Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

Charlotte Husom Grøder

IoT and Data Governance in Long-Term Environmental Monitoring

An interpretive case study on the use and governance of intelligent technologies in environmental work practice

Master's thesis in Computer Science
Supervisor: Elena Parmiggiani
June 2021

NTNU
Kunnskap for en bedre verden

Charlotte Husom Grøder

IoT and Data Governance in Long-Term Environmental Monitoring

An interpretive case study on the use and governance of intelligent technologies in environmental work practice

Master's thesis in Computer Science
Supervisor: Elena Parmiggiani
June 2021

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

NTNU
Kunnskap for en bedre verden

Abstract

The use of intelligent technologies, particularly IoT devices, in research has contributed to more accessible and available monitoring of the environment with the possibility of monitoring almost anything with low costs. Examples of intelligent technologies used in environmental research are sensors monitoring temperature, humidity or pressure, cameras, and acoustic devices like echo sounders. The devices are intelligent because they can automatically deliver data without human interaction. However, automatic collections of real-time data from sensor devices with heterogeneous quality and sensitivity lead to extensive raw data storage. Subsequently, researchers must ensure that the devices are calibrated and ensure that the data are trustworthy, readable, and meaningful to the context.

To make research feasible, research institutions – which can also be referred to as research infrastructures – sometimes are dependent on receiving funding from governing institutions such as ESFRI and the Research Council of Norway. In order to receive funding – and with an increased focus on open data sharing between research infrastructures, researchers must adapt to the governments' guidelines and requirements to document their research, influencing their data governance activities.

This research aims to contribute empirical insights into the use of intelligent technologies at research infrastructures for environmental monitoring and how researchers' work is affected by the increased use of technology in monitoring. Additionally, researchers' experience adapting to guidelines and requirements on data documenting is also investigated.

It is conducted a case study of selected research infrastructures for environmental research in Norway. The study is based on qualitative data collected from structured and semi-structured interviews and relevant documents from governing institutions such as strategy documents, guidelines, and roadmaps. The findings are based on information retrieved from interviews of data managers, environmental researchers, research coordinators, and research managers that daily work at research infrastructures for environmental monitoring.

The findings show that IoT-based environmental monitoring is enabled by data governance with the established processes to translate raw, often big data sets, into reliable, readable, and meaningful information to support future reuse and interpretation. The processes are affected by policies, procedures, and standards on collecting and managing the data that consequently affect researchers’ work practices. However, due to constraints in time, funding, and resources, the findings also show that it can be challenging to have the desired quality in researchers’ scientific work.

Sammendrag

Bruken av intelligente teknologier, IoT enheter, i forskning har bidratt til å gjøre arbeid med miljøforskning lettere og mer tilgjengelig, da det er mulig å måle nesten det en vil til en lav pris. Eksempler på intelligente teknologier er sensorer som måler temperatur, fuktighet eller trykk, kameraer og akustiske enheter slik som ekkolodd, og de er intelligente fordi de kan levere data automatisk uten behov menneskelig interaksjon. Automatiske innsamlinger av sanntidsdata, ofte fra enheter med ulik kvalitet og følsomhet, fører imidlertid til lagring av omfattende mengder med rådata. Som et resultat av dette må forskere kontinuerlig forsikre seg om at enhetene er kalibrerte, og gjøre dataene pålitelige, lesbare og meningsfulle med hensyn til konteksten.

For å gjøre forskning gjennomførbart, er forskningsstasjoner – som også kan bli referert til som forskningsinfrastrukturer – noen ganger avhengige av finansiering via styrende institusjoner som for eksempel ESFRI og Forskningsrådet i Norge. For å motta finansiering – og med et økt fokus på åpen datadeling mellom forskningsinfrastrukturer, stilles det høyere krav til datadokumentasjon i henhold til de standarder og krav som myndighetene setter. Dette er med på å påvirke forskeres datastyringsarbeid.

Forskningen som er gjort i dette prosjektet har som mål å bidra med empirisk innsikt i hvordan intelligente teknolgier brukes på forskningsstasjoner for miljøovervåkning, og hvordan forskeres arbeid påvirkes av økt bruk av teknologi i overvåkning. Det er undersøkt hvordan forskere opplever retningslinjene satt av forskningsinstitusjoner om hvordan dataene skal dokumenteres, og hvordan deres dataarbeid må tilpasses deretter.

Det er gjennomført et casestudie av utvalgte forskningsstasjoner for miljøovervåkning i Norge, og studien baserer seg på kvalitative data fra strukturerte og semi-strukturerte intervjuer og relevante dokumenter fra myndighetene slik som strategidokumenter, retningslinjer og veikart. Funnene er basert på informasjon fra intervjuer av informanter som til daglig arbeider på forskningsstasjoner for miljøovervåkning som dataledere, forskningssjefer, forskningskoordinatorer eller miljøforskere.

Funnene viser at IoT basert miljøovervåkning er muliggjort av datastyringsarbeid, med etablerte prosesser for å sikre at store sett av rådata blir gjort pålitelige, lesbare og meningsfulle for å støtte fremtidig gjenbruk og tolkning. Prosessene er påvirket av retningslinjer, prosedyrer og standarder for hvordan dataene skal samles inn og håndteres, som er med på å påvirke forskeres arbeidspraksiser. På grunn av begrensninger i tid, finansiering og ressurser, viser funnene imidlertid at det kan være vanskelig å ha ønsket kvalitet i forskernes vitenskapelige arbeid.

Acknowledgments

In collaboration with the Department of Computer Science at the Norwegian University of Science and Technology (NTNU), this thesis is written in the spring semester of 2021 as a final assignment to accomplish a Master of Science (MSc) in Engineering.

I wish to express my deepest gratitude to my supervisor during the research project, Elena Parmiggiani, Associate Professor at the Department of Computer Science at NTNU, who allowed me to do this exciting project on IoT and data governance in long-term environmental monitoring. You were always available during the project and provided insightful feedback on my work. Thank you also for giving me good advice and ideas for my writing.

A very big thank you also to Nana Kwame Amagyei, PhD Candidate at the Department of Computer Science at NTNU, for the educational collaboration in collecting qualitative data through interviews and for the professional conversations about our research. Thank you for always being available, helpful and for supporting my work.

Finally, I would like to thank all the informants who participated in the project through interviews. It would not have been possible to complete this thesis without your intervention, providing the research with valuable insights into environmental research at your research sites.

Charlotte Husom Grøder

Trondheim, June 2021

Contents

List of figures

List of tables

List of abbreviations

AI	Artificial Intelligence
DSS	Decision Support System
eLTER	European Long-Term Ecosystem Research network
IoT	Internet of Things
IS	Information System
ISO	The International Organization for Standardization
LTER	Long-Term Ecosystem Research
NTNU	Norwegian University of Science and Technology
RFID	Radio Frequency Identification
RI	Research Infrastructure
SDI	Stepwise Deductive Induction

1 Introduction

This thesis is the final part of a project consisting of a semester project in the autumn of 2020 and a master's thesis in the spring of 2021 investigating IoT and data governance in long-term environmental monitoring. The semester project was a literature review of scientific papers belonging to the information systems (IS) field about IoT and data management to investigate the advantages of using sensor devices in environmental research and identify the main challenges of data management (see Grøder, 2020).

This chapter is a continuation of the semester project, and it will identify the purpose for the project, the research questions, and the project's contributions. Finally, it will present the structure of this thesis.

1.1 Purpose of the Project

During the COVID-19 pandemic, researchers have monitored and identified surfaces to locate the virus to understand the risks humans are exposed to by traveling (see Klokk and Mikalsen, 2020). Understanding such situations as early as possible can be essential regarding future pandemics to place measures early (Klokk and Mikalsen, 2020).

Long-term environmental monitoring is essential for understanding and revealing phenomena in nature (Karasti, Baker and Halkola, 2006), and it can be arranged in networks of research stations that are a part of research infrastructures (RIs) for environmental monitoring. Each RI typically focuses on a few critical aspects of the natural environment like fresh- or marine waters, forest, agricultural, or alpine areas (LTER-Europe, 2017). The type of monitoring can depend on the investigated objective and the purpose of the monitoring.

Developments in technology have increased the use of Internet of Things (IoT) devices in research, such as sensors, cameras, and acoustic devices, which leads to new opportunities to observe and monitor processes and physical objects (Labonnote, Bryhni and Lech, 2021). IoT devices' automatic data collections allow researchers to gather real-time data more efficiently through, for example, sensor networks (Tan and Wang, 2010; Monteiro and Parmiggiani, 2019). However, obtaining real-time data from the research field can be challenging as this results in collections of large amounts of raw data (Boos *et al.*, 2013; Angelakis *et al.*, 2017). Often, sensor devices of heterogeneous quality and sensitivity are used in this process, which means that environmental researchers must constantly manage the data and ensure that the sensor devices are calibrated, and that the environmental data are made trustworthy, readable, and meaningful for its context. As a result, researchers must perform tasks such as minimize the data size, ensure that the data have a certain quality, and establish processes that handle the data through its life cycle – which all are part of data governance. Therefore, data governance is a concern for monitoring with IoT devices.

One way to form the data governance process is establishing a RI. RIs are becoming more regulated by international and national organizations that support RIs with funding, resources, and tools to conduct research, such as ESFRI and The Research Council of Norway. In order to receive funding, and with an increased focus on open data sharing between RIs (LTER, 2020), researchers must often adapt to guidelines and standardized ways to document their research (Karasti, Baker and Millerand, 2010). Making data fit existing standards was elaborated to be a challenge in the semester project (see Grøder, 2020) as research on local levels might differ from the guidelines. Moreover, in accordance with open science, a new mindset of data ownership should focus on researchers being stewards of the data instead of data *owners*. However, it still exists a mindset where data is best kept private instead of shared; for example, Johnston (2021) writes about public officials reluctant to share data through data dashboards, even during a crisis like COVID-19 (see Johnston, 2021).

This thesis presents a theoretical framework to investigate the complex socio-technical processes at RIs where humans and technology must interact to make research feasible. The research will investigate how IoT-based long-term environmental monitoring is affected by the established data governance processes. While the research will take a socio-technical view on monitoring with IoT devices, it will not contain careful descriptions of different technologies that are used in environmental monitoring, such as sensor types, software, or hardware.

To illustrate the framework, semi-structured and structured interviews are conducted to investigate how researchers adapt to the increased usage of IoT for environmental monitoring and how it affects their daily work and data governance activities.

1.2 Research Questions

The following questions are pursued to guide the research:

RQ1: How are environmental work practices affected by the use of IoT in monitoring?

RQ1.1: How is IoT used in research infrastructures?

RQ1.2: How is data governance arranged in research infrastructures?

1.3 Contribution

The IS literature contains research papers from several interpretive case studies about IoT for environmental monitoring (see, for example: Sung and Hsu, 2013; Parmiggiani, Monteiro and Hepsø, 2015; Govoni *et al.*, 2017; Truong, Dinh and Wahid, 2017). Previous research investigates data governance (see Otto, 2011; Parmiggiani and Grisot, 2020; Mikalef *et al.*, 2020), but there is a lack of published papers in the IS field covering data governance as a result of IoT-based monitoring. This research aims to contribute to the research domain with a socio-technical view of monitoring with intelligent devices at research infrastructures for environmental monitoring. The research project can contribute empirical insight into how IoT is used in RIs in Norway by improving the understanding of monitoring with IoT and data governance. A theoretical insight can also be a contribution while different concepts will be defined and discussed (Walsham, 1995).

1.4 Structure of the Thesis

The thesis is divided into seven chapters, and they are structured as follows:

Chapter 2 defines the research's theoretical framework by elaborating on previous research to define the main concepts relevant to this research.

Chapter 3 describes the case of the thesis.

Chapter 4 describes the research method and elaborates on the research process: The research strategy, data generation methods, and data analysis, and finally, the research paradigm.

Chapter 5 presents the findings identified from the empirical material collected from interviews.

Chapter 6 considers the research questions and discusses the findings presented in the analytical framework against the theoretical framework.

Chapter 7 will conclude the thesis and elaborate on limitations and propose future research directions.

Appendices contain two documents: The questions for the structured interview and the interview guide with interview questions, both in Norwegian and English.

2 Literature Background

This chapter investigates previous research in the IS field to provide a theoretical framework for the research. Academic literature and relevant documents about IoT, data governance, and research infrastructures have been reviewed to understand concerns related to sensor-based monitoring.

2.1 The Internet of Things

The use- and development of computer-based systems drive the discovery process. Humans are surrounded by thousands of *smart* devices every day, ranging from simple sensors to home appliances like smart TVs, pens, cameras, computers, and smartphones. Human interaction with multiple devices was named *Ubicomps' third wave* in late 1980 by Mark Weiser, the Computer Science Laboratory director at Xerox's Palo Alto Research Center. Today, ubicomp is known as the Internet of Things (IoT) (Dourish, 2016). Devices incorporated with sensors, capabilities, or actuators – an example of digital embeddedness – are considered smart when they can provide intelligent services, like communication and computation, without human intervention (Dourish, 2016; Rolstadås, Krokan and Dyrhaug, 2017; Silverio-Fernández, Renukappa and Suresh, 2018).

The International Organization for Standardization (ISO) have adopted a standardized definition of IoT and define it as:

> An infrastructure of interconnected objects, people, systems and information resources together with intelligent services to allow them to process information of the physical and the virtual world and react (ISO, 2014, p. 4)

Through unique addressing, things connected to the Internet can exchange information and interact with each other and the outside world. Hence, things can connect digital information to the physical world (Rolstadås, Krokan and Dyrhaug, 2017). IoT objects are mainly connected wirelessly, using, for instance, WiFi or RFID. RFID is an abbreviation for Radio Frequency Identification which means that radio waves are used to identify- and track items' location and status in real-time (Tan and Wang, 2010; Singh, Tripathi and Jara, 2014; Truong, Dinh and Wahid, 2017). An RFID tag uniquely addresses each device, which can be used to track the device. Accordingly, it is possible to access digital information about the state of objects from anywhere at any time (Boos *et al.*, 2013; Whitmore, Agarwal and Xu, 2015; Dourish, 2016; Rolstadås, Krokan and Dyrhaug, 2017).

Because of wireless connections and unique addressing, devices can communicate without human interaction through thing-to-thing communication, often called machine-to-machine (M2M) communication. Based on M2M communication, things can solve rudimental problems through automatic actions if needed – such as fault management in redundant systems or indicate dangerously high temperatures in rooms that automatically switch on cooling systems. Accordingly, IoT can improve everyday lives and simplify activities, also in domains of smart buildings and smart cities, providing citizens benefits of addressing their requirements and needs (Angelakis *et al.*, 2017). According to Tan and Wang (2010, p. 376), IoT will represent "the future of computing and communications".

The number of devices connected to the Internet exceeded the number of people on earth in 2008. By 2020, The European Commission predicted that over 50 billion devices would connect to the Internet (Tan and Wang, 2010; ISO, 2014; Dourish, 2016; Băjenescu, 2018) The IoT continues to grow, and devices become more intelligent and reliable every day (Rolstadås, Krokan and Dyrhaug, 2017; Shim *et al.*, 2020). Moore's law states that the number of transistors in each area doubles about every two years, which consequently reduces the area necessary for the required or corresponding processing power. Accordingly, devices are becoming smaller. Intelligent, reliable, and small devices, combined with advances in technology, have made electronics significantly cheaper, providing increased availability and usage (Dourish, 2016; Rolstadås, Krokan and Dyrhaug, 2017).

2.1.1 IoT for Environmental Research

The data and how they are collected drive the phenomena discovery process (Leonelli, 2019; Parmiggiani & Grisot, 2020). By designing IoT-based systems like sensor networks or real-time algorithm-based monitoring that can update monitoring plans according to the incoming data, it is possible to perceive information about the environment with increasingly good approximation, introduced as *synthetic knowing* by Monteiro and Parmiggiani (2019). Sensors are small and manageable and can act locally without complex installations to get good measurements. With their automatic collection of real-time data from their environment, they can provide information about the physical world, and if they observe changes in the environment, "corresponding things can make some responses if needed." (Tan and Wang, 2010, p. 377).

Rolstadås, Krokan and Dyrhaug (2017) give a relevant example of how the environment can connect to the Internet and how to benefit from the information that intelligent devices in a field provide:

> Farmers in Brazil already have access to technology that connects their crops to the Internet. Information about humidity, temperature, and nutrient content gets connected with weather forecasts and historical statistics. The information, linked with Artificial Intelligence (AI) that can learn from experiences and perform automated irrigation and fertilization, leads to the best possible utilization of expensive irrigation and fertilizer, resulting in larger crops with higher quality. (Rolstadås, Krokan and Dyrhaug, 2017, p. 24, translated from Norwegian to English by the researcher)

According to Parmiggiani, Monteiro and Hepsø (2015, p. 424), "[r]eal-time environmental monitoring involves developing a comprehensive network of measuring devices, sensors, communication lines, databases, and tools for analyzing and presenting environmental data". The growing usage of digital devices for monitoring and automation causes a continuous collection of real-time data, simplifying many everyday tasks for researchers. These tasks can include transmitting temperature, air quality, and pressure without human intervention (Boos *et al.*, 2013; ISO, 2014; Singh, Tripathi and Jara, 2014; Angelakis *et al.*, 2017).

Creating an IoT sensor network allows for data collection from different sources that can upload real-time data to a central station like cloud storage. Locating sensors in various places is called *sensor fusion* (Singh, Tripathi and Jara, 2014; Ahlers *et al.*, 2016; Truong, Dinh and Wahid, 2017; Zhou *et al.*, 2017). Multiple researchers within the IS field have investigated using sensor networks in environmental monitoring, assisting environmental researchers in their scientific work. For instance, Parmiggiani, Monteiro and Hepsø (2015) present a case of placing a lander on the seafloor equipped with a camera and sensors to monitor pressure, temperature, salinity, and turbidity for real-time information about coral reefs. Ahlers *et al.* (2016) investigate how creating wireless sensor networks could improve emission and produce more concise, accurate, and valuable information than provided by any individual data source. Truong, Dinh and Wahid (2017) look into how the design of an IoT-based system can provide easily accessible real-time local environmental data about fungus in crop fields. Moreover, Monteiro and Parmiggiani (2019) introduce a case using sensor devices integrated into networks to get measurable properties about oil leaks.

However, a world of intelligent devices poses a significant problem for management, not only in environmental monitoring but also in other sciences, for example, medical infrastructures, health care, and data science (see for instance: Ribes and Polk, 2014; Vassilakopoulou *et al.*, 2017; Passi and Jackson, 2018). Yang *et al.* (2017, p. 48) describe the complexity of IoT:

> Modern intelligent sensing systems generate huge volumes of sensing data. As a result, collecting, managing, and processing IoT big sensing data within an acceptable time duration is a new challenge for both research and industrial applications. The massive size, extreme complexity, and high speed of big sensing data bring new technical requirements including data collection, data storage, data organization, data analysis, and data publishing in real-time when deploying real-world IoT applications.

As the number of types and devices that become a part of- and connected to the Internet increases, the devices can "collect data in volumes that are many orders of magnitude greater than is possible today" (Angelakis *et al.*, 2017, p. 90). A promising solution to the challenges that huge volumes of data sets cause is using AI-based computer programs that can learn from extensive data sets through machine learning and make decisions to achieve specific goals (Singh, Tripathi and Jara, 2014). However, Passi and Jackson (2018) elaborate on *the problem of trust* in the knowledge and the results that data science produces. Parmiggiani and Grisot (2020, p. 24) explored the importance of data filtering and researchers' significant work in interpreting "the models produced by the algorithms" in an interview with an environmental scientist. For example, the output of a tree modeling algorithm produced inconsistent data that mistakenly could detect 50-meter-tall cranes as very high trees. The example shows in practice a problem of trust in data science and is one example of why human interaction is necessary to among other things detect unexpected results from algorithms.

2.2 Data Governance

Several terms in the IS literature deal with data and information, but "the definition of 'data' varies significantly across communities" (Parmiggiani, Monteiro and Hepsø, 2015, p. 426), and no formal definition exists about what a data set is (Renear, Sacchi and Wickett, 2010). Even though there is a shared perception about what the terms represent, Renear, Sacchi and Wickett (2010) argue that the absence of a precise definition can make it challenging to integrate digital, multi-disciplinary data from multiple sources (Renear, Sacchi and Wickett, 2010; Karasti, Baker and Halkola, 2006). The earlier focus where data was only a provider of information has shifted to a new focus that includes coverage, openness, and data quality (Zhang, Indulska and Sadiq, 2019), for instance, to ensure preservation (Zimmerman, 2008) and sharing of data.

Data management is a general term of the processes needed to organize, process, and store data to ensure data quality (Khatri and Brown, 2010). Data curation emphasizes fixing, assembling, and adjusting data to guarantee data quality and minimize size. (Yang *et al.*, 2017; Parmiggiani and Grisot, 2020). According to Karasti, Baker and Halkola (2006, p. 322), data curation is "critical in providing a substrate for the successful access, sharing and (re)use of data collections" which appears to be a crucial part of monitoring with sensors because it is impossible to have sensor-based monitoring without working with the data (Leonelli, 2019). Parmiggiani and Grisot (2020) investigated three leading practices of data curation: Achieving data quality, filtering the relevant data, and ensuring data protection. Supported by Leonelli (2019), these elements are crucial in data governance practices. Data governance concerns an organization's capability to ensure that data are made available, consistent, and usable through its lifecycle and who is made responsible for the data's decision-making in governing a system (Khatri and Brown, 2010; Otto, 2011; Micheli *et al.*, 2020; Parmiggiani and Grisot, 2020).

IS researchers sometimes describe data management as data governance (Otto, 2011; Mikalef *et al.*, 2020). Alhassan, Sammon and Daly (2016, p. 65) elaborate on the differences between "management" and "governance" and argue based on the research by Otto (2011) that management is influenced by governance:

> (…) governance refers to the decisions that must be made and who makes these decisions in order to ensure effective management and use of resources, whereas management involves implementing decisions. (…) Therefore, we can distinguish between the activities for data governance and the activities required for data management.

Khatri and Brown (2010) present a framework that the researcher finds interesting to study the different parts of data governance. The framework consists of five interrelated decision domains that the researchers argue should be considered for data governance: Data principles, data quality, metadata, data access, and data lifecycle. Figure 2-1 shows the way they are interconnected.

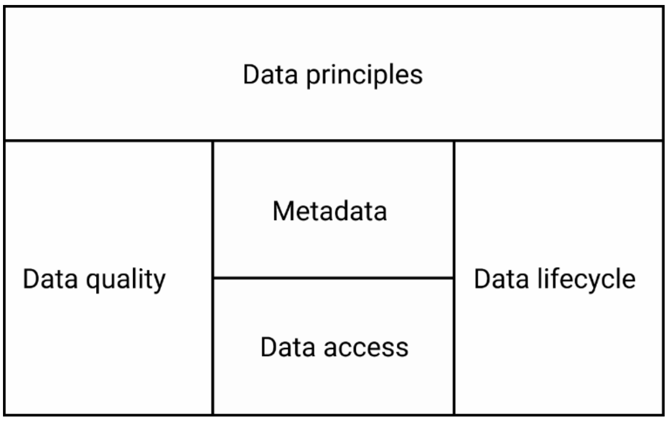

Figure 2-1: Decision domains for data governance. Figure retrieved from Khatri and Brown (2010, p. 149).

The data principles domain sets the boundary requirements for the intended uses of data and establishes the direction for the other domains. The principles address the standards for data quality – which depend on requirements such as completeness, accuracy, validity, consistency, and the ability to satisfy these (Zhang, Indulska and Sadiq, 2019).

The data quality, in turn, clarifies how to document data with metadata and how users can access them. Data access includes specifying access requirements for privacy and availability of the data to track who has access – and thus can modify the data. The data lifecycle domain includes the definition, production, retention, and retirement of data to make the data ready for use in the infrastructure (Khatri and Brown, 2010; Alhassan, Sammon and Daly, 2016).

Alhassan, Sammon and Daly (2016, p. 71) state that there is "a lack of maturity around data governance in practice". For this reason, they studied data governance activities in literature through the lens of the framework by Khatri and Brown (2010) to identify articles that explicitly mention data governance activities. They reviewed 59 articles and counted the frequency of different areas of data governance. The results showed that *definitions of the data roles and responsibilities, data policies, data processes and procedures*, and *data standards* had the highest frequency count. In contrast, *implement* and *monitor* had the lowest frequency count. Accordingly, Alhassan, Sammon and Daly (2016) show that the different domains rarely are discussed in practice, making it difficult to understand how data governance *actually* takes place in different infrastructures. Researchers studying data governance are struggling to identify how data governance takes place in the daily work, which arguably can result from different phenomena to study and infrastructures' various workflows and strategies in how to perform work due to different standards (Ribes, 2014). Arguably, the data governance framework by Khatri and Brown (2010) can be perceived as a "golden standard" with domains to perform "perfect" data governance.

Alhassan, Sammon and Daly (2016, p. 72) state that "(…) there is a lack of research that explicitly studies activities for governing data" and conclude that more research is needed in the data governance domain while research is growing in IS. Mikalef *et al.* (2020) substantiate a gap in research governance in practice, and Boos *et al.* (2013) state that research needs a closer examination of who is involved in making decisions about the data in infrastructures over time. With the emergence of IoT, and while sensors become more important for monitoring, there is a gap in research studying data governance and IoT in environmental monitoring (Karasti, Baker and Halkola, 2006; Parmiggiani and Grisot, 2020). Arguably, more research is needed into data governance practices in the case of environmental monitoring with IoT. The following section will expand on the framework by Khatri and Brown (2010) to explore data governance with IoT. The interaction between data governance and IoT will be investigated to fill in the IS research gap in this field.

2.2.1 Data Governance and IoT

The need for data governance increases due to the emergence of sensors and connected devices because of the large amounts of raw data the IoT causes. Mikalef *et al.* (2020) argue that data currently is one of the most valuable resources possible to monitor for organizations. Further, they state:

> In order to derive value from big data, firms must develop the organizational capacity to identify areas within their business that can benefit from data-driven insight, strategically plan and execute data analytics projects, and bundle the resource mix necessary to turn data into actionable insight. (Mikalef *et al.*, 2020, p. 2)

This substantiates the importance of establishing routines for how to turn data into meaningful information concerning data governance. Leonelli (2019, p. 320) explains that "extracting knowledge from data is not a neutral act", implying that identifying information from big data sets produced by the IoT can be challenging. To identify what to extract from the data and learn how to produce well-formed and relevant data, Parmiggiani and Grisot (2020) argue the importance of supporting users with continuous training and education on handling data sets. Training should also help researchers understand the ethical issues of their data curation work because those "[controlling] the infrastructure and own the data have a significant advantage in embedding their interests above others" (Angelakis *et al.*, 2017, p. 91; Parmiggiani and Grisot, 2020).

Previous research into the field (see Ribes and Bowker, 2009; Ribes, 2014) elaborates on the importance of data governance. Moreover, data governance extends beyond environmental monitoring (see Boos *et al.*, 2013; Whitmore, Agarwal and Xu, 2015; Haavik, 2017; Aversa, Cabantous and Haefliger, 2018; Xie *et al.*, 2020). Literature does not always recognize data work as data governance*;* for instance, Haavik (2017) discovers work with sensor data and digital representations through what he recognizes as *sensework.* The following subsection will elaborate on two cases in previous research in the IS literature that discusses IoT for real-time monitoring and data governance, showing that there are different ways to organize data governance in practice.

Examples of IoT governance

Aversa, Cabantous and Haefliger (2018) study a Formula 1 race that discovers how to extract knowledge from big data and use the information it provides during a race. They explain that each Formula 1 car is equipped with between 160 and 300 sensors to transmit live data streams during a race, which can be about 1 and 20 gigabytes of data collected in each race. A team of engineers and IT specialists analyze and run models on the real-time sensor data from the cars. The models simulate the data in real-time with historical data and assumptions to predict various factors, such as relative position to other cars and different race outcomes. The team sends back a selected portfolio of strategic options to the center at the racetrack. Then, "the chief race strategist has only a few minutes to cross check selected strategic options with the data (…) consult with the race engineer in charge of the team's cars, and make a decision" (Aversa, Cabantous and Haefliger, 2018, p. 224). This complex process involves an incredible time pressure to analyze and simulate data, elaborate, and make fast decisions based upon models and subsequently provide the information to the team and the driver. It is crucial to have a reliable real-time decision support system (DSS) with high performance, while Formula 1 races are described as "won or lost partly because of the processes of analysing the data (…)." (Aversa, Cabantous and Haefliger, 2018, p. 224). As a result, there is a yearly cost of between 150 and 500 million dollars in developing these systems.

The Formula 1 case can be an example of data governance in practice with sensor-based monitoring because of the work related to sensor-based monitoring. Formula 1 race and long-term environmental monitoring both involve monitoring with sensors collecting real-time data in one place and then sending the data to the cloud (Zhou *et al.*, 2017) for further interpretation by both systems (DSS and algorithms) and humans (engineers and environmental researchers). Haavik (2017) elaborates that human judgment must be made by someone engaged with the operational context, for instance, the chief strategist in Formula 1 and researchers in environmental monitoring.

IoT smart services on individual users that focus on an infrastructure consisting of devices connected over in-vehicle networks are called Internet of Vehicles (IoV) technologies. In China, a driver school offers driver students a smart simulation test based on IoV-technology to improve the formal driving test performance. The test is smart because it can provide feedback based on real-time data streams from sensors, which human coaches usually offer. Based on the lack of understanding of the user impacts of smart services and while literature focuses on technology evolution, Xie *et al.* (2020) investigated whether the technology improved performance on the formal driving test.

The IoV-based smart testing simulation service capture and analyze real-time data streams from sensors and cameras installed in vehicles. The collected data generate a report of errors that students have made during simulation. The report is provided to the individual student immediately after the test. During the simulation test, there is no need for a human coach. Such cases where complex socio-technical systems are replacing humans are discussed to trigger further work involving human judgment (see Haavik, 2017). The smart IoV-service, just like other smart IoT services, generates rich data streams that can be generated anytime due to the real-time nature of data streams that arguably triggers further work by humans. In this case, large data streams mean the generation of large quantities of feedback. To optimize the feedback in its form, timing, and quantity to fit the individual students' ability to process the information, Xie *et al.* (2020) argue a need for designing services and feedbacks *smarter* and more personalized. Designing solutions in an ad-hoc manner so that organizations can benefit from the large data streams is arguably what Mikalef *et al.* (2020) initially discuss in their statement, showing the importance of data governance. Arguably, the smart system only partially replaces human intelligence since data governance involves humans in making decisions (Alhassan, Sammon and Daly, 2016).

The IoT is expanding due to technology developments. Accordingly, the usage of intelligent devices in environmental research increases because of the benefits it provides researchers in efficiently collecting information about the environment. As the usage of IoT devices in monitoring results in big data sets with high volume, dimensions, and speed, it is essential to establish and achieve efficient workflows and data analysis to derive value from the data. A possible way to form data governance structures in environmental monitoring is to establish a research infrastructure.

2.3 Research Infrastructures

By defining, evaluating, and implementing strategies and tools, the European Commission provides Europe with sustainable research infrastructures (RIs). The European Commission presents the following definition for RIs: "Research Infrastructures are facilities that provide resources and services for research communities to conduct research and foster innovation." (European Commission, 2019b). Accordingly, infrastructures can provide resources and services like computing systems and collections, archives, scientific data, collaboration tools, and calibrated instruments to research communities to make research possible, easier, and faster (Ribes, 2014; European Commission, 2019a). The European Union established the European Strategy Forum of Research Infrastructures (ESFRI) in 2002 as a governing authority to regulate and support funding for RIs in Europe (ESFRI, 2019; Micheli *et al.*, 2020)

The Research Council of Norway is a government agency that funds Norwegian research and innovation and contributes to infrastructure access. It is a provider of the roadmap for RIs that identifies Norway's needs for updating RIs and provides tools and funding for research (Forskningsrådet, 2018). Forskningsrådet (2018, p. 5) explain that access to infrastructure is necessary to:

> (…) research high international quality, achieve a high degree of institutional co-operation and division of labor at a national level, increase international co-operation [and] achieve open access to the use and reuse of research data. (Translated from Norwegian to English by the researcher)

The Research Council of Norway and the European Commission further agree that research has become more data-driven; for example, the European Commission (2019a) argues that "activities go increasingly online and produce vast amounts of data". As a result, greater demands are placed on heavy rain capacity, increased storage capacity, and advanced tools. Supporting RIs with digital resources is essential for efficiently carrying out research (Forskningsrådet, 2018). Examples of digital resources are software, computers, and storage devices – which are part of the electronic infrastructure (e-infrastructure)(European Commission, 2019a). Arguably, the definitions of a RI from the European Commission and the Research Council of Norway describe RI as quite technical facilities – as infrastructures where research is conducted, with systems and tools necessary to provide and foster research and innovation.

Leonelli (2019, p. 319) explains that the resources needed to maintain global data infrastructures and related institutions have spread out as they have grown, "and in ways that do not fit contemporary regimes of funding, credit, and communication." To receive funding, RIs must often adapt to guidelines and standardized ways to document research. Monteiro *et al.* (2013, p. 577) elaborate that "[s]tandardisation deals with how the design, implementation and customization of a technology at one local site interacts with, and is constrained by, implementation of "the same" technology elsewhere". Standardizations can be beneficial while infrastructures generally have a history with different ways to document data, and ecological research deals with different types of data and non-standard datasets (Karasti, Baker and Halkola, 2006). For instance, Karasti, Baker and Halkola (2006, p. 334) discuss that "an unwritten rule is that each site manages their own data". However, researchers can also experience standards as challenging while ensuring that the data format and the metadata fit the standard (Leonelli, 2014).

According to Zhang, Indulska and Sadiq (2019, p. 576) it is the ones that need to use the data that need to complete the management activities "in an ad-hoc manner without following well-defined processes or guidelines". The collection, saving, and consistent documenting of data to preserve the data for future use is arguably a complex socio-technical process where the work conducted at RI also includes *the researchers*. The social aspects at RIs are not sufficiently recognized in the definitions from the European Commission and the Research Council of Norway that arguably finance *research* and not researchers' *work*. As a result, some researchers adopt an information infrastructure (II) perspective (Karasti, Baker and Halkola, 2006; Pollock and Williams, 2010; Monteiro *et al.*, 2013) to describe the complex socio-technical relationship present at RIs. Monteiro *et al.* (2013, p. 576) present a working definition of information infrastructures and explain:

> IIs are characterized by openness to number and types of users (no fixed notion of 'user'), interconnections of numerous modules/systems (i.e. multiplicity of purposes, agendas, strategies), dynamically evolving portfolios of (an ecosystem of) systems and shaped by an installed base of existing systems and practices (thus restricting the scope of design, as traditionally conceived)

This definition considers the socio-technical relation between researchers and technology present at RIs because it includes the resources to conduct research, the researchers, and their relationship. Thus, adopting an II perspective can contribute to recognizing researchers' work in documenting research. However, it may look like there is a governmental shift towards adopting an II perspective while a policy document by eLTER recognizes the socio-technical aspects of publishing data (see eLTER, 2019). The document evaluates data policies and governance issues and refers to the development of an *information infrastructure* to document and share data.

While research and innovation shape our future and contribute to achieving goals, the EU has established a key funding program called Horizon Europe. In 2021-2027 the program has several mission areas like adaption to climate change, cancer, climate-neutral, and smart cities. The program's primary goals are, among other things, to strengthen the EU's scientific and technological bases to offer competitiveness and jobs and deliver on citizens' priorities (European Commission, 2021). The program is said to strengthen research and innovation but arguably still lacks a social focus that recognizes researchers' work. As a result, it can look like the socio-technical perspective recognized by eLTER (2019) remains to be solved in practice. The need to illuminate the complex socio-technical process present at RIs due to monitoring with IoT devices is thus further strengthened.

3 Case Description

The project started with a literature review in the semester project, providing a basis for this interpretive case study investigating sensor data management activities at environmental research sites in Norway and how environmental researchers' work is affected by the increased usage of digital devices in monitoring.

Long-Term Ecosystem Research (LTER) is an essential component established for research sites for long-term ecological research to improve, among other things, knowledge and understanding of ecosystems and environmental pressures and threats through research and monitoring (Oggioni *et al.*, 2012; NINA, 2021a). The sites are divided into networks of research infrastructures. The eLTER network is a European-based umbrella network of the LTER, consisting of research stations in Europe. The Norwegian Institute for Nature Research (NINA) coordinates Norwegian participation in this network. NINA focuses on researching different aspects of nature, such as climate, environmental monitoring in water and land and species, and the interplay between nature and society (NINA, 2021a).

Arranging environmental research into networks of research infrastructures is beneficial as each infrastructure can focus on a few key aspects of the natural environment. In the last two to three decades, the emergence of smart technology in environmental research has increased dramatically. Today, numerous research infrastructures use smart technology in research. The technology can range from small, simple sensors to more advanced measuring instruments; pressure cells, precipitation-, temperature and conductivity meters, light loggers, GPS-senders, or full-scale weather stations. The emergence of smart technology has changed the demands of who is present at research infrastructures, where diverse disciplines work together to conduct research. The stakeholders are, but not limited to, data administrators, engineers, environmental researchers, environmental station managers, PhD students, project managers, and technology workers, which all form an essential part of environmental research with sensor technology.

While research infrastructures are more regulated by governing institutions – for example, ESFRI and the Research Council of Norway – to receive funding, the different disciplines must work together to document data concerning the guidelines and requirements set in roadmaps and strategy documents (Forskningsrådet, 2018; ESFRI, 2019; Labonnote, Bryhni and Lech, 2021). While these guidelines and requirements tend to be scientific and technical, focusing on conducting research and not work, researchers' work is accordingly affected and often not sufficiently recognized by these guidelines and requirements.

4 Research Method

This chapter elaborates on the project's research process, including the research strategy chosen for the project, how data are collected, methods for analyzing the data, and the recruitment process of participants. Finally, there is a discussion of the research paradigm.

4.1 Research Process

The research started with a literature review about IoT and data governance. Studying earlier research about the topics in the IS field provided a conceptual framework for the project. The literature review led to an experience that sensor-based monitoring is not fully automated because researchers must perform a lot of curation work to set up and calibrate sensors and clean data to make them reusable. While literature discusses the lack of research into data governance with IoT, this led to a motivation to fill this gap. Based on the literature review, the following research questions were addressed to guide the research (reproduced from 1.2 Research Questions):

RQ1: How are environmental work practices affected by the use of IoT in monitoring?

RQ1.1: How is IoT used in research infrastructures?

RQ1.2: How is data governance arranged in research infrastructures?

Figure 4-1 illustrates the research process.

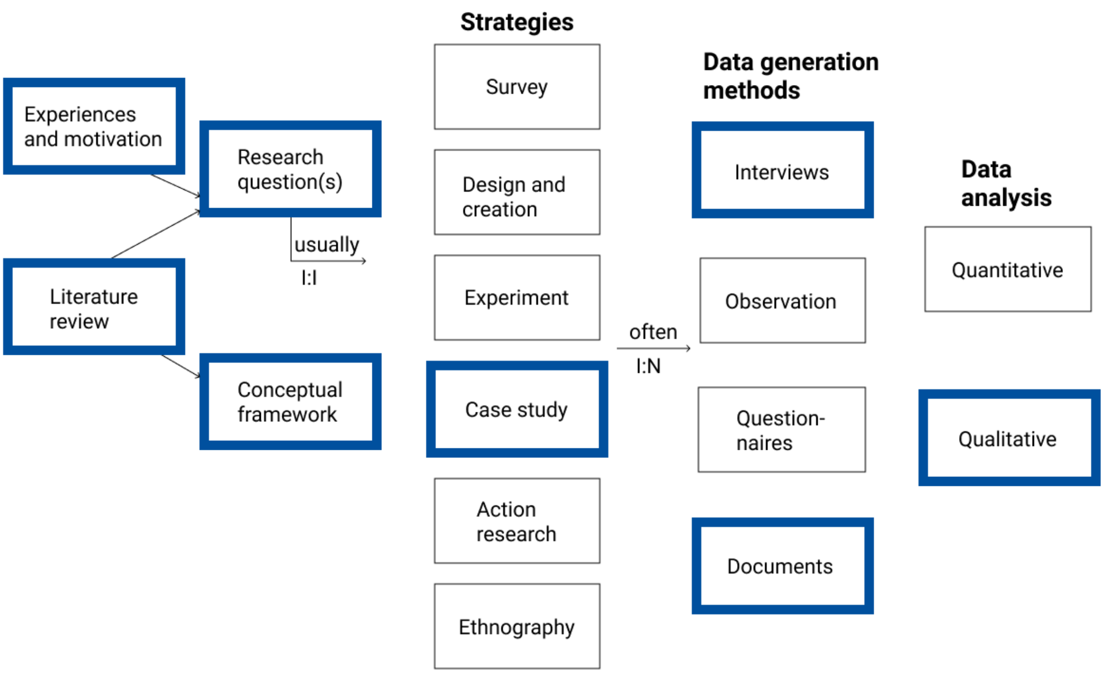

Figure 4-1 Model of the research process. The blue frames mark the path followed in this project. Figure retrieved from Oates (2006, p. 33).

4.1.1 Strategy

It is necessary with a thorough understanding of IoT and monitoring activities at environmental research infrastructures to answer the research questions. Since the socio-technical relationship between researchers and the technology at RIs is closely interconnected, this holistic process should be studied in-depth. Consequently, the research is dependent on collecting qualitative data, preferably through different methods, to answer the "how"-research questions (Oates, 2006; Baxter and Jack, 2015). A case study allows exploring phenomena in a natural context using multiple data generation methods. For this reason, the chosen method for data generation in this research is a case study.

Considering the survey strategy to generate data, it became clear that the strategy focuses on covering the breadth of research while simplifying the real world's complexity and not concerning details in the research (Oates, 2006; Baxter and Jack, 2015). Therefore, the survey strategy was omitted from this research.

An ethnography strategy requires the researcher to spend time in the field, contributing to understanding the environmental researchers' work in practice. This research is limited to one semester, and Walsham (1995) discusses that even though the researcher stays long in the field, there is no guarantee that valuable data will be collected. Arguably, it is not possible to conduct an ethnography study over one semester, and it is rather valuable to use a variety of data sources to collect data to "develop a comprehensive understanding of phenomena." (Carter *et al.*, 2014, p. 545); like in a case study.

4.1.2 Recruitment of Participants

Before initiating the empirical data collection, it was necessary to seek approval from the Norwegian Centre for Research Data (NSD). A consent form to participate in interviews was made based on advice from NSD to address ethical issues around data collection. Once the application was approved, the data collection was performed according to NSDs guidelines and rules. Accordingly, the interviews were anonymized and stored securely on NTNU's servers to safeguard the participants' privacy. While the project does not require gathering personal data, no personal information about the participants was stored.

The recruitment of participants started with contacting relevant individuals within research infrastructures for environmental research before following a snowball sampling approach. In this recruitment technique, current participants propose new potential participants. However, the recruitment process was identified by concerns about finding individuals willing to participate. My assumption is that people were busy and potentially affected by the ongoing situation around the COVID-19 virus. Therefore, other individuals that appeared to be interesting were also contacted. The participants were recruited from three different infrastructures for environmental research.

4.1.3 Data Collection

The data collection relies on qualitative data collected through semi-structured interviews, written *self-administered* structured interviews, and documents as the primary data sources. Using multiple methods for data collection is called *method triangulation,* a method that is frequently used in qualitative research to collect data from different sources about the same phenomena (Carter et al., 2014; Oates, 2006).

While participants report what they do when questioned in interviews, this can deviate from what they are doing in practice. Consequently, the original plan also included conducting field studies of sensor data activities at environmental research sites in Norway to observe how IoT affects researchers' daily data governance work. However, it became impossible to conduct field studies due to restrictions from the Norwegian government and the Norwegian University of Science and Technology (NTNU) to combat the COVID-19 virus. Therefore, semi-structured and structured interviews, supported by documents, are the primary data generation methods for this project.

The structured and semi-structured interview questions were made in collaboration with a PhD student. With very similar projects, we both were interested in looking into data work for performing sensor-driven environmental monitoring. Accordingly, we conducted the interviews together, taking place in April, May, and June. The structured interview questions and the interview guide are attached in the Appendices.

Structured interviews include, like questionnaires, pre-determined standardized questions that are asked to each participant. The questions for the structured interview were divided into three categories, containing open and closed questions to identify how data collection and analysis work are carried out in infrastructures for environmental research. At the beginning of the project, the document with questions was sent by email to 33 relevant individuals working in research infrastructures for environmental monitoring. They were requested to respond to a written, structured interview without the researcher being present, and five participants responded to the document. For this reason, it entails more accuracy referring to the document like a written, structured interview than a questionnaire – since questionnaires often collect data from many participants and thus provide answers from a large population.

Furthermore, the questions for the semi-structured interviews were divided into five sections to investigate and discover researchers' and other stakeholders' experiences with IoT and data governance within their research infrastructure. After conducting two interviews, some changes were made to the interview guide. For instance, we experienced that it was relevant to ask about research infrastructures in light of economics and discovered that asking about participants' job descriptions was less relevant as these were discovered to be quite open and not very specific.

The semi-structured interviews were valuable supplements to the written, structured interviews because we were able to ask the participants additional questions based on what they were reporting, which allowed us to investigate exciting topics in-depth. In addition, we were able to detect participants' emotions, experiences, and feelings about different topics, for instance, how they experience adapting to guidelines and standards. However, we were aware that interviews could not conclude a whole population while the responses varied and involved exploring the informants' personal experiences (Oates, 2006).

The data generation also included document analysis of relevant documents to corroborate and question data from the other data generation methods. These were documents from governing institutions and research organizations, for example, guidelines, roadmaps, and strategy documents. Hence, method triangulation is achieved since data about IoT and data governance activities are collected from structured interviews, semi-structured interviews, and documents. The data generation methods are presented in Table 4-1, together with the participants' roles and the disciplines they represent.

Table 4-1 Overview of the data generation methods and disciplines represented by participants.

Data source	Discipline	Participants
Semi-structured interviews	Environmental research	1 Research coordinator 1 Senior researcher 3 Research managers 1 Data manager
	IoT for environmental research	1 Research manager 2 Environmental researchers
Structured interviews	Environmental research	1 Data administrator 1 Environmental researcher
	IoT for environmental research	1 Research coordinator 2 Environmental researchers
Documents	ESFRI European Commission SINTEF The Research Council of Norway	

4.1.4 Method for Analyzing the data

The data analysis followed a stepwise-deductive induction (SDI) model for qualitative research (see Tjora, 2019). The model is based on inductive processing of qualitative data following a stepwise deductive analysis starting from the raw data generated from interviews, resulting in a quality-assured and manageable analysis of the data that can be used to develop empirical arguments.

An analysis of the documents led to the identification of codes in the raw data due to a two-step iterative process of empirical coding that preserved the details in the empirical material. For example, when the researcher found an interesting sentence in the raw data such as “sensors are sensitive to power outages", then the code could be “sensors are sensitive to power outages” because that would preserve the details in the empirical material. Subsequently, the codes would be labeled into conceptual categories, where “sensors are sensitive to power outages” could belong to the category “Digital data collection” because the code implies a challenge with digital data collections. Labeling codes into conceptual categories will help identify patterns within the collected data. In this regard, it is easier to determine overall themes. The computer-assisted qualitative data analysis software NVivo was used in the coding process, as shown in Figure 4-2 and Figure 4-3.

Table 4-2 shows the themes, conceptual categories, and excerpts found in this project. Each theme is a finding from the semi-structured and structured interviews that describe concerns about IoT and data governance.

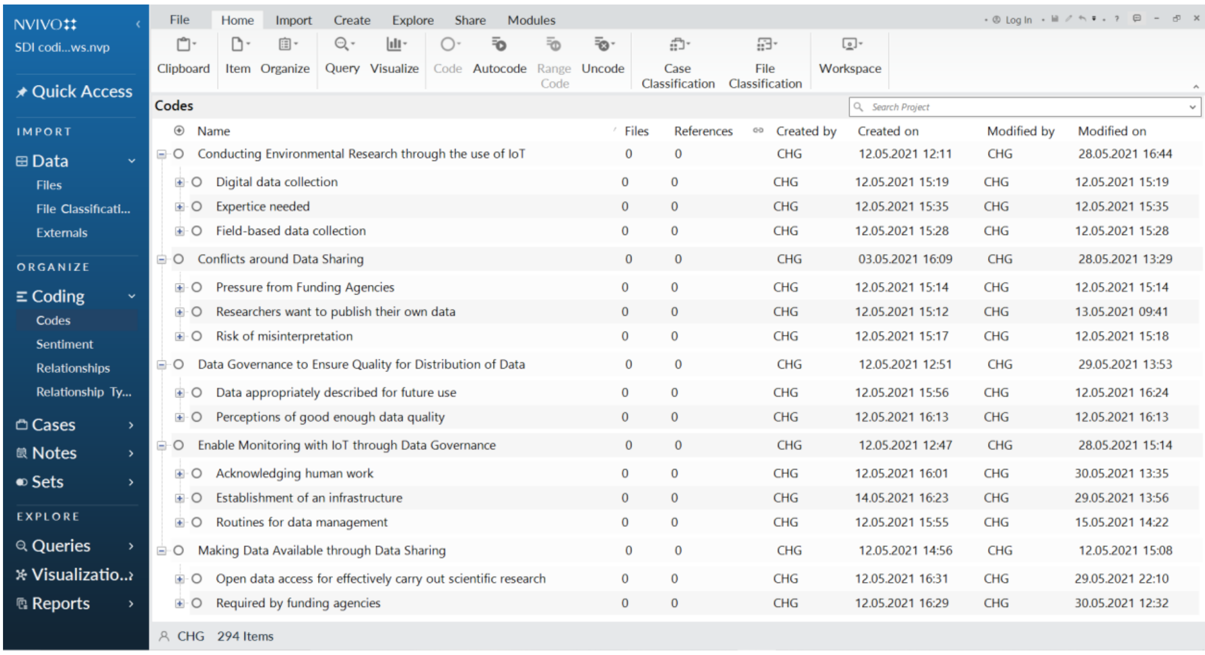

Figure 4-2 Overview of the coding process in NVivo. The codes are organized in five themes with belonging conceptual categories.

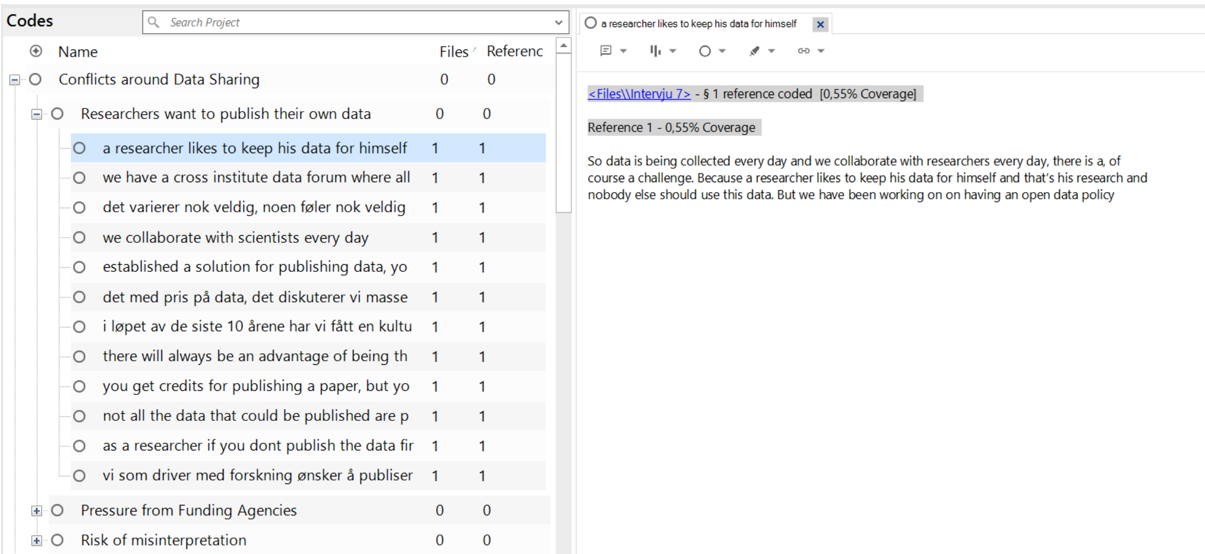

Figure 4-3 The codes appear below each conceptual category with a corresponding reference to the empirical material.

Table 4-2 The analytical framework that identifies relevant themes in the empirical material. The themes are divided into conceptual categories with relevant excerpts belonging.

Theme	Conceptual category	Excerpt
Conducting Environmental Research through the use of IoT	Digital data collection	"We use sensors and collect data through measurements in buildings." *Research manager, semi-structured interview*
	Expertise needed	"Often, younger people are good at technology and are used to such changes [technology changes], so we have to bring a lot of specialist expertise into the projects." *Research manager, semi-structured interview*
	Field-based data collection	"We have not seen that technology has the ability to obtain data of the type we are looking for." *Research coordinator, semi-structured interview*
Enable Monitoring with IoT through Data Governance	Acknowledging human work	"Even though the sensors are automatic and save us a lot of hours in the field, we spend a number of hours in the office afterward, at present." *Senior researcher, semi-structured interview*
	Establishment of an infrastructure	"The researchers and data scientists collaborated in deciding how we wanted this infrastructure to look like." *Environmental researcher, semi-structured interview*
	Routines for data management	"We need to find a way to distribute this [data], we cannot just put it on the Internet and start to download 500TB, that would have taken forever." *Environmental researcher, semi-structured interview*
Data Governance to Ensure Quality for Distribution of Data	Data appropriately described for future use	"That is my opinion, but in reality, it must be decided by future users." *Environmental researcher, semi-structured interview*
	Perceptions of good enough data quality	"The data that we produce from the monitoring programs and such, we assume – and several others that look at the results of some kind – that they maintain good enough quality that it is worth publishing." *Research coordinator, structured interview*

Theme	Conceptual category	Excerpt
Making Data Available through Data Sharing	Open data access for effectively carry out scientific research	"Open sharing of data makes scientific findings verifiable and has a preventive effect on research cheating." *Data administrator, structured interview*
	Required by funding agencies	"The funding agency will accept any kind of open repository that fulfills, say, international standards (…) and having a set of metadata that describes the data." *Senior researcher, semi-structured interview*
Conflicts around Open Data Sharing	Pressure from funding agencies	"There is also a pressure from funding agencies, let's say the Norwegian Research Council, for instance when they give money to research, they also require that these data should be made publicly available." *Senior researcher, semi-structured interview*
	Researchers want to publish their own data	"As a researcher, if you don't publish the data first, then you don't survive." *Senior researcher, semi-structured interview*
	Risk of misinterpretation	"When sharing data, it is demanding to ensure that the user gains sufficient insight into the data set's weaknesses." *Research coordinator, structured interview*

4.2 Research Paradigm

The qualitative in-depth case study about IoT and long-term environmental monitoring is concerned with understanding the socio-technical context of the researchers and the IoT technology. The thesis focuses on human interpretations and meanings, implying that the research adopts an interpretive perspective that characterizes a complex human process, aiming to enrich the understanding of how researchers experience IoT in environmental monitoring. The researchers are studied in their natural setting at research infrastructures for monitoring, investigated from the researchers' perspective.

The study will not provide one fixed explanation to the research questions since conducting research and monitoring the environment can differ across infrastructures. As a result, there cannot be right or wrong answers to the research questions – which also characterize an interpretive paradigm. Walsham (1995) discusses that preserving an openness to data and not strictly using existing theory is desirable in interpretive studies. Rather than seeking the “truth”, multiple interpretations within the collected data are investigated to identify which one could seem the strongest – aiming for *plausibility* (Oates, 2006; Walsham, 2006).

Acknowledging that the researcher will always affect the situation and shape the research by influencing informants' interpretations is also considered in the interpretive paradigm. Therefore, the researcher has strived to balance a neutral position, not being too passive nor over-enthusiastic while interacting with informants during interviews (Walsham, 1995; Oates, 2006).

5 Findings

Based on Table 4-2, this chapter will present the findings from the semi-structured and structured interviews. Each subsection addresses one theme from the analytical framework and will include the conceptual categories with relevant examples and excerpts.

5.1 Conducting Environmental Research through the use of IoT

In accordance with developments in intelligent technologies, our ability to perceive information about the environment has changed. In a few decades, the usage of IoT for environmental research has increased considerably. An environmental researcher elaborates: "(…) nature has not changed, but our ability to measure it has changed quite largely, I think" (Environmental researcher, semi-structured interview). However, due to limitations in technology, not all parameters in nature are possible (yet) to monitor with sensors or digital devices. Consequently, environmental monitoring can be categorized into two groups of data collection: *Digital data collection* and *field-based data collection,* and the data collection depends on the types of data to be collected:

> We have environmental sensors that measure temperature, oxygen, and, say, echo sounders that look at the vertical distribution of the fish. Cameras can be used to measure steaming speed (…). And we also have [a] manual inspection by collecting the fish [where we] note down what we see, and laboratory-based indicators like blood samples, plasma samples, and genomics (…) (Research manager, semi-structured interview).

Retrieving information about the environment fast and straightforward is one reason for using digital devices in research. *Digital data collection* involves collecting data through devices ranging from simple climate loggers monitoring, for example, temperature or moisture, to wild cameras, light loggers, GPS-senders, and sometimes very advanced measuring instruments. Moreover, the methods of perceiving information about the environment are continuously evolving: For example, a research coordinator elaborates on a new statistical data procedure using species' DNA in monitoring but that it is not yet used in their infrastructure:

> What one can imagine after a while (…) but which is not quite as suitable for exactly what we have been doing, yet at least, it is the kind of environmental DNA. That is, you take a sample, and run it through genetic analyzes and so on, automated, and get a result at the other end (Research coordinator, semi-structured interview).

The use of environmental DNA can be important in monitoring since automatic genetic analyzes can be taken from, for instance, water samples and provide results about species. However, even though the technologies are automated, the findings imply challenges using, for example, sensors for digital data collections. They can be summarized into four points: (1) Defective durability, (2) sensitivity to power outages, (3) sensitivity to weather changes – such as drought and frost, and (4) difficulties covering the area that the researcher wants to monitor. For example, a research coordinator elaborates: "It can be difficult to get sensors that cover the entire measurement spectrum we operate within." (Research coordinator, structured interview). Accordingly, combining methods can help researchers get the broader picture and more reliable data than collected by any individual source. An environmental researcher discusses the benefits of using observations with sensors and models on supercomputers to get information about the physical environment:

> An observation is only valid for one time and one location; you might not know how representative this observation is on itself, so you should combine the model result with the observation, so both may be wrong, but together they will be more descriptive of the actual situation than on their own (Environmental researcher, semi-structured interview).

Nevertheless, even though our ability to collect information has increased immensely, not everything in the environment is possible nor beneficial for researchers to monitor with IoT technology. Accordingly, environmental researchers must perform *field-based data collections* to retrieve this information, which requires the researchers to spend time in a specific area and collect data about a phenomenon. A senior researcher elaborates that collecting recordings through manual data collections is a prominent part of an ecologists' work. Moreover, an environmental researcher discusses that other than surveying areal and GPS-position to be in the right place, this type of research is often "directly from nature and into the human head, completely without instruments" (Environmental researcher, semi-structured interview). Accordingly, the parameters are sometimes written down on paper: "It [fish registration] is on a handwritten form because it goes pretty fast (…), and it can be a bit messy with spills and mucus, so therefore, it is written on plastic sheets that can withstand water" (Research manager, semi-structured interview). Other parameters collected through field-based data collections are species registrations, collections of biological material, collections of forest data like trees' diameter- and height growth, soil data collections, humus samples, and various geophysical variables.

A research coordinator elaborates the reason why they are not using technology in some of their environmental monitoring programs: "This [environmental monitoring program] has been going on for a long time, 30 years, we have probably not seen that technology has the opportunity to obtain the type of data we are looking for" (Research coordinator, semi-structured interview). In addition, some observation material is dependent on the researchers' expertise to judge the type of observation, which makes it difficult to use technology for that type of monitoring:

> Some species are so difficult to tell the difference, so they must enter the lab and be analyzed, for instance, chemically, to determine if it is one or the other species. (…) It is, in a way, only the researchers' expertise and the various aids that you have available to determine each individual species that is observed (Research coordinator, semi-structured interview).

Accordingly, field-based collections are dependent on the researchers' expertise to perform the monitoring. However, performing digital data collections also requires expertise, even though the data collection is automatic – operating with digital devices sets requirements for *expertise* and competence in technology. A research manager elaborates: "I would say that the human factors we struggle most with are lack of competence to handle and understand new software and systems" (Research manager, semi-structured interview).

Several informants agree on the need for competence in, for example, informatics, physics, ecology, and experts in machine learning to operate with technology in monitoring. For instance, a research coordinator says: "It requires special expertise in operating the sensors and the loggers" (Research coordinator, structured interview). As a result, researchers sometimes go into scientific collaborations with experts and personnel in different disciplines to perform environmental monitoring with IoT devices. A research manager elaborates:

> Far back in time when I actively myself used measurement data, you had to fix it in a way yourself (…) while now, you get help from people who are more professional in data and data management. (…) The technology has become a much bigger subject in itself (Research manager, semi-structured interview).

Arguably, monitoring with IoT gives an increased need for expertise operating with the technology, which differs from how monitoring earlier was arranged when researchers fixed their own equipment. Accordingly, new demands are set on data management to handle data from digital devices.

5.2 Enable Monitoring with IoT through Data Governance

In line with industries' need for information, researchers at environmental monitoring sites experience an increased need for expertise to handle the technologies. Accordingly, an environmental researcher argues that it requires a collaboration between the data scientist and researcher collecting the data to establish an infrastructure: "The data scientist would advise us on you know standardization. How can we make this comparable to international standards? How can we make sure that we fulfill the requirements for documentation?" (Environmental researcher, semi-structured interview). The *establishment of infrastructure* can help environmental research sites to form data governance structures. A data manager elaborates: "I wrote one proposal, and we had a pre-project in 2009, then we got [financing amount] from this Council to establish a national research infrastructure in Norway, funded for five years" (Data manager, semi-structured interview). Furthermore, the data manager elaborates that the data group in their infrastructure advises the data center on how the data should be managed. The routines are often stated in formal data management plans on how to manage the data through its lifecycle: Including how the data should be documented in a systematic way, how the data should be stored, and how the files should be formatted with columns and rows: "We have fairly specified procedures and method descriptions for how data is to be collected" (Research coordinator, semi-structured interview).

However, devices used in environmental monitoring have different levels of complexity, and the data they collect range from quantitative data like temperature and salinity to qualitative data like marine resources, fish, and models producing different types of results. The informants share the perception that various data need to be processed. A research manager elaborates: "The first challenge is getting the data into the right format" (Research manager, semi-structured interview). However, the data management will differ based on the type of device used in monitoring and the data type it collects, leading to various *routines for data management*. An environmental researcher elaborates:

> (…) there are many different systems – internationally recognized systems – for how to document the dataset; some are specific, some are targeted specifically towards certain disciplines. (…) Some of them are extremely complex, and

> sometimes they are so complex that it can almost be a bit of a barrier to watch for documenting your data (Environmental researcher, interview).

For example, small temperature sensors collecting just one type of variable can be straightforward to document since they can be standardized and analyzed easily, while cameras and acoustic devices, in contrast, can be more of a challenge: "[there are] large challenges, I would say, related to how to actually process that acoustic or optical information into ecological information" (Environmental researcher, semi-structured interview). In addition to difficulties documenting data according to standards, an environmental researcher elaborates on a challenge that arises when data are managed differently: "it is very difficult to compare the results in two different sensors. (…) I think the technology in itself, or the sensor technology is okay, but the rest is not sufficiently developed." (Environmental researcher, semi-structured interview). In addition, a data manager elaborates the need for educating data managers to handle the massive data sizes: "there is one thing that we miss, and that is someone to educate data managers" (Data manager, semi-structured interview). At present, people involved with data management with no experience are encouraged to look at how others manage data to do it in the best manner.

As the data sets often are massive, several informants elaborate on the challenges going through large amounts of data: "I would really like to have an interface to all this terabyte of models that can easily extract information from a single point of time series for instance" (Environmental researcher, semi-structured interview). Accordingly, a lot of time goes just to extract meaningful information from large data files, like stated in this excerpt: "It is very difficult to look for signals or phenomena in two petabytes of data, that will take some time to dig out the gold" (Environmental researcher, semi-structured interview). Even though data is stored in computer systems, part of the data management must be done manually. *Acknowledging human work* when using intelligent technologies in research is important in order to support researchers with necessary tools: "It is a lot of manual work still, and also laboratory work has a lot of manual parts so you cannot just put the sensor in the water, there is still a lot of human work." (Environmental researcher, semi-structured interview). Even though AI and machine learning are discussed to be solutions to the problem of extracting meaning from data, there is a shared perception among the informants that the data structures are in an early adoption phase in environmental research like shown in this excerpt:

> What is coming up now is artificial intelligence, machine learning, and the big numbers analysis, and if that might be a way to leap around the problem of turbulence and chaos, that would be interesting (…). So, I guess that is not up to me, it might be the next generation (Environmental researcher, semi-structured interview).

In addition, researchers must routinely check if there are any missing data: "That's something I typically do – run and check through the variables. You have to see if you have any missing data and try to find out why is there missing data there." (Research manager, semi-structured interview). Consequently, monitoring with IoT is not fully automatic while (1) the devices collect tremendous amounts of data that must be analyzed in order to extract meaning from them, (2) researchers must manually check if all the data have been retrieved by the device and (3) researchers must ensure that data are documented in a systematic way. While the primary goal of environmental research is to report data and make the data available for future reuse and interpretation, it is crucial that the data are understandable for others unfamiliar with the context – concerning the data quality.

5.3 Data Governance to Ensure Quality for Distribution of Data

An important part of data governance is to follow the data management plan and ensure that data are *appropriately described for future use.* Adding descriptions to data makes it easier for others to reuse and reproduce, in addition to preventing data from being misunderstood: "a lot of data are not readily understandable, and we don't know their background" (Research manager, semi-structured interview). As a result, the work of making data understandable needs to be done by someone familiar with the context, as shown in the following excerpt: «it has to be done by somebody who knows what you could expect from that particular locality» (Environmental researcher, semi-structured interview). The work of making data understandable includes adding descriptions of contextual information, metadata, to the data:

> Obviously, the purpose of metadata is that somebody else should be able to take that data set, understand enough about how that data has been collected, the kind of sensor used (…), so that they can use it again without the special knowledge of the person who actually collected it (Environmental researcher, semi-structured interview).

However, as metadata are connected to the collected data, it is not always easy to understand the context from the metadata only. A research coordinator elaborates: "dependent on the use, more basic explanation of how the data is collected, what they actually represent, is needed" (Research coordinator, semi-structured interview). Thus, researchers can follow internal guidelines like ISO standards – an informant refers to it as their *handbook* – for quality assurances. Moreover, some researchers use systems that report error messages that need to be solved and eliminated before publishing or sharing to ensure quality. Others manually go through the data and sort out data that are deviating from expected or historical trends:

> If something is unusually low or unusually high, there will usually be reanalyzed. Or, if there is unacceptable high deviance between replicates, they will automatically be realized. I would say that output is high-quality data (Research manager, semi-structured interview).

However, a research manager elaborates that to avoid others from manipulating the data, there are large restrictions on who has access and can change the data. In addition, the quality assurances can be time-consuming: "For example, if you spend 280 hours per year to ensure quality, then someone has to pay the 280 hours" (Research manager, semi-structured interview). Data documenting and data quality are affected by price competitions, which leads to generally underfunded agency-funded projects. A data administrator elaborates: "In general, it is difficult to get the operation costs covered to have the desired quality in work" (Data administrator, structured interview). Moreover, a senior researcher elaborates: "The resources we have received in this program, they should mostly go to data collection and the minimal reporting that the [funding agency] wants" (Senior researcher, semi-structured interview).

Paradoxically, even though participants elaborate difficulties to have desired quality in work due to minimal funding, there is a shared *perception of good enough data quality* among the informants, implying that they believe that the reported data gives lasting results for what it says something about:

> The data that we produce from the monitoring programs, we assume – and most others who look at the results of some kind – that they maintain good enough quality that is worth publishing (Research coordinator, semi-structured interview)

Even though funding agencies give researchers minimal resources to ensure data quality, it has become clear that data quality is an essential part of researchers' work to support future reuse of data and make the data ready for distribution.

5.4 Making Data Available through Data Sharing

The participants share the perception that open access to data is crucial to *effectively carry out scientific research*. A data manager elaborates: "there has been a development in kind of making institutions aware of that it is a need to share data" (Data manager, semi-structured interview). Based on the interviews, three main reasons are identified for why open data access is essential: (1) it gives new possibilities for analysis which can give society increased insight, (2) it makes scientific findings verifiable and has a preventive effect on research cheating and (3) the main principle for the data management *is* to share the data.

The infrastructures' policy can decide whether the data should be shared. However, sometimes, it is project-dependent: While data in privately funded projects often are kept within the infrastructure, researchers experience an increased focus on data sharing in publicly funded projects. Accordingly, researchers can be *required by funding agencies* to share the research data. A researcher elaborates on the reason for sharing data: "We want to have a sustainable industry, that is the goal [of] the government. So, our job is to give scientific evidence on how this sustainability actually is." (Environmental researcher, semi-structured interview)

As the data size has increased, data are today typically made available through infrastructures' web solutions: "Before, the data was not so big, then we could send them over, but there is less of it now, we invite in a way and open up the areas on our system." (Research manager, semi-structured interview). In addition, stakeholders interested in the data can also contact the infrastructure directly: "If someone gets in touch and says what they are going to use it [the data] for, then we will usually make that data available to them" (Research coordinator, semi-structured interview). However, an environmental researcher reveals that data sharing is more than just a way of making data public to others: It is also about combining data contributed by several researchers. This in turn shows the importance of having data in a format to make the combination of researchers' work more accessible.

Moreover, open access to data for scientific implementations can also be perceived as challenging. An environmental researcher elaborates: "it [open data sharing] requires facilitation and documentation (which is usually not covered by project funds)" (Environmental researcher, structured interview). Hence, open data sharing is not entirely without conflicts.

5.5 Conflicts around Open Data Sharing

Data sharing provides new opportunities for analysis that can give society increased insights. Even though there are different perceptions of open data sharing between researchers, the idea of where researchers make their research available to the public is getting more ground. Some RIs are required by their internal data policies and funding organizations such as the Research Council of Norway to make data available. Also, journals where researchers publish their research sometimes require that the data are made available along with the paper. With multiple agents pulling in different directions towards open data, conflicts around data sharing can be divided into three main categories: *Pressure from funding agencies, risk of misinterpretations,* and *researchers wanting to publish their own data.*

In cases where data is collected with public money with funding from agencies, the data should often be made available without further conditions. The *pressure from funding agencies* to make data available is elaborated on by a senior researcher in the following excerpt: "when they [funding agencies] give money to research, they also require that these data should be publicly available" (Senior researcher, semi-structured interview). Paradoxically, even though funding organizations require data to be made available, open access to data requires facilitation and documentation that usually are not covered by project funds, which the following excerpt shows: "it is very difficult to get funded, and especially on data management, that is, one that we got in 2012, that is kind of the first time we ever got any funds, external funding for doing data management" (Data manager, semi-structured interview).

Sometimes, parts of the data have limited transparency due to confidentiality, making it challenging to make data available within the duty of confidentiality. Accordingly, it can be challenging to describe the data context well enough for future use or for someone unfamiliar with the situation. Consequently, data sharing can allow for *misinterpretation of the data*: "Parts of the data set (which can run over 30 years) may have weaknesses, and when sharing the data, it is demanding to ensure that the user gains sufficient insight into the dataset's weaknesses" (Research coordinator, structured interview).

While researchers are experiencing pressure from funding agencies in making data available, this conflicts with some researchers' desire *to publish their own data*:

> [As a researcher] you have collected it; you do not just want to give it away. So, there is a balance between openness and securing that the research actually works and that researchers can survive from a career perspective (Senior researcher, interview).

However, it may look like environmental research has experienced a cultural change towards open data policy where people now, to a greater extent than before, are willing to share data. A data manager elaborates on the cultural change towards willingness for publishing data, putting a digital object identifier or ID on the data, so researchers are ensured that they will be credited for realizing the data:

> (…) you ensure that this data is referred to in the correct way when other people use it. And having this possibility also kind of gives a bit more trust to the researcher that the data will be shared in a good way (…). And we see that more and more other researchers actually want to publish their datasets (Data manager, semi-structured interview).

Even though people are willing to share data, there is an ongoing discussion on *how* to share the data. While environmental research sites often have invested much money to establish an infrastructure, researchers are not willing to just give away the data:

> We have spent an incredible amount of money investing in this infrastructure from which we now can retrieve a lot of data. So, the price of data… We discuss it a lot. (…) people cannot just come and get it in a way. You have to be involved in investing in what we do, they also have to pay a share of that data. (…) We want people to use it, but at the same time we can in a way not just give it away (…) (Senior researcher, semi-structured interview).

Accordingly, there is a problematic business model about how to price the data to avoid people using others' data without permission or without investing anything in the infrastructure. Even though data sharing sometimes conflicts with researchers' perceptions of owning the data and the funding institutions' lack of funding making data ready for sharing, there has been an awareness among infrastructures that there is a need to share data.

6 Discussion

Answers to the research questions were sought by conducting an interpretive case study of RIs for environmental monitoring in Norway that uses IoT in research. This chapter will link the findings to the research questions and elaborate upon the theoretical framework to investigate work practices related to IoT-based environmental monitoring and how data governance affects these.

Figure 6-1 will guide the discussion elaborating on the themes from the analytical framework and relate them to the theoretical framework. The direction of the arrows is a result of the following data analysis.

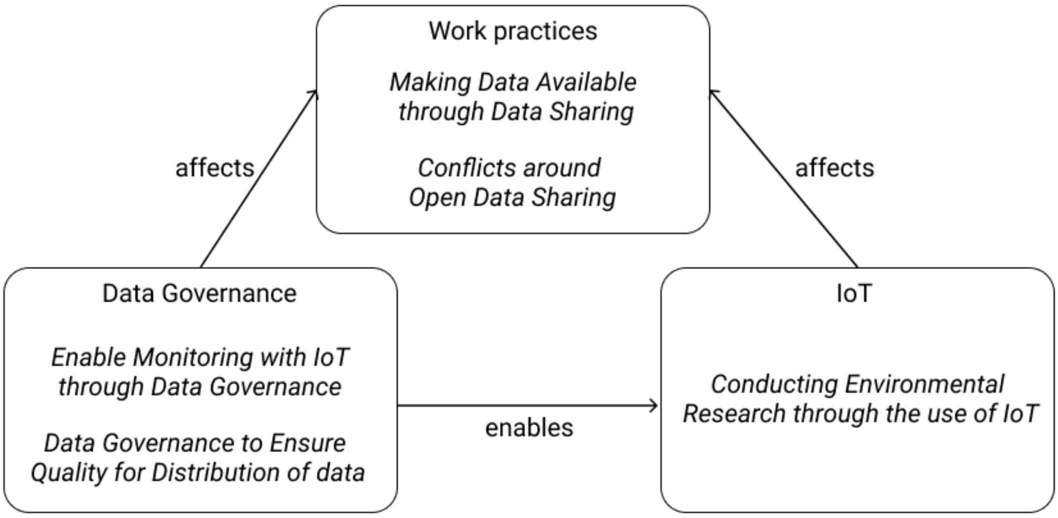

Figure 6-1 Overview of the themes from the analytical framework and how they are connected to concepts from the theoretical framework.

The research questions are addressed in separate subsections. The sub-research questions concerning IoT and data governance will be elaborated on first. Finally, they will be linked to the main research question that concerns how the environmental work practices are affected by IoT and data governance.

6.1 Digital Data Collections in an Early Adoption Phase

Rather than looking *at* the technology to predict performance, Orlikowski (2000) encourages looking at the *use* of technology to predict performance acts. Accordingly, to enrich the understanding of- and obtain insight into environmental monitoring with IoT, the following research question was addressed: *How is IoT used in research infrastructures?*

The literature discusses how creating sensor networks to perceive information about the environment can assist researchers in their scientific work (Ahlers *et al.*, 2016; Truong, Dinh and Wahid, 2017; Monteiro and Parmiggiani, 2019). The findings show a shared perception among informants that IoT supports researchers in their scientific work. IoT is used in RIs to research specific domains: For example, humidity in buildings is monitored by sensor devices, while cameras and GPS senders are used to research wildlife. Specific parameters in marine life are monitored using temperature-, pH-, oxygen-, carbon dioxide-, and salt loggers, while information about fish distribution is perceived using echo sounders. In addition, a new method of monitoring with environmental DNA was elaborated on by a researcher where samples from nature can be analyzed to reveal or study ecological phenomena. NINA recently used the method to identify how large the char population is at Svalbard (see Haugli, 2021). In addition, a podcast episode by NINA (see NINA, 2021b) discusses the method in genetic analyses of wolves. The method is perceived as an advancement in technology, enabling research without animals being affected (NINA, 2021b).

An informant elaborates that they have been using the same sensor type to monitor temperature for about 15 years due to its stable connections, which can imply that even though the IoT is growing and devices are becoming more reliable and intelligent (Shim *et al.*, 2020), new technologies are not necessarily used for environmental monitoring even though they are available. Arguably, one of the reasons for this may be the *problem of trust* in data science (Passi and Jackson, 2018). For example, a researcher elaborates that even though they use sensors to monitor temperatures in water, researchers still monitor the temperature manually every day just in case the sensor suddenly stops working. Manual temperature monitoring also provides researchers with an extra parameter if the sensor should monitor either too high or too low values. In addition, technologies are not used to register anesthetized fish because if the data systems suddenly shut down, there is no time to wait for the technology to work. Consequently, this part of work occurs by taking handwritten notes on papers that are vulnerable to be lost in case of a fire (Parmiggiani and Grisot, 2020) and sometimes can be difficult to decipher due to handwriting and mess.

Boos *et al.* (2013) discuss that time-consuming manual activities can be automated and monitored by IoT applications, and the data it provides can be used in improving processes. Truong, Dinh and Wahid (2017) investigate how the establishment of IoT systems can help managers facilitate better management to prevent fungal disease spread. The findings support the claims of IoT assisting researchers in their scientific work while enabling easy monitoring of, for example, wild- and marine life using cameras and loggers. In addition, this case study reveals that using IoT in specific research areas does not *replace* the researchers; on the contrary, the findings show that using IoT in monitoring places new requirements on *who* can use the technologies. For example, an informant elaborate that the technology has developed into a separate subject. With the continuous collection of data, the amount of data collected by smart devices are huge (Singh, Tripathi and Jara, 2014). Angelakis *et al.* (2017) question whether whoever can extract meaning from data and further that it is a scientific skill to give meaning to numbers. Accordingly, the perception of a need to bring special expertise in technology into the monitor projects to use IoT in environmental research is strengthened.

The findings of (1) using the same sensor in monitoring over a long period, (2) the technology not able to monitor what the researchers are looking for, and (3) researchers performing manual registrations on paper can imply that the use of IoT in environmental monitoring currently is in an early adoption phase. Flexibility in data governance to allow for the use of technologies to stabilize (Ribes, 2014) is essential. However, RIs' well-established routines and standards for making data available, consistent and usable can make flexibility difficult.

6.2 Managing Data through a Framework for Best Practice

Greater demands are placed on researchers to turn data into meaningful and accurate information (Khatri and Brown, 2010; Otto, 2011), while raw data from monitoring projects with IoT are not readily understandable. Accordingly, infrastructures enable exploring objects of research possible (Ribes, 2014), with established routines that are affected by policies, guidelines, and standards. There are inadequately published papers in the IS literature focusing on data governance due to monitoring with IoT, and Alhassan, Sammon and Daly (2016) state a need for research to investigate data governance practices. Accordingly, the following research question was asked: *How is data governance arranged at research infrastructures?*

Activities related to data governance are rarely discussed in practice, which can make it difficult to understand how data governance actually takes place (Alhassan, Sammon and Daly, 2016). The literature review showed the importance of data governance in other disciplines (Aversa, Cabantous and Haefliger, 2018; Xie *et al.*, 2020) and how data governance in relation to IoT needs a closer examination and needs to be investigated even more. The researcher found it interesting to investigate data governance activities at environmental research infrastructures based on the "golden standard" framework by Khatri and Brown, that take a step back from the daily decision making and focus on fundamental decisions, as shown in Figure 2-1.

The findings show a shared perception among the informants with updated routines and documents on how the data should be used and the purposes of producing data from monitoring projects. For example, an informant elaborates that it is clear what will be done in the projects and how the data will be used after the monitoring projects are finished, which leads to the argument of a well-established *data principles* domain.

Data quality can be difficult to determine as those who use the data are others than those who created them (Zhang, Indulska and Sadiq, 2019). However, the data quality points accuracy, consistency, completeness, and validity might give an indication. The findings show that due to lack of funding, it can be challenging to describe data accurately. For example, an informant elaborates that it is challenging to have the desired quality of work since the operation costs are not covered. In addition, data can be described by others unfamiliar with the research, such as technicians (Parmiggiani and Grisot, 2020). Combined with a requirement to follow well-established guidelines (Zhang, Indulska and Sadiq, 2019), standards (Ribes, 2014), and policies to describe data, it can be challenging to describe data sufficient for its context, affecting the data quality. In addition, factors as lack of time, resources and people to document data also affect the data quality.

Without special knowledge of the person collecting the data, it should be possible to use the data based on its *metadata* descriptions. The descriptions will vary based on the data type, and an informant elaborates how metadata could be standardized in a better way making the work of adding contextual information to the data easier. A research manager mentioned that there is limited access on who is allowed to modify the data in their infrastructure, which can imply established routines on who has access and thus are able to modify the data.

Finally, making data ready for use in the infrastructure concerns the *data lifecycle* domain. The findings show that there are specified procedures on how the data should be managed with an awareness of what happens to the data and how they will be used.

Several monitoring projects aim to report the findings to the funding agency, which leads to researchers following standards and guidelines for handling data through its lifecycle.

Focusing on fundamental decisions in data governance (Khatri and Brown, 2010), the analysis of findings imply that data governance routines are well established with guidelines and standards telling how to collect and manage data through its lifecycle. However, the established processes make it difficult to ensure, for example, sufficient data quality as the devices and the data types they produce vary. Accordingly, ensuring data quality is a complex process that often involves researchers and data managers making decisions. This in turn leads to the argument that researchers and data managers' day-to-day decision-making are important and often necessary in managing data generated using IoT, which apparently is omitted in current standards and guidelines for data governance practices.

Accordingly, greater demands should be placed on developments in data governance practices to recognize researchers' work in documenting research. An example supporting that claim is the development of the decision support systems in Formula 1. The systems were elaborated to be a way to form data governance because they handle real-time data from the Formula 1-cars equipped with sensors. Annually, developments of these systems are supported with between 150 and 500 million dollars (Aversa, Cabantous and Haefliger, 2018). The big investments result from Formula 1 races being dependent on reliable data from the systems; however, it can give a perspective to the data governance practices at RIs that are perceived inadequate for some purposes. Researchers and managers are performing much work governing data from IoT, and hence, IoT and data governance leads to changing work practices at RIs.

6.3 Changing Work Practices

A literature review of previous research in the IS field about IoT and data governance led to the experience that several papers cover IoT for environmental monitoring (Sung and Hsu, 2013; Parmiggiani, Monteiro and Hepsø, 2015; Truong, Dinh and Wahid, 2017) and data governance (Otto, 2011; Parmiggiani and Grisot, 2020; Mikalef *et al.*, 2020), but few published papers focus on data governance as a result of IoT based monitoring. Consequently, the following research question was addressed: *How are environmental work practices affected by the use of IoT in monitoring?*

The findings show how monitoring with IoT simplifies environmental monitoring tasks such as automatic monitoring of humidity in buildings through sensors or analyses of environmental DNA. This supports previous literature elaborating that IoT can provide intelligent services without human intervention, simplifying and improving everyday tasks (Tan and Wang, 2010; Angelakis *et al.*, 2017). However, perceiving monitoring with IoT from this perspective indicates that IoT facilitates environmental research entirely without human intervention, which does not entail complete accuracy. The findings show that researchers invest their time and energy in capturing knowledge from IoT data to make it fit established standards and guidelines.

Increased usage of technology and demands of knowledge into AI and machine learning in the future implies that researchers need to be prepared to use new technologies in the future. Through work with, for instance, identifying surfaces to locate the COVID-19 virus (Klokk and Mikalsen, 2020), it is apparent that researchers are prepared to support future work to give society increased insight – but insights are enabled only if there is an enhanced focus on open data sharing. Research conducted with funding from public funding agencies must often be shared openly: Some informants elaborate that reporting scientific evidence to society is the main reason for environmental research. However, open data sharing requires sufficient documentation and descriptions of data to make them understandable, work that often is not funded as funding agencies give money to *research.* The findings show how researchers experience a lack of time, resources, and people to conduct research and make data available, meaning that the work towards open data sharing also affects work practices. In addition, the work towards open sharing of scientific data requires work in changing mindsets as some perceive the research data as their own and are reluctant to share data (Johnston, 2021).

A deeper understanding of data governance in practice has led to the experience that greater demands are placed on the researchers to monitor with IoT. Consequently, using IoT in research leads to a need for establishing routines on how to handle data through its lifecycle and a need for personnel and expertise in technology to operate with IoT. Accordingly, this leads to the argument that both data governance and IoT affect the work practices and that monitoring with IoT is enabled by data governance, as shown in Figure 6-1.

7 Conclusion

Researching RIs as complex socio-technical systems where research is conducted has provided new empirical insight into the use of IoT in long-term environmental monitoring and how it affects work practices and is enabled by the established practices for data governance as shown in Figure 6-1. The analytical framework, as shown in Table 4-2, presents the case study's significant findings.

The findings show how digital data collections support researchers in their scientific work, providing fast and accessible information about the environment. An extensive part of the work in perceiving information about the environment is still performed field-based – and there are conflicting perceptions for that. Some informants discuss that it is necessary to perform manual monitoring due to a lack of trust in technology. There are also perceptions that technology is not yet developed sufficiently for specific purposes, especially when the research is dependent on the researchers' expertise.

There are different approaches to arrange data governance in practice. This thesis has elaborated on the use of decision support systems to support managers in Formula 1 races and using an IoV-smart service to enable fast feedback to students in driving tests. Due to funding allocated by the EU and other funding agencies, a common way to arrange data governance in environmental monitoring is to establish a research infrastructure. Monitoring sites arranged as a RI are provided with established standards and guidelines on how the environmental data should be documented to fulfill, for example, international standards on how to document data. Monitoring with IoT is enabled by the established processes that transform raw, extensive, and potentially meaningless data into meaningful information that can be shared and reused.

To a greater extent than before, researchers are encouraged by funding agencies to share data openly. Open data sharing requires data documenting and quality that is not covered by the funding from the agencies, leading to researchers performing work that is not sufficiently recognized as work. Even though a governmental shift recognizing the socio-technical aspects of data publishing was elaborated, the findings imply that the work still remains to be solved in practice. I do not claim that the study is finished as more research into data governance practices with IoT is needed. However, I hope to inspire further work to recognize researchers' work enabling IoT-based monitoring.

7.1 Limitations

Since it became impossible to conduct observations through field studies because of restrictions due to COVID-19, the case study of IoT and data governance in long-term environmental monitoring relies entirely on empirical data from interviews and documents. Consequently, it is difficult to say something concrete about data governance practices in RIs for environmental monitoring and how IoT is used in monitoring.

Moreover, since the project was limited by time, the research focuses on a few RIs in Norway only, which might not be representative for other RIs in Norway or for RIs in other countries. Accordingly, the findings are based on a limited amount of data from interviews and rely on data mostly from environmental researchers and leaders who did not do much research or data collection themselves on a daily basis. To really gain a holistic understanding of the socio-technical process at RIs, it would have been beneficial to interview representatives from all identified groups, such as PhDs, postdocs, and others involved in the practical work conducting research. Rather than generalizing the findings to a perception shared by the majority of stakeholders at RIs for environmental monitoring, the findings should be used as an indication to be investigated in the future by conducting several interviews and observations through field studies.

7.2 Future Directions

While still lacking a complete understanding of current practices around data governance at RIs for environmental monitoring, a more thorough analysis with more interviews and observations could contribute to awareness into how the governance practices are affected by the increased usage of IoT in monitoring. In that regard, it would have been interesting to investigate several RIs in Norway and RIs outside Norway to see whether work practices are the same.

The increased usage of technology in monitoring changes the demands for who can work with the digital devices as expertise is often required in monitoring with IoT devices. In an interview with a researcher, it became clear that RIs experience a lack of people able to handle data. An exciting path to investigate further is how knowledge related to data management will be in demand in the future while the methods of perceiving information about the environment are continuously evolving, and whether the education of data managers will have an impact on achieving the desired quality in work related to documenting research.

References

Ahlers, D. *et al.* (2016) A Measurement-Driven Approach to Understand Urban Greenhouse Gas Emissions in Nordic Cities. NIK.

Alhassan, I., Sammon, D. and Daly, M. (2016) Data governance activities: an analysis of the literature, *Journal of Decision Systems*, 25:sup1, pp. 64-75. Available at: https://doi.org/10.1080/12460125.2016.1187397.

Angelakis, V. *et al.* (2017) *Designing, Developing, and Facilitating Smart Cities : Urban Design to IoT Solutions*. Springer International Publishing : Imprint: Springer.

Aversa, P., Cabantous, L. and Haefliger, S. (2018) When decision support systems fail: Insights for strategic information systems from Formula 1, *Journal of Strategic Information Systems*, 27:3, pp. 221-236. Available at: https://doi.org/10.1016/j.jsis.2018.03.002.

Băjenescu, T.-M. (2018) From The Internet of Things (IoT) To Smart Cities (SC), *Journal of Engineering Science*, XXV, pp. 85-94. doi: 10.5281/zenodo.2557328.

Baxter, P. and Jack, S. (2015) Qualitative Case Study Methodology: Study Design and Implementation for Novice Researchers, *Qualitative report*. doi: 10.46743/2160-3715/2008.1573.

Boos, D. *et al.* (2013) Controllable accountabilities: the Internet of Things and its challenges for organisations, *Behaviour & Information Technology*, 32:5, pp. 449-467. doi: 10.1080/0144929X.2012.674157.

Carter, N. *et al.* (2014) The use of triangulation in qualitative research, *Oncol Nurs Forum*, 41(5), pp. 545-547. doi: 10.1188/14.ONF.545-547.

Commission, E. (2019a) *e-Infrastructures*. Available at: https://ec.europa.eu/digital-single-market/en/e-infrastructures.

Commission, E. (2019b) *What are Research Infrastructures?* Available at: https://ec.europa.eu/info/research-and-innovation/strategy/european-research-infrastructures_en.

Commission, E. (2021) *Horizon Europe - Investing to shape our future*. Available at: https://ec.europa.eu/info/sites/default/files/research_and_innovation/funding/presentations/ec_rtd_he-investing-to-shape-our-future.pdf.

Dourish, P. (2016) The Internet of Urban Things, 30(4), pp. 363-382.

eLTER (2019) *Governance and data policy for sharing and publishing of data*. Available at: https://www.lter-europe.net/document-archive/elter-h2020-project-files/d3-2-governance-data-policy (Accessed: 20.05.2021).

ESFRI (2019) *Strategy Report on Research Infrastructures Roadmap 2021*.

Forskningsrådet (2018) *Tools for Research. National strategy for research infrastructure 2018-2025*. Oslo. Available at: https://www.forskningsradet.no/contentassets/94079aa751f94b349c15ffed04d5a541/national-strategy-for-research-infrastructure-20182025.pdf.

Govoni, M. *et al.* (2017) An Information-Centric Platform for Social- and Location-Aware IoT Applications in Smart Cities, *EAI endorsed transactions on internet of things*, 3(9), pp. 153049. doi: 10.4108/eai.31-8-2017.153049.

Grøder, C. H. (2020) IoT and data management in long-term environmental monitoring (pp. 26).

Haugli, B. (2021) Ny metode skal kartlegge Svalbard-røya, *Svaldbarposten*.

Haavik, T. K. (2017) Remoteness and sensework in harsh environments, *Safety Science*, pp. 150-158. Available at: http://dx.doi.org/10.1016/j.ssci.2016.03.020.

ISO (2014) Internet of Things (IoT) - Preliminary Report, *ISO/IEC JTC 1 Information Technology.* Available at: https://www.iso.org/files/live/sites/isoorg/files/developing_standards/docs/en/internet_of_things_report-jtc1.pdf.

Johnston, R. (2021) Missouri faced 'roadblocks' on data literacy during COVID-19, *STATESCOOP*. Available at: https://statescoop.com/missouri-covid-19-data-driven-decisions/ (Accessed: 26.05.2021).

Karasti, H., Baker, K. S. and Halkola, E. (2006) Enriching the Notion of Data Curation in E-Science: Data Managing and Information Infrastructuring in the Long Term Ecological Research (LTER) Network, *Computer supported cooperative work*, 15(4), pp. 321-358. doi: 10.1007/s10606-006-9023-2.

Karasti, H., Baker, K. S. and Millerand, F. (2010) Infrastructure Time: Long-term Matters in Collaborative Development, *Computer supported cooperative work*, 19(3), pp. 377-415. doi: 10.1007/s10606-010-9113-z.

Khatri, V. and Brown, C. V. (2010) Designing Data Governance, *Communications of the ACM*, 53(1), pp. 148-152. doi: 10.1145/1629175.1629210.

Klokk, H. F. and Mikalsen, K. S. (2020) Virusjegerane, *NRK*. Available at: https://www.nrk.no/virusjegerane-1.15155270?fbclid=IwAR3Gdcombo4norfMF8Xm0bQ8UonFjSAPemEnDa49Ev4gF_ozP_AhYpz2dFM (Accessed: 26.05.2021).

Labonnote, N., Bryhni, A. and Lech, T. C. (2021) *Digital samhandling og datadeling i bygge-, annleggs- og eiendomsnæringen*. SINTEF.

Leonelli, S. (2014) What difference does quantity make? On the epistemology of Big Data in biology, *Big Data & Society*, 1(1), pp. 11. doi: 10.1177/2053951714534395.

Leonelli, S. (2019) Data - from objects to assets, *Nature*, 574(7778), pp. 317-320. doi: 10.1038/d41586-019-03062-w.

LTER-Europe (2017) *Building eLTER RI*. Available at: https://www.lter-europe.net/elter-esfri (Accessed: 13.04 2021).

LTER (2020) *LTER Network data access policy*. Available at: https://lternet.edu/data-access-policy/ (Accessed: 28.04.2021).

Micheli, M. *et al.* (2020) Emerging models of data governance in the age of datafication, *SAGE Publications Ltd*, 7(2). Available at: https://doi.org/10.1177/2053951720948087.

Mikalef, P. *et al.* (2020) Big data and business analytics: A research agenda for realizing business value, *Information & Management*, 57(1). Available at: https://doi.org/10.1016/j.im.2019.103237.

Monteiro, E. *et al.* (2013) From Artefacts to Infrastructures, *Computer supported cooperative work*, 22(4), pp. 575-607. doi: 10.1007/s10606-012-9167-1.

Monteiro, E. and Parmiggiani, E. (2019) Synthetic Knowing: The Politics of the Internet of Things. doi: 10.25300/MISQ/2019/13799.

NINA (2021a) *LTER Norway*. Available at: https://www.nina.no/english/Research/Projects/LTER (Accessed: 07.05 2021).

NINA (2021b) Naturligvis *Med ulven i reagensrøret*. https://nina.no/podkast: NINA. Available at: https://nina.no/podkast (Accessed: 30.04.2021).

Oates, B. J. (2006) *Researching information systems and computing*. London: Sage Publications.

Oggioni, A. *et al.* (2012) *Monitoring of Environmental Status through Long Term Series: Data Management System in the EnvEurope Project*. Unpublished paper presented at EnviroInfo Dessau 2012, Part 1: Core Application Areas.

Orlikowski, W. J. (2000) Using Technology and Constituting Structures: A Practice Lens for Studying Technology in Organizations, *Organization science (Providence, R.I.)*, 11(4), pp. 404-428. doi: 10.1287/orsc.11.4.404.14600.

Otto, B. (2011) Organizing Data Governance: Findings from the Telecommunications Industry and Consequences for Large Service Providers, *Communications of the Association for Information Systems*, 29(3). doi: 10.17705/1CAIS.02903.

Parmiggiani, E., Monteiro, E. and Hepsø, V. (2015) The Digital Coral: Infrastructuring Environmental Monitoring, *Computer supported cooperative work*, 24(5), pp. 423-460. doi: 10.1007/s10606-015-9233-6.

Parmiggiani, E. and Grisot, M. (2020) Data Curation as Governance Practice, *Scandinavian journal of information systems*.

Passi, S. and Jackson, S. J. (2018) Trust in Data Science: Collaboration, Translation and Accountability in Corporate Data Science Projects, *Proceedings of the ACM on Human-Computer Interaction*, 2, pp. 28. Available at: https://doi.org/10.1145/3274405.

Pollock, N. and Williams, R. (2010) e-Infrastructures: How Do We Know and Understand Them? Strategic Ethnography and the Biography of Artefacts, *Computer supported cooperative work*, 19(6), pp. 521-556. doi: 10.1007/s10606-010-9129-4.

Renear, A. H., Sacchi, S. and Wickett, K. M. (2010) Definitions of dataset in the scientific and technical literature, *Proceedings of the American Society for Information Science and Technology*, 47(1), pp. 1-4. doi: 10.1002/meet.14504701240.

Ribes, D. and Bowker, G. C. (2009) Between meaning and machine: Learning to represent the knowledge of communities, *Information and organization*, 19(4), pp. 199-217. doi: 10.1016/j.infoandorg.2009.04.001.

Ribes, D. (2014) The kernel of a research infrastructure *Computer supported cooperative work* (pp. 574-587): ACM. doi: 10.1145/2531602.2531700.

Ribes, D. and Polk, J. B. (2014) Flexibility relative to what? Change to research infrastructure, *Journal of the Association for Information Systems*, 15(5), pp. 287-305. doi: 10.17705/1jais.00360.

Rolstadås, A., Krokan, A. and Dyrhaug, L. T. (2017) *Teknologien endrer samfunnet*. Bergen: Fagbokforlaget.

Shim, J. *et al.* (2020) The Internet of Things: Multi-faceted Research Perspectives, *Communications of the Association for Information Systems*, 46:4. doi: https://doi.org/10.17705/1CAIS.04621

Silverio-Fernández, M., Renukappa, S. and Suresh, S. (2018) What is a smart device? - a conceptualisation within the paradigm of the internet of things, *Visualization in engineering*, 6(1), pp. 1-10. doi: 10.1186/s40327-018-0063-8.

Singh, D., Tripathi, G. and Jara, A. J. (2014) A survey of Internet-of-Things: Future Vision Architecture, Challenges and Services, *IEEE*, pp. 6.

Sung, W.-T. and Hsu, C.-C. (2013) IOT system environmental monitoring using IPSO weight factor estimation, *Sensor review*, 33(3), pp. 246-256. doi: 10.1108/02602281311324708.

Tan, L. and Wang, N. (2010) Future Internet: The Internet of Things, *IEEE*, 5, pp. 5.

Tjora, A. H. (2019) *Qualitative research as stepwise-deductive induction*. Abingdon, Oxon: Routledge.

Truong, T., Dinh, A. and Wahid, K. (2017) IoT Environmental Data Collection System for Fungal Detection in Crop Fields, *IEEE*, pp. 4.

Vassilakopoulou, P. *et al.* (2017) Building National eHealth Platforms: the Challenge of Inclusiveness, *Proceedings of the International Conference on Information Systems - Transforming Society with Digital Innovation, ICIS 2017*.

Walsham, G. (1995) Interpretive case studies in IS research: nature and method, *European Journal of Information Systems*, 4, pp. 74-81.

Walsham, G. (2006) Doing interpretive research, *European Journal of Information Systems*, 15, pp. 320-330.

Whitmore, A., Agarwal, A. and Xu, L. D. (2015) The Internet of Things - A survey of topics and trends, *Inf Syst Front*, 17, pp. 261-274. doi: 10.1007/s10796-014-9489-2.

Xie, W. *et al.* (2020) Living in a Simulation? An Empirical Investigation of a Smart Driving-Simulation Testing System, *Journal of the Association for Information Systems*, 21:4, pp. 843-863. doi: 10.17705/1jais.00622.

Yang, B. C. *et al.* (2017) Big-Sensing-Data Curation for the Cloud is Coming: A Promise of Scalable Cloud-Data-Center Mitigation for Next-Generation IoT and Wireless Sensor Networks, *IEEE consumer electronics magazine*, 6(4), pp. 48-56. doi: 10.1109/MCE.2017.2714695.

Zhang, R., Indulska, M. and Sadiq, S. (2019) Discovering Data Quality Problems, *Bus Inf Syst Eng*, 61:5, pp. 575-593. Available at: https://doi.org/10.1007/s12599-019-00608-0.

Zhou, J. *et al.* (2017) Security and Privacy for Cloud-Based IoT: Challenges, *IEEE communications magazine*, 55(1), pp. 26-33. doi: 10.1109/MCOM.2017.1600363CM.

Zimmerman, A. S. (2008) New Knowledge from Old Data: The Role of Standards in the Sharing and Reuse of Ecological Data, *Science, technology, & human values*, 33(5), pp. 631-652. doi: 10.1177/0162243907306704.

Appendices

Appendix A Structured interview questions

Appendix B Interview guide

Appendix A Structured interview questions

Norsk versjon

Del 1 Kort respons

1. Hva er din primære rolle?
2. Hvilken type anlegg jobber du på? (For eksempel ferskvann, skog)
3. Hvilke typer data jobber dere med? (For eksempel tekst, video, lyd, epost)
4. Har dere et formelt dokument som sier noe om hvordan data skal håndteres under et forskningsprosjekt og etter at prosjektet er fullført? (Ja/Nei)
   Hvis ja, hvor ofte oppdateres dette dokumentet?
5. Kreves det at dere gjør data tilgjengelig for bruk utenfor denne forsknings-stasjonen? (Ja/Nei)
   Hvis ja, hvem setter et slikt krav?

Del 2 Åpent langsvar

6. Samler dere data med digitale sensorer? (Ja/Nei)
   Hvilken type digitale sensorer bruker dere i så fall?
7. Hvordan samhandler dere med de digitale sensorene? (For eksempel visualiseringsprogramvare/berøringsskjerm/smarte innebygde nettbrett osv.)
8. Opplever dere noen utfordringer med bruk av digitale sensorer for miljøovervåkning? (Ja/Nei)
   Hvis ja, hva er noen av utfordringene?
9. Er det noen parameter som ikke kan overvåkes med sensorer? (Ja/Nei)
   Hvis ja, hvilke parameter?
10. Foregår innsamlingen/lagringen av dataene i perioder? (Ja/Nei)
    Hvis ja, hvor lang er denne perioden og hvorfor slutter den (hvis den gjør det)?
11. Beskriver dere data med kontekstuell informasjon? (Ja/Nei)
    Hvis ja, kan du beskrive mer detaljert om dette er lokal praksis eller formelle standarder?

Del 3 Uttalelse

12. Retningslinjene for prosedyrer for kvalitetssikring er tilstrekkelige

    1 - Helt uenig
    2 - Uenig
    3 - Ingen mening
    4 - Enig
    5 - Helt enig

    Kan du kort utdype hvorfor?

13. Dataene du jobber med er passende beskrevet for fremtidig bruk

    1 - Helt uenig
    2 - Uenig
    3 - Ingen mening
    4 – Enig
    5 - Helt enig

    Hvis du svarte 1- Helt uenig eller 2 – Uenig: Kan du kort utdype hvorfor?

14. Du har tilgjengelige de ressurser og det personell som kreves for å utføre arbeidet ditt

    1 - Helt uenig

2 - Uenig
3 - Ingen mening
4 – Enig
5 - Helt enig

Kan du kort utdype hvorfor?

15. Det er viktig å åpne tilgang til data slik at vitenskapelig forskning kan utføres effektivt.
1 - Helt uenig
2 - Uenig
3 - Ingen mening
4 – Enig
5 - Helt enig

Kan du kort utdype hvorfor?

16. Det er utfordrende å åpne tilgang til data for vitenskapelig gjennomføring

1 - Helt uenig
2 - Uenig
3 - Ingen mening
4 – Enig
5 - Helt enig

Hvis du svarte 4 – Enig eller 5 – Helt enig: Kan du kort utdype hvorfor?

English version

Part 1 - Short response

1. What is your primary role?
2. What type of facility do you work at? (e.g. fresh water, forest, etc.)
3. What types of data do you work with? (e.g. text, video, audio, email)
4. Do you have a formal document that says something about how data should be handled during a research project and after the project is completed? (Yes/No)
   If so, how often is this document updated?
5. Is it required that you make data available for use outside this research station? (Yes/No)
   Who makes such a claim?

Part 2: Open/long response

6. Do you collect data with digital sensors? (Yes/No)
   If so, what type of digital sensors do you use?
7. How do you interact with the digital sensors? (e.g. visualization software / touch screen / smart built-in tablets, etc.)
8. Do you experience any challenges with the use of digital sensors for environmental monitoring? (Yes/No)
   If so, what are some of the challenges?
9. Are there any parameters that cannot be monitored with sensors? (Yes/No)
   If so, what parameters?
10. Are the data collected / stored in periods? (Yes/No)
    If so, how long is this period and why does it end (if it does)?
11. Do you describe data with contextual information? (Yes/No)
    If so, can you describe in more detail whether these are local practices or formal standards?

Part 3: Opinion

12. The guidelines for quality assurance procedures are adequate
    1 - Strongly disagree
    2 - Disagree
    3 - No opinion
    4 - Agree
    5 - Totally agree

    Can you briefly elaborate on why?

13. The data you are working on is appropriately described for future use
    1 - Strongly disagree
    2 - Disagree
    3 - No opinion
    4 - Agree
    5 - Totally agree

    If you chose 1 - Strongly disagree or 2 – Disagree: Can you briefly elaborate on why?

14. You have available the resources and personnel required to carry out your work
    1 - Strongly disagree
    2 - Disagree
    3 - No opinion

4 - Agree
5 - Totally agree

Can you briefly elaborate on why?

15. It is important to open access to data so that scientific research can be carried out effectively
    1 - Strongly disagree
    2 - Disagree
    3 - No opinion
    4 - Agree
    5 - Totally agree

    Can you briefly elaborate on why?

16. It is challenging to open access to data for scientific implementation
    1 - Strongly disagree
    2 - Disagree
    3 - No opinion
    4 - Agree
    5 - Totally agree

    If you chose 4 - Agree or 5 - Totally agree: Can you briefly elaborate on why?

Appendix B Interview guide

Intervjuguide

Forstå dataarbeid i miljøovervåkning

Hva handler dette om?

Forespørselen om intervju vedrører personer som til daglig arbeider med praksiser for å håndtere data på forskningsstasjoner for miljøovervåkning. Dette kan innebære prosesser med å forberede, integrere og renskrive heterogene datasett. Vi har som mål å få empirisk innsikt i hvordan du på en slik forskningsstasjon arbeider med data, hvordan det samarbeides med andre, hvordan praksisene er påvirket av de teknologiene som brukes, og reguleringer for arbeidsplassen.

Hvem er ansvarlig for forskningsprosjektet?

Institutt for datateknologi og informatikk (IDI) på Norges teknisk naturvitenskapelige universitet (NTNU) i Trondheim er ansvarlige for forskningsprosjektet.

Hva innebærer deltakelse for deg?

Ved å delta i denne studien bidrar du til å hjelpe forskere i informasjons- og kommunikasjonsteknologi (IKT) til å bedre forstå hvordan arbeid faktisk er utført i miljøovervåkning ved å gi oss informasjon som kan hjelpe oss å adressere de riktige problemene.

Hvordan vil dine data bli brukt?

Vi vil ikke lagre sensitive data om deg. All informasjon om deg slik som navn og forskningsstasjon vil bli anonymisert, og notater vil bli tatt manuelt. Dersom det gis samtykke til lydopptak, vil dette bli slettet etter at intervjuet er transkribert. Alle notater vil bli lagret i en NTNU-godkjent sikker mappe, og vil slettes etter at prosjektet er ferdig.

Deltakelse er frivillig

Deltakelse til intervju er frivillig. Dersom du velger å delta, kan du trekke ditt samtykke når som helst uten å oppgi grunn for det. All informasjon som kan knyttes til deg vil da bli slettet. Det vil ikke være noen negative konsekvenser for deg dersom du ikke ønsker å delta, eller hvis du senere ønsker å trekke deg.

Personvernet ditt - hvordan vi vil lagre og bruke dine personlige data

- Vi vil kun bruke data fra deg til de formål som er spesifisert i dette informasjonsskrivet. Vi behandler personopplysninger konfidensielt i henhold til EUs databeskyttningslovgivning (den generelle databeskyttelsesforordningen (GDPR) og personopplysningsloven)
- Kun prosjektlederen, prosjektveiledere og medlemmer av prosjektet tilknyttet institusjonen som er ansvarlige for prosjektet, vil ha tilgang til personopplysningene.
- Navnet ditt og dine kontaktdetaljer vil bli erstattet med en kode. Listen av navn, kontakter og respektive koder vil bli lagret adskilt fra resten av den samlede dataen. Dataene vil bli lagret i en NTNU-godkjent sikker mappe.
- Deltakere vil ikke være gjenkjennelige i publikasjoner/rapporter sendt til NTNU.

Spørsmål

Del 1 – Om Deg

1. Hva er din primære rolle?
2. Hva arbeider du med på en daglig basis?

Del 2 – Data Curation Verktøy

3. Hvilke verktøy bruker du/brukes på din forskningsstasjon for å samle data?
4. Pleier nye teknologier å forandre seg hyppig? Med ny teknologi menes slik som ny programvare, sensorer, servere eller databaser for innsamling.
5. Tenk på en teknologiendring. Hvordan tilnærmer du deg til nye endringer i teknologi?

Del 3 – Samarbeidsaktiviteter

6. Utfører du ekstraarbeid på dataene med dine kollegaer dersom dere for eksempel er nødt til å rapportere data eksternt?
7. Hvilke utfordringer møter du på en daglig basis: Med tanke på innsamling, samarbeid, deling, det å gjøre data forståelig?
8. Hvilke målinger tar du for å forsikre at dataene er forståelig?
9. Hvilke målinger tar du for å forsikre at dataene er av god kvalitet? (Slik at de for eksempel kan brukes senere, være forståelig for andre utenfor forskningsstasjonen)

Del 4 – Reguleringer

10. Er du nødt til å gjøre data tilgjengelig til andre forskningsstasjoner? Hvem setter slike krav?
11. Hvordan er data delt med andre? (Innenfor forskningsstasjoner og utenfor)
12. Hvilke retningslinjer (f.eks avtalte praksiser, standarder) er brukt for kvalitetssikring og langtidsarkivering?
13. Hvordan er data lagret, plassert og åpnet for bruk/tilgang?
14. Utfører du noen ekstraoppgaver som går utenfor din jobb-beskrivelse?

Del 5 – Fordeler og ulemper

15. Beskriv et aspekt ved jobben din som du syns fungerer bra.
16. Hva kan bli gjort annerledes for å forbedre ditt daglige arbeid?
17. Er ditt daglige arbeid mer utfordrende enn jobb-beskrivelsen din?
18. Føler du at du er tilstrekkelig oppmuntret for eksempel til å dele data?

Interview Guide

Understanding Data Governance in Environmental Monitoring

What is this about?

This is an inquiry about the practices, work processes, information systems, and infrastructures adopted by participants who work in environmental monitoring research stations to prepare, integrate, and make sense of heterogeneous datasets on a day-to-day basis. The aim is to contribute with empirical insights on how participants in research stations work on data including how they work with each other; and how these work practices are affected by the technologies they use and the workplace policies on a daily basis.

Who is responsible for the research project?

The Department of Computer Science at the Norwegian University of Science and Technology (NTNU, Trondheim).

What does participation involve for you?

By participating in this study, you are contributing to help researchers in ICT to better understand how work is actually performed in environmental monitoring by taking an active role to provide information to help us address the right problems.

How will your data be used?

The research methodology (qualitative research) incorporates information that will be collected by interview. The information will be anonymized and recorded on paper and/or by audio recording if consent is given. Audio recordings will be deleted when the interview is transcribed. The notes will be taken manually and stored electronically in a NTNU-approved safe folder.

Participation is voluntary

Participation in this project is voluntary. If you choose to participate, you can withdraw your consent at any time without giving a reason. All information about you will then be made anonymous. There will be no negative consequences for you if you choose not to participate or later decide to withdraw.

Your privacy – how we will store and use your personal data

- We will only use your personal data for the purpose(s) specified in this information letter. We will process your personal data confidentiality and under EU data protection legislation (the General Data Protection Regulation and Personal Data Act).
- Only the project leader, project supervisors and members of the project affiliated with the institution responsible for the project will have access to the personal data.
- Your name and contact details will be replaced with a code. The list of names, contacts, and respective codes will be stored separately from the rest of the collected data. The data will be stored in a NTNU-approved safe folder.
- The participants will not be recognizable in publications/reports submitted to NTNU.

Questions

Part 1 – About You

1. What is your primary role?
2. What do you do on a daily basis?

Part 2 – Data Curation Tools

3. What tools do you use to collect data?
4. Do new technologies (new software, sensors, servers, databases) for collection often change?
5. Think of an example of technology change. How do you adapt to any changes in technologies?

Part 3 – Collaboration Activities

6. Do you perform extra work on the data with your colleagues if you, for example, have to report data externally?
7. What challenges do you face on a daily basis: In collecting, collaborating, sharing, making data understandable?
8. What measures do you take to ensure that data is understandable?
9. What measures do you take to ensure that data is of good quality?

Part 4 – Policies

10. Are you required to make data available to other research stations? Who makes these requirements?
11. How are data shared with others? (within research station and outside)
12. What guidelines (e.g., agreed practices, standards) are used for quality assurance and long-term archiving?
13. How are data saved, located and accessed?
14. Are there any extra tasks you perform outside your job description?

Part 5 – Benefits and Challenges

15. Describe one aspect of your work that you think works well.
16. What can be done differently to improve your day-to-day work?
17. Is your daily work more challenging than your job description? Why?
18. Do you feel that you are adequately incentivized (e.g., to share data)?

NTNU
Kunnskap for en bedre verden