Achieving Sustainability Through Geodata: An Empirical Study of Challenges and Barriers

MARIA S. ANDERSEN

DANIEL T. M. PETTERSEN

SUPERVISOR

Ilias O. Pappas

University of Agder, 2020
Faculty of Social Sciences
Department of Information Systems

Preface

This master thesis is the culmination of a two-year long master's program (MSc) in Information Systems at the University of Agder (UiA). The study was conducted and written by two master's students with common interests within emerging technologies and sustainability.

The purpose of this exploratory qualitative study was to identify challenges with geodata using a data lifecycle approach with sustainability dimension in mind. The study also identifies and proposes initiatives that can help mitigate these challenges. The choice of research area – geodata and sustainability – was originally proposed by Norkart, a company specializing in geographic systems and services. When we initially reached out to Norkart we were looking for a provident research area that could encompass some of our primary interests. We assessed the combination of geodata and sustainability as being highly relevant and worthy of investigation. The crossroads between Information Systems research and Geography research was a particularly interesting area to focus on.

We would like to thank our advisor, Ilias O. Pappas, at the Department of Information Systems at UiA, for encouraging us, providing great feedback, and guiding us in the right direction. We would not have been able to pull this immense task through without his help, especially considering the ongoing COVID-19 pandemic. We would also like to thank Alexander Salveson Nossum, from Norkart, for proposing and encouraging us to take this topic and direction in our research.

Finally, a special thanks to all the informants that contributed with their expertise and knowledge towards the findings of this study.

Kristiansand,
June 4th, 2020

Maria S. Andersen

Maria S. Andersen

Daniel M Pettersen

Daniel T. M. Pettersen

Abstract

Research within data management is often based on the elements of the data lifecycle. Organizations and businesses are also becoming more interested in data lifecycle management to leverage their data streams, compounded by an interest in geographical attributes within the data – referred to as *geodata*. Geodata provides a richer basis for analysis and is increasingly important within urban planning. Furthermore, the pressure to achieve sustainability goals calls for improving the data lifecycle. The challenge remains as to what can be improved within the data lifecycle – with geodata as an important input – to achieve sustainability dimensions.

This thesis aims to exploratively investigate challenges in the data lifecycle with geodata, as well as understand the impact of such challenges for the sustainability triple bottom line. The research questions to be answered are:

1) RQ1: How do challenges within the geodata lifecycle impact the achievement of sustainability triple bottom line?
2) RQ2: How do organizations meet the data lifecycle challenges, and what initiatives can be implemented for achieving increased sustainability?

The study engages the qualitative research approach and uses the qualitative survey – semi-structured interviews – to gather data about the phenomenon. The results are analyzed with a thematic approach, where the data lifecycle and sustainable value creation are primary theoretical understandings for analyzing the results.

The study identifies several challenges within the data lifecycle with geodata. While some challenges are more typical to the data lifecycle in general, some challenges pose a high level of uniqueness. For example, conveying geodata to decision-makers, such as urban planners, is experienced as challenging, as well as the complexity of parameterizing geodata for sustainability goals.

Finally, we discuss initiatives identified in the literature and through informants that can mitigate challenges and potentially lead to better sustainability achievement. Some mitigating initiatives are increasing competence around geodata, enabling sharing and collaboration of collected data, or defining sustainability parameters clearly.

Our main contribution through this study is shedding light on challenges with geodata from an Information Systems (IS) and sustainability perspective. Additionally, the identified challenges are also feedback to data management research and the data lifecycle.

Table of Contents

Figure and Table List

Figure list

Table list

This page is intentionally left blank

1. Introduction

Ubiquitous information systems across all levels of society continue to impact the way sustainable development is approached. The Norwegian government has highlighted the importance of harnessing data for sustainable development in their one year closure of their 2030 Agenda for Sustainable Development (Utenriksdepartementet, 2018). At the same time, the government places responsibility within both the public and private sectors to ensure that Norway can reach the 17 goals outlined by the United Nations. Both sectors can engage in leveraging data and further developing solutions that tackle important sustainability paradigms.

The importance of data and information systems for sustainable development is further compounded by recent trends, such as smart cities, where ubiquitous computing and vigorous data collection is supposed to allow for more secure, efficient, productive, and sustainable cities. That is, data and metadata that is actionable and provides decision making support (Kitchin, 2014). One perspective for ubiquitous technology in smart cities considers that a cohesive and overarching understanding is required to become more efficient and sustainable (Hancke et al., 2013). Practical implementations include data analysis that leads to more sustainable behavior. Some examples are more effective public transport routes based on sensors that collect passenger and timetable data, or controllable systems that involve sensors and actuators to regulate, for example, water or electrical grids.

The key takeaway from the relationship between sustainable development and information systems is the overarching perspective for understanding urban contexts. While much of information systems research focuses on Big Data and the data lifecycle, smart cities, and IoT as significant factors for sustainable development, there is a need to understand further and leverage a geographic understanding of urban development. Kuster et al. consider lacking semantics, such as GIS (Geographic Information Systems), within current schemes for assessing urban sustainability (Kuster et al., 2020). This is important to consider because most urban planning and development have important geographic assessments.

Based on qualitative interviews, this thesis aims to understand the challenges in the data lifecycle, with *geodata* as the primary input. Since geodata plays an essential role in urban development, we will also look at how specific challenges in the data lifecycle related to geodata can affect achieving the sustainability triple bottom line. The qualitative study looks at eight different public and private entities in Norway - spanning 13 interviews - with a heavy focus on understanding their work with geodata in a data lifecycle context and identifying challenges that may be unique to geodata.

1.1 Background and Context

Sustainability is, in general, a rather broad and controversial term that is hard to define. Brown et al. proposed already in 1987 to define sustainability from a global perspective, with the anthropocentric view that involves the indefinite survival of the human species, ensuring the quality of life beyond mere survival and persistence of the biosphere with no human benefit (Brown et al., 1987). Even so, Brown et al. acknowledge the three dimensions in which sustainability is often viewed as economic, environmental, and social - also known as the sustainability triple bottom line (Milne & Gray, 2013). In any case, there seems to exist a generally accepted macro-level view of sustainability, but there is disagreement on the intricacies.

Within Information Systems (IS) research, there are calls for working towards sustainability-oriented research (Green IS), despite the broadness and potential vagueness of the term. Gholami et al. (2016) argue that IS academics should engage in impactful research for solutions to threats to global sustainability. Motivation is hindered by incentive misalignment, the low status of practice science, and scoping issues, among others (Gholami et al., 2016).

However, government policy in many countries points towards data-driven sustainability, and organizations are becoming more interested in correlating their data with sustainability. Some practical examples are management of water and ecosystems, weather prediction, financial instruments, and value from data about the natural world, adopting "information systems"-style management because of the data-driven nature (Etzion & Aragon-Correa, 2016). The growing relationship between data management and sustainability, both in research and in practice, provides a basis for analyzing phenomena such as geodata from a management and sustainability perspective.

Indeed, geodata research shows a slew of data-driven approaches that tackle sustainability issues, usually through urban development. Participatory mapping has been used to empower citizens (Eitzel et al., 2018), while Geographic Information Systems (GIS) are becoming more common for water management, and geospatial information can be used to provide more precise Building Information Models (BIM).

In Norway, historically, digital government initiatives and extensive standardization provide a wide variety of open geodata at several government levels (Flåthen, 2007). Additionally, Norkart - the informally commissioning company for this study - presents today's scenario as the private and public sector cooperating to increase geodata efforts, usually to address urban development and sustainability issues.

The combination of 1) increased focus on Green IS research, 2) a data-driven approach to sustainability, and 3) geodata's potential to address sustainability issues serves as a background for this study.

1.2 Problem Statement

Despite the existing potential within geodata to address sustainability issues, it is still a research area that is generally confined to research fields outside IS. When assessing sustainability in an urban context, Kuster et al. notably identify GIS and BIM as missing semantic aspects that should be considered for assessing sustainability (Kuster et al., 2020). Furthermore, while smart cities and similar research fields are strongly focused on data management, geodata research has only recently embraced this approach (Shifeng Fang, Xu, Zhu, et al., 2014). Therefore, we have conducted a qualitative exploratory study that:

1) accounts for geodata within a data management approach, using the *data lifecycle*,
2) identifies challenges with geodata for achieving sustainability within the data lifecycle approach.

1.3 Research Question and Objectives

The purpose of this qualitative exploratory study is to investigate the challenges that organizations face when working with geodata and whether these challenges can act as barriers for achieving sustainability measures. In order to address this purpose, the following research questions have been developed:

(1) RQ1: How do challenges within the geodata lifecycle impact the achievement of sustainability triple bottom line?
(2) RQ2: How do organizations meet the data lifecycle challenges, and what initiatives can be implemented for achieving increased sustainability?

The research questions take a holistic approach, to understand the fundamental connection between working with sustainability and geodata and identify challenges that could be tackled in different ways. Therefore, the main objectives for this exploratory study are to:

1. Examine and better understand the relationship between sustainability goals and geodata.
2. Identify challenges when working with geodata, with a lifecycle approach. Challenges may act as barriers to reaching sustainability.
3. Propose possible solutions to challenges presented that can ultimately lead to increased sustainability.

1.4 Rationale and Contribution

The researchers' motivation behind this study is to contribute to IS research addressing sustainability, seen from a geodata perspective. This study identifies challenges that are specific to geodata with a data lifecycle approach and attempts to further our understanding of the relationship between geodata and sustainability. Challenges with geodata that are not addressed can reduce the ability of organizations to achieve sustainability goals. This is because geodata is often used in urban planning contexts, where environmental, social, and economic factors must be addressed.

Traditionally, research within geodata - including GIS, geoinformatics, cartography, among others - are approached within their respective research fields. However, we have identified a need to approach geodata challenges with a *data lifecycle* approach - an approach that is traditionally found within IS research.

Our empirical findings uncover challenges for organizations when working with geodata that may have consequences when attempting to achieve sustainability goals. We focus on organizations in the private and public sectors that work with geodata and reflect on how the challenges can be addressed.

1.5 Research Approach

This study applied a qualitative research approach, with an exploratory research purpose to uncover the phenomenon. We adopted the interpretive research perspective to investigate the phenomenon in depth within its context. The technique for data collection was primarily qualitative surveys in the form of semi-structured interviews, with some basic document analysis to cross-check data from interviews and information about the selected organization. We analyze the interviews based on the data lifecycle, comprising three main phases: acquisition, processing, and preservation (Sinaeepourfard et al., 2016). Furthermore, we also account for value creation in the data lifecycle by integrating the sustainability triple bottom line (Dyllick & Muff, 2016).

1.6 Limitation and Scope

This study considers challenges with geodata in the data lifecycle impacting the sustainability triple bottom line. The objects of the study are organizations in the private and public sectors that work with geodata, focusing mostly on organizations that can be connected to urban planning and development. Norkart, a Norwegian company specializing in geographic systems and services, is the informally commissioning company for this study. We reached out to them, asking whether they had any possible research areas they would like us to investigate. We were motivated to contact them because of their geodata perspective and cross-disciplinary work. First, the study receives a basis from a literature review carried out between January and March 2020. Then, interviews were carried out between March and April 2020. Informants are limited to the Norwegian context and originate from Norkart's professional network, as well as contacts that some informants could provide within their networks.

Due to the ongoing COVID-19 pandemic, at the time of this writing, interviews were delayed and had to be carried out online, which somewhat affects the interview dynamics. The delay also caused a shortened and more intense analysis period. The researchers hoped to carry out Action Design Research, doing workshops with stakeholders to produce impactful research. However, due to the circumstances, workshops could not be carried out and remained only a secondary plan, in case there was a realistic possibility to carry out such workshops. Even so, this qualitative exploratory study still provides essential insights about geodata seen from a data lifecycle perspective.

1.7 Thesis Overview

Chapter 1 – Introduction gives an overview of the problem at hand and describes the research question.

Chapter 2 – Related Research presents and describes related research that provides further understanding of the 1) data lifecycle and 2) sustainability, specifically with geodata. Chapter 2 is presented as a brief literature review.

Chapter 3 – Theory describes the theoretical foundation of the study. Data management is understood through a data lifecycle approach, sustainability is explicitly defined, and geodata is explained.

Chapter 4 – Research Approach explains the rationale for choosing a qualitative exploratory study with an interpretive approach. Furthermore, data selection and analysis methods are described.

Chapter 5 – Empirical Findings presents findings from the fieldwork. The findings are a thematic analysis of interviews carried out this semester.

Chapter 6 – Discussion explains the findings from the researchers' perspective and theoretical lens.

Chapter 7 – Conclusion summarizes the study and contribution to IS research. Concluding remarks and reflection on future work is briefly provided.

2. Related Research

Ahead of the qualitative interviews, we conducted a literature review to get an overview of the existing challenges within the research field. To carry out the mapping study, we followed Kitchenham's guidelines (Kitchenham & Charters, 2007). The objective of this approach was to get an understanding of the existing research and challenges that we would later examine through the interviews. Through the search string and criteria described in Appendix 1: Literature Review Search Process we ended up with 26 relevant articles, from which we extracted the presented challenges and mapped out which phase of the data lifecycle they were associated with. Additionally, we found some potential connections between these challenges and achieving sustainable advantages within the triple bottom line, consisting of economic, environmental, and social parameters.

2.1 Challenges Within the Data Lifecycle

Most of the challenges discussed in the articles apply to all phases in the data lifecycle. However, they may relate to the phases in different ways. Table 1 identifies the challenges that each article addresses.

Table 1: Challenges from the Literature Review

Challenge	**References**	**%**
Knowledge silos	(Nimmagadda et al., 2017), (Kuster et al., 2020)	7,7 %
Data intensity	(Yang et al., 2011), (Pijanowski et al., 2014), (Krämer & Senner, 2015), (Pons & Masó, 2016), (Jensen et al., 2015), (Lu et al., 2019), (Ardissono et al., 2017)	26,9 %
Computing intensity	(Yang et al., 2011), (Pijanowski et al., 2014), (Sharma et al., 2018)	11,5 %
Spatiotemporal intensity	(Yang et al., 2011)	3,8 %
Monitoring	(S. Fang et al., 2014), (Stefan et al., 2018), (Huang et al., 2019)	11,5 %
Cross-domain impact	(Athanasis et al., 2018), (Claramunt & Stewart, 2015), (Lu et al., 2019), (Shook et al., 2018)	15,4 %
Compatibility	(Granell et al., 2010), (Shifeng Fang, Xu, Pei, et al., 2014), (Sharma et al., 2018)	11,5 %
Accessibility	(Yang et al., 2011), (Granell et al., 2010), (Shifeng Fang, Xu, Pei, et al., 2014), (Sharma et al., 2018), (McGlinn et al., 2017), (Athanasis et al., 2018),	26,9 %
Standardization	(S. Wang et al., 2013), (Shifeng Fang, Xu, Pei, et al., 2014), (McGlinn et al., 2017), (Athanasis et al., 2018)	15,4 %
Complexity	(Shifeng Fang, Xu, Pei, et al., 2014), (Krämer & Senner, 2015), (Lu et al., 2019), (Stefan et al., 2018)	15,4 %
Modelling	(S. Fang et al., 2014), (Lü et al., 2019)	7,7 %
Interdependency	(McGlinn et al., 2017)	3,8 %

Security	(McGlinn et al., 2017)	3,8 %
Intellectual property	(McGlinn et al., 2017)	3,8 %
Quality	(Athanasis et al., 2018), (Hong & Huang, 2017)	7,7 %
Selecting datasets	(Hong & Huang, 2017)	7,7 %
Uncertainty of data	(Hong & Huang, 2017)	3,8 %
Error checking	(Pijanowski et al., 2014)	3,8 %
Management of multiple executions	(Pijanowski et al., 2014)	3,8 %
Integration architecture	(Pijanowski et al., 2014)	3,8 %

These challenges have been placed within the data lifecycle, as shown in Figure 1 below.

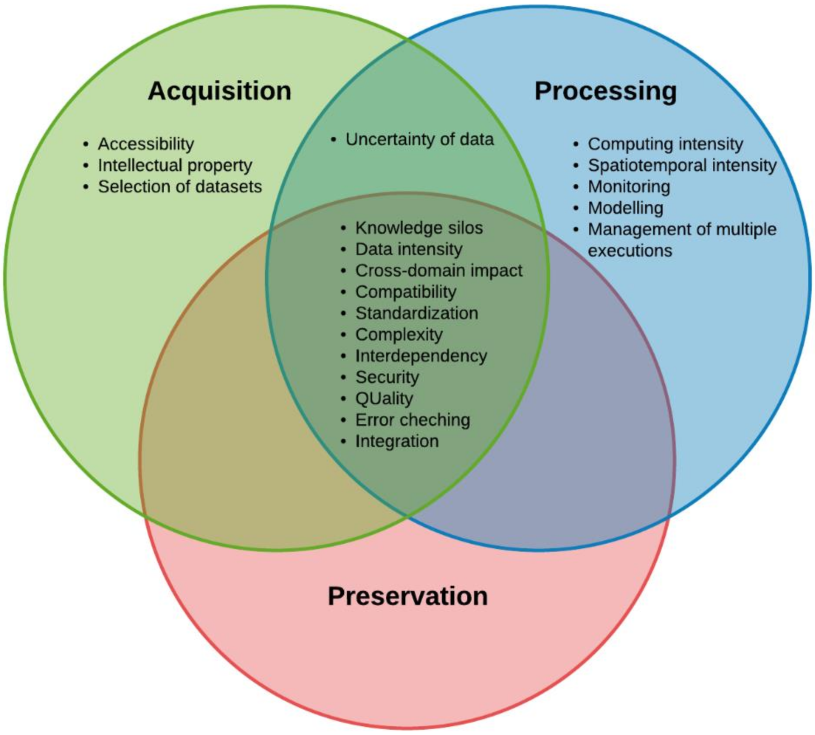


Figure 1: Challenges Within the Data Lifecycle.

2.1.1 Acquisition

Most of the challenges with data acquisition are applicable within data processing and preservation. Poor accessibility due to knowledge silos can prevent stakeholders from acquiring necessary data because of lacking collaboration and data sharing (Nimmagadda et al., 2017). Data intensity increases the complexity of the data because of the massive amounts and variety of data and makes it harder to collect the right data and discard noise (Krämer & Senner, 2015; Pijanowski et al., 2014; Yang et al., 2011). The cross-domain impact of working with geodata and the data lifecycle refers to the process impacting domains outside of only geodata related problems. This implies that competency is required across disciplines, which can be challenging. Compatibility issues can cause loss of some data sources due to incompatible formats compared to what the information systems can handle (Shifeng Fang, Xu, Pei, et al., 2014). Standardization is also an important aspect, closely linked to compatibility. Lack of standardization can lead to compatibility and communication issues, which emphasizes the need for standards, strategies, and policies to attain data (Barik et al., 2018). Some challenges mainly apply to the acquisition phase, such as accessibility of data, intellectual property issues, and difficulties with selecting datasets.

2.1.2 Processing

Processing is often problematized, and all of the articles from the literature review mention it to some degree. Data intensity, computing intensity, and spatiotemporal intensity make processing more complex, which increases the requirements of both processing software and competence in processors (Jensen et al., 2015; Krämer & Senner, 2015; Pijanowski et al., 2014; Pons & Masó, 2016; Yang et al., 2011). Monitoring processes allow for real-time insights on what is happening but requires efficient designs and architectures to reduce delays and obtain accurate monitoring within, e.g., disaster monitoring, resource allocation, and command decisions (Huang et al., 2019; Stefan et al., 2018). Modeling of geographic-featured environments and other rich data can be challenging due to the variety of formats and attributes (Lü et al., 2019). Several other challenges require complex information systems, processing tools and competence, such as standardization issues, integration, cross-domain impacts, compatibility issues and interdependencies (Athanasis et al., 2018; Barik et al., 2018; Shifeng Fang, Xu, Pei, et al., 2014; Shook et al., 2018; S. Wang et al., 2013).

2.1.3 Preservation

None of the reviewed challenges are exclusive to data preservation. However, several of them are uniquely adaptable to this block. Data can be stored long-term or with the intent to process it again. The original data format or the given format after processing the data needs to be compatible with the data storage platform, in order to keep all attributes and not compromise the quality of the data (Shifeng Fang, Xu, Pei, et al., 2014; Granell et al., 2010). Security also affects preservation, with both GDPR and an increased focus on protecting valuable data from unauthorized access.

2.2 Sustainability impact

The three sustainability dimensions mentioned initially are all affecting and affected by the challenges within the data lifecycle. Some of the challenges are directly preventing sustainable development, while some have minor impacts. We will elaborate on some of the examples that impact each of the dimensions of sustainability.

2.2.1 Economic Impact

Economic sustainability can include getting a return on equity, seeing stock market growth, increased market share, profitability, and new market opportunities through innovation (O'Brien, 2015). Businesses that are able to innovate or acquire substantial market shares can, therefore, obtain economic advantages. Several of the mentioned challenges can work as bottlenecks for innovation, such as lack of accessible datasets to build analyses upon, data complexity that requires more human resources or software to solve, low data quality, etcetera. If resources are used on solving problems connected to necessary operations, less is likely to be spent innovating and creating new products and services.

2.2.2 Environmental Impact

Environmental issues such as climate change have received much attention in recent years, and environmental monitoring, modeling, and management enable us to gain a deeper understanding of natural environmental processes (S. Fang et al., 2014). Environmental informatics has significantly improved since it originated, along with the development of environmental information systems (EIS), remote sensing (RS), geographical information systems (GIS), global positioning systems (GPS), etcetera (S. Fang et al., 2014). These developments allow for automated, rapid, and more efficient data acquisition, processing, and preservation. However, much of the effort is implemented to locate and analyze environmental problems rather than to solve them (S. Fang et al., 2014).

2.2.3 Social Impact

Social sustainability aims to ensure positive social benefits (O'Brien, 2015). Safety is of great importance within this dimension, and digital safety can be translated to cyber-security, which makes security-related challenges highly relevant to social sustainability.

Additionally, stakeholder engagement and accountability are aspects of this dimension that can be influenced by accessible data, cross-domain impact, compatibility, and complexity in different ways.

3. Theory

We will now present the theoretical foundation of this study. The data management perspective is addressed through the data lifecycle model, while sustainability is understood through the triple bottom line. Finally, geodata is accounted for through the understanding of geospatial data.

3.1 Data Lifecycle Model

The increased focus on creating value from data requires data management approaches where the whole data lifecycle is accounted for. Data lifecycle models provide a high-level framework to plan, organize, and manage all aspects of data during their life stages (Sinaeepourfard et al., 2016). We reviewed data lifecycle models to find an adaptable and basic one, that cover all the phases in the lifespan of data. We also need a model that allows for managing and organizing data and improving quality in all phases. The COSA-DLC model is the one that best fits this description, and is not tailored to any specific environment, but easy to be adapted to fit the requirements of any particular field (Sinaeepourfard et al., 2016). Additionally, it is created based on a review of other main DLC models and their limitations. The three main blocks are Acquisition, Processing, and Preservation.

3.1.1 Acquisition

Data is gathered through the Data Acquisition block, which collects data from different sources, assesses quality, and tags it with descriptions required in the model (Sinaeepourfard et al., 2016). The Data Acquisition block consists of four phases, namely, Data Collection, Data Filtering, Data Quality, and Data Description.

3.1.2 Processing

The Data Processing block is responsible for performing any data to information/knowledge/value transformation through analysis and/or analytical techniques (Sinaeepourfard et al., 2016). This block consists of the phases: Data Process, Data Quality, and Data Analysis.

3.1.3 Preservation

The Data Preservation block is responsible for data archiving, and the data can then be prepared for publication or dissemination. The data may be used by end-users or further processed (Sinaeepourfard et al., 2016). The Data Preservation block consists of the phases: Data Classification, Data Quality, Data Archive, and Data Dissemination.

3.2 Sustainability Triple Bottom Line

In general, sustainability is a rather broad and controversial term that is hard to define. As Brown et al. point out already in 1987, the term sustainability has gained traction within environmental policy and research and requires a more specific definition (Brown et al., 1987). In order to define the scale of the term sustainability, it needs to be contextualized with temporal and spatial factors in mind. Brown et al., proposes global sustainability, which carries an anthropocentric view, comprising the indefinite survival of the human species, ensuring the quality of life beyond mere survival and the persistence of the biosphere with no apparent human benefit (Brown et al., 1987).

Within the overarching sustainability perspective, there is a view comprising economic, environmental, and social sustainability, today known as the triple bottom line (Milne & Gray, 2013). It is argued that this perspective is primarily used by businesses for reporting purposes and that, in order to achieve the triple bottom line truly, there needs to be "ecological literacy." Furthermore, there is a potential contradiction between increasing profits and practicing sustainability (Milne & Gray, 2013).

For this study, it was, therefore, quintessential to contextualize sustainability. Primarily, we consider the idea of business sustainability (BST) (Dyllick & Muff, 2016) and the ability to manage the triple bottom line – BST 2.0. The ideal to be achieved is BST 3.0, which focuses on increasing positive impact instead of just minimizing the existing negative impacts of a business. We refer to sustainability through its three dimensions – originally from the triple bottom line – but BST is a more suitable contextualization for businesses. Additionally, motivation is found through Malhotra et al. (2013) & Gholami et al. (2016) to address sustainability within information systems research. Solution-oriented information systems research has the ability to address specific sustainability issues and create impactful research (Gholami et al., 2016). However, addressing all three dimensions of the triple bottom line simultaneously is still an insurmountable problem that information systems alone cannot tackle.

3.3 Geospatial Data

Awange & Kiema (2019) explains that "data is distinguished as geodata (or geospatial data) if it can be geographically referenced in some consistent manner using, for example, latitudes and longitudes, national coordinate grids, postal codes, electoral or administrative areas, watershed basins, etcetera." Mapping has traditionally been managed by state organizations. However, that changed towards civilian National Mapping and Cadastre Agencies in the 20th century, mainly responsible for land administration, infrastructure planning, or environmental monitoring (Heipke, 2010). The process of map-making has until recently been long, complicated, and resource-demanding. Goodchild (2008) elaborates that "accurate positioning required considerable skill in the use of photogrammetric techniques, expensive equipment for observation and analysis, and substantial investment in large-format printing. Today, however, almost all of these constraints have been removed." Heipke (2004) estimates that some 80 percent of our daily decisions rely on geospatial information. Geographic information systems (GIS), which facilitate acquisition, storage, manipulation, analysis, visualization, and dissemination of geodata, are, therefore, of prime interest to the society (Heipke, 2004). Lee & Kang explains the different forms of geospatial data:

> Traditionally, geospatial data can be categorized into three forms: raster data, vector data, and graph data. First, the raster data include geo-images typically obtained by unmanned aerial vehicles, security cameras, and satellites. The raster data is being provided by digital map services, e.g., Google Earth. Data analysts extract the tracks of moving objects or useful features from these raster data. Representative use cases include life pattern mining and change detection. Second, the vector data consists of points, lines, and polygons. The map data belongs to this form, and there are various data sources. Representative use cases include detection of hot spots and spatial correlation patterns. Third, the graph data mainly appears in the form of road networks. Here, an edge represents a road segment, and a node represents an intersection or a landmark. The trajectories of vehicles on the road network are represented by sequences of road segments (Lee & Kang, 2015).

4. Research Approach

The study aims to investigate challenges in the data lifecycle with geodata that can affect achieving the sustainability triple bottom line. To this end, the study proposes two research questions:

(1) RQ1: How do challenges within the geodata lifecycle impact achievement of sustainability triple bottom line?

(2) RQ2: How do organizations meet the data lifecycle challenges, and what initiatives can be implemented for achieving increased sustainability?

This chapter describes the research approach for the thesis (Figure 2). Additionally, a rationale is discussed for the research approach and the methodology of expert interviews. The data collection method is then presented, followed by the data analysis method. Finally, we conclude with issues of validity and ethical considerations.

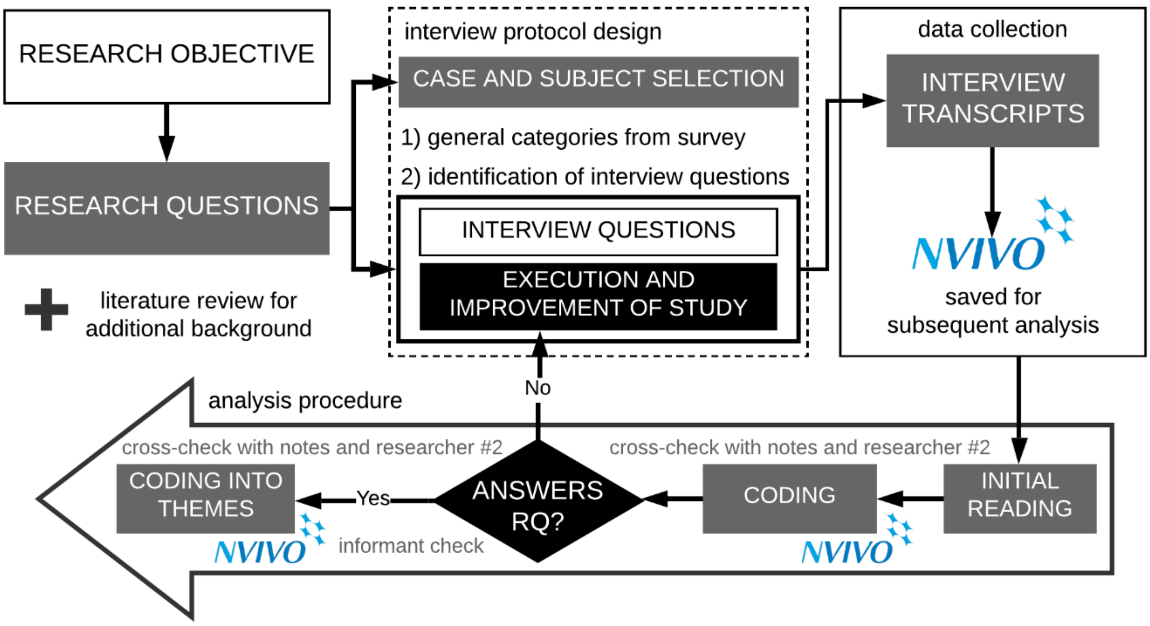


Figure 2: Research Process, based on Thomas (2006), Cruzes & Dybå (2011) and Berg et al. (2020)

4.1 Rationale for Qualitative Research Approach

Data lifecycle research within IS typically focuses on management and technology aspects, usually with a focus on processes. Research that focuses on these processes often dives into the context of the phenomenon and attempts to identify factors or model processes according to the empirical findings. Qualitative research aims to understand phenomena within their context (Robson, 2002). Depending on the knowledge requirements of the study, the approach can either take a deep dive within a few case studies or take a broader scope as a qualitative survey (Robson, 2002; Andersson & Runeson, 2002; G. Walsham, 1995). The main goal of the study is to identify challenges in the data lifecycle with geodata, within organizations that engage with geodata. Typically, these organizations work with urban planning, both in the public and private sectors. Since the data lifecycle approach is mostly uncharted territory – specifically regarding geodata – it is sensible to take a broad approach to identify challenges that may not have been identified in prior research. Furthermore, it is important for this study to understand the relationship between the identified geodata challenges and the ability to achieve sustainability goals. To this end, a qualitative survey is suitable. Within this approach, Robson categorizes four research purposes (Berg et al., 2020; Robson, 2002):

- Exploratory – understand what is happening; gain new insights.
- Descriptive – portraying a situation or phenomenon.
- Explanatory – seek an explanation of a problem or situation, mostly as a causal relationship, although not always.
- Improving – try to improve an aspect of the studied phenomenon.

The organizational and business context of our interviews, combined with the exploratory nature of our research questions, results in this thesis being an explorative study, researching several organizations, such as consulting companies, government agencies, local government, and research institutes, among others.

There are three different research perspectives to account for when designing the qualitative research approach: positivist, critical, and interpretive (Myers & Newman, 2007). The positivist perspective applies if there are quantifiable measures of variables, hypothesis testing, or evidence of formal propositions (Klein & Myers, 1999; Orlikowski & Baroudi, 1991). Fundamentally in positivism, reality exists independently from human experiences and in an objective manner (Chen & Hirschheim, 2004). Critical research encourages social change by critically exposing illusions and contradictions of social existence (Richardson & Robinson, 2007). Finally, the interpretive perspective argues that reality is a subjective, social construction by human actors (Geoff Walsham, 2006). The phenomenon in question, in other words, cannot be seen independently from social actors – including the researcher – in order to make sense of that reality (Orlikowski & Baroudi, 1991).

We adopt the interpretive perspective for this study, because we investigate a phenomenon – challenges in the geodata lifecycle – within an organizational context, integrated with sustainability goals in mind. We view the different elements of this study as interconnected and influenced by the social contexts in which they exist.

Qualitative Research Methodology

The qualitative survey is the most appropriate methodology for this qualitative interpretive approach. According to Myers (living version), the qualitative survey is suitable when the researcher is trying to study social and cultural phenomena (Myers, 1997). Some examples of data sources for qualitative research include observation, interviews, questionnaires, the researcher's impressions and reactions, documents and texts, among others (Myers, 1997). Many recommendations have been proposed over the years to address challenges with qualitative methods, such as clearly defining the researchers' role and involvement, overcoming social issues within interviews, ensuring accurate results, and addressing bias or interpretation issues (Geoff Walsham, 2006). However, there is some criticism regarding the degree of subjectivity in qualitative research, at least when considering the interpretive approach (Myers, 1997), although qualitative research can also be critical and positivist.

Regardless, the interpretive perspective in this qualitative research is one of many philosophical assumptions and guides the methodology of this study. Our research question calls for exploring a phenomenon that is influenced by its context, making a qualitative survey acceptable for diving in-depth into the problem at hand.

4.2 Informant Selection

Informants for the interviews were collected through a theoretical sampling, a sampling method closely related to grounded theory (Glaser & Strauss, 1967). A theoretical sampling, according to Glaser & Strauss, generates theory by collecting, coding, and analyzing data to decide where to collect data next (Glaser & Strauss, 1967). To have a starting point for data collection, we reached out to Norkart, the informally commissioning company for this study. Since Norkart specializes in geographic systems and services, they have a vast network with people who work with geography and GIS-related jobs. Additionally, we set some criteria to leverage Norkart's contact network:

- Organizations or businesses in the public or private sectors.
- Organizations or businesses that work with geodata.
- Organizations or businesses that are related to urban planning and development, or other societal aspects, relevant to geodata.
- Roles in the organizations or businesses that work with geodata or sustainability aspects within their contexts.

The sample in the survey, shown in Table 2 comprises eight organizations and businesses in the public and private sectors and 14 informants. The organizations in the public sector can range from local government to cartography or government agencies. Private sector organizations include consulting companies and research entities that have specializations in GIS and geodata, among others. Related to all organizations is the utilization of geodata for tasks further in the supply chain related to urban development, in some form or another. A government agency such as a taxation authority is, however, more interested in ensuring correct tax in any given area, meaning that kind of government agency has no particular direct effects on urban development. Still, it has significance for economic and social factors in urban contexts.

Table 2: Informant and Organization Descriptions

ID	Profession	Sector	Nº of employees	Area
1	Researcher	Private	50-100	Transport
2	Data Manager	Public	500-1000	Mapping
3	Researcher	Private	50-100	Transport
4	Data Manager	Public	8,000-10,000	Planning and Building
5	Data Manager	Public	8,000-10,00	Planning and Building
6	Engineer	Private	100-300	Consulting Company
7	Project Manager	Private	1,000-3,000	Consulting Company
8	Engineer	Public	5,000-8,000	Transport
9	Analytical Advisor	Public	8,000-10,000	City and Community
10	Enterprise Architect	Public	8,000-10,000	IT Unit
11	Analyst	Public	5,000-8,000	Taxation
12	Section Manager	Public	8,000-10,000	Mapping
13	Environmental Advisor	Public	8,000-10,000	Environmental Unit
14	Researcher	Private	50-100	Transport

In the case of local government, we required more than one interview within the organization to gain insights about sustainability aspects. This also means that across the organizations, we selected people with varied backgrounds, from more technical GIS-related roles to leadership and HR roles higher up in the organization. A total of 14 informants were selected across the eight organizations, spanning 13 interviews. Only one interview was carried out as a group interview (two informants simultaneously).

An important factor for the theoretical sampling is allowing for recommendations from our initial respondents, for possible organizations that we could interview. This method would prove useful to quickly identify highly relevant interviewees, as our initial informant baseline was general.

From the beginning, we decided to ensure anonymity with the organizations that we were in contact with. The aim of this study is primarily to identify challenges with geodata and analyze the sustainability effects through a thematic analysis. Therefore, none of the interviewed respondents or organizations are identifiable in this study. Later, some of our interviewees would also express their wish to remain anonymous, which for this study is perfectly acceptable.

4.3 Data Collection Techniques

This study primarily applies semi-structured interviews for data collection, although some document analysis is also performed to cross-check data from interviews and information about the selected organizations. Semi-structured interviews allow the respondents to provide additional insights that are not necessarily accounted for in the interview protocol. As Fontana & Frey point out, semi-structured interviews do not impose categories as opposed to fully structured interviews, allowing for new categories to be discovered and broadening the inquiry (Fontana & Frey, 2000; Myers & Newman, 2007).

4.3.1 Semi-structured Interviews

Interviews are commonplace in most interpretive studies and are essential for gathering interpretations from informants in the field (Geoff Walsham, 2006). Semi-structured interviews usually provide higher interactivity, enabling an understanding of how individuals react to their context (Iyamu, 2018; Tsang, 2014). Iyamu helpfully highlights prior research that shows the ability of interviews to be able to study complex phenomena and shared practices by exploring detailed data through interaction. Additionally, interaction and varied responses resulting from semi-structured interviews can reduce the risk of bias (Iyamu, 2018; Marshall et al., 2015). When trying to uncover and explain a phenomenon, an in-depth and detailed explanation becomes an essential priority for this study.

The semi-structured nature of the interviews require some predefined questions, but allow delving into unplanned follow-up questions and gain new insights. This approach also enables some predefined categories that will later be used for the results and analysis. However, many new categories will also be identified as a result of the interviews.

The interview opens with a short introduction to our study and practical information about data processing and consent (see Appendix 2: Consent Form). The questionnaire is originally in Norwegian but is translated to English for the adequate readability of this study. The following questionnaire guided the data collection process:

Section 1: General information about the respondent

1. Position or role in the organization, work experience in the organization.

Section 2: Introduction

1. What do you work with on a daily basis? Have you partaken in a relevant data management project?
2. What was/is the goal of the project?
3. Did you have any challenges?

Section 3: Data Management

1. What kind of data do you use in your work/project?
2. What kind of data sources do you use? Internal/external?
3. What kind of formats do you manage? Is the data structured/unstructured, semi-structured?
4. How is data collected? / Who do you communicate the data to?
5. How is the data saved?
6. How is data processed?
7. How is data analyzed?
8. How do you take advantage of the data?

Section 4: Socio-technical Aspects

1. In your organization, do you experience that you receive the information you require when you need it?
2. Do you experience being able to understand the available data? Yes/no, if so, why?
3. Do you see unused potential in the data that your organization has? If so, can you give some examples?
4. Do you have other potential challenges that you experience as a bottleneck?

Section 5: Sustainability

1. What relationship do you or your company have to sustainability goals or the UN's sustainability goals?

2. What do you consider the biggest challenges related to sustainability?
3. Which sustainability dimensions do you consider relevant for your organization?
4. Which parameters do you use to measure development/progress?
    a. OR: How do you follow up on your development/progress?
5. Do you see any relationship between taking advantage of data in your project and achieving more sustainability?

Section 6: Closing-up

1. Do you have any final remarks?
2. Follow up on documents or other potential respondents, and thank the respondent for the interview.

The questionnaire takes a data management approach because the study aims to identify challenges when working with geodata in a lifecycle context. Geodata lifecycle challenges may differ from traditional data management challenges, but typical IS processes, and IT infrastructure has become relatively commonplace in many disciplines. To most questions, we follow up by asking whether the respondents can identify possible challenges relative to – or to an approximation of – each phase in the data lifecycle. The semi-structured approach allows us to follow up at each phase or question if we identify something that requires a more detailed explanation or could contain some reasonable relevance to the study.

Both researchers were in direct contact with the subjects, allowing for a more demanding first-degree data collection technique. With both researchers present at every interview, it is possible to mitigate one single interpretation and allow for follow-up questions to cover the necessary ground and explore other directions (Runeson & Höst, 2008). To provide data for the analysis, all interviews have been recorded, to be transcribed for the analysis stage.

Important Considerations About Limitations with Interviews

There are some inherent limitations with interviews in general and the semi-structured interviewing technique. Myers & Newman summarize a wide variety of problems and pitfalls with interviews that can affect the quality of the study. Some social aspects that have a negative effect on the interview quality are (Myers & Newman, 2007):

- *Artificiality of the interview, lack of trust and time* – Respondents are strangers to the interviewers, which makes the social situation uncomfortable. Participants may not initially trust the interviewers, limiting their responses. Time pressure can cause incomplete interviews, or participants generating new but unreliable opinions.

- *Level of entry and elite bias* – Starting with interviewing people at lower levels may later make it harder to interview senior managers. Gatekeepers may inhibit access to information. Additionally, a researcher may interview only high ranking members in the organization, underrepresenting other views in other parts of the organization.
- *Hawthorne effect and constructing knowledge* – Interviewers are not invisible or neutral, and have the ability to interfere with people's behavior. Interviewers may also be collecting data that already exists, and respondents can construct stories to appear knowledgeable and rational.

Other pitfalls are also *language being ambiguous* and that *interviews can fail altogether*. These negative aspects have all been experienced to a greater or lesser degree by the researchers of this study. However, it is crucial to be aware of these pitfalls in order to manage them in the social situation that the interviewers and respondents exist within. These problems and pitfalls are limitations that arise, in addition to the limitations discussed earlier in qualitative studies.

4.4 Analysis

We followed an inductive approach for analyzing results data (Figure 3), specifically in the form of thematic analysis (Thomas, 2006). There are three purposes for a general inductive analysis approach (Thomas, 2006):

1. Condense raw data into a summary format
2. Establish clear links between research objectives and summary findings.
3. Develop a model or theory about the underlying structure.

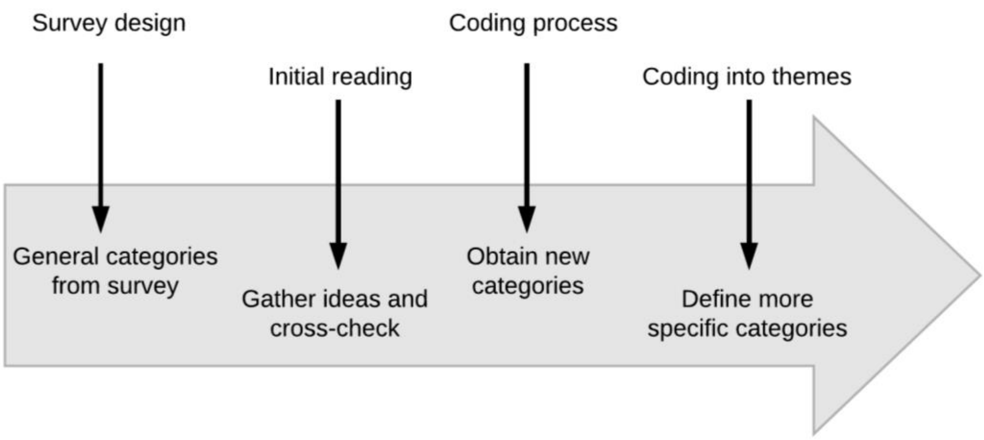


Figure 3: Analysis Process, based on Thomas (2006) and Cruzes & Dybå (2011)

There are some main principles for a general inductive approach, as described by (Thomas, 2006). First, the analysis is guided by the evaluation objectives, but multiple readings and interpretations of raw data satisfy the inductive component. The objectives provide focus, but not expectations about findings. Second, the analysis is primarily based on developing categories from the raw data into a model containing themes and processes identified by the researcher. Third, multiple interpretations are accomplished to arrive at the results, where findings are generally shaped by the researcher's perspective. The researcher must decide importance, or lack thereof, within the data. Fourth, different researchers may come to non-overlapping, distinct findings. Finally, the trustworthiness of such an analysis derives from other kinds of qualitative analysis (Thomas, 2006).

The inductive approach with a thematic analysis serves its purpose to answer the research question and provide a model of themes that describe challenges in the geodata lifecycle and its effects on sustainability. The Nvivo software, by QSR International, provides a simple way to systematically approach the analysis and coding process (Bazeley & Jackson, 2013). We will now explain each step of the analysis procedure.

Initial Reading

After the raw data files are cleaned, they are read by the researchers to find patterns and get some general ideas. Notes from the interviews are also a way of supplementing the initial reading. One of the researchers always took notes during the interviews as a backup measure and to have a cross-checking mechanism.

Coding Process

The study already provides some overarching codes – or categories – based on the data lifecycle. Within each main category, we used descriptive coding techniques (Saldaña, 2015) to obtain a new set of codes that can be subcategorized into the overarching categories. To begin with, the overarching categories are rather general, so they do not set any particular precedents or dictate the next steps of the coding process.

Coding into themes

The last step is to code into a subset of themes to gain more specificity and avoid requiring categories that could be too vague or general.

4.5 Validity

It is quintessential to consider the validity of this qualitative study to ensure that it is replicable and can be trusted. Four general types of trustworthiness are considered by Thomas (2006) regarding the inductive approach, and these types are described initially by Lincoln & Guba (1985) for qualitative approaches: credibility, transferability, dependability, and confirmability. Thomas (2006) primarily addresses validity based on Lincoln & Guba (1985) in the context of data analysis. However, the study also requires an overarching approach for validity across all research phases, to which Johnson proposes assessing descriptive, interpretive, theoretical, internal, and external validity (Johnson, 1997). We will now discuss assessing validity in our study following Thomas (2006) and Johnson (1997).

Analysis validity

Validity in the data analysis is approached by (Thomas, 2006):

- Performing stakeholder checks – allow participants or other members to comment or assess the findings, interpretations, and conclusions, addressing the *credibility* issue.
- Carrying out research audits – compare the data to the research findings and interpretations, addressing the *dependability* issue.
- Doing consistency checks – use independent parallel coding and category clarity checks by a second coder, addressing the *confirmability issue*.

To address transferability in interpretive research, it is nearly impossible to replicate the same results, and that kind of replication is not a goal for interpretive research. However, the transferability criterion can still be addressed by using the same methodology in other contexts to respond to the same phenomenon (Korstjens & Moser, 2018).

Qualitative research validity

Several strategies can promote qualitative research validity, such as research as a "detective," triangulation, peer review, reflexivity, among others (Johnson, 1997). Reflexivity is a notable strategy, where researchers must be self-aware and control biases. However, outside of the variety of strategies, five main validity types should be assessed:

descriptive, interpretive, theoretical, internal, and external validity. We address each form of validity as follows:

- *Descriptive validity*: whether the accounts are being accurately reported – we use investigator triangulation; multiple observers are present at the time of interview to cross-check observations and make sure that the interviewers and respondents are on the same page. One researcher takes notes that can later be compared with transcriptions.
- *Interpretive validity*: accuracy in the reporting of facts – we attempt to get into the minds of the respondents. Additionally, we carry out member checks to ensure that meaning is accurately portrayed.
- *Theoretical validity*: the degree to which a theoretical explanation from research fits the data, and is as a result, credible and defensible – we verify our findings against prior research. Also, we try to find patterns but importantly include a negative case. The negative case is simpler to identify with a healthy amount of interviews.
- *Internal validity*: the degree to which it is justifiable to conclude that an observed relationship is casual – we use data triangulation on a few occasions to get different perspectives within the same context. Additionally, possible causal relationships are addressed in the discussion.
- *External validity*: the ability to generalize from a set of research findings – although not a primary goal, a detailed methodology including information about the people in the study, contextual information, etcetera, allows for a certain degree of repeatability to identify similar patterns in later research.

4.6 Ethical Considerations

The study does not gather any personally identifiable information or sensitive data. The primary focus of this study results in a thematic analysis that does not require any identifiable or sensitive data. Respondents were informed about their rights to revoking their consent at any time. The researchers have also ensured a safe and encrypted cloud storage of interview recordings and transcripts. This cloud storage is provided by the University of Agder, using Microsoft OneDrive.

Due to the ongoing COVID-19 pandemic, all physical interviews had to be canceled. The GDPR legislation traditionally requires that interview recordings be done in an offline environment, with a recording device that has no network connectivity. However, the pandemic and social distancing rules forced us to find other solutions to carry out the study as planned. We used the University of Agder's Skype for Business service (provided by Microsoft) to carry out and record the interviews. Skype for Business has industry-standard encryption technologies and was deemed safe enough for our data collection.

5. Empirical Findings

The empirical findings are presented and categorized based on the data lifecycle model and the structure used in the interviews, as shown in Figure 4 below. All 13 interviews, with 14 informants have been coded according to the same structure, and the results are presented within the given sub-categories. The 14 informants are anonymously presented in Table 3 (next page), with their general profession and their respective sectors. Additionally, it shows which phase of the data lifecycle they relate to in their work. Whether they specified challenges related to that phase is color-coded, as explained below the table.

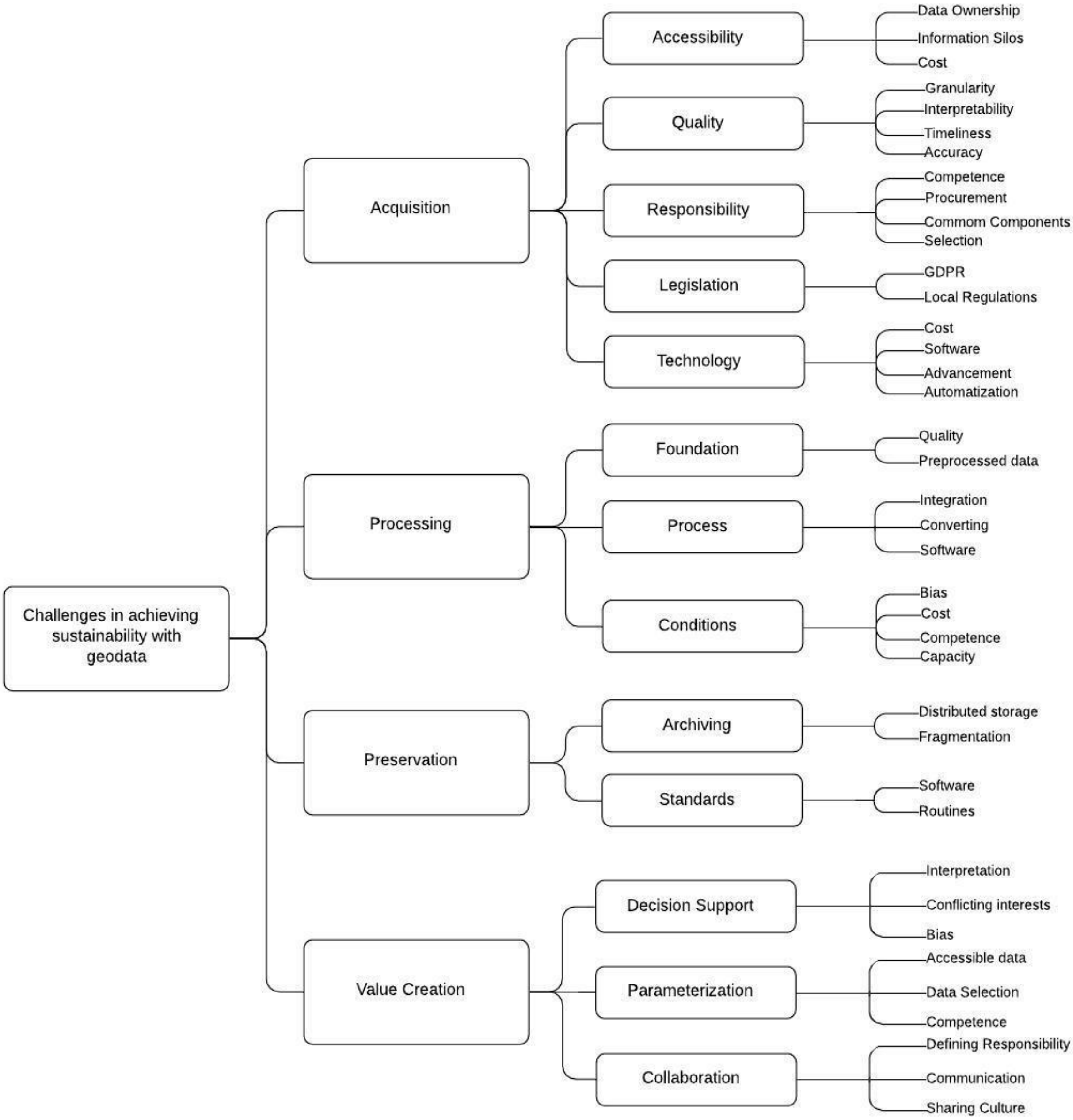


Figure 4: Thematic Map of the Challenges Gathered from the Empirical Findings

Table 3: Overview of the Informants.

ID	Profession	Sector	Employees	Area	Acquisition	Processing	Analysis	Preservation	Utilization
1	Researcher	Private	50-100	Transport	X	X	X	X	X
2	Data Manager	Public	500-1000	Mapping	-	X	-	X	-
3	Researcher	Private	50-100	Transport	X	X	X	-	X
4	Data Manager	Public	8000-10.000	Planning and Building	X	X	X	X	X
5	Data Manager	Public	8000-10.000	Planning and Building	X	X	-	X	X
6	Engineer	Private	100-300	Consulting Company	X	X	X	X	X
7	Project Manager	Private	1000-3000	Consulting Company	X	X	X	X	X
8	Engineer	Public	5000-8000	Transport	-	X	X	X	-
9	Analytical Advisor	Public	8000-10.000	City and Community	X	X	X	X	-
10	Enterprise Architect	Public	8000-10.000	IT Unit	X	-	-	X	-
11	Analyst	Public	5000-8000	Taxation	X	-	X	X	X
12	Section Manager	Public	8000-10.000	Mapping	X	X	X	X	X
13	Environmental Advisor	Public	8000-10.000	Environmental Unit	X	-	-	-	X
14	Researcher	Private	50-100	Transport	-	X	X	X	-

X = Work with, but no challenges mentioned

X = Work with, challenges mentioned

\- = Do not work with

5.1 Data Acquisition

Out of the 14 informants, 11 said they had a relationship to data acquisition in their work, and 7 of them experienced some challenges with acquiring data. All challenges mentioned in regards to data acquisition are categorized and resulted in five main groups, namely accessibility, quality, responsibility, legislation, and technology, as shown in Figure 5. Each group has subcategories that describe the challenges, which will be further explained in this section. Data acquisition methods vary based on the informant. Some manually collect data from respondents or through observations, while others use automated sensorics, photogrammetry, satellites, or lasers to collect geodata.

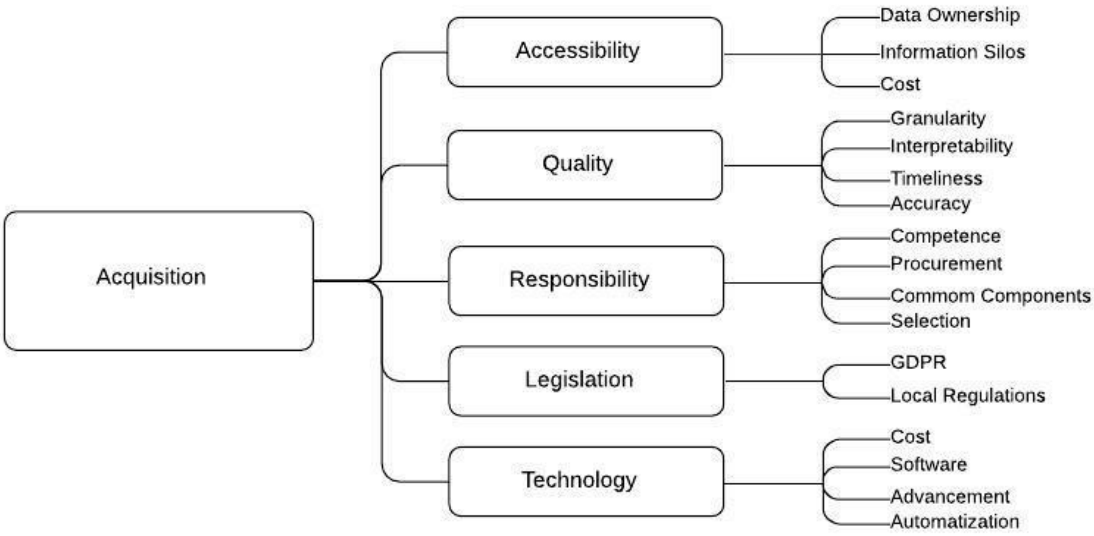


Figure 5: Thematic Map of the Challenges Within Acquisition.

5.1.1 Accessibility

This section is closely linked to responsibility due to variations of competence amongst the people who are responsible for accessing and selecting data. These human aspects will be accounted for in their respective sections. However, even without these factors, it is not granted that the data is available to the actors that need it. This can be due to data ownership, a variation of sources, information or knowledge silos, or cost of acquiring and sharing data. Accessibility can also be compromised due to data quality, such as poor descriptions and incomplete data, or vice versa. Perceived accessibility might influence the selection processes and compromise the data foundation and, thus, the quality.

If the desired data foundation is not available, less suitable data or data of lower quality can be the only option to use.

Data Ownership

Some actors are obligated to share their data while others are not, and some view their data as a competitive advantage. I1 explained that private Norwegian transportation actors, for instance, are not obligated to share their data, while the ones in Sweden are. However, demanding shared data requires someone to follow up on these data providers to ensure that they share what they are supposed to. Different owners and their frames for sharing data is, therefore, a recurring problem within accessibility. This is a problem of varying degrees amongst different sectors. I1 also explains that collecting data from private persons varies in difficulty and that different groups of people are more challenging to get data from, either due to distrust in authorities, lack of interest, language barriers, or not being able to reach the desired respondents.

Different data owners also describe a lack of incentives to share data, especially raw or unprocessed data, before it has been utilized. I3 also states that different datasets could be much better utilized if they were increasingly shared, regardless of who owns it. In some cases, data can also be confidential or licensed, and therefore only accessible for the data owners themselves. When it comes to ownership of map data, willingness to share can be connected to the cost of collecting it, which is described as very high. The cost of sharing data is also mentioned as a barrier for data owners to share their data by I1, I10, I11, and I12. For instance, cleaning, describing, documenting, and providing formats usable for others require resources from data owners, without necessarily yielding any benefits. In some cases, I3, I7, and I8 explain that data owners are not aware of what data they possess, either, which can reduce accessibility and utilization of data.

Information Silos

Out of the 14 informants, 8 have either experienced or been the cause of low accessibility of data due to decentralized data or information silos and view it as a problem. This can be internal files needed in their work, data needed for analysis or parametrization, and not necessarily within collection of raw data. I3 mentioned that they have a data bank, but that it needs to be updated. Additionally, the data bank in question did not contain a comprehensive amount of data, and decentralized data silos were still an issue. I3 also explained how this could impact information flow, and consequently, that data does not reach the right planners or decision-makers. Not being able to find data is a recurring challenge, and can be caused by data being stored on everything from local databases, memory sticks or in poorly described project folders. A data bank or common database with an intuitive and accessible folder structure is often lacking and is a desire that has been lifted from several informants. At the same time, I5 explained that:

"It is not like we can start to share our databases with others, and give them permanent access and rights to edit it, because we need full control of the data so that nothing wrong happens to it. And many probably think like that, because they are scared that the data might not be managed properly. That can be a barrier for sharing data."

Cost

Cost is described by 6 of our informants as the most significant barrier within data acquisition. According to I2, I3, I4, I5, and I7, many services that use geospatial data as a basis or at its core depend on as detailed maps as possible. Still, mapping requires expensive technology and is mainly collected by external companies and consultants. In some cases, the data is collected through collaborations with both public and private organizations and entities to share the cost of collecting data. Although being beneficial regarding cost reduction, there are still some issues with these collaborations. The desired detail level is one of them, as some actors are satisfied with less detailed and thus cheaper data material, while some require a higher level of quality. I4 and I5 have experienced problems with the data foundation due to this and explains the conflict between quality versus quantity. They would ideally like to have more frequent updates and more detailed height data. Still, they experience that this commonly is down prioritized due to conflicting interests, and in these cases, cost is usually the winning argument. In addition to the cost of collecting, there is the cost of sharing data. To clean, structure, and make data understandable for someone else, in a format they can use, can be resource consuming. I1, I10, I11, and I12 view this as an essential barrier for making data accessible and point to the lack of incentives for sharing data.

5.1.2 Quality

The challenges and barriers mentioned regarding quality have been divided into four categories: granularity, interpretability, timeliness, and accuracy. These represent the majority of the mentioned challenges within quality assessment.

Granularity

The granularity of data refers to the level of detail in raw, processed, or aggregated data. Many application areas for geodata require a high level of detail in order to provide analyses, mapping, or planning. However, as mentioned in regards to cost, the higher the granularity, the higher the price. Nine of our informants said providing satisfactory granularity in their data foundation as a challenge due to cost, competence, or technology. Different stakeholders in acquisition collaborations are described with different needs in regards to granularity and is also a challenge for the stakeholders with the highest requirements for detailed data.

Additionally, communicating the need for granularity to external data collectors or consultants is mentioned as challenging, and we received several examples of misunderstandings that lead to lesser quality than needed within acquisition projects. I10 also mentions challenges related to storing 3D files, because the storage platform does not support storing of all attributes, and flattens the data into PDFs, which compromises the level of detail in the data. Data formats and converting are also mentioned as factors that can decrease granularity. If the data is needed after being stored or converted, the quality might have dropped as a consequence.

Geospatial data can be used to trace movements, and coordinates are regarded as highly detailed in that sense. However, due to legislation or access, they are not necessarily possible to use. In some cases, addresses are an option. Even so, many of the projects mentioned by I3 and I10, commonly use Basic Statistical Units, which usually includes a couple of hundred people, but can stretch over several kilometers. In other cases, only data on a municipal level is available. Still, sometimes a general overview is all that is needed, and a high level of detail is not necessary for the purpose of the data foundation. Furthermore, lacking standards, signs, and explanations to describe granularity is also an issue that can make it hard to determine whether the granularity is sufficient.

Interpretability

Both raw and processed data often needs to be interpreted by stakeholders. Understanding, interpreting, and communicating the data can be problematic due to the wide variety of formats, terms, and complexity within geodata. I6 and I7 point to the importance of competence when working with this data, partially because they need to be able to use up to 300 different formats. To understand and interpret the data correctly is vital for both utilizing it and communicating it to employers, colleagues, politicians, decision-makers, and other stakeholders. Data quality plays a significant role in whether the data is easily interpretable, but is not solely the deciding factor. We found that the informants with an increased focus on visualization and competence within the data formats and complexity, experience fewer problems with being understood by other stakeholders when presenting data. However, some data types are less adaptable than others, raster compared to vector, for instance. Interpreting raw data such as aerial images, terrain models, and sensor data are also mentioned as challenges.

Timeliness

When it comes to timeliness of data, I4, I5, I6, and I7 refer to data as a perishable commodity, although this varies based on what kind of data. Constructions, cities, patterns, terrain, and roads can be outdated within days or years, depending on changes, while some places stay the same for decades. Whether the data is reliable or not needs to be assessed both based on data type and the purpose of using the data. I7 view this assessment as vital because it potentially can prevent unnecessary acquisition projects, or prevent the use of outdated data when newer data is required and propagate into processing, analysis, or utilization of data.

Accuracy

In many cases, accuracy within the data has a lot to do with granularity. The level of detail in geodata also refers to how accurate it is, especially in regards to localization. As mentioned, coordinates are referred to as highly accurate, but sometimes The Basic Statistical Unit, municipality, or even higher levels of data accuracy is possible to use. The characteristics of data can also impact the accuracy, as the attributes and properties might not always measure what was intended. I7 describes this problem within sensors, where error margins or unintentional adjustments can impact the outcome of the collected data.

I6 explains that low accuracy can be caused by inadequate properties, lacking or faulty attributes, and metadata and thus make the data unusable. Descriptions of intended use are also mentioned as a showstopper, because of its importance in using the data for different purposes.

Some datasets like the national road map have been described as being close to complete, although trails, smaller roads, and common shortcuts are lacking. I3 and I8 explained its consequences and the ripple effects it can cause within data analysis. Issues regarding repeatability of studies have also been highlighted by I1. Low response rates in surveys and lack of understanding from the respondents are some of the consequences that may compromise the accuracy of the collected data.

5.1.3 Responsibility

When acquiring data, responsibilities need to be determined. Eleven of our informants mentioned challenges related to who is responsible for providing or acquiring their data foundation. Competence and understanding is the most common reason for these challenges and can be lacking from internal resources, external consultants, employers, and customers. Defining responsibilities is not always straightforward, and it is not a given that the preferred competence is available either. I1, I8, and I10 talk about ongoing projects within different organizations, where the goal is to create a centralized data pool or data bank with data from all kinds of providers. Neither of the informants is sure of what sources will be included, or who will be responsible for delivering this data. Responsibilities in who own, manage, and update the data are also described as shifting, which can cause problems with the awareness and accessibility of data.

Competence

Competence within data availability is important to get the right data foundation but depends on who is responsible for selecting and collecting data. We found that consultants, to a greater extent, view competence and awareness of data availability as valuable expertise, which they capitalize on, compared to internal resources in public organizations. Due to the high number of potential data sources, formats, and complex information systems, knowledge on how and where to navigate in order to obtain necessary data is a field of competence that requires constant updating and growth.

Moreover, insight and judgment on what competence they have themselves, and what competence they need to obtain from others are crucial.

On the other hand, I1, I7, and I11 point to the lack of information regarding internal resources, such as their education, experience, and knowledge. I1 and I7 have access to a CV database, but with features that do not support efficient browsing of personnel, and are not described as user friendly. For example, I1 experienced that their colleagues of four years did not know their educational background, although there can be several responsible factors for that. The culture for sharing knowledge and competence is described somewhat vaguely, and only I3 and I12 have concrete platforms or forums for this, whereas one of them has been suspended.

Procurement

Within procurement, I12 mentions competence as a key factor in obtaining the right specification and, thus, the right services. At the same time, both I12 and I7 explain that there are both technical and legal aspects to such specifications, which may require several fields of competence. Defining and obtaining this competence has been problematized by the informants. I9 explains the potential consequences of miscommunication between employer and service providers:

"It is an important part of project management to clarify ambiguities as early as possible. But what we experience in regards to tendering, within railways and roads, for instance, there is some lawyer with the responsibility of public procurement that has no clue or bearing on railways or anything other than procurement. And when we ask questions of technical manner, they can not answer in a clear way, because they do not realize the importance of it, and do not use the right professionals for such projects."

This problem can also go the other way if technical personnel is hired to provide procurement specifications without the necessary knowledge on regulations for public procurement. Procurement can, therefore, be a cross-disciplinary task that requires careful considerations. I10 exemplified some of the consequences with poor procurement, spanning from bad data to being more or less stuck with systems because the data is hard to separate from procured systems.

Common components

Initiatives to coordinate different components within public entities was presented as a measure to prevent redundant work. One of which is different municipalities conducting the same development projects instead of reusing each other's work and services. According to I10, these are referred to as common components and exist on a platform available for all municipalities. Several of these services have been implemented and standardized according to the informant that introduced this subject. However, I7 pointed to examples where different public entities conducted the exact same data collection projects, although the data already was available.

Selection

The selection of data can be limited to what is accessible, but even if the data is available, selecting a data foundation can be just as challenging. Within our range of respondents, 5 mentioned challenges related to data selection and defining the scope of relevant sources. For instance, within the transport sector, the definition of what is actually transport data can be difficult to determine. Due to the variety of possible sources, both private and public, competence on data availability is important but hard to obtain. Within the three ongoing projects that are supposed to centralize data access through new platforms, there are uncertainties in what data should be added. Another issue is selecting data from different disciplines. I7 exemplified with construction projects, where terrain models, existing roadmap, and planned roadmap are essential sources. In addition to that, socio-economic data was vital when dealing with stakeholders with conflicting interests, such as neighbors.

Furthermore, environmental and socio-economic data can often be conflicting, which can create difficulties in using both simultaneously. I3 explained this issue and, consequently, how it impacts the prioritization of factors based on background and profession. Choosing indicators that reflect the desired performance or measurements is a profession on its own, and we found that this is a lacking field of competence within the range of our informants. Even when mentioning the importance of indicators and performance measurements, none of the informants can identify concrete indicators that are actually implemented.

5.1.4 Legislation

There are several local, national, and international legislations that may apply when working with data. Eleven of our informants find such legislations challenging themselves or have experienced this as a barrier for data collection, sharing, or aggregation. GDPR is the most common regulation that our informants experience challenges with, but other local regulations can also apply, and thus create issues.

GDPR

The most prominent challenge, when it comes to GDPR, is low competence within the regulations. People are unsure of what data they are allowed to collect, aggregate, utilize, and share. The fines for breaching the regulations are described as big and intimidating, and several of our informants exemplify with cases where organizations constrain themselves, sometimes even more than necessary, in order to comply with them. As a consequence, I1 mentions that organizations might choose not to share data that could have been shared, in fear of violating GDPR. Within data collection, different challenges were mentioned. Spatial data, for instance, can be collected on different levels. Data with detail level, accuracy, or attributes that could identify someone is prohibited from collecting without a permit and using without consent. However, a recurring issue is awareness and knowledge within these regulations. When using either coordinates or Basic Statistical Units, there are restrictions in combining datasets from other sources. Through the interviews, we got examples of how this can restrain utilization of data. There is also the aspect of using collected data for other purposes than agreed upon beforehand. For instance, I1 mentioned not being able to contact survey respondents to ask for follow-up interviews, because they forgot to ask for consent to do so. In that case, they were told that they could do it anyhow, because the surveys were sent to institutes and not specifically the people who answered, while others said they could not, as long as single people had responded to the survey.

Local Regulations

In addition to confusion and difficulties with GDPR, some challenges were also related to internal governance and regulatory conditions. For instance, informants being scripted or censored because they are not allowed to say anything that could misrepresent the organization or wanting to hide incriminating information. These and other confidentiality agreements can restrict information sharing and are often regulated by data owners or organization policies. The Cadastre Act and Spatial Data Act were also mentioned by I6, as examples of legislation that might lead to further restrictions. I6 also mentioned regulatory conditions revolving technology, that gives a certain leeway around what can be done with data. They explain that within geodata, there usually are clearly defined data owners that have their defined roles in managing this data. When it comes to sharing data, there are mechanisms established to exchange data between different actors. That causes variations between data owners, and how accessible their data is. Simultaneously, different legislations apply to different contexts that need to be accounted for when providing services. For instance, in automating services within the plan for land use, all plans are current, which causes problems when trying to automate services that have been established on several levels, because the old plans are legally binding, which demand complex architecture to account for. Additionally, long value chains can be affected by different legislations. For example, The Planning and Building Act can apply to applications, before The Cadastre Act determines how this data is supposed to be added to the municipal register, where The Archiving Act also applies.

5.1.5 Technology

Geodata is acquired in a wide variety of ways, using a wide variety of technologies. Cost of this technology, difficulties in using necessary or desired software, technological advancements, and automatization are the main groups of challenges mentioned in the interviews. Nine of the informants have experienced challenges within these categories. There is a wide variety of what level these challenges arise from, as some of the informants are far more advanced in technology usage than others.

Cost

The cost of sensors, planes, drones, lasers, and other technology for collecting data, imaging, or mapping is described as a barrier for data acquisition. However, our informants are optimistic that new technology will reach the market at a lower price within years and be available for commercial use. I7 explained how photogrammetry is done, and how both collecting the images and processing them is very resource demanding. Photos can only be taken after snowmelt, before leafing, in the right weather conditions with no clouds. The window for that is narrow, and one of the reasons it is expensive. Balancing sensor technology is also challenging due to them being more expensive with increased accuracy or resolution. However, lower accuracy can be acceptable if it is possible to interpret the data anyhow.

Software

Challenges related to accessing, selecting, and using different software have been mentioned in a variety of forms. Within acquisition, Application Programming Interfaces (APIs) are described by I10, I12, and I14 as an important enabler for making data accessible. However, lacking common platforms and information systems can prevent actors from sharing. As mentioned, I1, I8, and I10 described ongoing projects for creating such platforms, where information can be stored while being easily accessible for others. At the same time, problems regarding synchronizing, updating, and integrating the data arise.

Advancements

Several of our informants have been in the geodata game for decades, and have seen technology change drastically. We got examples by I2, I4, I5, I7, and I14 that span from manually creating physical maps to drone and laser technology that has, and will continue to revolutionize the collection of data. However, mapping data with planes, drones, lasers, or satellites are still resource demanding, and our informants say this part of the data lifecycle will go much effortlessly in a few years. We got forecasts on drones being able to take off, automatically photograph their desired landscape, or even use lasers to provide height models, light detection and ranging in a much more detailed level than available today. Advancements, or lack thereof, are mentioned as a barrier for collecting sufficient data for maps and terrain models.

Automatization

Automatization refers to both challenges with automating services and challenges with achieving the necessary quality from automated services. I6 and I7 mentioned the public's desire to automate processes or services to become more efficient, save resources, and prevent human error. Competence in how this can be achieved is much higher in the private sector than in the public sector, and consultants are often hired to help public entities improve that.

5.2 Data Processing

Within the selection of informants, 11 out of the 14 stated that they have some tasks related to data processing in their work, whereas 5 of them experience challenges. Due to data analysis being a particular profession and field of competence, we have separated the statistics for processing and analysis in Figure 6. This is mainly due to informants claiming to not work directly with processing while doing some analyses or vice versa. Out of the 14 informants, 10 informants analyze data as part of their work, out of which only two experience challenges. Most of the influential barriers that affect processing are due to the data foundation and not the processing itself. However, some technical challenges are causing problems when processing data. Additionally, we collected some conditions that affect this phase. Bias, cost, manual labor, competence, demand, and capacity are all mentioned as challenging factors within processing.

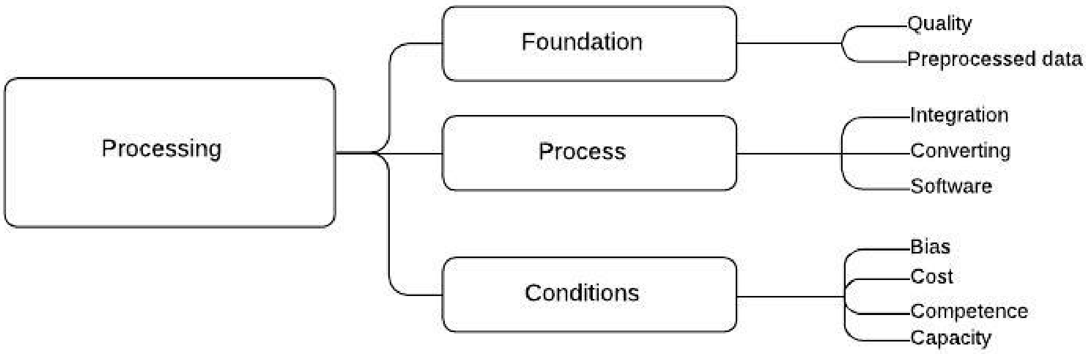


Figure 6: Thematic Map of the Challenges Within Processing.

5.2.1 Foundation

As mentioned, the most significant barriers related to processing and analysis are due to bad data foundation. Although important, being able to use complex systems and methods for processing is described as a secondary factor for succeeding and will not make a difference if the data foundation is inadequate. Seven of the informants mention challenges with bad data foundations, and three of them explicitly identify it as the most significant challenge within processing, including two of the private actors with high competence within data analysis.

Quality

The aspects of quality mentioned in 5.1.2, are applicable here as well. The acquisition process is supposed to provide a sufficient data foundation for further processing or utilization, and when that is lacking, it propagates into the processing and analysis phase. Whether the data is noisy, messy, unstructured, inaccurate, or outdated, the processing will be a result of the initial quality and will affect the outcome. I7, which is one of the informants with the most experience with managing data processing, said that bad data input will always result in bad data output. Nonetheless, the quality of data should be described to make it possible for others to assess whether the data is sufficient or to look for data elsewhere, which brings us over to challenges related to preprocessed data.

Preprocessed data

When using preprocessed data, interpretations, work practices, and unknown standards might have been used. Using datasets that have been processed in some way beforehand, without knowledge of the methods or standards used, it can be hard to interpret, use further, or continue to process. I1 highlights the importance of this and explains how different methods can give different numbers and results. They also express a deficiency of widespread standardizations to ensure the quality of both the data and the methods.

Without access to the raw data used, it can be hard to compare the processing method to other practices. In some cases, that means they have to trust the data, method, and competence of the responsible personnel without documentation if they want to use the data. The purpose, intention, and limitations of the data and the processing also needs thorough descriptions, because of the impact it has on the outcome. I1 also explains that access to different outcomes from the same dataset can increase the validity of the data, and enable reviewing of cleaning and analyzing methods. However, that is rarely available. They are also familiar with some European standards for describing metadata but are not sure which organizations use it, or if it is a common standard to use, as it is time and resource-demanding to follow. None of our informants mentions using a framework or standard like that.

As mentioned initially, the geodata field incorporates a wide variety of formats, characteristics, technical and socio-technical aspects, and jargon within the field can be difficult to understand for people from other fields of competence. When using internal or external sources, names on datasets and their descriptions are not always intuitive and easy to understand. In many cases, the informants are not under the impression that the data is adapted or meant for whoever. Descriptions and nametags only consisting of cryptic letters and numbers are sometimes impossible to interpret for others than the owners of the data. I7, I9, and of our informants mention this as a challenge.

5.2.2 Process

The processing aspect of this phase refers to the technical challenges highlighted by the informants. As mentioned, the data foundation plays a substantial role in the outcome. However, there are also challenges related to integrating, converting, and analyzing due to software, platforms, or information systems used in the process.

Integration

Integrating datasets with other sources of data has been highlighted as a challenge, both due to the different formats within the data and lack of compatibility in platforms. Mapping of wires, cables, and pipes is one of the examples we received, which often comes in incompatible formats for integration. Within address data, gathered from registers like the Brønnøysund Register Centre, there can be a lot of different ways to write street names. I14 explained how that could cause problems when trying to integrate with other datasets due to inconsistencies and deviations. Incomplete names and abbreviations are also a part of this problem. APIs mitigate some of the compatibility and integration challenges but are referred to as rare and competence demanding. Furthermore, integrating and aggregating datasets, even with the right competence and a satisfying data foundation, is not necessarily possible due to legislation.

Conversion

Some platforms demand specific formats, which require converting data in many cases. We got examples of some formats, such as DWG (from drawing) and DXF (drawing interchange format) files, where challenges in converting between them are common. However, I4 and I5 described that the formats have existed for decades, and should be manageable. At the same time, the data has different attributes, which makes it hard to pass on these properties into other formats. Another informant differentiates converting within projects, which can be handled ad hoc for one-time endeavors, from creating processes or more permanent solutions. Computer-Aided Design (CAD), Building Information Modelling (BIM), and Geographic Information Systems (GIS) introduce several challenges within both integration and conversion.

Software

Whether the data is being processed, cleaned, structured, moved, integrated, or converted etcetera, some kind of software is necessary. Only 2 out of the 14 informants mentioned using open source software, while others mostly use proprietary software. SPSS, R, Python, ArcGIS, AutoCAD, FME, QGIS, KOMTEK, and GISLINE are some of the platforms and systems used by the informants, with a variety of different challenges and characteristics. Some of the organizations have guidelines for what software to use, but most of them are free to use what they want. Some of the programs are better for handling large datasets, some require manual caching, and some are easier to create documentation in. Several use cases were presented, with these systems and workspaces' strengths and weaknesses. However, it is also mentioned that despite technical difficulties in some cases, the technology and software itself is rarely a barrier. Regardless of the platform or program being used, competence in using them correctly is viewed as rare and valuable.

5.2.3 Conditions

We gathered some conditions that were mentioned as challenges and barriers within data processing. Bias when selecting data foundation, methods or angle, cost of processing, competence within data and software, and the demand for data processing are described as important factors for succeeding, not conducting at all or failing to process data.

Bias

Three of the informants, I1, I3, and I12, mentioned bias as a challenge when processing data, and that traces and characteristics of the individuals processing and analyzing the data are often unavoidable. This is a more significant issue within qualitative data than quantitative but applies to both. Bias when choosing methods, data sources, and viewpoints also occur, even in more mathematical approaches. Another example mentioned by I3 is organizations choosing socio-economic analyses instead of environmental and already before starting the project have decided that economics are more important than other sustainable factors. Bias can also occur when interpreting the outcome of processing and analyses, based on the way the data is presented.

Cost

The cost of processing data shares several similarities with the cost of acquiring data. Often, human resources and competence are the highest expense, although licenses also make up significant costs. For individual projects and events, I11 mentions that costs are usually described as lower than implementing processes and services that require maintaining and managing over time. Additionally, the cost of making data accessible for others is high, compared to merely using it themselves. These costs are related to time and resources spent preparing and documenting the data, as well as making it technically accessible.

Competence

Competence is a factor that is relevant in all the other categories as well and can impact them in different capacities. Competence can mitigate many of the challenges but also have negative impacts, on cost, for instance. However, long-term benefits from competence can ultimately cut costs, which makes that impact inconclusive. Competence gaps and variations have been described, both from private to public entities and small to big organizations. Once again, I3 mentions the importance of assessing self-knowledge, whether their competence is sufficient or if external competence is needed.

Competence awareness comes into the equation here as well. I11 mentions challenges with not knowing which of their colleagues have the necessary competence or use certain types of data. They have initiated a project to assess that, and explains:

"The project revolves around assessing whether there is someone within the organization that understands the information we have in a geographic way. I know there are some, but not if this is done in a systematic way. Due to the number of employees we have, it is hard to get that overview of all the information we possess. There are a lot of different people that do exciting things. The problem is just to find them."

Demand

Without a demand for processing, analyses, and new services or processes, it is less likely that resources will be assigned. I4 and I5 mention that the demand for geodata is increasing and that it is a prerequisite for awareness in improving the production of it in the public sector. However, I11 explains that a balance between demand and actual usefulness is required, as not all notions are good enough to be put into production. Thus, prioritizations between cost, capacity, and demand are necessary, but can ultimately prevent projects from being initiated.

5.3 Data preservation

Out of the 14 informants, 12 of them have a relationship to preservation of data in their work, out of which 5 experience challenges. Some challenges are minor and do not impact utilization or compromise the quality. In contrast, others go on the expense of both accessibility and quality, and thus the utilization of different data sources. We have divided these challenges on whether they have to do with archiving or the standards surrounding it, as shown in Figure 7. The most prominent challenges are distributed storage such as information silos and fragmentation of data. Standards refer to the challenges within the software or the routines surrounding data preservation.

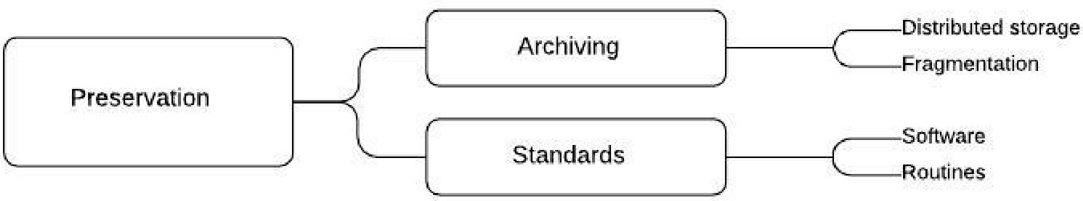


Figure 7: Thematic Map of the Challenges Within Preservation.

5.3.1 Archiving

Distributed storage

As mentioned in data acquisition, distributed storage or information silos can make data hard to access. This barrier ultimately stems from how data is preserved. As mentioned, 8 of the informants have either experienced or been the cause of low accessibility of data due to decentralized data or information silos, and view it as a problem. I1 explains that data often is stored in internal project folders and that others do not know it exists because they are often occupied with their own projects. A common databank in the organization is also lacking. I1 also mentions a project to create a common database for transport-related data. It is, however, unclear what data will be available in it and who will be responsible for adding the data. I4 and I5 are also familiar with this and explain that it can be hard to obtain data from other departments within the organization.

"People that possess data about things we might be interested in are used in their daily operations, and they are busy with their own work. So they are content with their management of it, whereas we would like to extract it and be able to combine it with other data, but I have experienced that as challenging." - I4

Fragmentation

That different kind of related data is spread across different storage units is also a problem when trying to assess what data exists and what data is relevant in different projects. It is tightly coupled with decentralized storage, but it relates more to data of the same kind or associated data. I9 explains that:

"A great challenge is that a centralized register within municipal services does not exist. Very often, each department has an excel file lying around with details about some activities, while another department has a file with other related activities and area details, and then another department has a file with public services. The fact that this kind of data is spread across different departments can be challenging. Then, I have to search through the entire organization to find this data."

5.3.2 Standards

Routines

There are considerable variations of the informants' experience and challenges with routines in regards to preservation. I8, for instance, has no issues when it comes to preservation, and explains that their routines are well established, and systems for storing are working as intended. I14 has experienced that different individuals store data in different ways and that their own routines are not standardized when it comes to structure and catalogs. I10, however, explains that they have a plan to separate case files and archive files. They would also like an automated process for transferring case files to historical files after a given time, but integrating data from different departments and sections within the organization makes this challenging. The main problem is a lack of standardization of such routines across the organization. Additionally, different routines for updating registers are also a problem that can cause inconsistencies. Further, I10 explains that:

"Out of the 400 systems we have, I think 100 of them contain personal data. Some of them have imported the National Registry once, and then they update themselves every time they get messages about changes. Others may be interconnected with the National Registry, and some get regular data cleaning. However, the National Registry is costly to use. Thus, the quality of much of the data is bad due to manual updating. As a consequence, the data might not be accurate due to different formulations and naming. We have had much information in our systems that is incorrect because some have been too nice or sloppy with changes."

Software

Some of the software used has problems with user-friendliness or technical incompatibilities. I2 explains that some of the systems they use are hard to navigate and that many people within the organization have problems with that. Additionally, some formats lose some of their attributes, such as 3D within the storage platforms, as mentioned by I6 and I10. I6 mentions a platform for synchronized geodata, but it only accepts 100% specified data. They further explain that:

"When set up [a platform for synchronized geodata], it can share information with subscribers and providers, for instance, governmental agencies, municipalities, and transport agencies. This way, everyone is connected, and the data will be synchronized and flow seamlessly. And then everyone has access to updated information so that if something gets built, it will show up in your system without having to do anything."

However, there is much information that is not a part of this solution, and that exists only in their own specialized systems.

5.4 Value Creation

The view on sustainability and value creation is very different within the selection of informants. Regardless of whether sustainability is a driver for them, a goal they have, or something that is just imposed by management, there are some similarities in their view of achieving it. For example, in order to achieve sustainability, the right decisions have to be made, the right parameters have to be used, and different units need to collaborate. Nine of the informants have a relationship to utilization of data, and all of them find it challenging. The different challenges related to these categories are presented in Figure 8.

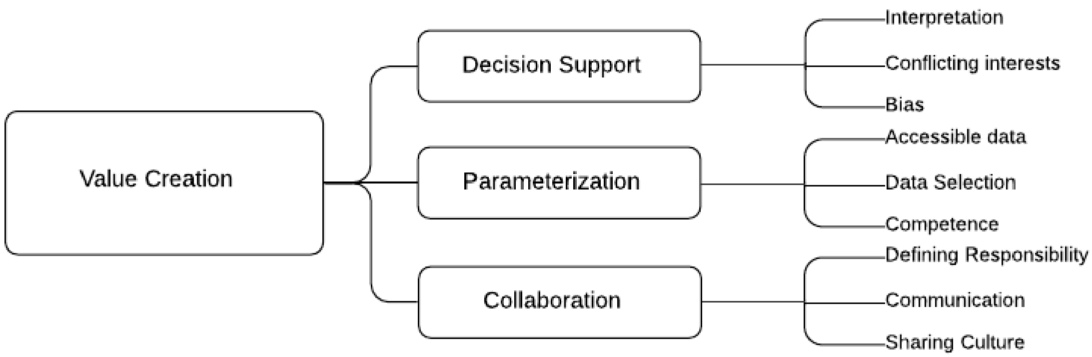


Figure 8: Thematic Map of the Challenges Within Value Creation

5.4.1 Decision Support

The ultimate goal of producing, processing, and analyzing data is often that it can be used for decision support. However, making data presentable and understandable is mentioned as a challenge. Decision-makers that interpret data differently than intended, conflicting interests and bias are viewed as some of the reasons why this occurs.

Interpretation

Depending on how the data is presented, whom it is presented to, and what case it is connected to, interpretations are necessary but often hard to anticipate. I1 mentioned that they rarely get notified when their reports or projects have been used for decision support. On one occasion, however, they found out randomly that the report or project had been. This can be connected to interpretation in the way that data collectors, processors, analysts, or researchers do not necessarily get the chance to express their intentions with the work. Neither will users and decision-makers get the change to express their desires for the data foundation. That leaves much more power to the interpretation. I9 also explains that strategists, planners, and overall politicians are the most frequent users of their data foundation and that they deliver it processed and visualized for them to use. However, I7 explains that the data is often misinterpreted when used, due to, e.g., engineers, architects, and analysts making non-understandable visualizations and drawings of phenomena because they take for granted that other people, decision-makers or customers understand it. Further, I7 explains how a focus on good visualizations can mitigate such misunderstandings through 3D models, movies, or estimates. I1 points out that it is easier to understand data they have collected themselves than data that is collected or processed by someone else.

Conflicting interests

When a decision is being made, there are often many considerations to take. Economic, social, and environmental aspects are a big part of that. I9 mentions that such conflicting interests can lead to their professional assessment and data foundation not being considered. Additionally, pre-existing assumptions, lobbyism, political gains are mentioned as conflicting areas to professional assessments. I3 mentions a case where this was an issue:

"I can only say as much as that we have been hired to do an assessment of something, and we ended up assessing it as pretty bad. We work within research, so we cannot censor ourselves, but there was a lot of back and forth and resistance before we got it through. We ended up going over budget because the people that hired us were not satisfied with what we found because it shed a bad light on them. So we had to spend extra resources on arguing on whether we would be censored, but we held our ground."

Different stakeholders often have conflicting interests. Within all of the building and planning projects mentioned, there is always a budget. Additionally, there might be social aspects that need considerations. Whether it is new roads, trails, bridges, railways, parks, industrial areas, etcetera, I7 mentions neighbors and local residents as a great source of complaints.

Bias

When it comes to bias in decision support, I3 mentions cherry-picking, which relates to choosing data based on pre-existing opinions instead of making decisions based on the data. It can even go to the point of suppressing evidence, intentionally or unintentionally. This is also a point made by I9, with deviations on what data they present and what decisions are made.

5.4.2 Parameterization

Although nine of the informants have a relationship to sustainability in their work, I7 says they experience that many stakeholders talk about sustainability just to talk about it and lack actual initiatives. To be able to use available data to measure phenomena, progress, or conditions is a desire amongst several informants, all of which find that challenging.

Accessible data

One of the challenges with parameters and indicators is what data is actually available to use. As mentioned in 5.1.1, there are several reasons for low accessibility, such as different human aspects, data ownership, a wide variety of sources, decentralized or distributed storage, or cost of acquisition. Additionally, poor data quality, lacking descriptions, and incomplete data makes it hard to access a satisfactory data foundation. Emissions are mentioned as the primary focus in many cases, but I13 expresses a desire for measuring other aspects of sustainability, such as organizational operations, consumption, and usage of supplies and goods. However, this is a relatively new way of measuring, and the necessary data foundation for measuring this is not yet accessible for actors like I13.

Data selection

Data selection is described as an eternal discussion by I3 due to conflicting interests and competence. The selection of data can also be based on requirements within projects; I7 mentions geographical points and distances, traffic requirements, and safety measures. I7 also points to sustainability goals such as education, recruitments, clean energy, developments, and more direct environmental work and initiatives that are hard to measure. Their internal strategy also consists of sustainability goals in regards to how they deliver services, how they prioritize projects, and being transparent of the dilemmas this can cause. I9 works with the adoption of 92 indicators from the United Nations (UN) to better measure economic, social, and environmental sustainability. However, this project is just getting started, and they are currently looking at what data can be used, including data from Statistics Norway (SSB) and internally in the organization. Electricity, dynamic public transformation systems, traffic monitoring, open data, air-pollution, fresh water consumption, local food production, electronic health records, life expectancy are amongst these areas. They view finding relevant and representative data to connect such indicators as challenging. I10 also mentions this as an aspiration but has not yet started on anything specific related to it. Additionally, I13 expresses a desire for better indicators in regards to sustainability, but also experiences it as challenging to translate high-level goals into local indicators.

Competence

In order to create and follow up parameters related to sustainability, competence within both data acquisition and sustainability impacts are beneficial. As mentioned in 5.1.3, I1, I7, and I11 point to the lack of information regarding internal resources, such as their education, experience, and knowledge. This is viewed as a barrier to combine necessary competencies that can work towards sustainability goals. I3 expresses that consulting companies possess much of the relevant competence in regards to data foundations. As a consequence, public entities need to outsource these assessments in order to obtain good indicators and data sources. Additionally, I4, I5, and I8 agree that technical staff with competence on the data in question rarely look for potential in regards to sustainability within the data and that actors that work with sustainability do not possess the technical competence to see the opportunities with available data fully.

5.4.3 Collaboration

As mentioned, different stakeholders and actors often possess data or competence that can be of value to others, without being aware of it. These synergies demand collaborations in order to be detected and utilized. However, defining responsibilities and range, establishing communication, and achieving a sharing culture is described as challenging but crucial to get beneficial collaborations.

Defining responsibility

Defining what competence is needed is a big part of defining responsibilities within collaborations. Data often needs to be processed or aggregated in order to create value, but I6 points to lacking competence in how to aggregate data:

"I think few people know what data they can combine. To break into something new, people from different fields of competence need to come together and connect their perspectives. You often see that people choose areas where they can realize things

quickly, but I think there is much more to gain if we give it some time and broaden the view, because there are enormous amounts of good datasets that can be utilized in a lot of contexts. But they demand some sort of processing or composition, or to be aggregated."

Several of the informants, including I2, I4, I5, I8, and I14, said that creating value from the data was outside of their responsibility and that their main focus is to acquire, manage or preserve data. As mentioned, these informants agree that technical staff with competence on the data in question rarely look for potential in regards to sustainability within the data and that actors that work with sustainability do not possess the technical competence to see the opportunities with available data fully.

Communication

I3 experience it as challenging to get a dialogue with people from other fields and backgrounds. I7 mention several channels for communicating within the organization that enables cross-domain collaborations. I8 gave an example of a case of poor communication:

"I have had some experiences with networks for biking etcetera, where things have been completely stagnant due to not succeeding in connecting the ones that have experience with data processing with the ones that are concerned with biking."

Sharing culture

Sharing culture refers to both sharing competence or experiences as well as data. Lack of sharing has been described as intentional because of the value of the data or available resources. However, the lack of sharing can also be unintentional due to awareness or abilities. I5 explains that:

"People that possess data about things we could be interested in might be used in their daily operations, and people are busy with their own work and are content with adding it and using it in their data management. While at the same time, we would like to extract them and be able to use them with other data, so I have experienced that as challenging."

In regards to sharing competence, I8 mentions that they try to promote initiatives to achieve that in order to detect more synergies that are undiscovered. However, they do not believe that small initiatives that create awareness amongst one or few people at a time will be revolutionary within the current sharing culture.

6. Discussion

The findings from this study identified what challenges different representatives of a variety of organizations experience in using geodata and achieving sustainable development. The challenges were divided into phases according to the data lifecycle to assess where the weak spots and potential bottlenecks occur. In this section, we will discuss the results of the study, how they answer the research questions, and existing literature within the field. The goal of this study is to answer the initial research questions:

(1) *How do challenges within the geodata lifecycle impact achievement of sustainability triple bottom line?"*

(2) *How do organizations meet the data lifecycle challenges, and what initiatives can be implemented for achieving increased sustainability?"*.

First, we will discuss how the challenges affect achieving sustainability, responding to RQ1. Then we will discuss how organizations face some of these challenges that can positively affect sustainability, responding to RQ2.

6.1 Impact on Sustainability Achievement

In this section, the challenges within the phases of the data lifecycle and their impact on achieving sustainability will be discussed to respond to RQ1. In the findings, we identified and presented several challenges and barriers that occur within the geodata lifecycle. By discussing these with the informants, subsequently analyzing them and by viewing them in the context of existing literature, we have assessed their impact on achieving economic, social, and environmental sustainability dimensions. These three dimensions are also known as the sustainability triple bottom line.

6.1.1 General Challenges

To begin with, sustainability is a broad term that is often hard to define. It is most often viewed at a macro-level, emphasizing the survival and quality of life of the human species, as well as ensuring the endurance of habitats that do not have a direct human benefit (Brown et al., 1987). Sustainability comprises three dimensions, that are economic, environmental, and social sustainability, also known as the triple bottom line. This model is criticized from a business perspective because businesses tend to use these dimensions for reporting purposes, and often lack the holistic approach towards the three dimensions (Milne & Gray, 2013). We have identified specific geodata challenges that affect the data lifecycle as a whole and can have sustainability impacts.

Some of the impacts are relevant regardless of the sustainability dimension and might be repeated to some degree. Bias, for instance, was highlighted as a challenge that could impact data foundation, decision support, and parameterization. This phenomenon can be linked to selection bias, and studies where subjects showed severe confirmation bias by selecting only evidence that could corroborate, but not falsify their hypotheses. Several subjects even retained their hypotheses when presented with clearly falsifying evidence (Beattie & Baron, 1988).

Some respondents highlighted *information silos* as being a challenge in the geodata lifecycle. These information silos generally pertain to stakeholders not having access to information that could otherwise be proven useful to achieve specific goals. Within data management and information systems research, information silos are a widely analyzed challenge, and solutions typically aim to reduce costs and time to market, increase process efficiency and production quality (Shahrokni & Söderberg, 2015). Furthermore, Scholtz et al. (2014) point to lacking system integration, inconsistent data, and lacking information integrity, causing "sustainability silos," where organizations focus on reporting but lack a strategy for approaching sustainability issues (Scholtz et al., 2014).

Pertaining to the data lifecycle and geodata, information silos has a more significant impact on the lifecycle as a whole. For example, there may be a lack of awareness about the required data from other organizational units that can assist in achieving specific goals. In a worst-case scenario, this could cause the relevant stakeholders to waste resources by gathering the same data, having direct and unnecessary economic consequences. Lacking data on flora & fauna and natural habitats can lead to negative environmental impacts in urban development contexts. Social consequences could also arise when urban development decisions are made with a lacking data foundation.

Running new analyses on shared datasets may also pose scientific issues, as these new analyses, by definition, cannot be pre-planned and may, therefore, suffer from being guided by the data (Barbui et al., 2016).

Impactful geodata and IS research points to an integrated approach that generally requires a cross-disciplinary perspective, breaking down these information silos. Some examples of these integrated approaches are GIS water management systems (Fernández et al., 2016), flood forecasting (Shifeng Fang, Xu, Pei, et al., 2014), social empowerment through participatory mapping (Eitzel et al., 2018), and smart city design using GIS in a Big Data framework (Lu et al., 2019).

6.1.2 Impact on Economic Sustainability

Public procurement of information systems is a common activity to cover data management needs but is a challenging task that can have an economic impact. For example, without the necessary knowledge for procurement, local governments could implement systems that do not suit their needs and have a negative economic impact. Furthermore, organizations may have difficulties in transitioning to better overall systems because of the general wish of maintaining the status quo (Moe & Päivärinta, 2011).

Organizations also walk the balance between how much internal competence they require weighed against the usage of consulting services. Private sector actors within geodata specialize in data collection and costly analysis methods. The public sector must assess how much consulting services they can use against what internal competence would cost. Too many consulting services can have a negative economic impact and can lead to knowledge gaps within the organization leveraging consulting services (Stendal & Westin, 2018).

Turning sustainability goals into measurable parameters also proved to be challenging. Bias towards specific goals can cause a disregard for economic sustainability factors. However, this depends on organizational strategy, and the opposite is true when considering that the economy is often prioritized above other factors (Oberhofer & Dieplinger, 2014).

Political incentives for additional reviews that convey a different message and prioritizing contradictory measures – such as increased public transport combined with additional capacity for private drivers – have negative economic impacts.

6.1.3 Impact of Environmental Sustainability

Also, with public procurement, giving the suppliers too much autonomy for sustainability measures may have an environmental impact. If the local government or other entities do not place demands for sustainability measures, they could end up with poor data for these measures. These environmental impacts can also pertain to the supplier's production practices if that is required.

Parameterization, or lack thereof, can prove an organization unable to measure environmental impact. Several informants described the need to have more explicit measurable parameters that they could integrate into their workflows to create sustainable value. A possible explanation for lacking environmental measures is that organizations are often influenced by economic factors (Oberhofer & Dieplinger, 2014).

Low-quality data can result in an erroneous analysis that impacts environmental sustainability. Particularly significant errors – outliers – are not restricted to specific zones, but are instead related to poorly drawn roads. Incompletion is also a serious problem since a missing road produces a decrease in quality; even unnamed roads may have an impact on the positional quality aspect. It is concluded that, at least in the studied zones, more attention should be put in the thematic quality aspect (Castro et al., 2019)

Lacking data can stop analyses and projects altogether. In April 2016, for the first time in more than two decades, the European Parliament adopted a set of comprehensive regulations for the collection, storage, and use of personal information, the General Data Protection Regulation (GDPR) (Goodman & Flaxman, 2017). The GDPR raises the question of how companies can ensure that operations conform with external data processors according to the regulation (Kurtz et al., 2018). Several respondents highlighted that they were unable to combine geodata with other data sources due to privacy concerns. One analysis that could not be carried out was, for example, tracking people's usage of hiking areas and parks. More detailed tracking of this data would have otherwise provided valuable insight for planning and maintenance of these green areas.

Other kinds of data, such as topographical data, can be proven valuable within crisis management. However, the data may not be detailed enough or available because of the sheer amount of data or complexity in gathering such data. Organizations must prioritize the detail level that they require from geodata, meaning that some insights are lost in the process.

6.1.4 Impact of Social Sustainability

Parameterization is also a challenge for social sustainability. Bias can once again displace social factors in favor of other sustainability goals.

Data collected for transport research or by governmental entities tend to be stored in distributed data silos, with different ownerships and data formats, which can cause difficulties when cataloging, finding, accessing and using research data (Adesiyun et al., 2019)

Studies on open data initiatives also point to challenges with the fragmentation of data and a large variety of stakeholders, legal and technological issues, lack of skilled experts, and funding in regards to data accessibility (Palts et al., 2019). This is currently being addressed in an international collaboration funded by the EU commission. The objectives of the Be Open project are to create a shared understanding of the practical impact of Open Science and to identify and put in place the mechanisms to make it a reality in transport research (Adesiyun et al., 2019).

6.2 Practical Initiatives to Address Challenges

Not all the mentioned challenges are possible to eliminate at the current state of technology or other factors outside of the informants' control. Some can be mitigated, and some of the informants in this study already have plans in place to make that happen. The ones that are viewed as most influential by the informants and through the analysis are elaborated in this section. To answer RQ2, the challenges are connected to relevant mitigating initiatives for practitioners.

6.2.1 Improving Sharing Culture and Infrastructure

In regards to sharing data, the software, routines, culture, incentives, and benefits for sharing are mentioned as lacking. Initiatives to create sharing platforms and common repositories for geodata can mitigate problems related to the software, routines, and culture. However, it will not have an impact on incentives and benefits for the data providers, unless they get access to other data in return and view that as a benefit. Ultimately, what different actors consider incentives and benefits vary—other fields of research experience the same problem. Within health and medical research, e.g., there are seemingly no evidence-based incentives that increase data sharing, despite the current shift towards more open data (Rowhani-Farid et al., 2017). There are some incentives, such as Open Data Badges that have a documented effect on openness, accessibility, and persistence of data and materials (Kidwell et al., 2016). This, however, is restricted to research data, whereas geodata comes from a wider variety of sources. Some argue that there is an ethical obligation to share data generated by some participant groups or observations (Barbui et al., 2016), but that might not be an incentive for all actors.

An open repository for geospatial data is mentioned by three of the informants as ongoing projects. Some platforms already exist, such as OpenStreetMap, perhaps the most successful project from those that produce Volunteered Geographic Information (Castro et al., 2019). Such platforms can enable sharing, but mainly include participants with already cooperative relationships. Organizations, especially with potential competitive relationships, might refuse to share their data due to the worry that data sharing improves competitors' competitiveness. To accommodate this, an iterative model for data sharing has been proposed (Guo et al., 2018). In order to work, competitive actors need to trade data as a currency with each other through a system for assessing the value of the data. Other initiatives, such as Be Open, are also attempting to make scientific processes and results more transparent and accessible to everyone (Adesiyun et al., 2019). To do so, it is proposed to develop a framework to establish a common understanding of operationalizing transport data, to map existing open science resources and to provide the policy framework and guidance for open science implementation in transport (Adesiyun et al., 2019). Geodata is more than just transport, and all aspects of geodata could benefit from such initiatives.

6.2.2 Increasing Intelligibility

A key characteristic of geodata is its potential for diverse and multiple applications (Awange & Kiema, 2019). As a consequence, geospatial data and its applications may be highly complex and of great variety. Stakeholders not understanding the data they possess, the data they are presented with, or data needed in analyses is highlighted as a barrier for utilizing data and detecting synergies. That makes it a goal to be able to convey geodata and make it understandable, and simultaneously less open for interpretation. Some have tried to organize a coherent data basis, like structuring it as shown in

Figure 9 into general text information, data about the author, the temporal structure of the storyline, and the spatial objects (Reuschel et al., 2013). Such an overview can provide an overview and an understanding of the different components of the data foundation. However, for some of the stakeholders without knowledge on geodata, it needs to be more comprehensive than that.

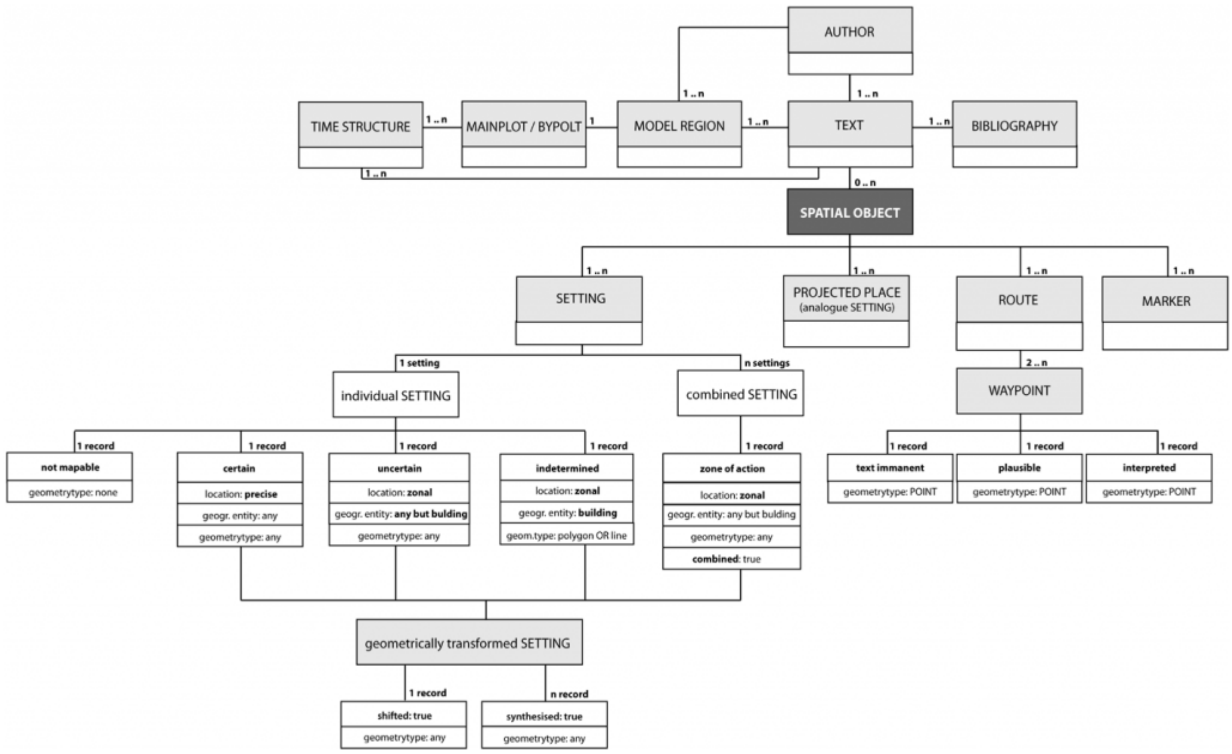


Figure 9: Data model breaks down the space of a literary fiction into individual spatial objects for the Literary Atlas of Europe

One key to understanding these complex phenomena is the representation of not only space and time but also of the interactions of different participants (Grigoropoulos et al., 2019). Visualizing this kind of data was mentioned as a success factor for interpretability and intelligibility in the interviews. However, little research reviews the impact on visualization and modeling on comprehensibility and intelligibility. That being said, we would still recommend increased focus and resources towards visualization and modeling to reach stakeholders, and especially decision-makers, with comprehensible material.

If there is baring to the fact that 3D and rich visualizations are a prerequisite for understanding amongst politicians, decision-makers, or other stakeholders, incompatibilities within archives could be a barrier, especially in regards to reusing data.

Therefore, another initiative to preserve intelligible data would be to acquire compatible storing and sharing platforms so that the data does not lose their attributes.

6.2.3 Creating Collaboration Platforms

In addition to systems for sharing data, platforms to share competence and encourage collaboration could tighten the gap between stakeholders. For instance, one of the objectives in open data projects such as Be Open is to engage a broad range of stakeholders in a participatory process (Adesiyun et al., 2019). Supply chain collaboration can be used for that purpose and refers to the extent to which participants try to standardize processes and to interchange the electronic document between trading partners (Kang & Moon, 2015). This collaboration encourages all players of the supply chain to engage in planning, forecasting, replenishment, information sharing, resource sharing, and incentive sharing. It is considered to be an effective strategy for improving its collaborative advantage and supply chain performance (Kang & Moon, 2015).

At the same time, organizational competence is mentioned as a critical factor, which the public sector often needs to outsource to consultants. Progress in projectification of the public sector creates an increasing need for developing competencies for public sector project managers. However, very little attention has been paid so far to the distinctive features of public sector project managers' competences (Jałocha et al., 2014). Thus, there is a need for both increased competence and a platform for sharing such competence amongst the stakeholders within geodata. This can ultimately contribute to detect synergies and combine points of views that could be mutually beneficial.

6.2.4 Parameterization

One of the greatest potentials with geodata within sustainable development is to be able to use it to measure the impact of different phenomena. Parameterization techniques and frameworks are used in a wide variety of cases to validate, find implications, predict, measure, and ultimately implement supervision or actions to meet them. On the other hand, convective parameterization is challenging for many stakeholders (Kain, 2004). There is little literature regarding parameterization of geospatial data within sustainable development. There is, however, a significant amount of research on evaluation, measurements, and parameterization within businesses, where there is a greater focus on both critical and non-critical performance indicators (Parmenter, 2015). Parmenter

(2015) explains that there are many misunderstandings when it comes to developing indicators and that it requires a lot of preparation and an environment where the indicators can operate and develop.

Initiatives to indicate sustainable development, e.g., as mentioned by I9, are in place but still in the starting phase. The one referred to by the informant, is supposed to provide indicators that will enable cities to measure their progress over time, compare their performance to other cities and through analysis and sharing allow for the dissemination of best practices and set standards for progress in meeting the Sustainable Development Goals (SDGs) at the city level (Smiciklas et al., 2017). There are a total of 92 indicators, of which we have included a sample in Figure 10 below. Some are more advanced than others and require a less accessible data foundation, and as mentioned, many of them are hard to adapt to a local setting.

Dimension	Sub - Dimension	Category	KPI	Type	Type
Environment	Environment	Air quality	Air pollution	Core	SUSTAINABLE
			GHG Emissions	Core	SUSTAINABLE
		Water and Sanitation	Drinking Water Quality	Core	SUSTAINABLE
			Water Consumption	Core	SUSTAINABLE
			Fresh Water Consumption	Core	SUSTAINABLE
			Wastewater Treatment	Core	SUSTAINABLE
		Waste	Solid Waste Treatment	Core	SUSTAINABLE
			EMF Exposure	Core	SUSTAINABLE
		Environmental Quality	Noise Exposure	Advanced	SUSTAINABLE
			Green Areas	Core	SUSTAINABLE
		Public Spaces and Nature	Green Area Accessibility	Advanced	SUSTAINABLE
			Protected Natural Areas	Advanced	SUSTAINABLE
			Recreational Facilities	Advanced	SUSTAINABLE
	Energy	Energy	Renewable Energy Consumption	Core	SUSTAINABLE
			Electricity Consumption	Core	SUSTAINABLE
			Residential Thermal Energy Consumption	Core	SUSTAINABLE
			Public Building Energy Consumption	Core	SUSTAINABLE

Figure 10: List of KPI's on Environmental Dimension

6.2.5 Holistic Approach

Some of the informants mentioned that conflicting projects often get funding from the same organization, or that projects with the same objective are conducted and end up with the same, and thus redundant results. A holistic view and management of projects could mitigate the redundant and conflicting work done towards sustainability. “Research findings also indicate that the use of integrated product development methods increases performance compared to traditional methods in contexts of complex problem solving” (Sommer et al., 2014). As shown in Figure 11, project governance is central in this, and are affected by human resources, organizational and external factors. The goal of such an approach includes parallelism in activities, standardized processes, and process integration.

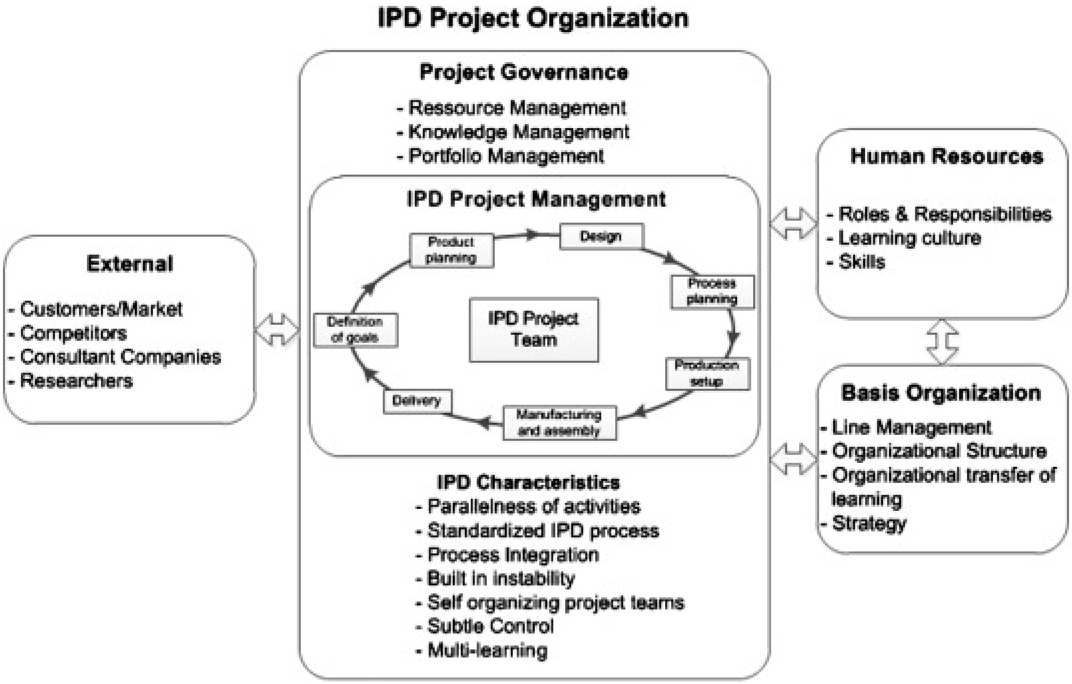


Figure 11: Proposed holistic framework for integrated product development.

However, not all projects are conducted by the same organization, thus creating a different image of how to integrate these projects. At the same time, the public sector has different entities on different levels and could still benefit from a high-level coordination effort across entities.

7. Contribution and the Way Ahead

The aim of this study was to investigate the use of geospatial data within sustainable development by increasing the focus on its challenges and potential. To do that, we started by conducting a literature review to get an understanding of what literature exists, what challenges are most prominent in the existing research, and to prepare for the qualitative approach following the literature review. Our empirical data collection consisted of 13 qualitative semi-structured interviews with 14 informants, which were thematically analyzed and categorized based on the data lifecycle model. We used them, in addition to existing literature, to answer our research questions.

7.1 Summary of Related Research

The systematic mapping study, as background for this thesis, has resulted in 26 papers identified papers. The review primarily provides 20 specific challenges that are identifiable when working with geodata and the data lifecycle. Furthermore, these challenges have been grouped into specific data lifecycle steps, according to which challenges affect different lifecycle steps.

First, the identified challenges provide a way of understanding how these challenges impact being able to achieve sustainability. Second, the approach towards attempting to solve each challenge in the literature inherently assumes that becoming better at the data lifecycle should also provide a sustainability benefit.

Our mapping tool (Figure 1, page 8) within each main phase of the data lifecycle – acquisition, processing, and preservation – aims to provide an understanding as to which challenges affect which phase. Then, it is possible to identify both the role of the sustainability dimensions – dimensions that primarily function as guiding tools for the geodata task at hand – and what challenges impact organizations in achieving greater sustainability.

7.2 Summary of Empirical Findings

The empirical findings identified 13 higher-level categories of challenges within the geodata lifecycle and value creation. The geodata lifecycle is simply the data management perspective of working with data within a lifecycle. Within the 13 higher-level categories, a further 39 specific challenges within each general category were identified. Given each data lifecycle phase, some identified general categories were aspects such as data accessibility, quality, process, archiving, and collaboration, among others. Within each general category in the lifecycle, aspects like data ownership, privacy, cost, software, and interpretation – among others – were identified as specific challenges to each general category at each phase. Some challenges could even act as barriers to achieving sustainability dimensions. We will now discuss the theoretical and practical implications of the empirical findings.

7.3 Contribution to Theory

We have identified some significant theoretical contributions, considering this study. These contributions impact the identified theoretical factors that serve as some of the background for this study: the *data lifecycle* (Sinaeepourfard et al., 2016) and the *sustainability triple bottom line* (Milne & Gray, 2013). The study also presents a new term that consolidates the data lifecycle approach and geodata: the *geodata lifecycle*.

The data lifecycle – consisting of acquisition, processing, and preservation – is a data management framework proposed by Sinaeepourfard et al. (Sinaeepourfard et al., 2016), specifically identified as the Comprehensive Scenario Agnostic Data LifeCycle, or COSA-DLC for short. The COSA-DLC is proposed to generate added value, simplify data management, prepare data for end-user access, provide high-quality data, identify sequences for essential activities, and help system designers create sustainable and efficient software (Sinaeepourfard et al., 2016). Therefore, we used the data lifecycle as a basis for categorizing the challenges, to account for all the phases of the lifecycle of geospatial data. Within the original data lifecycle framework, value creation is a part of every phase through efficiency in the different dimensions. However, we identified the need to add value creation as its own phase, analyzed through a sustainability lens. We added it on behalf of stakeholders who do not directly deal with the data but are assigned to create value from it or make decisions based on it.

The sustainability triple bottom line – consisting of economic, environmental, and social sustainability – is considered by this study as an element for value creation and an area that can be impacted based on the challenges identified in the data lifecycle. While sustainability is often seen from a global perspective (Brown et al., 1987), we identified the need to scope sustainability into the triple bottom line and consider business sustainability 3.0 (Dyllick & Muff, 2016) as an adequate ideal for organizations attempting to approach the triple bottom line. This study proposes to use the sustainability triple bottom line to weigh in on value creation in the data lifecycle. Primarily, we identify the need to understand data in a sustainability context. Geodata is a broad term that comprises a wide variety of data types that are especially relevant to urban planning and development. Therefore, geodata has the potential for sustainable value creation, but challenges or other factors may hinder this ability.

Finally, the geodata lifecycle is a new term that comprises the idea of the data lifecycle as a framework and a specific data domain pertaining to geography. This proposed term rises from the identified need to engage in cross-domain research to address challenges that society faces today (Gholami et al., 2016), as well as lacking semantics for assessing sustainability surrounding terms such as BIM and GIS (Kuster et al., 2020). As Jagadish et al. (2014) point out, creating value from data is a multistep process. Although that conclusion is based on Big Data processes, it shares many similarities with geospatial data. Acquisition, information extraction and cleaning, data integration, modeling and analysis, and interpretation and deployment are phases mentioned as fragmented in many discussions. As a consequence, discussions of these phases often only focus on one of two of them, ignoring the rest (Jagadish et al., 2014). The goal of defining a geodata lifecycle is to achieve a holistic view of its entire lifespan and what phases are relevant for researchers to take into consideration.

7.4 Implications for Practitioners

The results from the interviews and literature show a need for an enhanced focus and awareness of the abilities, challenges, limits, and possibilities of geodata amongst practitioners. We have uncovered that many stakeholders only have a relationship to some of the phases of the geodata lifecycle, without considering the rest. By increasing awareness of the remaining phases and respective challenges included in the geodata lifecycle, practitioners can better understand how their processes impact other practitioners' operations. If the practitioners are familiar with the impact of the challenges within other geodata lifecycle phases, they might have a better foundation to mitigate their own impact. Additionally, we received input on the unused potential in geodata, caused by a lack of definition of responsibilities and collaboration among geodata stakeholders. To find synergies and utilize geodata on a grander scale, we propose enabled sharing, cross-domain collaborations, and initiatives to make the data understandable.

There are restrictions that apply to the geographical localization, such as collection and fusion of datasets that could identify individuals. However, GDPR provides little if any technical guidance to entities that are obliged to implement it (Politou et al., 2018). Through the interviews, we found that this creates uncertainties in the legislations revolving all of the phases of the data lifecycle. In some cases, practitioners even limit themselves more than necessary in fear of violating the regulations. Consequently, an increased focus on competence within GDPR and other limiting regulations could allow practitioners to collect and combine geodata to the maximum extent of the legislation.

We also point to redundant, overlapping, or conflicting projects and investments, and propose a holistic view on project management within organizations (Elonen & Artto, 2003). Lastly, we hope that our perspective on the potential within geodata in sustainable development can inspire geodata practitioners to look for sustainable ways to utilize their data.

7.5 Limitations and Implications for Further Research

The study has been conducted as a qualitative study, but with an initial intent to perform workshops with some of the informants after the interviews. We considered this as a well-suited research strategy but did not get a chance to conduct the workshops as planned due to the COVID-19 pandemic outbreak and the following restrictions. The workshop could potentially give some more content regarding mitigating initiatives, and the impact of current solutions on sustainable dimensions. In addition to that, our sample did not consist of a satisfactory number of decision-makers. We wanted to be able to compare the findings and look for inconsistencies amongst data providers and their initial intent with the data, and decision-makers and their actual use of the data.

Nonetheless, we found many significant and relevant challenges related to the use of geodata and how it impacts sustainable development within our limited sample. A more extensive study with more actors, different stakeholders, and more decision-makers from additional organizations would be interesting to conduct. A quantitative study to assess the extent of these challenges could also provide a more generalizable foundation on their scope and impact.

It would also be interesting to see whether the ongoing projects, such as sharing platforms and parameterization, actually get implemented and what impact they have. In reference to lacking research on the impact of visualization and modeling of geodata on intelligibility and comprehensibility amongst decision-makers, more research is needed to close this research gap. The power of visualization is described as influential in this study, and we recommend further research on the subject. The same goes for parameterization based on geodata within sustainable development, where more general approaches to parameterization are dominant.

Furthermore, the sustainability term is hard to define and ambiguous, which makes it difficult to place the challenges within the given categories. In addition to being a self-report study, the sustainability terms are subjective, to some degree, which makes our interpretation of it open for discussion. Looking at geodata through an information systems lens, and in the context of sustainable development can also have an impact on the results. However, we recommend more research of this kind in other environments to mitigate the risk of results being influenced by bias.

7.6 Conclusion

To answer the first research question: *How do challenges within the geodata lifecycle impact achievement of sustainability triple bottom line?,* we started by researching challenges within the geodata lifecycle. Some of these challenges apply to all three dimensions, such as bias, parameterization, information silos, sustainability silos, procurement processes, and redundant, conflicting, or overlapping projects. Some challenges exclusively impact economic sustainability, such as balancing external and internal resources in regards to cost of competence, systems, and methods. Within environmental sustainability, challenges in controlling the sustainable extent of the organization, and external collaborators has one of the most substantial impacts. Lacking data can also impact the ability to analyze environmental factors. The impacts on social sustainability revolve around work conditions and culture, although it is the least frequently mentioned dimension of impact.

After obtaining an overview of the challenges and their impact, our second research question was formulated: *How do organizations meet the data lifecycle challenges and what initiatives can be implemented for achieving increased sustainability?*. To answer that, we evaluated the challenges and barriers, to assess which of these are possible for the informants to mitigate and to what degree those initiatives could have an impact. Some initiatives, e.g., increased sharing, parameterization, plans for new acquisition technologies, or projects to parameterize data towards sustainable factors, were mentioned by the informants. However, with most of the challenges mentioned by the informants, they did not see a solution or mitigating initiatives. To better answer RQ2, we turned to the literature to find potential solutions. That resulted in some implications on sharing culture, as well as open repository projects that could inspire beneficial projects. We also propose modeling techniques and levels as an initiative to increase intelligibility and supply chain collaboration and standardization of processes to better enable collaboration. Additionally, we presented parameterization principles and examples to better measure impact and a holistic approach to prevent conflicting projects and investments.

Many of the challenges we discovered in the interviews can be validated by existing literature. However, we uncovered some that are not accounted for, and some with a lacking literature basis; one of which is the challenges related to conveying geodata to decision-makers, and for the decision-makers to make sustainable decisions based on that data. Additionally, challenges in parameterizing geodata to achieve sustainable progress impacts sustainable benefit realization, but are not explicitly addressed in the literature. Still, mitigating initiatives can be linked to existing literature without the geographical aspect. When it comes to sharing and collaborating, there are initiatives under development, but none that are implemented with a proven impact. We also uncovered competence gaps within our selection of informants. We found that the actors who deal directly with the geodata do not actively look for potential application areas. At the same time, decision-makers and other stakeholders do not understand the data well enough to find synergies and utilize it to its full potential. Thus, there is a need for competence-increasing initiatives and collaboration platforms. Lastly, we introduce a set of recommendations for further research to increase value creation of geospatial data in the context of sustainable development.

References

Adesiyun, A., Tromaras, A., Almeras, C., Antunes, A., de Lurdes Antunes, M., Barateiro, J., Erdelean, I., Jasiūnienė, V., Manola, N., Mavropoulou, E., Milenkovic, M., Fleten Nielsen, A., Olsen, S., Paliukaitė, M., Pereira, A., Prändl-Zika, V., & Ratkevičiūtė, K. (2019). *Open/FAIR data, software and infrastructure in European transport research* (European Forum and OBsERvatory for OPEN Science in Transport) [Work Package]. FEHRL.

Andersson, C., & Runeson, P. (2002). Verification and validation in industry—A qualitative survey on the state of practice. *Proceedings International Symposium on Empirical Software Engineering*, 37–47. https://doi.org/10.1109/ISESE.2002.1166923

Ardissono, L., Lucenteforte, M., Mauro, N., Savoca, A., & Angioletta, V. (2017). *Personalised community maps*. 156–178. https://doi.org/10.1504/IJEG.2017.084648

Athanasis, N., Themistocleous, M., Kalabokidis, K., Papakonstantinou, A., Soulakellis, N., & Palaiologou, P. (2018). The Emergence of Social Media For Natural Disasters Management: A Big Data Perspective. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLII-3-W4*, 75–82. https://doi.org/10.5194/isprs-archives-XLII-3-W4-75-2018

Awange, J., & Kiema, J. (2019). Geodata and Geoinformatics. In J. Awange & J. Kiema (Eds.), *Environmental Geoinformatics: Extreme Hydro-Climatic and Food Security Challenges: Exploiting the Big Data* (pp. 17–27). Springer International Publishing. https://doi.org/10.1007/978-3-030-03017-9_2

Barbui, C., Gureje, O., Puschner, B., Patten, S., & Thornicroft, G. (2016). Implementing a data sharing culture. *Epidemiology and Psychiatric Sciences*, *25*(4), 289–290. https://doi.org/10.1017/S2045796016000330

Barik, R. K., Lenka, R. K., Sahoo, S. R., Das, B. B., & Pattnaik, J. (2018). Development of Educational Geospatial Database for Cloud SDI Using Open Source GIS. In K. Saeed, N. Chaki, B. Pati, S. Bakshi, & D. P. Mohapatra (Eds.), *Progress in Advanced Computing and Intelligent Engineering* (pp. 685–695). Springer. https://doi.org/10.1007/978-981-10-6872-0_66

Bazeley, P., & Jackson, K. (2013). *Qualitative Data Analysis with NVivo*. SAGE.

Beattie, J., & Baron, J. (1988). Confirmation and matching biases in hypothesis testing. *The Quarterly Journal of Experimental Psychology Section A*, *40*(2), 269–297. https://doi.org/10.1080/02724988843000122

Berg, V., Birkeland, J., Nguyen-Duc, A., Pappas, I. O., & Jaccheri, L. (2020). Achieving agility and quality in product development—An empirical study of hardware startups. *Journal of Systems and Software*, *167*, 110599. https://doi.org/10.1016/j.jss.2020.110599

Brown, B. J., Hanson, M. E., Liverman, D. M., & Merideth, R. W. (1987). Global sustainability: Toward definition. *Environmental Management*, *11*(6), 713–719. https://doi.org/10.1007/BF01867238

Castro, R., Tierra, A., & Luna, M. (2019). Assessing the Horizontal Positional Accuracy in OpenStreetMap: A Big Data Approach. In Á. Rocha, H. Adeli, L. P. Reis, & S. Costanzo (Eds.), *New Knowledge in Information Systems and Technologies* (pp. 513–523). Springer International Publishing. https://doi.org/10.1007/978-3-030-16184-2_49

Chen, W., & Hirschheim, R. (2004). A paradigmatic and methodological examination of information systems research from 1991 to 2001. *Information Systems Journal*, *14*(3), 197–235. https://doi.org/10.1111/j.1365-2575.2004.00173.x

Claramunt, C., & Stewart, K. (2015). Special Issue on Spatio-Temporal Theories and Models for Environmental, Urban and Social Sciences: Where Do We Stand? *Spatial Cognition & Computation*, *15*(2), 61–67. https://doi.org/10.1080/13875868.2015.1004178

Cruzes, D. S., & Dybå, T. (2011). Recommended Steps for Thematic Synthesis in Software Engineering. *2011 International Symposium on Empirical Software Engineering and Measurement*, 275–284. https://doi.org/10.1109/ESEM.2011.36

Dyllick, T., & Muff, K. (2016). Clarifying the Meaning of Sustainable Business: Introducing a Typology From Business-as-Usual to True Business Sustainability. *Organization & Environment*, *29*(2), 156–174. https://doi.org/10.1177/1086026615575176

Eitzel, M. V., Mhike Hove, E., Solera, J., Madzoro, S., Changarara, A., Ndlovu, D., Chirindira, A., Ndlovu, A., Gwatipedza, S., Mhizha, M., & Ndlovu, M. (2018). Sustainable development as successful technology transfer: Empowerment through teaching, learning, and using digital participatory mapping techniques in Mazvihwa, Zimbabwe. *Development Engineering*, *3*, 196–208. https://doi.org/10.1016/j.deveng.2018.07.001

Elonen, S., & Artto, K. A. (2003). Problems in managing internal development projects in multi-project environments. *International Journal of Project Management*, *21*(6), 395–402. https://doi.org/10.1016/S0263-7863(02)00097-2

Etzion, D., & Aragon-Correa, J. A. (2016). Big Data, Management, and Sustainability: Strategic Opportunities Ahead. *Organization & Environment*, *29*(2), 147–155. https://doi.org/10.1177/1086026616650437

Fang, S., Xu, L. D., Zhu, Y., Ahati, J., Pei, H., Yan, J., & Liu, Z. (2014). An Integrated System for Regional Environmental Monitoring and Management Based on Internet of Things. *IEEE Transactions on Industrial Informatics*, *10*(2), 1596–1605. https://doi.org/10.1109/TII.2014.2302638

Fang, Shifeng, Xu, L. D., Zhu, Y., Ahati, J., Pei, H., Yan, J., & Liu, Z. (2014). An Integrated System for Regional Environmental Monitoring and Management Based on Internet of Things. *IEEE Transactions on Industrial Informatics*, *10*(2), 1596–1605. https://doi.org/10.1109/TII.2014.2302638

Fang, Shifeng, Xu, L., Pei, H., Liu, Y., Liu, Z., Zhu, Y., Yan, J., & Zhang, H. (2014). An Integrated Approach to Snowmelt Flood Forecasting in Water Resource Management. *IEEE Transactions on Industrial Informatics*, *10*(1), 548–558. https://doi.org/10.1109/TII.2013.2257807

Fernández, P., Santana, J., Sánchez, A., Trujillo, A., Domínguez, C., & Suárez, J. P. (2016). A GIS Water Management System Using Free and Open Source Software. In C. R. García, P. Caballero-Gil, M. Burmester, & A. Quesada-Arencibia (Eds.), *Ubiquitous Computing and Ambient Intelligence* (pp. 383–394). Springer International Publishing. https://doi.org/10.1007/978-3-319-48799-1_43

Flåthen, K. (2007). Norway Digital—The most comprehensive administrative co-operation in Europe. *Roczniki Geomatyki - Annals of Geomatics*, *5*(6), 41-45 47-52.

Fontana, A., & Frey, J. H. (2000). *The Interview: From Structured Questions to Negotiated Text*.

Gholami, R., Watson, R., Molla, A., Hasan, H., & Bjorn-Andersen, N. (2016). Information Systems Solutions for Environmental Sustainability: How Can We Do More? *Journal of the Association for Information Systems*, *17*, 521–536. https://doi.org/10.17705/1jais.00435

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*.

Goodchild, M. F. (2008). *Assertion and Authority: The Science of User- Generated Geographic Content*.

Goodman, B., & Flaxman, S. (2017). European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation." *AI Magazine*, *38*(3), 50–57. https://doi.org/10.1609/aimag.v38i3.2741

Granell, C., Díaz, L., & Gould, M. (2010). Service-oriented applications for environmental models: Reusable geospatial services. *Environmental Modelling & Software*, *25*(2), 182–198. https://doi.org/10.1016/j.envsoft.2009.08.005

Grigoropoulos, G., Keler, A., Kaths, H., Spangler, M., & Busch, F. (2019). *Using Geodata for Simulating Urban Traffic—Current Research in the Field of Traffic Engineering and Control* (pp. 45–49).

Guo, B., Deng, X., Guan, Q., Tian, J., & Zheng, X. (2018). An Incentive Mechanism for Cross-Organization Data Sharing Based on Data Competitiveness. *IEEE Access*, *6*, 72836–72844. https://doi.org/10.1109/ACCESS.2018.2882233

Hancke, G. P., Silva, B. D. C. e, & Hancke, J. (2013). The Role of Advanced Sensing in Smart Cities. *Sensors*, *13*(1), 393–425. https://doi.org/10.3390/s130100393

Heipke, C. (2004). Some Requirements for Geographic Information Systems. *Photogrammetric Engineering & Remote Sensing*, *70*(2), 185–195. https://doi.org/10.14358/PERS.70.2.185

Heipke, C. (2010). Crowdsourcing geospatial data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *65*(6), 550–557. https://doi.org/10.1016/j.isprsjprs.2010.06.005

Hong, J.-H., & Huang, M.-L. (2017). Enabling smart data selection based on data completeness measures: A quality-aware approach. *International Journal of Geographical Information Science*, *31*(6), 1178–1197. https://doi.org/10.1080/13658816.2017.1290251

Huang, D., Chen, H., Huang, S., Lin, Y., & Ma, Y. (2019). Application of Internet of Things and GIS in Power Grid Emergency Command System. In G. Wang, M. Z. A. Bhuiyan, S. De Capitani di Vimercati, & Y. Ren (Eds.), *Dependability in Sensor, Cloud, and Big Data Systems and Applications* (pp. 304–313). Springer. https://doi.org/10.1007/978-981-15-1304-6_24

Iyamu, T. (2018). Collecting qualitative data for information systems studies: The reality in practice. *Education and Information Technologies*, *23*(5), 2249–2264. https://doi.org/10.1007/s10639-018-9718-2

Jagadish, H. V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R., & Shahabi, C. (2014). Big data and its technical challenges. *Communications of the ACM*, *57*(7), 86–94. https://doi.org/10.1145/2611567

Jałocha, B., Krane, H. P., Ekambaram, A., & Prawelska-Skrzypek, G. (2014). Key Competences of Public Sector Project Managers. *Procedia - Social and Behavioral Sciences*, *119*, 247–256. https://doi.org/10.1016/j.sbspro.2014.03.029

Jensen, M., Gutierrez, J. M., & Pedersen, J. M. (2015). Vehicle data activity quantification using spatio-temporal GIS on modelling smart cities. *2015 International Conference on Computing, Networking and Communications (ICNC)*, 303–307. https://doi.org/10.1109/ICCNC.2015.7069359

Johnson, R. (1997). Examining the Validity Structure of Qualitative Research. *Educ*, *118*.

Kain, J. S. (2004). The Kain–Fritsch Convective Parameterization: An Update. *Journal of Applied Meteorology*, *43*(1), 170–181. https://doi.org/10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2

Kang, S., & Moon, T. (2015). Impact of Organizational Competence on Supply Chain Performance through Supply Chain Collaboration. *Indian Journal of Science and Technology*, *8*. https://doi.org/10.17485/ijst/2015/v8i12/68356

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T. M., Fiedler, S., & Nosek, B. A. (2016). Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. *PLOS Biology*, *14*(5), e1002456. https://doi.org/10.1371/journal.pbio.1002456

Kitchenham, B., & Charters, S. (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering*.

Kitchin, R. (2014). The real-time city? Big data and smart urbanism. *GeoJournal*, *79*(1), 1–14. https://doi.org/10.1007/s10708-013-9516-8

Klein, H. K., & Myers, M. D. (1999). A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems. *MIS Quarterly*, *23*(1), 67–93. JSTOR. https://doi.org/10.2307/249410

Korstjens, I., & Moser, A. (2018). Series: Practical guidance to qualitative research. Part 4: Trustworthiness and publishing. *European Journal of General Practice*, *24*(1), 120–124. https://doi.org/10.1080/13814788.2017.1375092

Krämer, M., & Senner, I. (2015). A modular software architecture for processing of big geospatial data in the cloud. *Computers & Graphics*, *49*, 69–81. https://doi.org/10.1016/j.cag.2015.02.005

Kurtz, C., Semmann, M., & BÃ¶hmann, T. (2018). Privacy by Design to Comply with GDPR: A Review on Third-Party Data Processors. *AMCIS 2018 Proceedings*. https://aisel.aisnet.org/amcis2018/Security/Presentations/36

Kuster, C., Hippolyte, J.-L., & Rezgui, Y. (2020). The UDSA ontology: An ontology to support real time urban sustainability assessment. *Advances in Engineering Software*, *140*, 102731. https://doi.org/10.1016/j.advengsoft.2019.102731

Lee, J.-G., & Kang, M. (2015). Geospatial Big Data: Challenges and Opportunities. *Big Data Research*, *2*(2), 74–81. https://doi.org/10.1016/j.bdr.2015.01.003

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic Inquiry*. SAGE.

Lü, G., Batty, M., Strobl, J., Lin, H., Zhu, A.-X., & Chen, M. (2019). Reflections and speculations on the progress in Geographic Information Systems (GIS): A geographic perspective. *International Journal of Geographical Information Science*, *33*(2), 346–367. https://doi.org/10.1080/13658816.2018.1533136

Lu, Y., Xie, H., Arshika Zain, S., & Xu, Z. (2019). Geographic Information Systems and Big Data Driven Framework for Planning and Design of Smart Cities. *2019 4th International Conference on Information Systems Engineering (ICISE)*, 6–10. https://doi.org/10.1109/ICISE.2019.00009

Malhotra, A., Melville, N. P., & Watson, R. T. (2013). Spurring Impactful Research on Information Systems for Environmental Sustainability. *MIS Quarterly*, *37*(4), 1265–1274. JSTOR.

Marshall, C., Brereton, P., & Kitchenham, B. (2015). Tools to support systematic reviews in software engineering: A cross-domain survey using semi-structured interviews. *Proceedings of the 19th International Conference on Evaluation and Assessment in Software Engineering*, 1–6. https://doi.org/10.1145/2745802.2745827

McGlinn, K., Debruyne, C., McNerney, L., & O'Sullivan, D. (2017). Integrating Ireland's Geospatial Information to Provide Authoritative Building Information Models. *Proceedings of the 13th International Conference on Semantic Systems*, 57–64. https://doi.org/10.1145/3132218.3132223

Milne, M. J., & Gray, R. (2013). W(h)ither Ecology? The Triple Bottom Line, the Global Reporting Initiative, and Corporate Sustainability Reporting. *Journal of Business Ethics*, *118*(1), 13–29. https://doi.org/10.1007/s10551-012-1543-8

Moe, C. E., & Päivärinta, T. (2011). Challenges in Information Systems Procurement in the Norwegian Public Sector. In M. Janssen, H. J. Scholl, M. A. Wimmer, & Y. Tan (Eds.), *Electronic Government* (pp. 404–417). Springer. https://doi.org/10.1007/978-3-642-22878-0_34

Myers, M. D. (1997). Qualitative Research in Information Systems. *MIS Quarterly*, *21*(2), 241–242. http://www.misq.org/supplements/.

Myers, M. D., & Newman, M. (2007). The qualitative interview in IS research: Examining the craft. *Information and Organization*, *17*(1), 2–26. https://doi.org/10.1016/j.infoandorg.2006.11.001

Nimmagadda, S. L., Reiners, T., & Rudra, A. (2017). An Upstream Business Data Science in a Big Data Perspective. *Procedia Computer Science*, *112*, 1881–1890. https://doi.org/10.1016/j.procs.2017.08.236

Oberhofer, P., & Dieplinger, M. (2014). Sustainability in the Transport and Logistics Sector: Lacking Environmental Measures. *Business Strategy and the Environment*, *23*(4), 236–253. https://doi.org/10.1002/bse.1769

O'Brien, W. (2015, June 15). *Sustainability is the Key Driver of Innovation | Will O'Brien | TEDxWPI - YouTube* [Online Video]. https://www.youtube.com/watch?v=jl1Er3w_Beg

Orlikowski, W. J., & Baroudi, J. J. (1991). Studying Information Technology in Organizations: Research Approaches and Assumptions. *Information Systems Research*, *2*(1), 1–28. https://doi.org/10.1287/isre.2.1.1

Palts, K., Kindler, K., & Campillo Borras, E. (2019). *Open Science framework, terminology and instruments* (European Forum and OBsErvatory for OPEN Science in Transport, p. 68) [Work Package]. German Aerospace Centre (DLR).

Parmenter, D. (2015). *Key Performance Indicators: Developing, Implementing, and Using Winning KPIs*. John Wiley & Sons.

Pijanowski, B. C., Tayyebi, A., Doucette, J., Pekin, B. K., Braun, D., & Plourde, J. (2014). A big data urban growth simulation at a national scale: Configuring the GIS and neural network based Land Transformation Model to run in a High Performance Computing (HPC) environment. *Environmental Modelling & Software*, *51*, 250–268. https://doi.org/10.1016/j.envsoft.2013.09.015

Politou, E., Alepis, E., & Patsakis, C. (2018). Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions. *Journal of Cybersecurity*, *4*(1). https://doi.org/10.1093/cybsec/tyy001

Pons, X., & Masó, J. (2016). A comprehensive open package format for preservation and distribution of geospatial data and metadata. *Computers & Geosciences*, *97*, 89–97. https://doi.org/10.1016/j.cageo.2016.09.001

Reuschel, A.-K., Piatti, B., & Hurni, L. (2013). Modelling Uncertain Geodata for the Literary Atlas of Europe. In K. Kriz, W. Cartwright, & M. Kinberger (Eds.), *Understanding Different Geographies* (pp. 135–157). Springer. https://doi.org/10.1007/978-3-642-29770-0_11

Richardson, H., & Robinson, B. (2007). The mysterious case of the missing paradigm: A review of critical information systems research 1991–2001. *Information Systems Journal*, *17*(3), 251–270. https://doi.org/10.1111/j.1365-2575.2007.00230.x

Robson, C. (2002). *Real World Research: A Resource for Social Scientists and Practitioner-Researchers / C. Robson.*

Rowhani-Farid, A., Allen, M., & Barnett, A. G. (2017). What incentives increase data sharing in health and medical research? A systematic review. *Research Integrity and Peer Review*, *2*(1), 4. https://doi.org/10.1186/s41073-017-0028-9

Runeson, P., & Höst, M. (2008). Guidelines for conducting and reporting case study research in software engineering. *Empirical Software Engineering*, *14*(2), 131. https://doi.org/10.1007/s10664-008-9102-8

Saldaña, J. (2015). *The Coding Manual for Qualitative Researchers* (3rd ed.). SAGE.

Scholtz, B., Connolley, A., & Calitz, A. (2014). Enterprise Architectures for Addressing Sustainability Silos. In B. Funk, P. Niemeyer, & J. M. Gómez (Eds.), *Information Technology in Environmental Engineering: Selected Contributions to the Sixth International Conference on Information Technologies in Environmental Engineering (ITEE2013)* (pp. 141–154). Springer. https://doi.org/10.1007/978-3-642-36011-4_12

Shahrokni, A., & Söderberg, J. (2015). Beyond Information Silos Challenges in Integrating Industrial Model-based Data. *BigMDE@STAF.*

Sharma, R., Kamble, S. S., & Gunasekaran, A. (2018). Big GIS analytics framework for agriculture supply chains: A literature review identifying the current trends and future perspectives. *Computers and Electronics in Agriculture*, *155*, 103–120. https://doi.org/10.1016/j.compag.2018.10.001

Shook, E., Vento, D. D., Zonca, A., & Wang, J. (2018). GISandbox: A Science Gateway for Geospatial Computing. *Proceedings of the Practice and Experience on Advanced Research Computing*, 1–7. https://doi.org/10.1145/3219104.3219150

Sinaeepourfard, A., Garcia, J., Masip-Bruin, X., & Marín-Tordera, E. (2016). Towards a Comprehensive Data LifeCycle Model for Big Data Environments. *2016 IEEE/ACM 3rd International Conference on Big Data Computing Applications and Technologies (BDCAT)*, 100–106.

Smiciklas, J., Prokop, G., Stano, P., & Sang, Z. (2017). *Collection Methodology for Key Performance Indicators for Smart Sustainable Cities* (No. 978-92-61-25231–1; p. 123).

United for Smart Sustainable Cities (U4SSC). https://www.unece.org/fileadmin/DAM/hlm/documents/Publications/U4SSC-CollectionMethodologyforKPIfoSSC-2017.pdf

Sommer, A. F., Dukovska-Popovska, I., & Steger-Jensen, K. (2014). Barriers towards integrated product development—Challenges from a holistic project management perspective. *International Journal of Project Management*, *32*(6), 970–982. https://doi.org/10.1016/j.ijproman.2013.10.013

Stefan, M., Lopez, J. G., & Olsen, R. L. (2018). Exploring the Potential of Modern Advanced Metering Infrastructure in Low-Voltage Grid Monitoring Systems. *2018 IEEE International Conference on Big Data (Big Data)*, 3543–3548. https://doi.org/10.1109/BigData.2018.8622154

Stendal, K., & Westin, S. (2018). IMPLEMENTING CLOUD BASED BIG DATA PLATFORMS – A CASE USING MICROSOFT AZURE. *26*. https://uia.brage.unit.no/uia-xmlui/handle/11250/2596543

Thomas, D. R. (2006). A General Inductive Approach for Analyzing Qualitative Evaluation Data. *American Journal of Evaluation*, *27*(2), 237–246. https://doi.org/10.1177/1098214005283748

Tsang, E. W. K. (2014). Case studies and generalization in information systems research: A critical realist perspective. *The Journal of Strategic Information Systems*, *23*(2), 174–186. https://doi.org/10.1016/j.jsis.2013.09.002

Utenriksdepartementet. (2018, July 11). *One year closer 2018* [Rapport]. Regjeringen.no; regjeringen.no. https://www.regjeringen.no/no/dokumenter/closer_2018/id2607226/

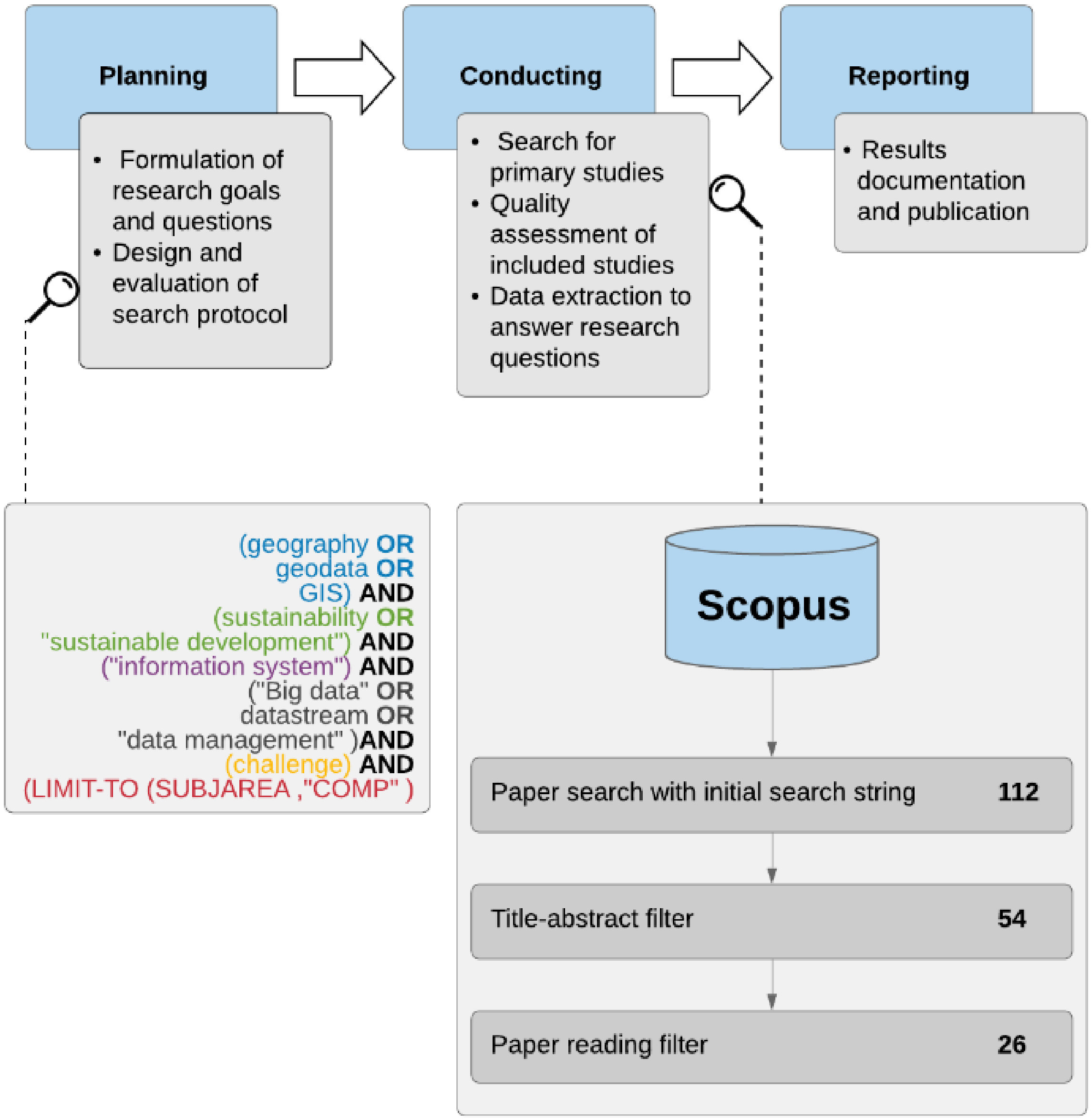
Walsham, G. (1995). Interpretive case studies in IS research: Nature and method. *European Journal of Information Systems*, *4*(2), 74–81. https://doi.org/10.1057/ejis.1995.9

Walsham, Geoff. (2006). Doing interpretive research. *European Journal of Information Systems*, *15*(3), 320–330. https://doi.org/10.1057/palgrave.ejis.3000589

Wang, S., Anselin, L., Bhaduri, B., Crosby, C., Goodchild, M. F., Liu, Y., & Nyerges, T. L. (2013). CyberGIS software: A synthetic review and integration roadmap. *International Journal of Geographical Information Science*, *27*(11), 2122–2145. https://doi.org/10.1080/13658816.2013.776049

Yang, C., Goodchild, M., Huang, Q., Nebert, D., Raskin, R., Xu, Y., Bambacus, M., & Fay, D. (2011). Spatial cloud computing: How can the geospatial sciences use and help shape cloud computing? *International Journal of Digital Earth*, *4*(4), 305–329. https://doi.org/10.1080/17538947.2011.587547

Appendix 1: Literature Review Search Process



Appendix 2: Consent Form

(Only in Norwegian)

Vil du delta i forskningsprosjektet ***"Dataforvaltning ved bærekraftig utvikling"*?**

Dette er et spørsmål til deg om å delta i et forskningsprosjekt hvor formålet er å avdekke utfordringer knyttet til dataforvaltning. I dette skrivet gir vi deg informasjon om målene for prosjektet og hva deltakelse vil innebære for deg.

Formål

Vi ønsker å se hvordan bedrifter utnytter data for å oppnå mål ifm. effektivisering, redusert ressursbruk eller andre bærekraftsmål. Vi fokuserer på prosessene som inngår i dataforvaltning, som omhandler alt fra innsamling av data til analyse og utnyttelse. Hensikten er å avdekke hvilke utfordringer som foreligger, og hvordan disse kan imøtekommes. Prosjektet er en masterstudie ved Universitetet i Agder.

Hvem er ansvarlig for forskningsprosjektet?

Universitetet i Agder er ansvarlig for prosjektet, i samarbeid med Norkart.

Hvorfor får du spørsmål om å delta?

Utvalget i studiet er trukket ut gjennom instituttets og veileders nettverk, samt aktører vi anser som relevant innenfor dataforvaltning og bærekraftig utvikling. Kontaktopplysninger kan dermed være innhentet gjennom Universitetet i Agder eller veileder i Norkart.

Hva innebærer det for deg å delta?

Hvis du velger å delta i prosjektet, innebærer det å svare på spørsmål stilt i intervjusetting, og eventuelt deltakelse på workshop. Vi samler inn navn, adresse eller telefonnummer i de tilfellene det inngår i kontaktinformasjon, i tillegg til lydopptak som transkriberes. Bakgrunnsopplysninger som vil kunne identifisere en person kan også fremkomme, men vil slettes og anonymiseres før publisering.

Det er frivillig å delta

Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du når som helst trekke samtykke tilbake uten å oppgi noen grunn. Alle opplysninger om deg vil da bli anonymisert. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg.

Ditt personvern – hvordan vi oppbevarer og bruker dine opplysninger

Vi vil bare bruke opplysningene om deg til formålene vi har fortalt om i dette skrivet. Vi behandler opplysningene konfidensielt og i samsvar med personvernregelverket.

- *Prosjektgruppe bestående av to masterstudenter, samt veileder ved Universitetet i Agder vil ha tilgang til opplysningene vi samler inn, før de anonymiseres.*
- *For å sikre at ingen uvedkommende får tilgang til personopplysningene, vil de lagres på instituttets godkjente datalagringstjeneste: OneDrive. Navn og kontaktopplysninger vil erstattes med en kode om lagres på egen navneliste adskilt fra øvrige data.*

Deltakerne vil ikke kunne gjenkjennes i publikasjon, resultatene anonymiseres.

Hva skjer med opplysningene dine når vi avslutter forskningsprosjektet?

Prosjektet skal etter planen avsluttes *04.06.2020. Ved prosjektslutt destrueres opptak og lagrede personlige data, og gjengis kun i anonymisert form.*

Dine rettigheter

Så lenge du kan identifiseres i datamaterialet, har du rett til:

- innsyn i hvilke personopplysninger som er registrert om deg,
- å få rettet personopplysninger om deg,
- få slettet personopplysninger om deg,
- få utlevert en kopi av dine personopplysninger (dataportabilitet), og
- å sende klage til personvernombudet eller Datatilsynet om behandlingen av dine personopplysninger.

Hva gir oss rett til å behandle personopplysninger om deg?

Vi behandler opplysninger om deg basert på ditt samtykke.

På oppdrag fra *Universitetet i Agder* har NSD – Norsk senter for forskningsdata AS vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Hvor kan jeg finne ut mer?

Hvis du har spørsmål til studien, eller ønsker å benytte deg av dine rettigheter, ta kontakt med:

- *Universitetet i Agder* ved Ilias Pappas, veileder for masteroppgaven (+47 381 41 449, ilias.pappas@uia.no).
- Vårt personvernombud: Ina Danielsen (+47 381 42 140 ina.danielsen@uia.no).
- NSD – Norsk senter for forskningsdata AS, på epost (personverntjenester@nsd.no) eller telefon: 55 58 21 17.

Med vennlig hilsen

Prosjektansvarlig	Kandidatene:	
(Veileder)	Maria S. Andersen	maria.sa2910@gmail.com
	Daniel Pettersen	daniel.m.pettersen@uia.no

Samtykkeerklæring

Jeg har mottatt og forstått informasjon om prosjektet Dataforvaltning ved bærekraftig utvikling, og har fått anledning til å stille spørsmål. Jeg samtykker til:

☐ å delta i intervju
☐ å delta i workshop – hvis aktuelt

Jeg samtykker til at mine opplysninger behandles frem til prosjektet er avsluttet, ca. 04.06.2020

(Signert av prosjektdeltaker, dato)