described in Section 7.4. Even for intermediaries that are on a higher level of openness in the schema among other barriers they need further motivation in changing their business model to add value to their data, which is why a collection of successful case studies like the ones mentioned in Chapter 5, would be useful to show the benefits from data sharing, even if it is not the goal of the entire company, but some of their datasets or projects. Imagem could use the suggestion from Fugro about showcasing their technology of OARS, if they could find a dataset that was collected during other projects and they are the owner, to use it as a showcase. Tensing and Geojunxion could open a successful dataset as open that could be reused with the license CC By, which would mean that their company would be advertised and mentioned constantly. Those are companies that were identified as intermediaries in level 2 and they are closer in reaching level 3.

### 7.5.1. Recommendations for the government

For the identified providers that are closer to reaching level 2 of the schema, to achieve data sharing with external trusted parties, first there are more technical, legal and economic barriers that need to be resolved. To solve the problem of data standards and lack of use of licences with terms and conditions the European Interoperability Framework (EIF) [European Commission, nd], could be an opportunity for the government in collaboration with the private sector to explore synergies on existing open data standards. This could create a Legal and a Technical Interoperability layer of the framework that was created for public administration that takes into account the needs of the private sector. This could be used to ensure coherence between the different companies, by screening their existing contract and licensing and trying to integrate them into a framework, in that way the framework will be developed with the needs of the private sector in mind, rather than forcing existing frameworks that might not be working. Also, that could help clear the different licensing frameworks that were developed historically and might be overlapping. Something similar can be done also for the Technical Interoperability, although it is more costly to change ICT systems that were developed historically, it might be useful to create a catalogue of the available open data standards, with specifications on which every company should invest based on their characteristics and role in the data sharing system. This is also an opportunity to educate the private sector on what open data is with a clear description because for Tensing open data is free data, for ESRI the Living Atlas of the Earth is open data, for GeoJunxion anything with a CC licence is open data.

In continuation of the recommendations that the government or public administrations could do, during the interviews a common barrier that was mentioned was the lack of data science skills or all employees with the needed data science skill being only in one department of the whole company. That created delays in the process of validation and processing of geospatial datasets, referring to the users of existing open data, like Architects, Designers or Marketing. There was a suggestion from the interviewee from SWECO to create a 5 star system that is visible in the geoportal as part of the metadata of the dataset,

to help even users with limited data science skills in the reuse of existing open datasets. That could also be the economic barriers of companies that had problems with their lack of human resources.

### 7.5.2. Recommendations for an improved model

At the moment it is not mandatory for the private sector data to share their data as open data, that is why this research was also an opportunity to see what the private sector wants, or under which conditions to participate in data sharing. It was stated both from ESRI Nederland and Imagem NL that it is not a linear process to move a dataset from level 0 to level 3, which is why the current model does not fit in the way a dataset is shared as open data. Every dataset according to its characteristics, may have a different reason that was created or a different goal on the conditions of how it should be shared. It was noted that some datasets should remain at level 0 for security reasons, those could be datasets that where created for the Ministry of Defense and could affect National security, as well as raw datasets of the track system in the Netherlands could be prone to affect National security, according to Asset Insight, the measurements are so detailed that could show the weak points of the track system.

#### Culture and Strategy layer

To proceed with some of the recommendations changes need to be made in the schema of the multi-model that was used during the interviews and the analysis of the results to identify the levels of openness. As it was mentioned before the existing assessment model focused on technical, legal and economic barriers. After the end of this research and the interviews with the private sector there are some suggestions for the layer of culture and strategy that were left empty on purpose so the private sector can give their opinion more freely. The strategy layer in Figure 7.3, was created based on the examples that are active right now from ESRI, Tensing, SWECO and GeoJunxion. ESRI has the Department of Products and Service, with the responsibility to research new trends and available open geodata and inform the department and in continuation the rest of the company through monthly knowledge exchange sessions. Tensing, gave the role of the ambassador for geodata on the manager of their three teams that work in geospatial data, ArcGIS, FME, data science, the ambassador is called "Evangelist" which uses hours from there work week in research about trends and geospatial data, to actively seek for innovation and then inform the team. SWECO on the other hand has the GIS department as a hub for geospatial data knowledge that all departments can turn to.

In the same manner, the culture layer (Fig. 7.3) was developed during the interviews. Starting from level 0 and the case of Asset Insight, were the Volkerail department is considered competition, so the sharing of data is not permitted. In level 1 most companies realise that internal sharing of data can bring benefits to the department, such as Cyclomedia, Fugro. In level 2, ESRI described the slow process of trying to persuade the director when an opportunity to share data appears. Level 2.5 it is a new level and will be explained in the following section, but it would be an opportunity to turn that process into a standarized process, because of the clear benefits of data sharing. Finally, level 3 will be at the level of open culture were the open data sharing can bring only benefits, and there is not hierarchy process needed to share open data. The layers of culture and strategy can remain the same


Figure 7.3.: Culture and Strategy layer model after interviews

when identifying the level of openness, regardless if the company is a provider, intermediary or user, because they need a harmonized culture and strategy to further proceed with their technical, legal and economic barriers and how to solve them to move to the next level.

### New level 2.5

It was mentioned by Fugro, Imagem NL and Cyclomedia, that level 3 might not be the end goal for private companies. This has to do with the compensation of the companies that were willing to move forward with an innovative idea in the geodata community, that would discourage them from attempting that in the future without incentives (Imagem NL). During the interview with a representative of Fugro US about the Seabed 2030 case study, the need of a public-private collaborative funding that everyone can take advantage of the datasets created, maintained in that environment, was suggested as an idea that they are trying to implement now with their partners from Fugro that own the data, that Fugro collected. In that way using investment from both public and private sector for the development and maintenance of shared datasets would be more appealing for the private sector, considering the economic benefits.

Imagem also mentioned the failed attempt of the involvement of the private sector in the Galileo project with not benefits for the private sector, that discourages the future involvement of the private sector, without clear benefits in open data sharing. The proposal so far, was the creation of a level 2.5 as presented in 7.4. The Regime will be to maintain the business model by generating internal and external performance. This could be possible through the private-public collaboration with a shared investment, so the end data with be used both for economic (private) and social (public) benefits, as suggested by Microsoft US for the Seabed 2030 and Imagem NL.

The Find section is maintained as the previous level 3 as the data and services created can be used as marketing solution for the companies being part of the collaboration to generate more profit, based on the suggestion of Tensing with their Digital Twins marketing example. The Play section is maintained as in level 3 of the schema, with the exception of free of charge. All 9 companies mentioned that even if not entirely, they need to maintain their business model for some of their projects, or datasets, which is why the free of charge is not possible for now and it gives no initiative to the companies. The alternative suggestion, of the data being free for the members of the collaboration that also provide funding is a better option for now. Finally, for the Share section, it is suggested to keep licensing associated with

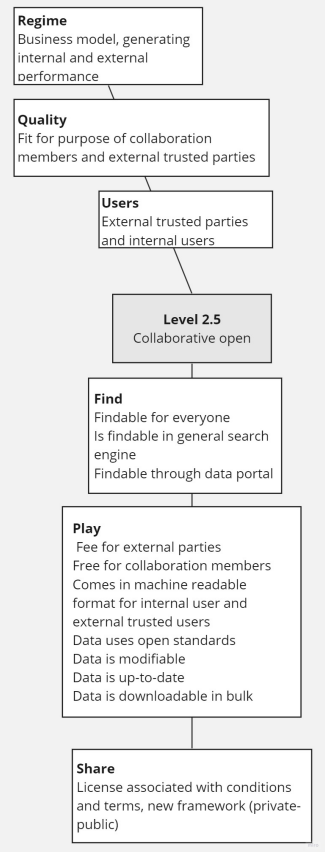


Figure 7.4.: Level 2.5 as part of the new model

conditions and terms, but it is better to follow a framework that is created in collaboration of the private and public sector, which is why the European Interoperability Framework is an opportunity to create it taking into account the needs, barriers and suggestions of the private sector.

### 7.5.3. Recommendations for the private sector

To give specific recommendations for the companies based on the current level of openness they are and how to move forward. The recommendations have to be divided based on their role, as it was mentioned before, the companies that were identified as providers of data appear to be mostly on level 1 going towards level 2. For that transition to happen they need to first solve most of their technical, legal and economic barriers as mentioned in Table 7.1. Specifically, technical barriers that rise from the lack of data standards to follow and software that is still not open, or does not always support open standards, as the standards are stated in every project contract, based in the needs of the client. This creates problems in the internal software interoperability that created delays in processing and validation as stated by Asset Insight and Cyclomedia and need to be resolved first to avoid economic damage from the delays. Also, it is needed to hire more people with high data science skills, or educate more people from the current personnel, as this knowledge is limited to less than 10 people in the company as stated by Cyclomedia, which created delays in the projects.

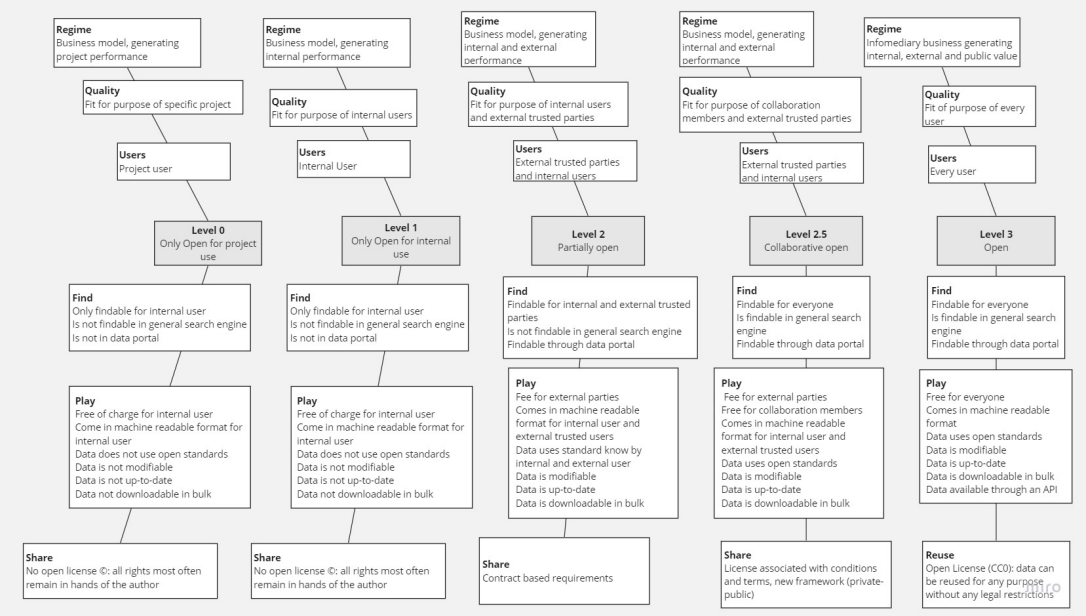


Figure 7.5.: New model with 5 levels only, Find, Play and Share

The companies that were identified as intermediaries are mostly in level 2 trying to move to level 3. For these companies, they have a strategy that is not about open data, but contributes in a way to that. This strategy helps with the minimization of technical, legal and economic barriers. One suggestion would be to use existing open data for their products and services as SWECO does, to minimize the legal issues that come with the ownership and licensing of data, and are provided already for free. They need to focus more on creating a more standardized process of use case identification, as ESRI, Tensing and Geonjunxion appear to have a fragmented process that yields results, but takes too much time for new uses cases to go through the hierarchy. This standardized process, will help identify uses cases, as the ones mentioned in Chapter 5, that could be part of the business model of the company, as ESRI does with maintaining open datasets and having income from software related products.

# 8. Conclusion, Limitations, Future Work

## 8.1. Conclusion

This research started with the intention to answer two important RQs "What are the barriers to private sector data sharing of geospatial data in the Netherlands?" and "How can those barriers be approached to reach (open) data sharing?" In order to do that an initial literature review of private sector data sharing barriers was conducted, an assessment framework to openness was developed. This was validated and further developed by 9 interviewing representatives of 9 companies, in addition to 3 open data best practices from the private sector.

The barriers that were identified for the private sector are different in relation to the role of the company, the level of openness. Companies that provide data in the data sharing system, such as Cyclomedia, Fugro and Asset Insight are mostly identified in level 1 of the schema moving towards level 2. These are the companies that also, have the most barriers to overcome to move to more open data. Specifically, technical barriers that they are not aware of data standards to follow, not an interoperable software to accept open formats, and limited personnel with the needed level of data science skills. Legal barriers that are highly connected to contract restrictions, as they are still working on a contract basis with specific requirements and not a licensing framework. All those barriers, translate into economic barriers, through delays in processing, validation of the data that leads to delays in results and often when interoperability is needed, software, data formats to use open data, this creates more economic barriers related to lack of human and technical resources.

The companies that were identified as intermediaries appear to be more prepared to share open data, as they are mostly identified in level 2 of sharing with external users to moving towards level 3 of open data. These companies are GeoJunxion, ESRI, Geodan, Tensing, Imagem, SWECO, while the do not have a strategy specifically for open data, the have strategies that contribute to data sharing, such as provide more data to their customers (ESRI), help other make their data open (Geodan, SWECO, Tensing), create open source software based on OGC standards (Imagem). Those existing strategies, contribute to minimize their technical, legal and economic barriers, that are only now focused on ownership of the data, support of all open standards is not feasible (Imagem, Geodan), so they have to choose, and maintaining a business model that is not entirely open but allows for some datasets to become open (ESRI, Tensing). They had established a department or ambassador that had to research open geodata and inform the rest of the company about current trends. They were also, more prepared in developing open data standards and open source software, but at the previous category the also did not use as a base a licensing framework, but the requirements were stated on a contract basis.

Those levels of openness for each company can change if the company is identified as a user of existing OGD, in Figure 7.2 Cyclomedia is in between level 0 and 1 as a user and in between level 1 and 2 as a provider. In the attempt to go more into detail, it became

known that the level of openness can change even on a dataset-based, project-based and department based level. Making it possible to share only a dataset as open data, while maintaining your business model for the rest to avoid income loss, a successful example was the Seabed 2030 case study, while ESRI is attempting something similar with the selling software, while providing open data.

Moving forward, to achieve open data sharing, not only the private sector but also the government should contribute, by making changes in their open data metadata to help users with lower level data skill, like the five star system that was mentioned by SWECO. Also, because of the lack of knowledge of the existing licensing frameworks (CC) and open data standards (OGC), they government needs to collaborate with the private sector in developing technical and legal interoperability frameworks and make those frameworks more robust with constant feedback from the private sector in the process the European Interoperability Framework is an opportunity for this collaboration especially for the legal and technical layer, to achieve an solution with the suggestion of the private sector minimizing objections and unwillingness to participate, but also with a more cost effective way, trying to integrate the existing systems.

## 8.2. Limitations

As this is an exploratory research, to identify the existing problem in private sector data sharing in the Netherlands, future research is needed to go more into detail of the role of the companies with the creation of models specified to their role. As this research was carried out in the timeline of a Master Thesis, the time frame was limited. For further research it would be suggested to have more companies being interviewed to have a better sample, but focused only in one role at a time. Also, only one interview was possible for every company, it would be suggested to have multiple interviews for each company, and the interviewees should be from different departments of the company, as every department will have expertise on a different category of barriers, especially for the strategic barriers, a higher level position will be needed to have an overview of their strategy, that does not always translate on the lower levels. In that way the bias of the interviewee would be removed with the different interviews for every company.

Also, the final model, is based on the literature review, the case studies and the results of the interviews and the proposed changes are only based on this analysis. The limitation of the model is that it did not go through another layer a validation, by the same interviewees that made those suggestions.

## 8.3. Future work

This whole research was geared towards the barriers to private sector data sharing, the next step is to also focus on identifying the benefits of it in a more standardized manner to create a catalogue for companies of the value added to their data through data sharing, because they would not have to use resources to do research on the matter themselves, and it would be a bonus in persuading them.

To develop further openness assessment frameworks that take into account the role of the company, the current model is geared towards providers. To create a model for users that are

the most overlooked in this process and it was harder to identify like Tensing and SWECO, we need further research on how the private sector uses existing OGD, in what occasions and how they deal with it in relation to their clients needs.

To do all that further research is needed on identifying the role of the company prior to the interview, to have a more focused research according to their role.

In the process of developing a technical and legal interoperability framework, more research is needed towards, how to involve the private sector in this collaboration and what tools and incentives may be used for their participation in it.

To move forward it was mentioned that the current model is linear with geared towards open data as the end goal, that may not always be the way for the private sector. The next step is to create a model that has a constant interaction with its actors, and those could be the private and public sector in the collaboration to form mutual investments pools for the research of the value added use cases, connected to the the development of the technical and legal framework.

The existing model at the end of this research as a next step should go through validation, from the interviewees that took part in the research to have direct feedback about level 2.5 and how much it comes in relation with their needs and barriers. That could happen through a workshop, where all 11 interviewees could give their opinion and interact with the model in real time. After that, it would be important to scale up the research, by getting validation for the model from at least 30 companies in the Netherlands that contribute to geospatial data sharing. And use this feedback to modify the model.

# A. Reproducibility self-assessment

## A.1. Marks for each of the criteria

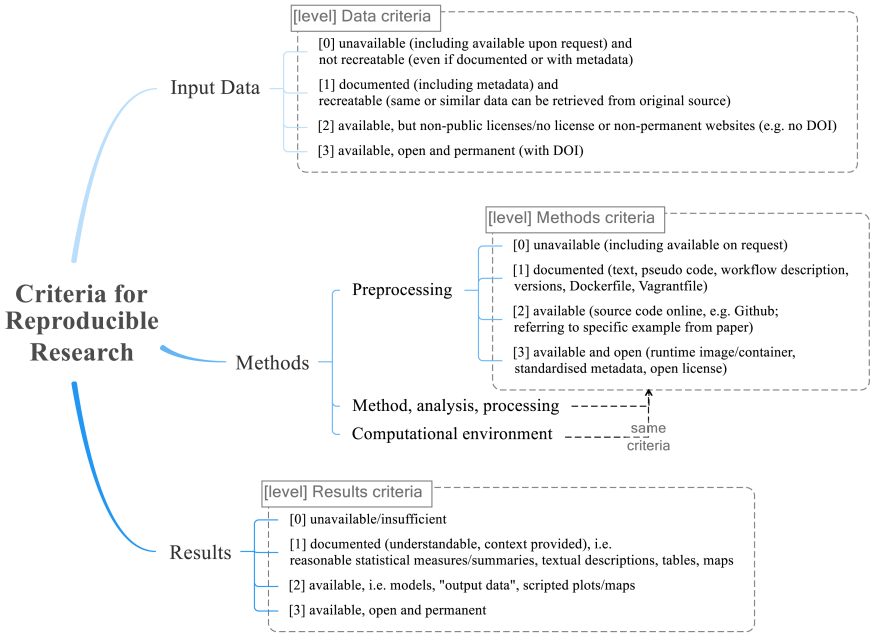


Figure A.1.: Reproducibility criteria to be assessed.

| category | criteria | grade |
|---|---|---|
| 1. Input Data | Barriers of data sharing<br>Levels of Openness<br>Interview transcript | 3<br>3<br>1 |
| 2. Methods | Literature Review<br>Case Studies<br>Semi-structured Interviews | 2<br>2<br>2 |
| 3. Results | | 1 |

Table A.1.: Evaluation of reproducibility criteria

## A.2. Self-reflection

**Input data**

The input data for this research is the literature review that was conducted with scientific papers available on Google Scholar and Scopus. Some of the papers were not open access and they were accessed only with the TU Delft log in credentials, but they are findable through a search engine.

The 11 semi-structured interviews that were conducted are documented in a transcript that is uploaded on the TU Delft repository with this report, after being reviewed by the interviewees and by giving their consent for them to be uploaded. The information of the interviewees is available on Appendix B for future reference. Those input data were not available prior to this research.

**Methods**

The methodology of this research is described in detail in Chapter 4.

The schema and interview questions are available in Chapter 4.3. As it is a semi-structured interview, the questions are open ended and it is not guaranteed that if the interview is conducted again with the same interviewee, the same answer in its entirety will be given.

The case studies are documented in Chapter 5. The information for the case studies were possible with the addition of interviews, because not all information was available through research on the internet. The same interview format was used that is presented in Chapter 4.3.

**Results**

The questions used in the interview have a temporal effect as they refer in the Past, Current (April 2023) and Future situation, which means that the results will be affected also by the moment in time when the interview will be conducted again, based on the state of the company at the time and the knowledge of the interviewee on the topic. Similar results are possible if the methodology described and the interview questions are followed with the same people.

A different interviewee, schema and interview questions will not yield similar results to the ones presented in this research.

The results are documented and described in Chapter 6 and Chapter 7 in the form of a model, graphs and tables.

# B. Interviewees Information

| Company | Name | Role in the company |
|---|---|---|
| Asset Insight | Marieke Bang | Project Coordinator of the Track Department |
| Cyclomedia | Peter Joosten | Consultant on Positioning System that is part of the Research and Development department |
| Fugro Americas | David Millar | Seabed |
| Fugro NL | Julinda Dyli | Production Manager Rail |
| Tensing | Antria Christodoulou | GIS Developer |
| Imagem NL | Patrick de Groot | Commercial Director |
| Geodan | Peter de Graaf | Business Consultant |
| ESRI Nederland | Niels van der Vaart | Manager in Product Management and Innovation |
| GeoJunxion | Catalina Gaete | Sourcing Manager |
| SWECO | Henri Veldhuis | Head of the GIS and ICT Department |
| Microsoft | Remko de Lange | Sr. CSA Environmental Data and AI |
| Knowledge Centre Open Data, Delft University of Technology | Dr.ir. Bastiaan van Loenen | |

Table B.1.: Interviewee information

# C. New Model

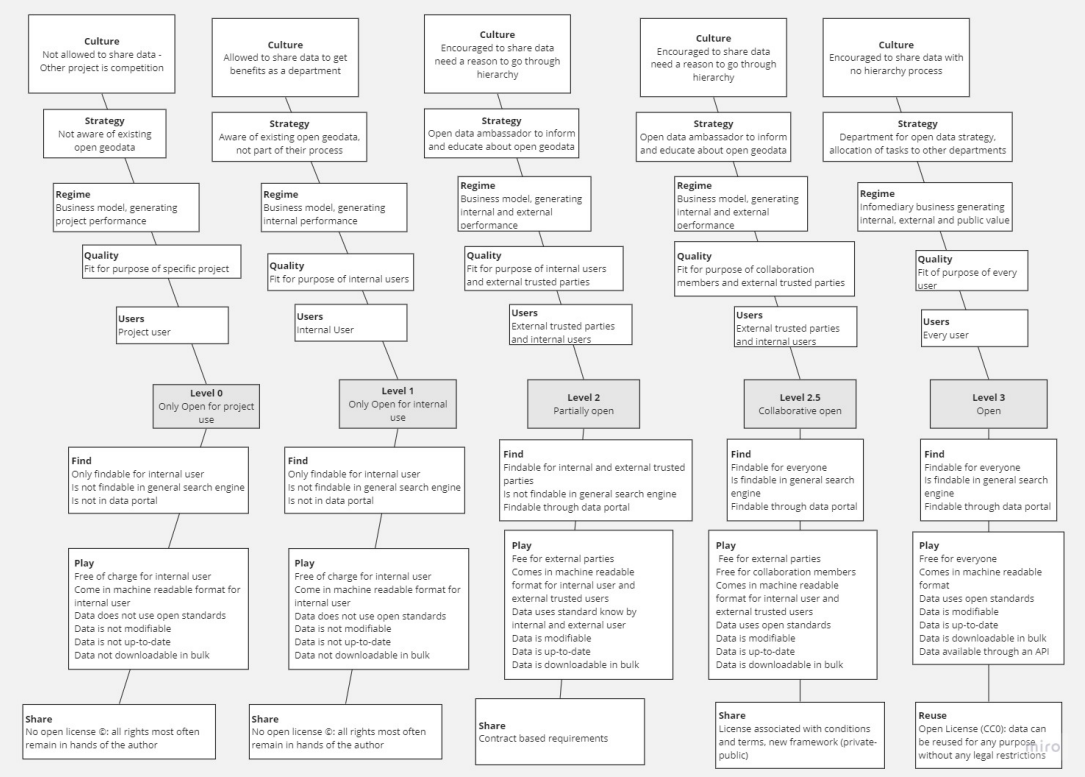


Figure C.1.: New model

# D. Interview transcripts

## D.1. SWECO

**Where do you think you are in this schema**

Level 2 and 3. It depends on each department. It depends on the collaborators.

**How did you arrive at this level in the schema?**

With the development of WFS, web services, to avoid the growth of a shapefile as a closed format. Created services and products on those web services. Also, developed the dg BGT, to retrieve topographic maps through web services.

**What are the barriers encountered?**

No strategy for open data, GIS viewer developed with ESRI. Avoid the famous shapefile format, by maintaining WFS service, didn't have to export the data into a shapefile, Accuracy, reliability, skills, data science knowledge, aware of metadata. Skill issue, learn within the department from superiors, make a star system for datasets. Rely on the reliability of the data for PDOK. Own open data platform, not developed (PDOK). No private sector partnership. SWECO to pay for the data, Dutch Kadastre. OSM as a user of data. Datalab create app to change format from open standards to CAD files. Culture

**How did you resolve/overcome them?**

The weren't aware that there were open data available. Inform them about availability (pdok). Showed examples that they already developed to persuade other departments. The applications were created from existence of data, now used by municipalities, companies.

**What are the barriers you expect to move to the next level?**

-

**How could you overcome those barriers to move to the next level?**

Need a five star system for the quality of open data sets to overcome the lack of data skills, for marketing or architects department.

**What is needed to move to the next level?**

If you want to publish open data you should differentiate them from your business model. Other companies and the government could use the data, to create innovative services.

## D.2. Asset Insight

**Where do you think you are in this schema?**

The level depends on a specific project Level 0 to 2. Level 3 will never happen, because we do not collect the data for ourselves, we are collecting that for another company and it is their decision what to do with them. Some that are in level 2 can be level 3 for different reasons, it depends on each department. It depends on each department. It depends on the collaborators (Prorail, Volkerail) assign them to collect data, a commercial reason to collect the data for maintenance of the tracks for Volkerail, to plan their maintenance efforts, they have an obligation to their client if they are not up to maintenance to Prorail. For volkerail it is level 0.

**How did you arrive at this level in the schema?**

By trying to fulfill our contracts with our clients.

**What are the barriers encountered?**

No strategy for open data, GIS viewer developed with ESRI exchange Prorail wants to have this data be used more widely within the company. Complex data formats, sensor data raw data could create more questions not always, open format different formats, their internal viewer can accept this format. Geotiff format, specific quality accuracy, general format by prorail. Reference model made by prorail. IT department separated. Possible to share into a private cloud or open cloud environment. Need to renew contract. Not possible to change their pipeline.

**How did you resolve/overcome them?**

Try to share information with other departments when possible for other uses than just to fulfill the contract, 2018 was first mentioned, 2020 started to develop it.

**What are the barriers you expect to move to the next level?**

Not allowed to share – contract restrictions, within company even National security, of the infrastructure. Allowed to keep source code – develop software Open data not reliable, liability to the company. Law restrictions on the collection of data (GDPR), video data.

Expertise, knowledge – provide to competitors not wanted to share it, information is knowledge, knowledge is money. The investment on the tools used for measurements, need a return on investment.

User of open data, validation process (Train spotters, hobbyists)

**How could you overcome those barriers to move to the next level?**

-

**What is needed to move to the next level?**

Need the ownership of the data, need to ask owners. Need funding to compensate for the resources we used.

## D.3. ESRI Nederland

**Where do you think you are in this schema?**

The geodata both level 2 and 3, that depends on the specific dataset. Some are offered level 2 some level 3. They do not have geodata that could be level 1, example the location of all their customers, not a dataset that is maintained. the data and services team responsible to provide data to the Living Atlas, producing and maintaining over 600 datasets level 2 and level 3.

The postcode dataset made it completely open source and the model on how to create it from other datasets, generated dataset, that it's level 3. Other datasets topographic maps aerial photos, free of charge to be used by organizations that use ESRI software. Free under conditions. Other datasets that are offered for a specific fee, true orthophoto, in order to make this orthophoto rent server capacity, process imaginery, and create this product, the costumer has to pay a yearly subscription to use it, level 2, because fee for a specific dataset.

**How did you arrive at this level in the schema?**

Three categories, completely open, open for ESRI customers, for a fee, people can buy data, pay yearly to get access to a dataset, or a paper use model. Commercial application that uses ESRI basemaps, large amount there is a paper use model, for each 100.000 views you should pay a couple of cents. Similar to how google, Microsoft, amazon have their mapping services. So they are different business models within the company.

In the past there used to be the topographic map of the Netherlands, had to ask it from Kadastre, and you get cds. When all those organizations are spending time maintaining this dataset the topographic maps of the Netherlands, why not do it for our customers and provide it through web services, started for datasets that are not specific to a use case but can be used for broader GIS analysis, topographic map, aerial photos, satellite images, boarders (municipality, province, country) changes every year, postal code zones. They didn't move from 1 to 2, they didn't open it for themselves first, but immediately for their customers.

**What are the barriers encountered?**

Postal code zones barriers Puting effort in creating it, why not ask for a fee, it is a strategic decision what is better for us, it is a business decision that they have to make. What are other organizations are doing, they also provide data, selling the same dataset for 1000 or 2000 per year, not so much for business, but we are offering it for free, we don't want to disturb their source of income. ESRI is a big organization who can afford to spend a couple of weeks to make a dataset and make it open source, but this may be a large investment for smaller parties, so we have to consider that.

**How did you resolve/overcome them?**

If something like this comes up, the department makes a proposal and the directors of Esri decide, might console a partner in the process, gave them a notice to anticipate it. Other organizations, used the free models and started selling the results. They do not have a specific open data strategy, they have a specific strategy for their Products and Services team, which translates into doing something with open data. The strategy is to make data as accessible to their customers, within this mission they have to decide, in which direction they want to go, spend man hours in adding 500 new datasets or 5 new datasets that are more complex.

**What are the barriers you expect to move to the next level?**

Quantity or quality, every dataset that they add, they have to maintain it, how much time do they want maintenance of data.

Difficulties to maintain, technical. They are working with TNO dutch geological survey subsurface models open data, contain 100.000 of 3d data points that is more work to provide as webservices, than doing a client map, other considerations may cost, if they want to make other aerial data as open data, to host it, need to pay for the hosting of terabytes of images, cost money out of the pocket to host this in the cloud.

Closed data, when working with customers, Ministry of Defense, they are not allowed to share data within different departments. As consultant, they send someone from Esri to the location of the customer. Not the owner of the data, can't make decisions on the data.

**How could you overcome those barriers to move to the next level?**

How to we want to get a return on investment? Is it marketing value if we share is open or is it euro value if we sell it and how does it affect the partner network, and the competition network, it might be an incentive for them to open it for free.

Postal code zones do not exist in real life, they are points, the polygon representation is used for analysis but it is not fixed, different ways on how to define them, if there is a river do you use it as a border. Create tools that generate postal codes zones of of the addresses and other topographic features. Without restrictions, because they wanted to allow organizations to reuse it and make their own things off of it. Also, there are many different sets available, if you make it open, more people will pick it up and make it better. If they get feedback from customers, they try to make changes based on them. The corrections of customers are for the underlying data, like a name written wrong, they will feed it back to their source, they don't fix it, they redirect them to the source of data (Kadastre).

For 99 per cent of the data they provide are already open somewhere, but they make it available as webservices. Sentinel processing together with Microsoft for landcovers. Provide an app which allows to interact with the data. There is a diclaimer with the sources of the data and the terms of use for each dataset in the Living Atlas. Not liable, because they do not edit the data, it is not CC 0 because the original sources are CC 0, that may be CC BY. If some else makes a map with their data, they make sure that get the same attributions. ESRI customers can host their data on the portal, add it to the map, but also share it with each other for different organizations. Prorail provides their data to the Living Atlas. Better to have it as webservices, than provide it as a gml that people have to download from their website. Construction companies that are partners with Prorail, they can directly use the data as webservices and they are up to date.

Future situation move from level 2 to level 3 : Two reasons underlying sources allow for cc0, does it benefit us in anyway, they have to find out. Still require from organizations to use their software. No technical barriers, they are prepared for open source, open formats. The expectations of the end users, to maintain it and keep it up to date, customers care more for the high quality than paying the small fee. Most customers perceive level 2 as open data. No sign in barrier, technically open, legally have Esri software.

**What is needed to move to the next level?**

Changes to model, not all organizations move from level 0 to level 3. ESRI does not produce geoinformation for their own use, their focus is on building software, all the data they

provide is to make it easier for the users of Esri software to start using gis. Main purpose to make it as easy to start.

Changes to the model, partially open dataset that is available for their customers, offer it as webservices, also offer bulk download, data is not up to date, data is not modifiable, is not applied for every case. The level 3 postal code is less up to date than other datasets in level 3, so it does not apply to every case.

## D.4. Fugro

**Where do you think you are in this schema?**

Track department Most of the time level 2 for the rail department, partially open, findable for internal and external trusted parties, Fugro employees, clients, when they get a project, they have specified in the contract which parties within Fugro can use the data, and the purpose for using it.

**How did you arrive at this level in the schema?**

-

**What are the barriers encountered?**

A lot of time restrictions from the contract to use or share the data outside of the department, that is a request from the client, because they consider the data highly sensitive, they do not want many people to have access to is for security reasons, you know where the tunnels are, where the switches are, where are the weak points of the network system.

The second one is findable in a general portal, the client is not willing to share it to an external data portal. Fugro has two internal portals, one owned by Fugro, they upload data from the project and the client asks for access to the portal. It's the same portal for all countries where Fugro is active. The data are in an editable format, csv, shp, image, video and it is downloadable. The are collecting and processing the data, Prorail in NL, they have the right to put it in general search engine or an open geoportal but it is their right not Fugro.

Access to the portal is through a list of users, sign in option, not access to all prorail employs 4000 employs, access to limited users. The client is paying for the collection of processing of data, contract clouse to cost to put on the cloud and maintain data for 5 to 7 years. If the project is finished in year one there are two ways of delivering it, via their data portal, or directly to their database, some clients may have an online transfer service, after the delivery they have to maintain it for the next years, and Fugro is allowed to use it, the data portal has a database on it's back of it, to store and visualise the have to pay a monthly fee to amazon.

For those 5 years, in some cases they ask from the client a small fee to store the data on the cloud. It comes in a machine readable format and the format is specified in the contract client requirements, formats like dgn or shp depend on the type product. The data uses standards, based on the rail standards, measure report it with required aquirashed base on the use case, fit for purpose. High quality of the data is needed, when doing allignment, in mm accurancy, the high speed and weight of the train can affect the security of the track, use case class detaction, when ordering a train you have to make sure that is fits on the

existed tracks, tunnels, bridges, this has different standards, on measurements, accuracy requirements, these do not require mm accurancy, for the volumes. Absolute or relative, sometimes the client does not case where the data is located in the world, but the 3d within the data, local distance, x, y.

The clients are allowed to modify the data, although we would not prefer it, that might have an effect on the accurancy of the data, and Fugro loses track of what they measured. The data are up to date on the monitoring projects, those are the projects that are continuing and they have to monitor the track, you do data collection with an interval based on the requirements, collect the data today and within 2 weeks the data is up to date for the client. There are projects where there was a data collection in 2020, and in 2022 there is an accident or changes to the track, then the data is not up to date, you need a new collection of data to be up to date.

The data that are collecting are quite big in storage from 1 tb collected per date. It is a lot of data to manage and delivering those data can be challenging, delivery online service transfer of the data, or through the portal. Used to sent physical discs, it changed rapidly during covid. The conditions on how to deliver are also specified by the client. Deliverable for point cloud .LAZ file of 25 x 25, it is all depending on the project.

She was not familiar with the cc licenses, but they are not allowed to resale the data. If they have data from client A, and after a few years client B needs the same data, they need written permission from client A to share the data. If they want to share the data with a university they need to ask for permission by the client again. Most of the time clients are not hesitant in allowing Fugro to resell or reuse the data for research purposes it has never been an issues, but the have to inform the client. TU Delft collaboration to provide LIDAR data, there has never been a problem. There are some restrictions, with whom inside Fugro they can share. If there a power department which has the same project in the same country, and they want to use their data, they have to inform the client, because they have to be transparent about how they collect, process and re use of the data.

They do not normally use data from other companies, because they are a data collector and provider, and processing, delivering and consultancy for the data to their client, they do not use other third party data. They use provided data to compare if it is requested by the client. Technical barriers in processing data outside the country, only for countries outside EU not relevant for now.

**How did you resolve/overcome them?**

Produce dataset for research purposes, they are the owner of this dataset. Use it only internally to check if the system is calibrated and validated properly. If a client asks data for a specific area and they have collected those data for internal research, then it was shared externally. How did they arrive in that level? The technology is changing rapidly, people want more digital deliverable, and to have access to them whenever they want, based on advance on technology, experience and time. A few years back the technology did not exist to provide the digital deliverables, even if they wanted. Sensitivity of data that relates to the high accuracy of the data, that is also the expertise of the company.

**What are the barriers you expect to move to the next level?**

Countries outside EU where it is not allowed for the data to leave the country, so all processing has to happen within the country, for security purposes, and restriction by country laws.

Can't move to level 3, because they are not the owner of the data, it's the owners decision, the Netherlands is open does that, but other countries, might have restrictive laws for that. Reselling data is still a profitable business model for them.

**How could you overcome those barriers to move to the next level?**

Future scenario to move to level 3 barriers. Clients restriction, because Fugro would like to advertise their portal to show their expertise to potential clients,how fast they can do it, how accurate they can be. The fear of misuse of the data from third parties, the data can end up on a search engine where everyone, with or without experience, facing a lot of questions, not that the company is hding something but that how it is created, when it is created, maybe the dataset is not up to date, because the dataset is out of the control of the company, someone oute there is using data that is not up to date and the company is liable for that. The mentality should change about the sensibility of the data, what is free or open data, the data are on a portal that every time you download there is a fee to Amazon about the service, that would effect the company, they do not have their own storage platform, they are dependant to Amazon. A benefit would be for a client to have viewer access of the available data to their open portal and in the future require something more. The cost of data collection and processing is really high. Policy to provide data for free in the case of a natural disaster like the Turkey earthquake 2023.

**What is needed to move to the next level?**

Fugro is huge and each department may have different conditions on how they share their data. This interview was only for the rail department within Fugro NL. Fugro makes sure that they are accredited for the data, and the data are trustable.

Strategic layer Huge amount of data within company, need to find the person in each department that knows about the dataset you need. There is an internal organogramma available to everyone to know you to contact for the data set that you are looking for. Contact the manager of the department that can give you a contact list to ask for the data. Project with over crossings of tracks and roads, collaboration with the road department. There is a database, each department has a bucket with their datasets, is used to be one bucket for all datasets, the division happened to maintain it better and not mix based on project. They do not have access to another departments bucket automatically, but they can ask with a reason why they need access. 10000 employees, no restriction can create problems, they are working on putting an internal search engine into it but it is not done yet to make internal sharing easier.

Culture (last 4 years) open on receiving or requesting data with different departments and Fugro in other countries, people are proud of the data they collected or processed and eager to share their results. High trust within the company.

## D.5. Cyclomedia

**Where do you think you are in this schema?**

They are not familiar with the CC licenses framework for data. Comparable with software licenses. The level of Cyclomedia, they identified it between level 1 and 2.

Do you think the change in the level is based on a dataset, a project or a department? They answered that it is just the next step in the flow to have a data analytics department that

uses the data from the other department. By checking the criteria for level 2 , quality fit for purpose, users are internal, only open for internal use, it is findable by internal users and every paying customer, they are using the same portal, because they are using the same data for different use. They can sell the license to use the data, but they can also use the same data to do data analytics. Obviously free for internal users, we don't transfer money between departments. They follow open standards, by following OGC data standards. The data are not modifiable, because the data analytics team should extract information from them, not modify them. Maintaining their data up to date, depends on the project, in the Netherlands they are always up to date within 14 months, that's what they promise their customers. Data is downloadable in bulk. No open license. External users/ customers, they can find only what they paid for. For example a municipality can see only the data that are within their borders, they don't see what else the company has. They can search within their borders based on the street name or coordinates. Internal users can have access to everything on the portal.

You work with municipalities, do you also have some private companies or universities that you are working with? Collaboration with Centric, in the Netherlands for tax assessment, in the Netherlands you have to pay if you have an advertisement in front of your shop, so they are working with Centric about that, with location data, so they are a partner. that would be level 2 it is images and location of the images so it is geospatial data. The small fee does not apply, you have a contract with the customer, and then share the income, we don't use common licenses, we have contracts negotiated between our legal department and their legal department.

We also have partnerships with universities, quite often with delft, they can use our data for certain research, usually no problem, but we create a contract, limiting the use to the specific purpose, free of charge, the use is limited in the contract.

**How did you arrive at this level in the schema?**

Cyclomedia, collect data with cars on the road, those can be LIDAR, point cloud data or images with very good positioning that is the product that they sale as a license to the customers. So they always own the data. Provide data to customers via internet, web application, they can view and use the data with a license but they don't sale the data. Some costumers use the data to do data analytics, might want to make a database of traffic signs, they measure the surface of the rooftop to see how many solar panels they can install. In that sense the sharing of data is at level 0 as they are only used for that specific project.

Since 10 years ago, Cyclomedia started doing data analytics themselves. So customers can request to have access to data analytics that they do. For example they created a database of traffic signs themselves that customers can have access to. Then that is level 1 because they are using the data that was created in the product/processing department. They have some external partners that they are working with, so that could be level 2, that is why they think they are between level 1 and 2.

**What are the barriers encountered?**

You said that you are at level 2 for some datasets, can you walk me through the process?

We have defined a way of providing data through a web portal or an API, where our partners can create code to retrieve the data. we have documentation on how our API works, and we only provided to our partners, the data follows a structure that we defined ourselves, we tried to follow OGC standards, level 2 because of open standards, never fee, only sometimes for universities.

Why did you change to OGC standards?

it was kind of gradual process , we started to do that, because not only our partners, but also our clients wanted to access the data through an API, and you have to describe how to do this, and OGC already did a good job at describing it. when you start distributing your data you have to explain what is the data format is a jpeg or a geotiff, how to access it and retrieve it, how to find it, what is the coordinate system.

Are those explained in your metadata?

No in the API documentation. of you have a lot of questions refer to the support department, if there are too many questions we have to improve our documentation. it;s not you run into a barrier and you solve it immediately,

They changed to OGC because they expected that the clients and partners would have a better understanding when you follow a certain standards, for the internal process it would not be necessary, but it makes sense with an eye on the future and who else might want to use the data. When you work with companies like ESRI who wants to add possibilities to view your data in their software, they already have standards, so you make adds on for the arcgis software in cooperation with esri and you are following standards that exist already, so that esri users can see our data if they have a license.

**How did you resolve/overcome them?**

It's a discussion you talk to customers who want to use your data, and they identify a certain problem and then gradually you remove the barriers, it requires usually a lot of discussion, what is the orientation of the images, why is that not in the metadata, because they are horizontal with the horizon in the middle. on a higher level it's easy but you run into small barriers, after a discussion, and a few hours changing your software.

**What are the barriers you expect to move to the next level?**

Level 3 make it open and freely accessible that is what we don't do.

Do you have plans to move some of your datasets to level 2, because you said most of them are in level 1?

It is in our interest to have more customers. our business model is actually record and create the data once and sale it to multiple customers, so far that only works in the netherlands, we work in germany and usa, what we do there is drive on a commission from a municiplaity, so that is why we do not sale the data, we license it, so we can lisence it also to another client. for the same dataset transport department and telecom company. so the vision in the future is to do that more often, in the netherlands we sometimes have ten customers for the same dataset.

On one hand it's completely open just you need a license and you need to pay for that, that is what it maintains your business model.

Internal geo portal. Only there you can be informed of the available datasets, in terms of variation our data is very limited images and pointclouds, for pointclouds we follow the LAZ standards (open standards) image format is open jpeg and we have the metadata described. they have their own metadata standards time of collection, location, laz (positional info), images coordinate of the camera, and the accuracy, it is pretty simple. most people are not interested in more detailed metadata.

**How could you overcome those barriers to move to the next level?**

What would you need to move to level 3?

in theory it would be possible if the government provided funding and require us to open up the data, but we have to buy fuel for our cars, pay our employs, and buy staff, completely opening up is not working, it is quite an open business. financial barrier.

There was an occasion that we opened up a dataset for a specific period of time so people can get a feeling of what is possible, they can get a free license for a couple of weeks. the only way to go to level 3 is if someone is paying the bills.

**What is needed to move to the next level?**

Do you use open data? yes sometimes, mostly internally, we don't use them a lot. for example in the us you have available aerial LIDAR data, but when you are in a city, urban canyon, GNSS GPS does not work very well, so our street level data positioning is not very good there, so we use the aerial data to connect to and in the us this is free. For the Netherlands, we don't use open data in our standard processes, I do use it for myself,the AHN data, use it as verification, only internally.

Why do you not use them in your standard processes? what would be the reason? basically it did not come up yet, it is not a requirement, we also have our own aerial data, in a similar quality, for production purposes we use that, internally, we use open data for verifications in the Netherlands. Because they are following open standards there is no technical barrier. Especially in the Netherlands if we make our data open we lose a lot of customers, telecom, electricity, insurance companies, solar panel installers. we lose a lot of many.

can you think of any benefits that could come up if you share your data openly? it is useful to use geographic data for society, some times people don;t use our data because it is too expensive, good for society , not great for the company. For the funding a private public partnership might be a solution, that was an option when they started building Galileo, there was really not option for the private sector to get involved because they would lose money. If you wanted to share data, how easy it would be to go through the company hierarchy and how comfortable you would feel to do that? For a small scale, for example universities that would be easy. on project basis keeping a good relationship with universities is for us also important because you need high quality employs and that is where you get them.

We do not have a gis department. r and d, production, sales and product management. I am part of r and d make the systems, design them and improve them, and systems it is not only the whole processing that comes after it. within r and d we have responsible for creating the base product, the images and the point clouds, data analytics they do research in machine learning on getting more information of the data, we have a platform team that makes sure that everything runs smoothly on a large scale, our database is enormous, we have 5 petabytes of data in ESRI, make 500000000 images per year, and the algorithm have to run on a large scale. front end api portal development, gis knowledge is spread, the knowledge on high accuracy staff is really limited. 4 people that know about geometry, high accuracy gps, 10cm or better, a little better than gis.

You try to guess might want in the next couple of years and then define a certain strategy, now we are running in the issue that customers want more control on where their data is stored, customers in the us do not want their data to be stored in western europe, we store everything in the microsoft esri cloud, but they have data centers all over the world, some customers in the us they do not want their data on an esri cloud in western europe they want it in the us, i do not understand why, but it is their request. then it makes the software

more complicated. If you do a request for a certain image, you have to first find out where is it.

## D.6. Geojunxion

**Where do you think you are in this schema?**

Understanding sharing as giving data for free. Geojunxion has developed apps to use based on specific types of data, just congestion zones/ eco alert zones. It is different from their base map or general geo data.

Confusion in the schema with the strategic layer. Is the goal for the company to become open or to use open data?

She thinks those are completely different things. Even if the company is not open, they can use open data, because of the licenses, terms and conditions that those datasets have. In fact in the company to create products that are in Level 0 or Level 1 they use open data , except when the copyright is CC SA, because they cannot do it commercially, and then they will have to share it openly, and they cannot do it because they are creating products that are going to another company, that is gonna be used in a closed environment. When it is SA we don't use open data. We are only becoming open when we are forced to do it, for example when we want to use SA data.

Talking about their customers, they don't know if it is level 0 or level 1, we don't know the difference, because the customers buy their service and they don't know after that, how they are gonna use it. They know the use case of the product, we are b2b they share it with their customers, if the contract says that they are the owner of the data, they can actually do something different with the data.

Level 0 and 1 identified as one level. They believe that GeoJunxion has been in all levels as on 25 years of existing. The have projects with other companies, partner companies, other mapping companies, for that they are at level 2. At the early stages of OSM they shared part of their data for the netherlands that is why they were at level 3. For the most part they are at level 0 or 1, create a product for a customer, that they use within their company.

**How did you arrive at this level in the schema?**

A couple of years before (2020), we had a change in the branding, and restructuring of their business model. Before the focus was on creating digital maps of whole areas and trying to keep those up to date. All the efforts of the company were focused on the digital maps projects.

Now it is more a costumer based company, companies approach them to ask for specific layers, not a general map of a country, but just an aspect, like kindergartners in a country, or areas were cars are not allowed to circulate. They have created many different types of layers, those products that were created are not the same anymore, now they are more based on customers needs. The licenses with the companies are not data licenses anymore, but they are service licenses. The company hires them to research what is already out there, make reports on how the system works and then deliver a dataset, but this is on top of an actual map.

In that process do you create every layer of the map from scratch for the customers? or do you share some of the datasets? Do you sale the data and not use it again or do you let them access it for a certain amount of time or it is different? It depends with every contract, because when it is a data license, they are the owners of the dataset and they can do what they want with it. But when it is a service license they are not allowed to sale the same data to another company, because that was created only for that company.

Some companies are not even aware of open data, that's why the focus changed. GeoJunxion uses open data, they have a Sourcing team that is constantly looking for information, whether this is open commercial, for purchase, they are looking for data customer based, when a customers comes and asks for something we are doing our research and all the information and data that are found during that research, it is stored so we know what is out there, but we are also, looking for information on a regular basis, let's say I found this dataset about topographic maps, let's put that in the folder so we are aware that it exists. Then they are keeping track of the copyrights, and the urls/links in case the copyrights change. For them the copyrights and licences are one of the core values, to avoid in any case breaching any licensing condition. And they are always following the requirements that come with the license, if it cc by we are always attributing the source, they just follow what the source says.

**What are the barriers encountered?**

When talking about where their own company is in the schema. They answered that the schema is confusing on whether it describes the company or their partners. When talking about them and whether they created a dataset for themselves that it is in level 0? They answered that they create maps for themselves, specifically, they have a base layer that they created years ago and that they are still updating for other companies, but they also use that base map layer, and on top of that they create their other products. But they are no longer delivering that base layer, they use it internally to create the geometries, to create their datasets, but they are not delivering that base layer. Some customers have access to it, and they have a data license with them.

There is not one line for the company (that it why it is difficult to abide by the schema). For the base map layer they are at level 2. A company can never be put into one box. For the customers in level 0 and 1 they know that they can not resale the data in most cases, not to a company with the same case study, but they know that the company is giving the data to their customers, which are just users, car owners most of the time.

Are you following open stardards for the format of the data? The customer specifies it every time. Are you capable for every format that the customer would ask. When we use open data and it is cc by, they inform the customers that they need to have a attribution section, or somewhere where we can cite the source, and the customers says that the format will look like that.

On the technical part I am gonna share my layer as an .shp or a web-service, do they have to specify it in the contract, that depends on the project and what the customers wants to, no we don't have a standard for it. To be honest I am not sure about the technical side of the work, because I work with licenses. She is the sourcing manager so that it why they have a better command on the license, she is looking for open data, commercial data. This is core values, we make sure that licensing is really clear, we make sure that we comply with all that is needs from us and is required of us.

**How did you resolve/overcome them?**

-

**What are the barriers you expect to move to the next level?**

She thinks that they are not becoming open, not because they are lacking the resources or because they don't know about open data, "we are not becoming open, because we are a commercial company". This does not mean we are not embracing the OSM community. The company was part of the community in the very early stages. The whole crowd sourcing data, we believed it at the beginning and we still believe in it, we know that it is important. When we are required or when we think and believe that this is important we are sharing information. Actually, during covid they also developed some datasets and some solutions that were not customer based. It was an idea within a company and we made it public, but they don't know in which stage it is right now, if they are updating it, since covid is not the main focus now.

We believe in crowd sourcing public and open information, but at the same time, our change in becoming a btob company and a customer based company, requires us to be focused on commercial data, and not on making data available publicly, because most of the times the data is not ours. We make the data we deliver it to the customer and we are not the owner of the data.

They combine different datasets, processing of the datasets, they do the sourcing to find datasets, but they don't provide them as it is. Electrical Vehicles charging points, they combine datasets to do that. For that they can find open datasets but they may also have to buy some datasets, because they realize in some area or region that the open data are not really good, they combine and refine datasets, make the attribute according to the requirements of the customers (not open standards). They are not redelivering open data, they are always enriching an creating something new. They are not resellers of data, they never resell data as it is. The do not have a contract of reselling data, they are always creating a product.

**How could you overcome those barriers to move to the next level? What is needed to move to the next level?**

Let's say for one of your more popular datasets, you would know better which, at some time you would find the incentive to make it fully open level 3. What would the barriers in that process would be? When we create a product that we know it is good and is relevant, and important, we want to make it the best, one of the products that we are very focused right now, because of out focus in sustainability, it's the eco alert zones, that project started in 2020, a few weeks before the pandemic, we are still developing it, we are still creating the best way of keeping it up to data. Because it is based on pollution information, based on regulations of the countries, that changing monthly or daily basis.

There is a regulation in Spain right now that says that all the cities in Spain above a certain population by 2023, will have to create zones where cars do not circulate and those are 400 cities, we don't know when those cities will come up with their own regulations, so we need to monitor and keep up to date. that is the research part and the sourcing part of the project, we are constantly looking for information in the media, in the gazettes of the countries to see if something comes up. In the development of that product I don;t think we are at that point were we could say that is the best quality to give it away, we are still creating, we are still working on it we are doing field tests, to know if it is accurate, if the time of the activation of zones is right or not, we are doing that with our customers, and the customer is leading our work, because it is a service contract, we are tied to that, but as long as we can create a product that becomes better and better, and that we see that the relationship

with our customers becomes more open, then we are happy to do it, but we are a customers based company. The customer leads the way.

It's a service license, they paid for the whole process, not only for the dataset. That is the business model that they decided to move to, because they realised it is more impactful, that has more real application, than just creating a product that is gonna be there, that someone may use, now we create something that we know that it is gonna be used, and it will be given a end user.

**AND - OSM**

Past

Level 3 because they are shared in OSM. How do you think you arrived at that stage and why do you think you decide to do that? This was at the early stages of the whole crowd sourcing idea, of sharing data and together creating something better. AND entered into that world in 2010 , one of the first companies in the Netherlands that believed in that project. At the beginning it was hard to "sale" this project, where you as a company, put your data that you commercially sale into an open. The goal at that point was that the company together will make the data more up to date, we'll make it better. if AND shared a dataset of addresses, they expected that a couple of years later or a couple of months later, the same dataset will be improved with information from other companies or other users. That's what we believed, we still believe in the OSM foundations, we have a relationship with the community, we are not outside of it, we believed in start ups and what they can do, we were also a start up, we wanted to collaborate with the whole start up geo scene.

Present

What happens now, we saw that when you deliver data to OSM, you don't know what is gonna happen with it, it is the community that decides how things are going to develop. The dataset that was put there 10 plus years ago, it is not actually deleted when it is not useful anymore, when things are not there anymore, people are adding layers, which is great, because people are getting more up to date, with people adding new information. What is out there is not being curated, revived, checked, that was disappointing, they expected that this would be a more precise, accurate system. Of course we are part of the system we are one of the companies actors within the system. If we are not improving that dataset no one else is doing it. We still embrace OSM, we still are in level 3 because of those data, we believe in the whole sharing information, but of course our focus is on selling our products.

In the process of giving the street map dataset to OSM, was there an agreement that specified, if it was your responsibility to keep the data up to date, or was it OSM's responsibility?

I not sure if there was a contract or an agreement for that, I know that there where a lot of conversations about it, there was also a conference, AND was part of all the geo conferences, that happened at that time in the Netherlands, even hosting other companies in AND about conversation about the OSM data. Nothing clear though, but I think the agreement was not that the one who shares the data is in charge of updating it, we thought/understood that that was a community based work. It is not entirely a disappointment, but it is something that changed our expectations. That was the expectations of all companies/ actors that were participating at that time, we had the expectation that this could be something that all companies would be able to benefit from, in terms of having up to date maps everywhere, we also realised that is not the case in every country, in every region. That led to the realisation

that we need to provide the best quality for our customers, usually OSM is not the place to go.

Was there a conversation at that point on how this sharing of data could affect your business model at that time? More than something affecting us, we really had the commitment to it, this was part of the new way of doing things, crowd sourced information and companies sharing. Of course all companies have a commercial goal, but this was part of our core values, and we still have these values, it is not something we completely got rid of, because during the pandemic we also provided data, and we work with governments, helping governmental organizations to improve their data. It is something we do at the same time, I wouldn't say that the levels are stages that we move one after the other, they are happening at the same time, all the time.

## D.7. Tensing

**Where do you think you are in this schema?**

Consultancy they don't own their data, they go to other clients to do what they ask as to do. have some internal projects related to a specific business model, for example digital twin they are creating 3d data for a certain area and provide to clients to show their expertise and based on that the client will ask them. They cannot identify level based on providing data, but they can identify level based on using open data between level 2 and 3.

**How did you arrive at this level in the schema?**

We have so many consultants working in different companies, so we are able to collect knowledge from all those different partners. We share this knowledge with our employees. for example I was working at the municipality of breda try to identify vegetation areas and build a process that calculate the vegetation areas and compare this through the years and we used data from PDOK, made a presentation to show to my colleagues that this exists, we are in a high level at using data, and that comes from the exchange of knowledge internally. It is a standard process, at the end of every project, no they have knowledge sessions, but also ask others this is interesting would you like to present it.

**What are the barriers encountered?**

We have a data portal for data, just for development processes, we don't use it to share data. it's not our data. Internal portal for development purposes not to share data.

**How did you resolve/overcome them?**

By having an extended network of partners and previous clients and being informed about the new market trends on geospatial data. Helpful as a strategic level specification, they are a level 2 or level 3 user of data because of knowledge exchange between partners and their extended network. At the end of every month a knowledge session. Channels in Microsoft teams for data science FME ArcGIS, and when everyone finds something new they share it there. They have a knowledge ambassador for the three teams. They get 2 days of their work to focus on research about data trends and what is out there, to inform their colleagues "evangelists", about open data or open applications, even a newspaper about geospatial data.

**What are the barriers you expect to move to the next level?**

The barriers that they expect are connected to the geospatial data science skills of their clients and partners. That they are not the owner of the data to make the decision to go open.

**How could you overcome those barriers to move to the next level?**

To find benefits and value to connect to those open data.

**What is needed to move to the next level?**

They would need to be the owner, to find a business model that fits on sharing the data as open. For now that is not needed with their current business model.

## D.8. Imagem NL

**Where do you think you are in this schema?**

You said through our email that maybe those questions do not apply to you, because you are not the owner of the data that you produce.

Well it is more than that, we do not produce that much data. What we do is we work a lot with geospatial data, but that data comes from the customer, they purchase that data, or we purchase data for them, or we use already existing open portals like PDOK and NSO. Then what we usually do is process the data, so for example we take raw imagery and we put an algorithm on them and on the other side comes out a specific result from that calculation and we provide that back to the customer.

What we do not do is tell them to give us your data and now it is ours to hold, the data will always go back to the customer. In some cases where we created a very prominent model, algorithm, that has very promising results, we would want to share the workings of that algorithm with the outside world, we usually have a conversation with the customers we created the model for. Is it okay to use your raw data to showcase the workings of the model to others, they are free to decide what to do. With the government almost in all cases it is fine, the data is open to begin with so it is fine for you to use it. In a commercial setting that usually is a lot different, and I think that is understandable.

**How did you arrive at this level in the schema?**

Everything that you are explaining in the schema is geared towards a company that produces a lot of data.

I made some notes on culture and how we feel about data. In essence we are a software development company. In that development culture there is a very strong advocacy for open standards, we feel like every time we produce something, the value of our software solutions should come from what the software is doing, and it should not come from the fact that we are able to enclose data and people have no other choice than to work with our software if they want to use the data. We see that happening a lot where data is transformed into a geodatabase or a Revit format, for BIM kinds of things, if you don't have a license to use the software then you are stuck. We believe that data we take in should be open, we do stuff with them and also data that comes out should be open.so other people, aca others systems should be able to use that data and you create a flowline of processing steps. Example: take in satellite images, put an object detection algorithm on it and the outcome is a format that is mbtiles or geopackage, anything that can be used by QGIS or a system that is also adherent to open standards.

You want to be open, us as a company we do a lot of work for the Ministry of Defense, we are obliged to secure data. Prove to us that you can keep our data secure and not make it available to the public. In that case it is not that we don;t want it to be open, it's just we are not allowed to, by law. The Ministry of Defense has three levels of classification where they say that some of us, even me, are not allowed to even see that data, cause I don't have the proper level of clearance to look at that data, other people from the company do.

**What are the barriers encountered?**

This is coming from the software creator, supplier, the barrier we encounter is never about culture, because we want to support open standards, the barrier that we see as a commercial company is to choose which standards we will support and at what stage to support. If money and time was infinite we would support all standards, OGC, ISO any other standards that is around. The reality is that our resources are not infinite, we need to choose at any moment in time which standards is most important for our clientele to be supported at this time. Because it takes time, you need to start creating the standard or support for the standard, put it through approval, do retesting and at one point get it certified if you do your job well, and from that point you support the standard.

**How did you resolve/overcome them?**

We talk to our clients, we follow general market trends. We spend a lot of time with our clients and the market and general organizations like GEONOVUM and Geobusiness Nederlands to constantly assess where we need to be and where we need to prioritize. This is really difficult, because everytime you talk to someone they would always put emphasis on the standard that is most important to them. Municipality, Water Board, Ministry of Defense three clients, they will have different standards, and they are not using each other's standards. That is not entirely true, they use some OGC standards, some NATO standards, they require from the company to support both, that is a challenge. You sit down with them and start to peel, what if we support this standard first, sometimes supporting one standard can help in speeding up the process of supporting another standard. That is not always the case. We are right now on a specific maritime standard that is only applicable to the Navy, but it is so important to them that it will hold investment that goes into technology which goes into hundreds of thousands of euros, so that is also a factor we need to take into account to prioritize.

Putting it like this we cannot always go for the ideological argument saying, we feel like that data should be open, we have to take into account what it mostly needed and what is needed for our company too.

**What are the barriers you expect to move to the next level?**

I do not expect that many new barriers, because there would be new barriers if our culture would be different, we really do not like the idea of open data. That is not the case, only to choose the correct data, we don't have to battle anyone. We do not see new barriers, so we do not need a strategy to overcome those barriers

**How could you overcome those barriers to move to the next level?**

Do you have a strategy for data, it does not have to be about data, or a department that allocates responsibilities based on that strategy?

No we do not, not a detailed strategy, because data is not our prime business model. If we talk in general terms, of course we have policies that support store once use many, store data in a centralized location, make sure that is secure but accessible.

How are you informed about trends, since you are working with Digital Twins, and that has been a trend the last few years?

Because we spent a lot of time with our clients and ask them what will be the forefront of your focus for the next few years. Of course, we are connected to a worldwide vendor called Hexagon, they cover their needs, and they share it with us. We are putting a lot of investments for Geospatial Asset Management and the third is the general organizations like geobusiness, events like the geobuzz, conferences, we have a lot of antennas into the market to decide where we fit best in, and where to invest in.

**What is needed to move to the next level?**

From your questions, I was contemplating and I might have read it wrong. What I noticed is that there is the drive behind what you are trying to achieve is to move as much as possible to the level of open data. I do not think in all cases it is fair to strive towards full openness. They are companies that spend a lot of time and a lot of financial means to create specialized data. Example aerial flight companies that operate an aircraft that spend hundreds of thousands of euros for sensors, I do not think it is fair to ask them to share those data as open to the public, they would have no means to get return on investment from their efforts.

I think it's okay when data is created using public money, then it should also be publicly available. I think that's a one on one. But when it comes to people that take the leap of faith to do research and innovation on some sort of new data, I think they should be allowed to create a business model around it where they get back some of their investment. And they would only do it if they saw that they could generate more money coming back in than they originally spent. If that is not the case then long term there would be no incentive for them to do so, because they are spending time but they are not growing their business. "Should it always be the case that we strive towards full openness of data" I do not think this is the case.

We have a specific business partner SpotInfo. They made a business model around harvesting open data sources, PDOK and the Key Registers, they spend time and effort to combine those in very comprehensive datasets, so people will only have to access one dataset and then they will get everything from that. And they make sure to keep it up to date, that it is aligned. This is a company that works with data, and they bring added value to them and they market that as a commercial product.

I do not think in any case it should always be a goal to move to the most open level, it should be really distinct in which cases that is something to try. When it comes to public data that is created with public funds, that should always become open. It should never be the case that data that has been created as open once should move back to not open. When it comes to data that was not created publicly, if you want to, you should always put in place for a framework for anyone that has put in the investment to make sure that investment has returns, to at least get back the cost that they put in creating the data. Because the companies will not do it.

Example the aerial photos of the Netherlands, it used to be a very lucrative market and now it is like a government regulated program and people are fighting to get contracts at the least possible cost. As a result, they also try to minimize their cost. It can be debated whether this is the way to get the best possible product. Then it is open, but is the quality of the data as good as it could be? Or is it what people can afford it to be, because they have to produce it as open data.

We sometimes purchase satellite data from a commercial vendor for their process the Water Board, it will become their data, we do not keep them, so we basically work as an intermediate, they say you guys know all about what to ask for, please purchase the data for us, and we always make sure that the license will go to that company. In that case we will put the data on a portal, but restrictive, because we are simply not allowed by the license that we purchased for the costumer, to share that data with the whole world. It would be different if the end user, in this case the Water Board, they are free to decide that they wanna share the data with somebody, but even then it could be that the satellite data provider, and they have done it in most cases, they restrict that, because if they allowed that, they could not resale those satellite images to anyone else. But this is not for Imagem to decide, we just stick to what we are supposed to do.

In that cases purchasing the data for the Water Board, did you try by specifying in your contract, I would like to at the end of the project to keep the data for the company and use it to develop another product?

We did do that in some cases, not with aerial picture data, in that situation, we can use the data internally, the restriction will be that we can never disclose that data, we can use it to further do research development, and see how that data react to the model and showcase the results that come out of it, but we can never share that data again.

Geospatial Webshop

There is an interesting thing going on, that we are trying to see if it will work, to maybe have companies move from level 0 to level 1 or level 2, we have developed a solution for a geospatial webshop. That allows for engineering companies that make laser scans for projects, drone data for that project, we are trying to see with them, if there would be a market to resale that data, through such a portal. Whenever they do a project in the North Holland or the South, they put it in the portal, and someone says hey I now have to do a similar project there or I can make use of that data for my project, which is a different use case, but the basic data is already there, because otherwise I would have to collect that data again. We are trying to see if there is a need for that. Project based data, if we put that out there is there a market, or use case to reuse that data .

## D.9. Geodan

You know we are a company so our primary role is to make money. We work for our clients, and our clients are divided into 2 big groups, the public domain, the government, the municipalities, the water board, or counties and we work for other companies. Our company is not gaining data directly. We are not working on the field and we are not collecting data, we are not working with satellite data, not primary data. Usually we start using data that has been collected by others or the government, and we analyze data to make products, the core business of our company. Open data is not really a business goal for us, but is an important trend, mostly the government but also some companies are asking advice for use of open data. The electricity companies in the Netherlands have an open data policy, they believe if they publish their data openly, it will benefit their process and their business.

**Where do you think you are in this schema?**

If you want me to put Geodan in the schema, it is a really open company, we really believe in open data. We do not work with vendor lock-ins. Interoperability is not always easy with ESRI software, because they do not follow the OGC standards strictly.

**How did you arrive at this level in the schema?**

10 years ago they changed their business model from selling data to licensing products, that was because of the change in open data legislation 2005.

**What are the barriers encountered?**

We do not have a lot of primary data of our own. If we use data it is usually from our clients, and it depends on what our client wants, to publish or not, and we have to make agreements on that, so they decide. On the other hand Geodan uses some open data sources and we do some analysis on them to make a product and sell that product. If it comes to the awareness culture we are at the top of the schema, we believe in the benefits of open data and advise our clients on that. What data can we make open, on what conditions, or what are the benefits of open data.

The data provided by their clients are internal data, we are allowed to use it or ask to use it and make new data products and it is up to them what they do with that. At the end of the project we can not sale the product in a different way, because we do not own the data property. Even for public data, we have to check the licenses, and if they are not classified as open data, we always have to ask. In the Netherlands there are also laws and regulations about data that specify that. Even internally, they have to ask permission for data from other projects.

**How did you resolve/overcome them?**

Our developers are working with developers from the whole world in developing open source but that is not open data.

**What are the barriers you expect to move to the next level?**

Data property, is the data ours or not. It is easy to publish open data, but is hard to maintain open data, is it the last version, is it the latest model. It is the end of a project, but it is actually the beginning of a process. It is treated as a project and then it is over, so there is old open data on the internet. Privacy is a big issue in the Netherlands, it is legislation, but if we have data on the location of people, that is under strict legislation, you have to anonymize them, we need a driver, a contract with the company or the government otherwise we are not urged to publish public data, it costs. We don't spend time on things that do not benefit us. On the technical part, they are following open standards and using open software, they are even contributing to the OGC standards.

Within a project internal sharing of the data is allowed, always within the borders of this project. They have project portals for data. Even within departments the data can be on demand from a different project. We are pretty careful with data, especially if they are not our data, because we would run into a big problem if those data end on the internet. Security policy for data that has been certified, they have been trained in security issues. Access on data and systems is regulated.

**How could you overcome those barriers to move to the next level?**

**What is needed to move to the next level?**

Do you use open data?

Yes a lot of data, we know what is available PDOK, NGR. If there is a need to use data, and we need that open data exist, we use it under the conditions of their license.

They are working with public undertakings for electricity or infrastructure or railway, that make their dataset as open data for INSPIRE, and extra datasets that they think is good for the business. We would need a driver, such as a contract with the government or a company.

## D.10. Fugro Seabed 2030

**Where do you think you are in this schema?**

This is a very complicated landscape and I think you did a very good job of capturing the complication in the various tiers. I wanted to kind of just refer to that as I describe what we are doing and why we are doing that. Fugro is a multinational corporation that collects all kinds of data, but the data we collect is almost always owned by our clients. They hire us to collect the data, when we collect the data it is almost always under a contract with another entity, and almost always that other entity owns the data. We are simply collecting the data, maybe providing some analysis and advice on the data, but the ownership of the data remains with our clients. It is really not our data to share, even though we are operating around the world with a lot of people and assets and equipment and platforms and collecting the data, it is not our data to share.

**How did you arrive at this level in the schema?**

Knowing that, in the bathymetry data context that you first asked for, there is an opportunity here for Fugro to convince our clients to be more open with data and share their data, and we've been working on that for several years, but we recognise that there are variations from country to country and from culture to culture, and within company culture as well. It is a very broad and challenging landscape that will take some time to address and sort.

But we did recognise, okay so when does Fugro have ownership of data, and it is really between projects. When we are on a project, we are hired by a client we can't share the data without their permission, but if we move to the next project, then we have the ability to collect data that we would own, and that is how the bathymetric data collection and sharing under Seabed 2030 evolved. We saw an opportunity to collect data from our ships between projects, when we had ownership of that data. We started relatively small with one ship and tried to get systems and processes that were fairly automated so it wasn't a very big investment without much manual handling. We wanted to make a positive impact, but we didn't want to spend a lot of money doing that.

The Fugro culture was open to that, we as a publicly traded company we value sustainability and have a sustainability program and we are trying to do the right thing from a corporate social responsibility stand point, and we saw that this is the way to be doing the right thing in a fairly low cost but impactful way. So that's were that started, and it has grown, now there are more vessels involved and a lot of data has been contributed over several years.

**And on why are we doing it?** We see that generally having more information, and a better understanding of our oceans is good for business is good for the planet it's good for the climate, all of those things. So this was a way that we could help with those efforts without spending a lot of money. So the motivation I think, first and foremost, the purpose of Fugro

is to create a safe and livable world, we saw this aligned with our purpose, we are focused on sustainability and ESG goals, and this is again kind of consistent with those, it is good for the brand, it's good to show Fugro is the leader in this area and being proactive. That was the motivation.

**How did you resolve/overcome them?** At the same time what was going on we embarked on a series of conversations with our clients were they own the data, we started trying to inform them of the opportunity of data sharing and would they be receptive to this and making this data, historical data, even at some lower resolution, not necessarily at the full resolution that it was collected, if those were made available.

I would say in general there was receptivity in those conversations, and we had some successes in terms of getting data contributed to the public domain. But in general while positive discussions there was not much action there, we recognised that it was challenging for Fugro to make this request, by that time the UN Decade of Ocean Science for Sustainable development has started and it was trying to encourage private sector participation and getting public access to privately owned and held data, we saw and opportunity for Fugro to help bridge that desire of the UN OD and the apparent interest of some of our clients, as end owners and users of the data.

That maybe if this no longer became a Fugro ask but it was a UN ask, and if it wasn't bilateral, but there was a committee of a working group that came together to address this topic as a collective, then looking at the challenges and barriers that may each be encountering in several parts of the world, or in certain sectors of the ocean, then maybe collectively they could address some of those challenges. That is where we are at now, we just recently, with the UN OD SS launched a corporate data group, and its function is to look at what you are looking at right now in terms of this environment and try to identify the barriers and challenges and try to work together, to try to address and mitigate those so that we ultimately move from the left side of your schema to the right.

**What are the barriers encountered?**

It is not an easy process, I am actually quite optimistic about this group and the members, it is not a large group, it is by invitation only, this should be very forward leaning companies on sustainability, they have to own ocean science data, and they have to be interested in trying to achieve this outcome. I am cautiously optimistic that by working together across all those various sectors, progress can be made. The goal is not to have only some companies donate data, but create this framework with mechanisms that would allow and facilitate this to happen.

You are referencing Creative Commons Data Licenses, is that known by all the end users of these data in all those sectors? If not then there is an education process requirement. There is an IP concern.

The goal of this group is to identify the major barriers, that are kind of consistent across data types and across geography and then start tackling them one at a time, an creating these mechanisms and frameworks that incetivise and motivate others to follow those best practices.

Part of this is historical data, so it data that has been collected, and could be made accessible, and part of it is the way we do business going forward. We see this does not have to be purely philanthropic data contributions, there could be a mechanism for coinvestment, public-private partnerships, where it is both in the interest government and industry. Because at the moment you have government who is collecting data, for regulatory and policy

purposes and you have industry who is collecting data to design, engineer and construct an infrastructure, they are not sharing the data currently, when in fact they could be using the same data to support multiple interests. Sure there can be some data that can be philanthropicaly donated, but what we really want is an environment where both public and private sector recognise and appreciate the need for both parties interest in the data, and have a mechanism, a coinvestment that could allow more, if they are working together.

These are big challenges to implement globally for sure, we had some successful public private partnerships with government in the US, would those mechanisms translate to other countries. Hopefully, with this group we can come up with those frameworks and mechanisms that other parties can use, at least have a model that we can adapt on a country or region basis. There is no standard or consistency globally, maybe regional or national basis, but it will take some time.

**What are the barriers you expect to move to the next level?**

That is why I appreciate your schema because it resonates with were we are were we want to be, and our desire to move towards level 3 in the schema. And to answer the question that was in the pdf. Fugro is largely in the left, where it is our clients data not even our data, but we are trying to move us and our clients and the industry in general to the right. Not neccesarily just by giving away data but creating incentives and motivation for this collaboration. If it is insdustry, there is profit requirements to the shareholders at the end of the day, so this has to be done consistently, with delivering a some level of profitability, but a positive outcome, for the business, the industry, society, all those things that corporations are expected now to deliver, it is not just profic, it is people and planet.

**How could you overcome those barriers to move to the next level?**

Fugro decided at some point that your visions is towards sustainability, and you are trying to achieve that, what was your motivation behind it, because I had conversations with other companies, that are so focused on maintaining their business model that they are not seeing the big picture? I think part of it is our CEO, our leadership, my experience is that our CEO is very sincere in his interest to deliver on the companies purpose on creating a safe and liverable world, this is always on his mind and he sees this iniciative and perticuraly, the UN OC , as an important mechanism to improve the way business is done as well as the future of our planet. I really believe that Fugro is a company led by our CEO, that is very sincere in this motivation, is why maybe Fugro has prioritized this when other companies haven't. He didn't have to, these are opportunities that were presented to him and the executive leadership team and they were embraced as important components of our sustainability program, but also as important elements of our strategy and culture. So I think at the end of the day, you can have your website with your words, but when the rubber hits the road, with your CEO and what they are willing to do and what priority they put on this versus other things. At the end of they day he has to still deliver profit it is a publicly traded company. I see that the company adopted this from very humble beginnings and embracing from the leadership of the company.

To the point were this corporate data group under the OD that I mentioned, he is personally co-chairing that group, he is personally involved along with the executive secretary of the intergovernmental oceanographic commision, to chair the meeting of that group. Fugro has loaned expertise or to the intergovernmental oceanographic commission of Unesco where they are sucondant to that secreteriat in Paris to organise and administer this working group and another data coordination group at Fugro's expense. they are working 100 per cent and being directed by the IOC and not by Fugro, in delivering on this wwork of the OD, that is

a financial investment that Fugro made beyond the collection of data and sharing that. That we think will hopefully improve the environment for public access to private data.

Speaking only for the bathymetry dataset: was it difficult in the process of making it open? No, the reason why this was quite successful was because there was already a place for it to go, that is another challenge that we have globally, in general there are a lot of silos for different types of data, those can be national or regional, there maybe a repository for surface current data in this country but not globally, or the repository of habitat exists in the EU but not globally. There is not really great standardization amongst data types and geographies, but in the world of Bathymetry there is a longstanding global organization called geBco, GENERAL BATHYMETRIC CHART OF THE OCEAN, that organization is 120 years old, and this organization has been the consolidator for public global bathymetric data. There was already a mechanism a place to put the data and then with Seabed 2030, which is a project of GEBCO and the Nippon Foundation, that was made to accelerate that data collection and compilation within that database. It was quite successful as it moved from 6 per cent now to 24 per cent global coverage.

**What is needed to move to the next level?**

The reason that I believe this worked out pretty well is there is already a structure in place thers was a mechanism even to get the data, the International Hydrographic organization, has the data center for digital bathymetry were anybody can upload bathymetric data and it ends up in this central GEBCO database. So it was easy for us to send our data to one place, not 12 or 50 places around the world, one portal and that portal serves the entire globe when it comes to bathymetry data. That database is mature.

## D.11. Microsoft Building Footprints

The building footprints are developed in the US, there is an AI model behind it, it is quite a resource effort so you need to let it run for a month. Maybe not accurately. It started for Africa, they had that model trained, and then re trained for different areas in the world, and it was retrieved for the USA, I don't know the reason why it started, invest in sanitation, access to clean water for everyone, first you need to know where it is, and with no maps that was not possible, disease come from regions where no people were leaving, of course there were houses, but that was not on the map. Then it became part of Microsoft research, that is the philanthropic part of the company where researchers can apply for a grant to do research. That is the thing with that company, you see things coming up, but you don't know where they come from and towards where they are going.

**Where do you think you are in this schema?**

There is a huge initiative to be mentioned, at first I looked through your chart, and I think that depends very much on which product or service you are talking about. I think we can depict a couple of them in different sections of this schema, but to focus more on the open side in the right side to level 3, let's check if it is level 3.

For the schema with Planetary computer we are at the right side level 3, we do have other services, through Azure space you connect your satellite fleet, but we don't have that many companies in the Netherlands, a satellite operator can communicate directly with a satellite through a service, you have the ground station that communicates to the satellite those can be earth observation satellites, but can also be communication satellites. Those companies

can be commercial or open and you can directly connect to them and retrieve the data and do the processing yourself, and you can get a little differently processed data or earlier that the regular dissemination of data.

We have a contract with multiple vendors that acquire the data, those are not open data, but you have to put an order but then we make it available in your own environment for these it is level 1 or 2 or even level 0 because it is not findable for anyone other than yourself. So it really depends on all those cloud environments which are customer specific, and customers can do anything with their data so then they are at the left side. They have data that are commercialized, they have a collaboration with TomTom with the traffic data this has been a service so it is not open, because this service is closed, you can pay for it and have it and then some money flows back to TomTom.

**Planetary Computer**

There is the initiative of the Planetary Computer, that is the investment that we made, originally with bridges from our sustainability goals, and one of them is about ecology and biodiversity, and you can't put your API on that topic, because it is quite hard, for example, we want to reduce the extinction to a number, because you don't know how many species that are extinct. The first assessment of biodiversity, the report more than 1000 pages, was delivered last year, which took 10 years of research. With the pace that we are moving with climate change on the effects on biodiversity, we can not wait for the next 10 years for the next assessment, we need to take action now with insight.

The Planetary Computer facilitates anyone, that can be researcher, NGOs, companies, governments in bringing together datasets, those are open datasets, not commercialized, with APIs on top of them to access and retrieve the data that you need, and that is quite difficult in the past, when you have a portal but then you need months to download all the data that you need. Besides the API, there is also the Planetary Computer Hub, this is a computer engine where you can do calculations against the data, in the same cloud environment, and it is free for use, that is the most open we can be, but we don't own the data, the data come from NASA, LISA, from organizations that share their data as open, and if it breaks down, organizations can not come to us and complain, we don't provide new data there. A year ago it was 5 petabytes, the idea was that if bringing these data together, facilitated with compute, anyone can make use to come to external insights than through other matters that you need to retrieve data yourself and download them, there are not that many organizations that can store that much data, and you don't need it all.

The homepage of the Planetary computer, datasets name, categorized by categories, biodiversity, climate, the dataset has the bands in which it was processed, provide an API, the whole thing is open source, you can see the data format that it was stored in, cloud processed .geotiff, you can browse the data catalog. Explore, like a geospatial portal, you can visualize the data, open need sign in with email.

With the option explore results in the Hub, you can have the python code, with the generated query of the data that has been selected here, and you can use it on your machine. You can ask for access with your email, and you will be granted access in 2 to 3 days, it can be used in open software, open source also R, QGIS, Python. It starts a compute engine for you and a Jupyter notebook for you to do the processing of the data.

**How did you arrive at this level in the schema?**

30 years ago Microsoft collaborated/ contracted ESRI to enable the spatial awareness of SQL database. Even that knowledge came from ESRI, always collaboration on a win win

situation, we have databases they have data why don't make it compatible, we also have competitors like oracle, you can also have the data in an oracle database, so it is a win win if they use our database, to make use of ESRI software, so they use our technology and connect to the technology.

**What are the barriers encountered?**

We are talking about petabytes. It is quite hard to ship it and to get it in the cloud at all actually, it is quite hard to get the data to the cloud, they shipped the hard disks to the data center, you need to have collaborations with these data providers. On one hand NASA and ECMWF, all those organizations are customers of us as well, so we have already had conversations with them, how they can improve their processing. If they are not customers, we started conversations with them on how to make this possible, years ago, on the documentations of some dataset, the names of other organizations, so we contracted ESRI to ingest the data, do some pre processing, so not all was done with technology of microsoft, but also technology of other organizations, then we need to have the license and agreements of ESRI to use the software.

**How did you resolve/overcome them?**

Nowadays, on the cloud, we need the processing capacity that the ESRI software enables, so we contacted them to make this possible, so there was a direct win for ESRI to help us. There is an ESRI SAS, the living atlas is also available on the planetary computer. The more data you have the more interesting the planetary computer is so Microsoft and Esri profit from that. That was about ESRI corp.

For ESRI Nederlands, they have communication with their dutch customers. ESRI Netherlands is a real independent company, with a license to sell ESRI licenses, Microsoft Netherlands is not an entity, the Microsoft Netherlands department focuses on the dutch customers. They have mutual customers and it is often a question how to integrate the ESRI software in the cloud in the best possible way, so it is for both a win win situation. If it runs remotely on the cloud they start using the cloud more, other organizations that already moved to the cloud, they ask how to deal with ESRI software and geospatial data.

I am personally focusing on digital twins as a topic, and it is more the business value that you drive, where do you wanna use digital twin, what is the usage, also more societal and sustainablity topics that you address, so those are the different layers that we have to consider.

**How did you learn about digital twins and if you had to persuade your department about it? What are the barriers you expect to move to the next level?**

Digital Twin is a hybrid, we have a digital twin service, this is an answer from the industry 4.0 market, how do you operate your factories and machines more efficiently. I saw the development there and I saw the development on the planetary computer and the sustainability topics, which is based on the philanthropic side, those two come together and now we have a service product azure space, and there it all comes together, the open data and the commercial data, so we have a contract with all the satellite companies that you can purchase data, and if you do it through our service it directly comes to your cloud environment and you can do your processing immediately after. Also, internally people are talking about digital twins and sustainability, and I tried to bridge that from the industry 4.0 it is kind of personal drive, I see a lot of things move in the market move internally, connecting the dots.

We are based on open source, portals had the difficulty that you can browse data, view it and download it takes time. Now luckily there are more standards on API level, and we implemented our stuck API, that you can query and search your data more efficiently, the API is open source and other organizations are implementing this API. you can use the same code to retrieve the same data with only changing the url, that is open source and they also invested on stack API.

**How could you overcome those barriers to move to the next level?**

We have organizations that need to create their data platform, has multiple departments, each departments needs to find data from other departments, the data platform design that is called a mesh, it is a federated way of disputing data, each department can produce a data product, and that will be on a data catalog, if they are interested in it they can ask for access to it .

Planetary computer, there are some national datasets for the US, I don't think there are any Dutch datasets, there is kind of a qualification to make available through the Planetary Computer, and I am not sure how much effort it will be for us, if it is a static dataset it is not that much work, but if it is Landsat, or things that need to continuously obtain data, then it is a lot more work to have the whole archive available on the cloud, but you can do a request on datasets, and then qualify through a team whether we think there a large use case for a turnout, and national dataset have smaller use cases than global datasets. We help other organizations like Rijkwaterstraat and kadaster to task to distribute their datasets, in the end it will be their design out of their ESRI environment, if they are using it, we help them make a solution so they can do that, but it is not us from Microsoft that we publish that, but each individual organization. Regarding your question about proprietary commercial data it is not much that we do, it is from our customers we are not publishing them. With the exception of the collaboration we have with a couple of them with a large weather company accuweather, tomtom and a couple of other interesting data creators were we think yes this is a value for our many, and then we build a service around it, then we open up this propriety data, but it is a service so it is not open, it is accessible for a fee.

**What is needed to move to the next level?**

I known that there have been individual conversations with some of the organizations. Also, very practically, who will pay for the download from their data center, each of these organizations we need it to pay, because of their efforts on their end, so there is a contract and with contracts there are all sorts of things to comply to and you agree to.

**Are you familiar with the license in level 2? Did they come up in your processes, are you following this framework or are you specifying it in each contract?**

Yes but barely. Microsoft does that a lot, me myself is not that familiar with them but I see it everywhere, in this case it depends on the organizations where it came from, and we need to apply the same, some times it is the MIT, sometimes it is the CC, but we need to put that in our own documentation, and our own agreements, on the github, on the code, that we publish . There are so many levels it can be on the code on the data on the service, I am far away from product development, we are a large company.

**In the process, it is important for the data to be interoperable, did you have difficulty with the different formats from the companies, are they following open standards?**

It depends on which organization, or department. here in the Netherlands we organized quite well the geo standards on the exchange level, and these are rules so for the formal

dataset this has been implemented well, and the exchangeable, but for many others, especially on the realm of digital twins it is far from open standards, we tried to help there with communities, to work through communities to come to standards, we are also part of the OGC, we are quite active there, also for the cloud optimized .geotiff for example, so how can you standarize your format so you can retrieve it quickly. In the industry the data standards are not standardized at all, in the construction.

The Rijkwaterstraat is building a bridge and the constructions offered a digital twin next to it, but what is the data model that it should be in, no one knows, there are many data models, but there are too many and no one agreed upon them. There are too many standards. We tried to optimize it to work together with communities, but it is the industry itself that needs to address that. and if they address it then our services get more used.

## D.12. AND - OSM

AND Interview Questions Aim of the interview: Identifying the barriers and incentives of the donation of the Netherlands street map to OSM.

**What was your role/involvement in the AND street map donation?**

As researchers in open data, we were attending many open data events in the timeframe of AND announcing their donation to OSM. We have discussed informally with AND employees their motivation for the donation.

**What was the legal background from the government side at the time of the street map donation to OSM?**

Around the time of the donation, 2007, there were within national government movements and discussions at the tactical administrative levels towards more open geo-data. Especially Rijkswaterstaat as part of the ministry of Infrastructure and Environment (infrastructuur en milieu, IandM) was a frontrunner on this. The Kadaster, as provider of the national topographic dataset, was in the same ministry responsible for the coordination of GI in the Netherlands. Rijkswaterstaat was sued for opening the NWB (national road atlasdatabase) by a company that had invested in developing a similar database and viewed publishing the NWB as unfair competition. but Rijkswaterstaat won the case as the company concerned could or would not provide an insight into their costs and lost income. The Court also ruled in favour As didof the ministry of IandM for opening the postcode dataset. As a result of the agreement on the INSPIRE directive in 2006 (enacted in 2007), the ministry was discussing with its primary stakeholders steps towards more open datasets including the topographic dataset. If this would be the case, then AND's dataset would have a competitor in the freely available topographic dataset that probably would result in the end of this part of AND's business model.

**How did AND react to the law about the topographic map TOP10NL?**

Rather than fighting the upcoming impact of the Act, AND responded pre-empted by providing their dataset under a CC-BY-SA to OSM. They had did only have had few a small number sales on of their dataset, and by donating the entire country-wide covering dataset, they obtained some free advertisement for their other products in the national press. In addition, and competitors using the AND dataset always have to give credit to AND and provide the new product also under CC-BY-SA from which AND may benefit.

**What where the incentives for AND to donate the street map to OSM?**

See above

**What were the barriers in the process of AND sharing the street map, based on your discussions with them?**

None it may even have saved them money as they no longer had to promote their database as a commercial product, thus reallocating those funds to other products and services.

*OSM: OpenStreetMap is built by a community of mappers that contribute and maintain data about roads, trails, cafés, railway stations, and much more, all over the world. **TOP10NL: TOP10NL is a digital object-oriented topographic file that underlies the topographic chart series 1:10.000 and 1:25,000 and is frequently used in various GIS and CAD systems for subsurface, analysis, and management and planning activities.

# Bibliography


(nd). Seabed 2030. https://seabed2030.org/. Accessed on March 23 2023.

Adair, A., Berry, J., Deddis, B., McGreal, S., Keogh, G., and Key, T. (1998). Barriers to data sharing in the surveying profession: implications for the commercial property market. *Journal of Property Research*, 15(4):331–346.

AIS (nd). AIS eLibrary. https://aisel.aisnet.org. Accessed on February 13 2023.

Awada, R. and Kiringa, I. (2015). Data Sharing and Exchange: General Data-Mapping Semantics. In *Beyond Databases, Architectures and Structures*, pages 159–169. Springer International Publishing.

Barry, E. and Bannister, F. (2014). Barriers to open data release: A view from the top. *Information Polity*, 19(1,2):129–152.

Berners-Lee, T. (2009). Linked Data, Is your Linked Open Data 5 Star? https://www.w3.org/DesignIssues/LinkedData.html. Accessed on February 13 2023.

Blogs, M. B. (2018). Microsoft releases 130 million building footprints in the usa as open data. https://blogs.bing.com/maps/2018-06/microsoft-releases-125-million-building-footprints-in-the-us-as-open-data. Accessed on April 29 2023.

Boone, F. and van Loenen, B. (2022). Identifying and overcoming the barriers towards open data of public undertakings∗. $Interdisciplinary Description of Complex Systems, 20(2) : 96 - -111$.

Boone, F. A. (2021). Third wave open data in the Netherlands: Identifying and overcoming the barriers towards open geographical data of public undertakings. mathesis, Utrecht University. https://studenttheses.uu.nl/handle/20.500.12932/39863.

Britannica (nd). Public sector. britannica.com/topic/public-sector. Accessed on March 1st 2023.

Cambridge Dictionary (nd). Private sector. https://dictionary.cambridge.org/dictionary/private-sector. Accessed on February 14 2023.

CCN (2019). Microsoft's stunning 44-year journey from ms-dos to the $1 trillion club. https : //www.ccn.com/microsofts - stunning - road - to - becoming - a - member - of - the - 1 - trillion - club/. Accessed on April 12 2023$.

Commission, E. (2018). Commission staff working document: Guidance on sharing private sector data in the european data economy. Brussels, 25.4.2018 SWD(2018) 125 final.

Commission, E. (2020). From the public sector information (psi) directive to the open data directive — shaping europe's digital future. https://ec.europa.eu/digital-single-market/en/public-sector-information-psi-directive-open-data-directive. Accessed February 14 2023.

Commission, E. (2022a). Data act. https://digital-strategy.ec.europa.eu/en/policies/data-act. Accessed on January 17 2023.

Commission, E. (2022b). European data governance act. https://digital-strategy.ec.europa.eu/en/policies/data-governance-act. Accessed on May 15 2023.

Cyclomedia (nda). About cyclomedia. https://www.cyclomedia.com/en/over-ons/about-cyclomedia. Accessed on April 14 2023.

Cyclomedia (ndb). Cycloramas. https://www.cyclomedia.com/en/producten/data-visualisatie/cycloramas. Accessed on April 14 2023.

Cyclomedia (ndc). Lidar point cloud. https://www.cyclomedia.com/en/producten/data-visualisatie/lidar-point-cloud. Accessed on April 14 2023.

Cyclomedia (ndd). Street ortho. https://www.cyclomedia.com/en/produkte/visualisierung-der-daten/street-ortho. Accessed on April 14 2023.

DIRECTIVE (EU) (2013). Directive 2013/37/eu of the european parliament and of the council of 26 june 2013 amending directive 2003/98/ec on the re-use of public sector information 2013(3), 1–8. https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:175:0001:0008:EN:PDF. Accessed on February 14 2023.

DIRECTIVE (EU) (2019). Directive (eu) 2019/1024 of the european parliament and of the council, on open data and the re-use of public sector information. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/
?uri=CELEX:32019L1024from=EN. Accessed on February 14 2023.

Durst, N. J., Sullivan, E., Huang, H., and Park, H. (2021). Building footprint-derived landscape metrics for the identification of informal subdivisions and manufactured home communities: A pilot application in Hidalgo County, Texas. *Land Use Policy*, 101:105158.

Edelstein, M. and Sane, J. (2015). Overcoming Barriers to Data Sharing in Public Health: A Global Perspective. *Chatham House*.

ESRI (nda). Arcgis living atlas of the world. https://livingatlas.arcgis.com/en/home/. Accessed on March 15 2023.

ESRI (ndb). Esri products. https://www.esri.com/en-us/arcgis/products/index. Accessed on March 15 2023.

European Commission (nd). The european interoperability framework in detail. https://joinup.ec.europa.eu/collection/nifo-national-interoperability-framework-observatory/european-interoperability-framework-detail. Accessed on July 10 2023.

Fassnacht, M., Benz, C., Heinz, D., Leimstoll, J., and Satzger, G. (2023). Barriers to Data Sharing among Private Sector Organizations. In *56th Hawaii International Conference on System Sciences*.

Fugro (nda). Our history and timeline. https://www.fugro.com/about-us/history. Accessed on March 21 2023.

Fugro (ndb). Our services. https://www.fugro.com/our-services. Accessed March 21 2023.

Fugro (ndc). Sustainability. https://www.fugro.com/about-us/sustainability-and-society/seabed-2030. Accessed on March 21 2023.

Fugro (ndd). What we do at fugro. https://www.fugro.com/about-fugro/group-overview/fugro-at-a-glance. Accessed March 21 2023.

GEBCO (nd). What we do. https://www.gebco.net/about$_u$s/*overview*/.*AccessedonMay*252023.

Geobusiness.nl (nd). Member overview. https://www.geobusiness.nl/ledenservice/ledenoverzicht. Accessed on February 27 2023.

Geodan (nd). Expertise. https://www.geodan.nl/expertisegebieden/. Accessed on April 30 2023.

GeoJunxion (2020a). About us. https://www.geojunxion.com/about-geojunxion/. Accessed on April 5 2023.

GeoJunxion (2020b). LOCATION INTELLIGENCE SERVICES. https://www.geojunxion.com/location-intelligence-services/. Accessed on April 5 2023.

Huang, X., Wang, C., and Li, Z. (2019). High-resolution population grid in the CONUS using microsoft building footprints. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Geospatial Humanities*. ACM.

I., C. and M., L. (2006). Semi-structured, narrative, and in-depth interviewing, focus groups, action research, participant observation. https://www.healthknowledge.org.uk/public-health-textbook/research-methods/1d-qualitative-methods/section2-theoretical-methodological-issues-research. Accessed on February 14 2023.

IEEE Explore (nd). Ieee explore. https://ieeexplore.ieee.org/Xplore/home.jsp. Accessed on February 13 2023.

Imagem (nda). About us. https://www.imagemnl.com/about-us/. Accessed on April 26 2023.

Imagem (ndb). Data automation from raw data to information. https://www.imagemnl.com/data-automation/. Accessed on April 26 2023.

Insight, A. (nda). Get the maximum out of your assets. https://assetinsight.nl/. Accessed on March 27 2023.

Insight, A. (ndb). Measuring train sherloc. https://assetinsight.nl/nl/diensten/spoor-en-wegmetingen/meettrein-sherloc. Accessed on March 27 2023.

Insight, A. (ndc). Our history. https://assetinsight.nl/nl/over-assetinsight/onze-historie. Accessed on March 27 2023.

Insight, A. (ndd). Our services. https://assetinsight.nl/nl/diensten. Accessed on March 27 2023.

Internet Archive Wayback Machine (nd). And and openstreetmap join forces to create digital map. https://web.archive.org/web/20070927051801/http://www.and.com/. Accessed on April 6 2023.

J., M. (2022). An introduction to exploratory research. https://www.scribbr.nl/research-methods/explorative-research/. Accessed on February 14 2023.

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

Laia, P. P. and Jonathan, W. (2020). The stickiness of scientific data. In *ICIS 2020 Proceedings*.

Liang, Q. and Wang, Z. (2019). Uncovering dominant social class in neighborhoods through building footprints: A case study of residential zones in massachusetts using computer vision.

Loenen, B. V., Zuiderwijk, A., Vancauwenberghe, G., Lopez-Pellicer, F. J., Mulder, I., Alexopoulos, C., Magnussen, R., Saddiqa, M., de Rosnay, M. D., Crompvoets, J., Polini, A., Re, B., and Flores, C. C. (2021). Towards value-creating and sustainable open data ecosystems: A comparative case study and a research agenda. *JeDEM - eJournal of eDemocracy and Open Government*, 13(2):1–27.

Maps, B. (nd). https://www.bing.com/maps/?cp=51.935375Accessed on April 12 2023.

Martin, S., Foulonneau, M., Turki, S., and Ihadjadene, M. (2013). Risk Analysis to Overcome Barriers to Open Data. *Electronic Journal of e-Government*, 11:348–359.

Microsoft (nd). The netherlands. https://www.microsoft.com/nl-nl/overnederland. Accessed on April 30 2023.

Nederland, E. (nd). Who we are. https://www.esri.nl/nl-nl/over-esri-nederland/wie-we-zijn. Accessed on March 14 2023.

OECD (2019). Oecd ourdata index: 2019. https://www.oecd.org/gov/digital-government/ourdata-index-netherlands.pdf. Accessed on February 21 2023.

Open Data Institute (nd). The data spectrum. https://www.theodi.org/about-the-odi/the-data-spectrum/. Accessed on February 14 2023.

Open Knowledge Foundation (nd). What is open? https://okfn.org/opendata/. Accessed on February 14 2023.

OpenStreetMap (2004). Openstreetmap provides map data for thousands of websites, mobile apps, and hardware devices. https://www.openstreetmap.org/about. Accessed on March 25 2023.

OpenStreetMap (2007). And donate entire netherlands to openstreetmap. https://blog.openstreetmap.org/2007/07/04/and-donate-entire-netherlands-to-openstreetmap/. Accessed on March 25 2023.

OpenStreetMap (nda). AND data OSM Services. https://wiki.openstreetmap.org/wiki/AND. Accessed on April 6 2023.

OpenStreetMap (ndb). And data/roadmap-integration-nl. https://wiki.openstreetmap.org/wiki/AND. Accessed on April 6 2023.

Parliament, E. (1997). Public undertakings and services in the european union, economic series w-21. https://www.europarl.europa.eu/workingpapers/econ/w21/sum-2$_e$*n.html*. *AccessedonFebruary*142023.

Richter, H. (2021). The law and policy of government access to private sector data ('b2g data sharing'). In *Data Access, Consumer Interests and Public Welfare*, pages 529–572. Nomos Verlagsgesellschaft mbH & Co. KG.

Safarov, I., Meijer, A., and Grimmelikhuijsen, S. (2017). Utilization of open government data: A systematic literature review of types, conditions, effects and users. *Information Polity*, 22(1):1–24.

Scopus (nd). https://www.scopus.com/search/. Accessed on February 13 2023.

Serena, C., Michael, G., Bruce, M., Mohamed-Ghouse, Z. S., Greg, S., Alexander, K., Nadine, A., Josef, S., Andreas, W., and Devarajan, T. (2021). Towards a sustainable geospatial ecosystem beyond SDIs. *EUROGI*.

SWECO (nda). Dgbgt. https://bgtsoftware.nl/over-bgt/. Accessed on March 26 2023.

SWECO (ndb). Obsurv. https://obsurv.nl/. Accessed on March 26 2023.

SWECO (ndc). Our Strategy. https://www.swecogroup.com/about-us/our-strategy/. Accessed on March 26 2023.

SWECO (ndd). 'state of your street': A climate label for every street. https://www.sweco.nl/portfolio/staat-van-je-straat/. Accessed on March 26 2023.

SWECO (nde). We are sweco. https://www.sweco.nl/wij-zijn-sweco/. Accessed on March 26 2023.

Tenopir, C., Palmer, C. L., Metzer, L., van der Hoeven, J., and Malone, J. (2011). Sharing data: Practices, barriers, and incentives. *Proceedings of the American Society for Information Science and Technology*, 48(1):1–4.

Tensing (nda). About us. https://www.tensing.com/en/about-us. Accesed on April 21 2023.

Tensing (ndb). Industries. https://www.tensing.com/en/industries. Accessed on April 21 2023.

van Loenen, B. (2018). The development of open data in the netherlands. In *Open Data Exposed*, volume 30 of *Information Technology and Law Series*, pages 215–232. T.M.C. Asser Press.

van Panhuis, W. G., Paul, P., Emerson, C., Grefenstette, J., Wilder, R., Herbst, A. J., Heymann, D., and Burke, D. S. (2014). A systematic review of barriers to data sharing in public health. *BMC Public Health*, 14(1).

van Veenstra, A. F. and van den Broek, T. A. (2013). Opening Moves – Drivers, Enablers and Barriers of Open Data in a Semi-public Organization. In *Lecture Notes in Computer Science*, pages 50–61. Springer Berlin Heidelberg.

Verhulst, S., Young, A., Zahuranec, A., Calderon, A., Gee, M., and Aaronson, S. A. (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. *SSRN Electronic Journal*.

W3C Recommendation (2008). SPARQL Query Language for RDF. https://www.w3.org/TR/rdf-sparql-query/. Accessed on February 13 2023.

Waterman, L., Casado, M. R., Bergin, E., and McInally, G. (2021). A Mixed-Methods Investigation into Barriers for Sharing Geospatial and Resilience Flood Data in the UK. *Water*, 13(9):1235.

Welle Donker, F. and van Loenen, B. (2016). How to assess the success of the open data ecosystem? *International Journal of Digital Earth*, 10(3):284–306.

World Wide Web Foundation (2017). The open data barometer. https://opendatabarometer.org/barometer/. Accessed on February 13 2023.

Zuiderwijk, A. and Janssen, M. (2014). Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly*, 31(1):17–29.

## Colophon

This document was typeset using LaTeX, using the KOMA-Script class `scrbook`. The main font is Palatino.