Master's thesis

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

Steinar Kornelius Selebø Kollerud

IoT and Data Curation in Long-Term Environmental Monitoring

An Interpretive Case Study

Master's thesis in Master of Science (MSc) in Computer Science
Supervisor: Elena Parmiggiani
June 2020

NTNU
Norwegian University of Science and Technology

Steinar Kornelius Selebø Kollerud

IoT and Data Curation in Long-Term Environmental Monitoring

An Interpretive Case Study

Master's thesis in Master of Science (MSc) in Computer Science
Supervisor: Elena Parmiggiani
June 2020

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

NTNU
Norwegian University of
Science and Technology

Abstract

Research infrastructures for environmental monitoring has become more regulated by governing bodies, such as ESFRI, through policies and requirements to receive financing. These policies have a technical focus primarily, while the complex socio-technical relations present at research infrastructures often get neglected. Previous research shows that these policies cause tension for the researchers, as they have to continually balance the interests of different actors in their daily work. Furthermore, data collection methods evolve with the development of new technologies, becoming more sensor-based and automated. Previous articles address the increased importance of data curation, as the research infrastructures become more technical and automated.

This study aims to contribute with new empirical insight into how the researchers are affected by the adoption of new technologies, by looking at the use of IoT technology at different research infrastructures, and how it impacts the data curation, the increased focus on data sharing, and the researchers' daily work. This thesis adopts an information infrastructure perspective to view research infrastructures as continuously evolving and complex socio-technical infrastructure.

This interpretive case study relied on qualitative data: Interviews, observations during seminars, and documentation to answer the research questions, and included environmental researchers and other participants of interest working at environmental research organizations. The findings describe how data curation could improve the use of IoT, support distributed data sharing, and the importance of data curation to ensure that scientific data is trustworthy. Furthermore, this thesis identifies the extra work required by the researchers, and how they are affected by the lack of incentives, time, and resources.

Sammendrag

Forskningsinfrastrukturer for miljøovervåking har blitt mer regulert av styrende organer, som ESFRI, gjennom retningslinjer og krav for å motta finansiering. Disse retningslinjene har hovedsakelig et teknisk fokus, mens de komplekse sosio-tekniske relasjonene mellom forskerne og den tekniske infrastrukturen ved forskningsinfrastrukturer ofte blir neglisjert. Tidligere forskning viser at disse retningslinjene skaper spenninger for forskerne, ettersom de kontinuerlig må balansere ulike interesser fra forskjellige aktører i det daglige arbeidet. I tillegg, er arbeidsmetodene for datainnsamling under kontinuerlig utvikling. Utvikling av ny teknologi har ført til at arbeidet med datainnsamling har blitt mer sensorbasert og automatisert. Tidligere forskningsartikler tar for seg den økte betydningen av *data curation*, som følge av at forskningsinfrastrukturene blir mer tekniske og automatiserte.

Denne studien har som mål å bidra med ny empirisk innsikt om hvordan forskerne blir påvirket av å ta i bruk nye løsninger, ved å se på bruken av IoT-teknologi ved forskjellige forskningsinfrastrukturer, og hvordan det påvirker *data curation*, det økte fokuset på datadeling, og det daglige arbeidet til forskerne. Denne oppgaven tar i bruk et *information infrastructure*-perspektiv for å se på forskningsinfrastrukturer som komplekse sosio-tekniske infrastrukturer som er under kontinuerlig utvikling.

Denne casestudien er basert på kvalitative data: Intervjuer, observasjoner under seminarer og dokumentasjon for å svare på forskningsspørsmålene, og inkluderte intervjuer fra miljøforskere og andre aktuelle informanter som jobber ved forskningsorganisasjoner for miljøforskning. Funnene beskrev hvordan *data curation* kunne forbedre bruken av IoT, støtte distribuert datadeling og viktigheten av *data curation* for å sikre at vitenskapelige data er pålitelige. Videre identifiserte studien ekstraarbeidet som kreves fra forskere, og hvordan de påvirkes av manglende insentiver, tid og ressurser.

Acknowledgments

This thesis is the author's master thesis as part of the fulfillment of a master of science (MSc) in computer science at the Norwegian University of Science and Technology (NTNU). This thesis is written in collaboration with the Department of Computer Science at NTNU, in the spring semester of 2020.

First, I would like to thank my supervisor associate professor Elena Parmiggiani of the Department of Computer Science at NTNU, for your guidance and feedback during the work with this thesis and semester project in the fall semester of 2019. You always stayed positive and supported my work, even when things were tough.

I would also like to thank all the informants and research organizations that contributed with valuable insight and information about the environmental research field and the use of IoT.

A very special thank you also goes to all the informants that helped me recruit new informants and gave me access to internal seminars at a research organization.

Trondheim, June 2020

Steinar Kollerud

Contents

List of Tables

List of Figures

List of Abbreviations

CSW Computer-supported cooperative work

eLTER European long-term ecosystem research network

IoT Internet of Things

LTER Long-term ecosystem research

NTNU Norwegian University of Science and Technology

RI Research infrastructure

SDI Stepwise-deductive induction

1 Introduction

This chapter describes the purpose and motivation for this project and elaborates on the research questions and contributions. It also explains the limitations and scope of this project and the structure of this thesis. The description of the purpose for this project and research questions is a continuation of the project preceding this thesis (Kollerud, 2019), that has been amended with a discussion of new articles that have become available after the project.

1.1 Purpose of this Project

Research infrastructures (RIs) for environmental monitoring, and the work conducted by environmental researchers to collect, process, and analyze data is an essential part of the process to produce trustworthy and understandable knowledge about the environment. Environmental research contributes to our understanding of the environment, provides insight into changes or disturbances of the environment, documents the impact of human influences, and provides decision-makers with timely and accurate information.

RIs for environmental research consist of research stations and infrastructures that measure and observe diverse environmental aspects, e.g., rivers, forest, sea, and air. Also, each RI uses various methods for collecting and handling the data through the data life cycle. Sensor devices of different quality and sensitivity are utilized, and environmental researchers need to ensure that sensors are continuously calibrated. Managing the data to make sure that it is trustworthy, readable, and meaningful is also an essential part of the work (Karasti, 2009).

One of the main challenges in this process described in previous articles by Karasti et al. (2006), Kaltenbrunner (2017), and Karasti (2009) is that RIs are becoming more regulated by supernational (EU via ESFRI) and national (The Research Council of Norway) bodies. These institutions are a vital source for research funding, and adhering to their policies and guidelines is an integral part of the work conducted by researchers. The work required to balance the interests of different actors can cause tension for researchers.

Since the beginning of the 21st century, there has been an increased focus on technological adoptions for environmental monitoring. Using sensors and more extensive sensor networks has become more common to satisfy the increased demand for large datasets in ”next-generation” science, more commonly known as e-Science (Lord and Macdonald, 2003; Hey and Trefethen, 2003). Karasti et al. (2006) describe e-Science as follows:

> The vision of e-Science (or Cyberinfrastructure) with interest in large-scale science carried out through distributed global collaborations brings forward the issues of access to and sharing of scientific data collections together with the supporting technologies of networks and computing resources (Atkins et al., 2003; UK Research Council e-Science definition, 2001). (p. 321)

Also, reduced costs for making processing units contributed to making sensors, sensor networks, and technical infrastructure cheaper and more accessible (Dourish, 2016). These sensor networks, more popularly known as the Internet of Things (IoT) networks, allow researchers to collect large datasets more efficiently and provide continuous data. However, they also create new problems related to accountability and data curation (Boos et al., 2013; Yang et al., 2017).

LTER-Europe has a leading position in Europe when it comes to establishing a ”next-generation” large-scale research infrastructure. LTER-Europe is an umbrella organization that consists of a distributed network of national long-term ecosystem research (LTER) sites all over Europe. It is an initiative funded by ESFRI to establish a network of RIs that make use of leading-edge technology to conduct environmental research. LTER-Europe was also motivated by e-Science to promote the sharing of data and collaboration between distributed RIs (LTER-Europe, 2019).

Previous research projects into this field look at the importance of data curation at LTER-RI to make the data understandable (Karasti, 2009; Karasti et al., 2006). Other articles address how to make data understandable for distributed collaborators, and how other researchers can reuse it (Zimmerman, 2008; Borgman et al., 2012; Arzberger et al., 2004b). Kaltenbrunner (2017) focuses on policies by governing institutions on data sharing, and how it affects the researchers. Articles by Monteiro and Parmiggiani (2019) and Parmiggiani et al. (2015) look at the process of establishing an IoT network for monitoring marine life in Norway. These articles all have a socio-technical perspective, but the majority of research articles with a focus on IoT and environmental monitoring have a technical focus, where the researchers' perspective often gets neglected. Wen-Tsai and Chia-Cheng (2013), Truong et al. (2017), and Park and Seo (2017) look at the use of IoT in different research fields of environmental monitoring, but they are limited and only focus on the technology.

There is a gap in the research regarding how the adoption of new technologies, such as IoT, and changes in policies and regulations, impacts the daily work of the researchers, and the work required to comply with these policies and implement new technology successfully. ISO/IEC JTC 1 (2014) defines IoT as an infrastructure of interconnected objects and people that is combined with an information processing unit. This definition of IoT is socio-technical and includes both technical devices and people. Therefore, it is natural to adopt a socio-technical perspective when looking at this technology.

This thesis contributes to the pool of socio-technical research when looking at IoT, data curation, and distributed sharing of research data within environmental monitoring.

1.2 Research Questions

This master thesis is the second part of a two-phased project that looked into the use of IoT for environmental monitoring by focusing on the research questions described below. The first part consisted of a literature review that aimed to provide a more comprehensive understanding of the field by looking into previous research (Kollerud, 2019).

This thesis consists of a case study into IoT for environmental monitoring by looking at the following research questions:

- RQ1: How is IoT used at different environmental research infrastructures?
    - RQ1.1: How do researchers assess the trustworthiness of data?
    - RQ1.2: How does IoT affect the daily work of the researchers?
- RQ2: How is the increased focus on data sharing affected by IoT?

The original research plan for data generation methods in this case study, consisted of performing field studies at multiple RIs in Norway, to make observations and conduct interviews. Due to restrictions and other challenges caused by the Covid-19 virus, the data generation methods had to be changed to only include remote interviews. Chapter 4. Research Methods contains a more detailed description of the research methods.

1.3 Contribution

This research project aims to contribute with new empirical insight into the use of IoT technology for environmental research at RIs in Norway. The focus is on the researchers, and how their daily work is impacted, and the work needed to ensure that the data is trustworthy and understandable. It also contributes with insight into how IoT affects the increased focus on data sharing in modern science. The themes and conceptual categories from the findings can be used as a framework for future research into the field (Walsham, 1995).

1.4 Limitations of the Scope

In this thesis, the author adopts an information infrastructure perspective and view RI as a complex socio-technical system (Monteiro et al., 2013). Therefore, this thesis does not contain in-depth descriptions of different technologies and their specifications, e.g., specific sensors, software programs, or hardware infrastructure. Instead, the technical infrastructure used for environmental monitoring is viewed as part of an information infrastructure.

The findings described in chapter 5. Findings were also limited by the constraints on the data generation methods explained in chapter 4. Research Methods, and the time constraint of the master thesis. The data collection was limited geographically to only include environmental researchers working in Norway.

1.5 Structure of this Thesis

This thesis is structured as follows.

Chapter 2 defines key concepts and discuss literature about the concepts addressed by this thesis. It also outlines the overarching theoretical framework.

Chapter 3 describes the case for this thesis.

Chapter 4 describes the research strategy, elaborates on the methods for analyzing the data, and defines the research paradigm for this thesis.

Chapter 5 presents the findings in the form of themes and conceptual categories derived from the research data.

Chapter 6 discusses the findings by looking at theory and research questions. It also includes other observations of interest from the findings.

Chapter 7 contains the concluding remarks for this thesis and suggestions for future work.

Appendix contains an example of the interview guides used in this case study.

2 Literature Background

This chapter contains the literature background and outlines the overarching theoretical framework for this thesis. It starts with defining infrastructure. Then it introduces the concepts data curation and Internet of Things. It also outlines the difference between IoT, sensors-based, and non-sensor-based data collection. It continues with the topics distributed sharing and reuse of data. Lastly, it describes the impact of policies by governing institutions.

2.1 Defining Infrastructure

State of the art research into different perspectives for looking at research infrastructure (RI) was reviewed in the project preceding this thesis (Kollerud, 2019). This thesis provides an improved understanding of the theme.

RI is a broad term. The European Commission (2019) defines it as; ”facilities that provide resources and services for research communities to conduct research and foster innovation.” This definition is limited and quite techno-centric. Instead of viewing RI as any facility where research is conducted, this thesis adopts an information infrastructure perspective (Monteiro et al., 2013; Karasti et al., 2010; Pollock and Williams, 2010). The author proposes this definition because it helps to foreground the social aspects of the work conducted at the RI. Monteiro et al. (2013) define information infrastructure as follows:

> As a working definition, IIs are characterised 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). IIs are also typically stretched across space and time: they are shaped and used across many different locales and endure over long periods (decades rather than years). (p. 576)

This definition of information infrastructure describes a complex socio-technical infrastructure that is continuously evolving and adapting to manage the implementation of new technologies, change in work procedures, and meet the users' demands (Pollock and Williams, 2010; Karasti et al., 2010; Mongili, 2014). Looking at RI from an information infrastructure perspective provides a framework for studying the socio-technical relationships between the researchers conducting work at the RI and technological devices present at the RI. These socio-technical relationships are an essential part of the daily work undertaken at an RI for environmental monitoring, and changes to the technological infrastructure often affect the researchers. For example, the endorsement of data sharing and large-scale collaboration is not only a technological challenge. It also impacts the researchers who have to balance external pressure and multiple interests, such as the extra work to create descriptions and metadata (Karasti, 2009; Arzberger et al., 2004a; Zimmerman, 2008).

In addition to looking at the socio-technical relations at RIs, this thesis also focuses on the cyberinfrastructure present at RIs. The term cyberinfrastructure emerged in the early 21st century to describe the initiatives for large-scale science facilitated by the sharing of research data between distributed RIs (Atkins et al., 2003; Ribes and Lee, 2010; Karasti et al., 2006; Parmiggiani, 2015). A cyberinfrastructure includes repositories for storing and querying data, computational systems for analyzing and manipulating the datasets, communication infrastructure for interconnecting all parts of the data infrastructure, and an interface for researchers to access and share data (Stein, 2008; Leonelli, 2013).

The efforts to establish an extensive network of distributed RIs for collaboration and data sharing led to an increased focus on data curation and how to make data trustworthy and usable when shared between distributed locations. These topics are addressed in the following subsections.

A key difference between cyberinfrastructure and information infrastructure is that cyberinfrastructure only includes the technical infrastructure present at a RI, while information infrastructure includes the technical infrastructure, the researchers, and their socio-technical relations. IoT technology that is the focus area of this thesis belongs to both the cyberinfrastructure as part of the technical environment and information infrastructure that looks at it from a socio-technical perspective, for example, the work needed by researchers to install, operate, and maintain them.

2.2 Internet of Things

Research articles and relevant background material about the origin and definition of IoT were reviewed in-depth in the project preceding this thesis (Kollerud, 2019). This is amended with a discussion about articles addressing the adoption of IoT that have become available.

The term Internet of Things (IoT) traces its origin back to the late 1980s. Computer Science Laboratory director Mark Weiser at Xerox's Palo Alto Research Center (PARC) initiated a research project that he named ”Ubiquitous Computing” (Weiser, 1991). His research project challenged the common practice of how computers were utilized at the time, ”one person one computer,” where each user had their own personal computer that served their needs (Dourish, 2016). Weiser proposed a new way of computing where:

> A single person would interact with tens, hundreds, or thousands of devices – some large, some small, some visible, some hidden. Rather than being devices that we would need to sit down and use, these devices would surround us – they would be embedded in our everyday environment. (Dourish, 2016, p. 28)

This change towards multiple devices per user is driven by numerous factors, such as processing units becoming more affordable, which has made it feasible to implement it into everyday objects at a low cost. Another factor is its potential to monitor and generate data about our surroundings by deploying big networks of sensors. Also, areas such as ”Smart cities” and ”Big Data” make use of IoT to satisfy their need for large amounts of data (Boos et al., 2013; Sundmaeker et al., 2010).

IoT is often viewed as a technical infrastructure that consists of multiple technological devices that monitor their surroundings or automate a process. A modern definition of IoT is: ”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/IEC JTC 1, 2014, p. 4). This defines IoT as a socio-technical technology where the people operating and using the IoT network is an essential part of the technology.

Smart cities is a field where IoT technology has been widely adopted. Cheap sensors capable of transmitting data are placed in everyday objects and provide a continuous feed of data that is processed and used for multiple purposes, e.g., tracking free parking spots, monitoring your heart rate, or controlling the thermostat (Băjenescu, 2018; Angelakis et al., 2017). Research into the use of IoT for smart cities has also uncovered challenges such as security and privacy problems (Romero et al., 2016; Zhou et al., 2017), and the need for an infrastructure to handle the continuous flow of data (Govoni et al., 2017).

IoT is becoming increasingly popular within environmental monitoring. Truong et al. (2017) and Park and Seo (2017) look into the use of IoT for monitoring fungus, and weather conditions for agricultural crops. Parmiggiani et al. (2015) look into the work to establish an IoT network for monitoring sea life around oil rigs. Borgman et al. (2012) describe how LTER-RI uses IoT for environmental research. Similar to the smart cities articles, they address the need for an infrastructure to transmit the data and in-depth plan for processing large amounts of data. This increased adoption of IoT for environmental monitoring makes understanding and creating knowledge about how IoT affect the socio-technical relationships present at RIs increasingly relevant.

2.3 Different Methods for Data Collection

This thesis mainly distinguishes between three different data collection methods when looking at how environmental researchers work to collect data. The three different data collection methods are as follows: Non-sensor-based, sensor-based, and IoT sensor networks.

Non-sensor-based data collection, also referred to as hand-sampling, relies on a researcher going into the field to collect samples that are brought back to the lab and analyzed. The researchers can also make physical observations that are documented, e.g., soil samples, measuring the size of fish, blood samples, marking animals, water samples, and document habitat types. Sensor-based data collection describes the process when a researcher utilizes a sensor device to collect data, e.g., temperature loggers, humidity sensors, GPS trackers, acoustic sensors, air pollution sensors, and wildlife cameras (Borgman et al., 2012; Karasti et al., 2006; Hobbie et al., 2003). IoT sensors networks make use of sensors capable of transmitting the data, as described in chapter 2.2. Internet of Things.

The different groups of methods for data collection are not static, but continuously evolving with the development of new sensors and adoption of technology. Data collected through manual sampling at the time of this project can be sensor-based or partly sensor-based in the future.

Resources for data collection, such as satellites for remote sensing, as described by Kwa (2005) and drones, can be challenging to place in a specific category, but are commonly used for data collection in environmental monitoring.

2.4 Data Sharing for Distributed Collaboration and Reuse of Data

Distributed sharing of scientific data and reuse of data, gained traction with the vision of e-Science. The concept of e-Science focused on large-scale science carried out by a distributed RIs through global collaboration. It led to an increased focus on issues such as sharing and access to datasets, and the technologies needed to facilitate this global collaboration, e.g., network infrastructure, computing resources, and storage capacity (Atkins et al., 2003; Karasti et al., 2006). Sharing of scientific data and cooperation has been at the heart of research within computer supported cooperative work (CSCW) (Greif and Sarin, 1987). Still, it mostly consists of collaboration on a smaller scale (Birnholtz and Bietz, 2003; Chin and Lansing, 2004; Karasti et al., 2006).

Environmental organizations such as LTER-Europe has adopted this vision of distributed collaboration and sharing of data, and work to encourage and facilitate the sharing of data between the RIs in the network (LTER-Europe, 2019; Karasti, 2009). Research councils and other governing bodies that fund research are also encouraging data sharing through policies (Kaltenbrunner, 2017). The impact of the policies by the regulating institutions is discussed in chapter 2.6. Policies and Governing.

The increased pressure to share and encourage reuse of data conflicts with existing research practices and culture, and causes tension among researchers (Hilgartner and Brandt-Rauf, 1994; Brown, 2003). Zimmerman (2008) describes some of the problems with data sharing as follows:

> A host of problems make the benefits of data sharing and reuse remarkably difficult to realize. These challenges include issues of data ownership, a lack of incentives for scientists to share; technical hurdles related to incompatible hardware, software, and data structures; and costs to document, transfer, and store data. While scholars have yet to fully grapple with these problems, especially in small sciences such as ecology, other authors have pointed to a lack of standards as one of the major impediments to the sharing and reuse of scientific data. They assert that successful systems of data sharing depend on various kinds of formal standards, including those related to data collection, description, storage, and quality control (e.g., National Research Council [NRC] 1995, 1997). (p. 632)

Only sharing the raw data is insufficient for researchers wanting to reuse the data. Knowledge about the social context of how the data was created and processed is also necessary, e.g., configuration and calibration of sensors, methods for collecting the data, and methods for processing the data. Metadata containing these details is essential to make it understandable and trustworthy for distributed collaborators and researchers that want to reuse the data (Zimmerman, 2008).

The importance of high-quality metadata that describes the social context of how the data was created has also gathered momentum within environmental monitoring (Hobbie et al., 2003; Baker et al., 2000). Karasti et al. (2006, 2010) further expand upon the requirement, and the work required by researchers to facilitate distributed data sharing with the introduction of data curation. This term is discussed in the following chapter.

2.5 Data curation

The increased focus on data curation for RIs emerged alongside the vision of e-Science around 17 years ago, that involved establishing advanced infrastructures with an exponential increase in the availability of scientific data, and conduct research and share datasets across distributed RIs as discussed in the previous subsection (Karasti et al., 2006; Hedstrom, 2004; Lord and Macdonald, 2003).

Data curation involves a wide range of data stewardship activities, such as processing and cleaning, assembling, transmitting, storing, and deleting the data (Leonelli, 2016). Karasti (2009) and Karasti et al.

(2006) further extend upon data curation to involve an established routine for handling the data at RIs through the data life cycle, which includes the technical resources required to manage the data. Data curation is an essential part of the work to ensure that datasets are understandable and trustworthy (Hobbie et al., 2003).

Lord and Macdonald (2003) describe data curation as ”the care of the record within scientific context and environment” (p. 45). Newman et al. (2003) and Helly et al. (2002) argue that knowledge about the social context of how and where datasets were created is critical for the successful reuse of and sharing of data. Therefore, the data curation process needs to extend beyond a single RI to facilitate distributed sharing of scientific data.

Establishing standards for the data curation process across multiple heterogeneous RIs with a focus on metadata, and fitting the data into templates has proved challenging (Karasti, 2009). Attempts to implement best practice standards for data curation at RIs in the LTER-Europe network has shown that standardization causes tension with the flexibility needed in such a diverse research environment (Karasti et al., 2010). The tension between standardization and flexibility is not unique to data curation; research into this topic for infrastructures has highlighted the difficulty of finding a balance between standardization and flexibility (Hanseth et al., 1996). The tension between standardization and flexibility is discussed in the following subsection.

The importance of data curation extends beyond environmental monitoring, Ribes and Polk (2014) address the use of data curation at medical infrastructures, and Passi and Jackson (2020) look at data curation in data science. Bossen et al. (2019) and Millerand and Baker (2010) argue that the people working with data curation contribute to shaping the data and infrastructures through their daily work. This view on data curation as an integral part of the socio-technical environment present at RIs is a vital part of the theoretical framework for this thesis (Parmiggiani and Grisot, 2020).

2.6 Policies and Governing

An in-depth review of relevant background material and previous research about the impact of policies and guidelines by governing institutions was carried out in the project preceding this thesis (Kollerud, 2019). This subsection is based on the document analysis from the preceding project.

The data that is generated and processed at a RI may hold value to other researchers around the world or next generation's researchers. Research conducted into the value of data sharing and reuse of old datasets has highlighted the benefits it provides for scientific advancement and potential for saving cost, but there are also some challenges connected to the quality of the data, and data curation (Jirotka et al., 2005; Arzberger et al., 2004b; Zimmerman, 2008). Traditional research facilities with individual scientists, where datasets are only shared among close collaborators are changing for a new era that endorses data sharing (Karasti, 2009). This increased focus on data sharing is not exclusively beneficial, it also creates new problems related to data curation..

One of the main driving forces behind the shift towards data sharing and large-scale collaboration is that RIs are becoming more regulated by supernational (EU via ESFRI) and national (The Research Council of Norway) bodies (Karasti, 2009). These institutions create guidelines for a foundation that research projects and RIs must follow to receive funding. Balancing the pressure and interests from different actor groups: funders, researchers, and policymakers, and ensuring the research project fits the mold they outline is an important part of the work to create and maintain RIs (Kaltenbrunner, 2017). According to Kaltenbrunner (2017) one of the motives behind the European Commissions's policies is:

> European initiatives, I argue, are based on a more centralizing, technology-driven vision of digital infrastructure that serves the European Commission's policy goal of integrating national research systems in institutional and epistemic terms. (p. 275)

Research contributions by Karasti et al. (2006, 2010) and Ribes and Lee (2010) have looked into the policies and governance of RIs and attempted to explicitly combine it with a constructivist approach. For example, Ribes and Lee (2010) have criticized research in Computer-supported cooperative work for a narrow focus on individual RIs. They argue that this narrow focus obscures policies and broader institutional frameworks that impact the everyday work of establishing new RI and conducting research at these facilities. Similar observations by Karasti et al. (2010) show that much research in CSCW neglect long-term issues about funding cycles and institutional policies on technology development, and focuses more on short-term problems about infrastructure design (Kaltenbrunner, 2017; Kee and Browning, 2010).

The policies put forth by governing institutions (e.g., ESFRI) cause tension similar to the tension between standardization and flexibility described by Hanseth et al. (1996). They argue that standardization is often obstructed and interleaved with processes that require the standards to be easy to change and flexible. Another issue they address is the problem of change once a standard has been implemented. Once a standard has been implemented across a large population of organizations, the cost and effort of making changes increases with the size of the population. The top-down pressure caused by these policies causes tension when the heterogeneous environment of RIs attempts to fit the mold outlined by the policies (Karasti, 2009). The work conducted at a RI is a complex socio-technical process that interleaves both the technical infrastructure and the researchers, and it is difficult to predict how a change to one part affects the socio-technical relations.

3 Case Description

This research project makes use of a two-semester (About 30 weeks) long interpretive literature review and case study (Walsham, 2006). The unit of analysis was the use and adoption of IoT for environmental monitoring in Norway, with a focus on the practices for handling the data, and the work required by environmental researchers to make use of the technology. Also, IoT's effect on the increased focus on data sharing in modern science is a crucial part of the unit of analysis.

The recruitment of informants was driven by pragmatic concerns of finding environmental researchers willing to participate in the case study, while also focusing on the Norwegian part of the European Long-Term Ecosystem Research Network (eLTER) and the primary research organizations in Norway.

The eLTER network is an umbrella organization that consists of national LTER-RI all over Europe. It is a diverse research network with highly heterogeneous RIs that observe different objects of interest, e.g., rivers, birds, sea life, and terrestrial. eLTER's long-term mission is to monitor and create knowledge about the effects of local, regional, and global changes to ecosystems and how it impacts society and the environment. To fulfill this mission, eLTER focuses on supporting cutting edge science with an increased focus on distributed data sharing between RIs, and policies for data curation (LTER-Europe, 2019; Willig and Walker, 2015; Parmiggiani and Grisot, 2020).

To receive financing from ESFRI or The Research Council of Norway to establish a RI for environmental monitoring, the researchers have to comply with the policies and guidelines for RIs described by strategy documents and roadmaps. These documents focus on the scientific and technical parts of the RI, and it is essential to be at the forefront of the technological developments within the research field to receive funding (Forskningsrådet, 2018; ESFRI, 2018).

The Norwegian node of the eLTER network consists of diverse RIs operated by the primary research organizations in Norway through funding from the research council. These RIs are an essential part of the process to create understandable and trustworthy knowledge about the environment, and they were the primary focus of this research project. This research project was not only limited to environmental researchers working at eLTER sites, other environmental researchers of interest, and people that work with IoT and data sharing at the primary research organizations in Norway were also recruited.

The primary research organizations for ecological research in Norway aim to provide high-quality research about the environment for decision-makers, environmental stewardship, and value creation. They spearhead the ecological research field with a focus on environmental monitoring, preservation of natural resources, and impact assessment. These organizations are often nonprofit foundations that rely on financing from the research council, other governing institutions, and private companies to conduct research.

4 Research Methods

This chapter elaborates on the research methods used in this thesis. First, it describes the research strategy and the changes made to the research strategy during this project. Next, it explains the process of recruiting informants and data collection. Lastly, it goes into detail about the methods for analyzing the data and research paradigm.

4.1 Research Strategy

The research questions in chapter 1.2. Research Questions aims to contribute with new empirical insight into the use of IoT for environmental monitoring, how it impacts the daily work of the researchers, and the effect it has on the work to make data trustworthy and data sharing. The socio-technical relationship between IoT technology and researchers is a holistic process. In order to study this holistic process in-depth, a case study was chosen as the strategy for data generation. Another reason for choosing a case study is that it allows the examination of the case in its natural setting and openness to multiple sources for data. This thesis was limited to one semester. Therefore, a short-term contemporary study was chosen to focus on the use of IoT at the time of this project. Another option for the research strategy that was considered was a survey. Due to surveys' limited possibility of going into detail on research topics, a case study was preferred (Oates, 2006; Baxter and Jack, 2008).

The original plan for the case study was to use multiple methods for data generation. It involved field studies at multiple research stations in Norway, to conduct observations about the daily-work and use of technology. Due to restrictions on physical meetings imposed by the Norwegian University of Science and Technology (NTNU) and the Norwegian government to combat the Covid-19 virus, it was not possible to conduct field studies.

Interviews were mainly going to be utilized in the early phase of this project, to create an overview of the current situation for IoT in environmental monitoring, standard practices, and challenges. Since conducting field studies with observations no longer were possible, interviews became the primary method for data generation. Interviews provided detailed insight from the informants' perspective on topics such as the use of IoT for environmental research, problems and benefits with current methods for data collection, and the work required to process the data. A key benefit of utilizing interviews is that they provide multiple views on the topics.

The data generation also included document analysis of relevant documents (e.g., strategy documents, roadmaps, and guidelines for data sharing) by governing institutions and research organizations.

4.2 Recruitment of Participants

The recruitment process of participants for the case study started with two key contact persons that were the first interview objects for this project—using a snowball sampling approach that involved asking the current participants about suggestions for new potential participants. Also, other participants of interest identified during this project were recruited.

Before the recruitment process and data collection could begin, the following ethical issues were addressed: Applying to the Norwegian Centre for Research Data (NSD) to get permission to start the data collection, and establish a process for the participants to consent to partake in the interviews either orally or written. All interviews were anonymized and stored securely on NTNU's servers to safeguard the identity and security of the participants.

The participants were mainly recruited from two different primary research organizations for ecological research in Norway. However, one of the participants worked at a private company that develops digital sensors for environmental monitoring. The recruitment process was driven by pragmatic concerns of finding researchers with a relevant background that were willing to participate.

4.3 Data Collection

The data collection relied on qualitative data: Interviews, observations during seminars, and documentation. Eight individual structured interviews and one semi-structured group interview were conducted during this project. The choice between structured and semi-structured interviews was decided in conjunction with the informants to accommodate the challenging work environment caused by the Covid-19 virus. An example of the interview guide can be viewed in Appendix A.

Interview Guide. Some minor changes to the questions were made between interviews based on the informant's background, but the main topics remained the same.

The informants consist of environmental researchers, software developers, and system engineers that work at different environmental research organizations. It was decided to include software developers and system engineers that work with facilitating the use of IoT for environmental research and develop frameworks for data sharing, to capture the perspective of both the environmental researchers and the people working with the technologies. A detailed overview of all the data sources for this project is provided in Table 1.

In addition to conducting interviews, one of the participants provided the opportunity to attend a couple of internal seminars at a research organization and take notes about the topics presented at the seminars. A seminar on the use of IoT for environmental research and a seminar presenting a framework for data sharing were chosen based on their relevance to this project. The first seminar addressed the current use of IoT at the research organizations, presented existing solutions, and discussed some challenges and benefits with the technology. The other seminar was a presentation of the research organization's new platform and framework for data sharing and described the platform in-depth, discussed the work required by the researchers to enable the platform's success, and addressed the benefits of the new platform.

The document study consists of multiple strategy documents from the research organization eLTER and the governing bodies ESFRI and The Research Council of Norway. The strategy documents include roadmaps, policies for sharing scientific data, guidelines and requirements for RIs, policies for the use of standards, and guidelines for receiving research funding. The document study was used to provide context and background for this project and corroborate the data collected from interviews and seminars (Bowen, 2009).

Table 1: A detailed overview of the data generation that displays the interviews conducted and recorded seminar notes. They are grouped by the domain of the participants, data generation method, and the working field of the participants. The final row contains the documents studied as part of this research project.

Domain	Data source	Participants
Environmental research	Semi-structured interviews	2 Environmental researchers
	Structured interviews	6 Environmental researchers
IoT for environmental research	Structured interview	1 System engineer
	Structured interview	1 Environmental researcher
	Seminar notes	2 System engineers
Data sharing framework	Structured interview	1 Software developer
	Seminar notes	1 Software developer
Policy Documents	Strategy documents by ESFRI (e.g., Roadmaps, guidelines for research funding, and guidelines for establishing research infrastructures)	
	Strategy documents by The Research Council of Norway (e.g., Roadmaps, guidelines for data sharing, and policies for research funding)	
	Strategy documents by eLTER (e.g., Data sharing policies, development of standards, and guidelines for data collection)	

4.4 Method for Analyzing the Data

The data analysis followed a stepwise-deductive induction (SDI) model for qualitative research, as described by Tjora (2019). The SDI approach focuses on generalized and detailed data analysis through the development of concepts. The SDI model for analyzing qualitative data is based on the inductive principle that starts with raw data and develops towards concepts and theories through a two-phased process.

Phase one begins with raw data from the interview notes and seminar notes, and the researcher reads through all the data material and identifies and labels concepts in the raw data. The coding of the data sources in phase one focuses on capturing and maintaining the original content from the raw data. When the researcher is satisfied with all the concepts, phase two begins.

Phase two consists of incremental deductive feedback loops where the concepts identified in step one are grouped into new conceptual categories. This cycle is repeated for the conceptual categories to develop themes and move towards theory. The number of cycles needed depends on the data material. For this project, three cycles of grouping concepts and conceptual categories were used. Table 2 shows a detailed overview of the top-level themes, conceptual categories, and examples of raw data belonging to each conceptual category.

The computer-assisted qualitative data analysis software HyperRESEARCH was used to identify and label concepts in the raw data, create new conceptual categories, and group concepts and conceptual categories into higher-tier conceptual categories or themes. Figure 1 contains a screenshot of the coding process in HyperRESEARCH. The results from the data analysis are presented in detail in chapter 5. Findings.

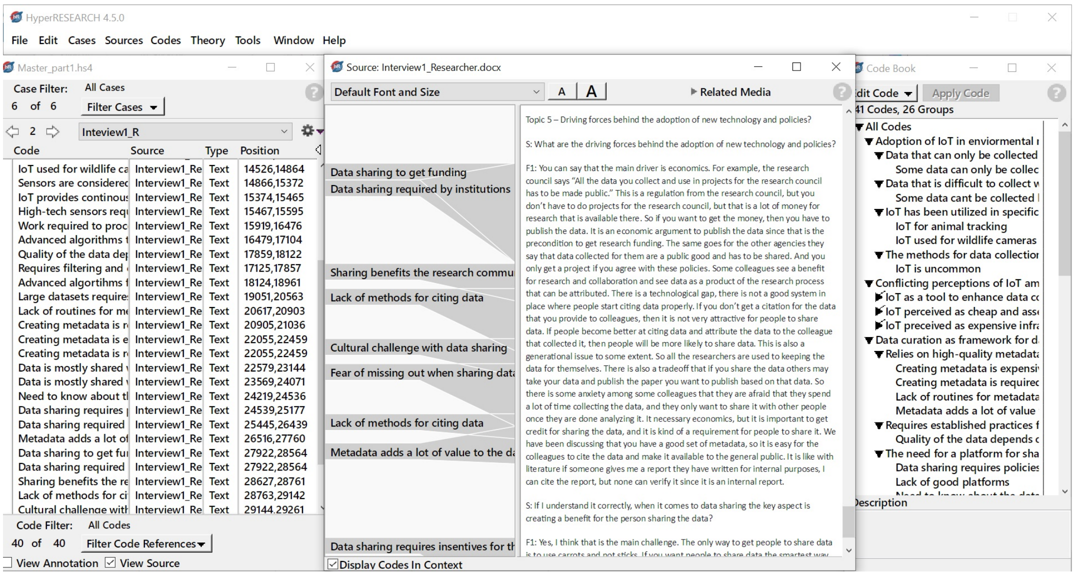

Figure 1: A screenshot of the coding process in HyperRESEARCH. The view on the left displays an overview of all the codes for an interview or seminar notes. The center view contains the interview or seminar notes with the codes on the left margin, and the right view shows an overview of all the codes and code groups for the data analysis.

Table 2: The table contains the analytical framework for this case study with an overview of the themes, top-level conceptual categories, and excerpts and examples from the data material.

Themes	Conceptual categories	Examples and excerpts
Adoption of IoT for environmental monitoring	The methods for data collection are continuously evolving	"It is unfeasible to do everything manually, and there is where remote sensing comes in to play. We go to selected spots and collect field data at some places. Then we use remote sensing to fill the gaps in a timelier fashion" (Environmental researcher, interview).
	Data that is difficult to collect with sensors	"As soon as you need an interpretation, that is something that requires human interaction. At least for plants when you determine species, you need to have humans that collect data" (Environmental researcher, interview).
	Data that can only be collected by sensors	"In contrast, research areas like wildlife tracking rely on sensors to be effective" (Environmental researcher, interview). "However, with stuff like temperature, you need some sort of temperature measuring devices anyways" (Environmental researcher, interview).
	IoT has been utilized in specific research fields	"There are possibilities with wildlife camera traps that send the data. You can insert a SIM-card, and it keeps sending the data" (Environmental researcher, interview).
Conflicting perceptions of IoT among researchers	IoT as a tool to improve data collection	"If none of the data must get lost, then people would use techniques where they send it over" (Environmental researcher, interview).
	IoT perceived as cheap and accessible devices	"Sensors, access points, and the infrastructure required have also become much more affordable" (System engineer, interview).
	IoT perceived as expensive infrastructure	"Other devices like with fresh water the sensors that are available there are expensive, and you don't necessarily put out a lot of those" (Environmental researcher, interview).
Data curation to facilitate the use of IoT	Requires a thoughtful strategy for the data curation process	"[With] large amounts of data, storage is a challenge. So you have to build good routines to avoid storing data that you don't need. [...] Some people have routines where they fetch data from specific times they are interested in, process it, and deletes it afterward. Then they keep the results to minimize the storage requirements" (Environmental researcher, interview).

	Relies on advanced algorithms for processing the data	”With environmental DNA, you can have algorithms that require so much RAM that even on a big server, you run out of memory quite fast. They have algorithms that require almost a terabyte of RAM, and huge amounts of data” (Environmental researcher, interview).
	Requires infrastructure to transmit, process, and store data	”I think that some other challenges are technology-related. When you start using sensors-based data, then you need to have the technology to store the data and analyze the data, that can prevent you from using sensors” (Environmental researcher, interview).
Data curation as a framework for data sharing	Requires established practices for handling the data	”The researchers are required to follow some best practice guidelines on how to store the data, process it, and create high-quality metadata” (Software developer, interview).
	Relies on high-quality metadata	”It is a good practice when you store your data to document what data it is, for what purpose it was used” (Environmental researcher, interview).
	The need for a platform for sharing of scientific data	”The other part of the system takes the data stored internally and makes it available on the search engine. Our goal is to make this process as automatic as possible” (Software developer, interview).
Data sharing conflicts with the present research culture	Lack of incentive for the researchers	”If you don't get a citation for the data that you provide to colleagues, then it is not very attractive for people to share data” (Environmental researcher, interview).
	It has been predominantly required by research councils	”All the data you collect and use in projects for the research council has to be made public” (Environmental researcher, interview).
	Balancing the interests of different actors	”When researchers see the usefulness of the new framework for sharing data, and how it benefits the research community and make data accessible to more people” (Software developer, interview).
Driving forces behind the adoption of IoT technology and policies	Adoption to receive funding	”One of the main drivers is economics; the research council requires that all the data you collect and use in projects for the research council has to be made public” (Environmental researcher, interview).
	Adoption to solve a problem	”IoT makes it possible for us to collect data over a longer period at a reduced cost” (Environmental researcher, interview).

4.5 Research Paradigm

This research project was a qualitative case study that aimed to provide new empirical insight into the use of IoT for environmental monitoring, how it affects the researchers, the work required to make the data trustworthy, and IoT's impact on data sharing. This study adopted a socio-technical perspective to focus on the relationship between IoT technology and the environmental researchers that utilize it in their daily work. An interpretive perspective was selected because research belonging to the interpretive paradigm is concerned with understanding the social context of information systems and other technologies (Oates, 2006; Klein and Myers, 1999).

Other aspects of this research project that made it suitable for an interpretive approach was that the data collected might be open for multiple interpretations, and the interviews conducted during this project focused on understanding the impact of IoT on the day-to-day work of environmental researchers. In an interpretive research project, the perspective of the observations is essential, in this project, the focus was on the environmental researchers' point of view and the view of the software developers and system engineers that facilitated the use of the technology (Oates, 2006; Walsham, 1995).

Researcher reflexivity is another crucial issue that was addressed—recognizing that researchers are not neutral and acknowledging how the researcher influences the research and the informants, is essential in interpretive research. The author focused on taking a passive role when interacting with informants and conducting the interviews to avoid influencing their views. However, the topics and direction of the interviews were guided by the author through the interview questions (Oates, 2006; Klein and Myers, 1999).

5 Findings

This chapter describes the findings of the case study by summarizing the results for each theme in Table 2 and includes examples and excerpts from the interviews. The themes have their own subsection that describes the conceptual categories belonging to each theme. Figure 2 provides an overview of all the themes. The themes are grouped based on their relations to the overarching theoretical framework. It also illustrates the relationships between different groups of themes by mapping the connections discovered in the data analysis.

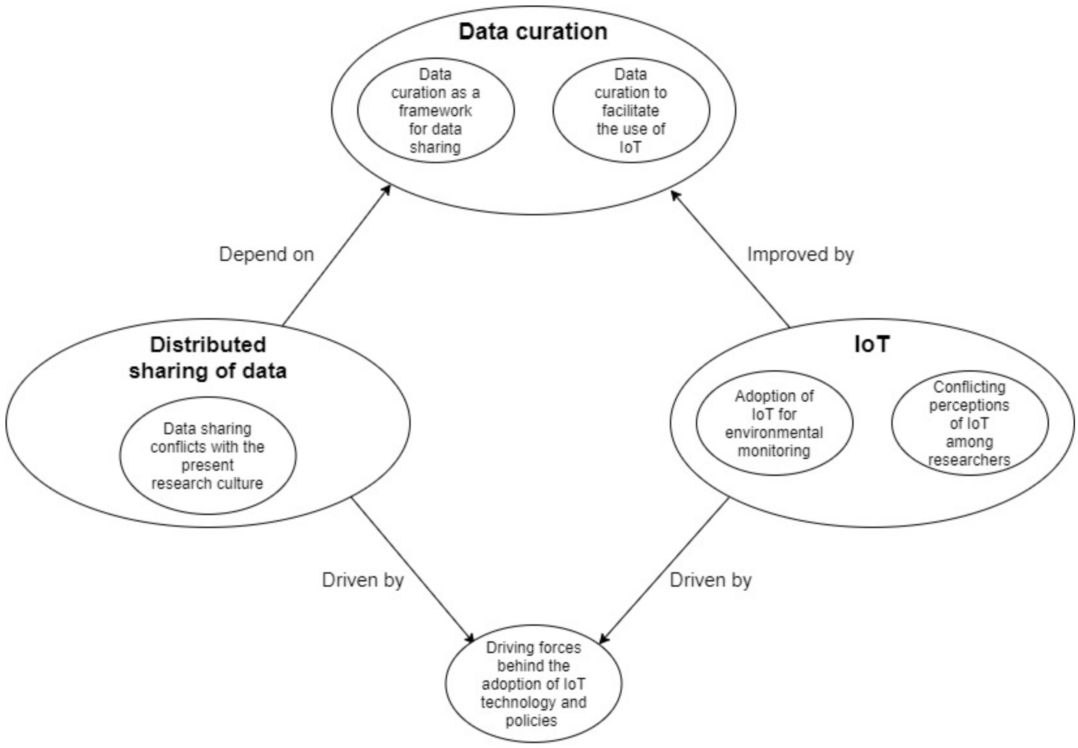

Figure 2: Overview of the themes from the analytical framework in Table 2. The themes are grouped based on the overarching theory, and the figure also illustrates the relationships between themes.

5.1 Adoption of IoT for Environmental Monitoring

The informants described the use of IoT for environmental monitoring as uncommon or early on in the adoption phase, as illustrated in this excerpt: ”We are currently in an early adoption phase, where we work on presenting the technology and help establish pilot projects" (System engineer, interview). One recurring argument among the participants for why IoT has had a low adoption rate was that the environmental research field is heterogeneous, and environmental researchers observe a large variety of objects of interest. Therefore, the data collection methods were limited by the existing sensors because sensors or IoT networks can not collect all types of data.

The different types of data that environmental researchers collect were categorized into two different groups; *data that can only be collected by sensors* and *data that is difficult to collect with sensors*. *Data that can only be collected by sensors* were described as observations requiring the use of sensor devices to collect data, e.g., temperature loggers, barometers, and humidity sensors. *Data that is difficult to collect with sensors* involved observations that could not be observed by existing sensors due to limitations in the technology, types of data that required the physical collection of samples, or as described by one environmental researcher:

> The more interpretation during data collection is necessary, the more likely there is that this is human collected data. For example, habitat types, if you want to know what kind of specific habitat types a specific location belongs to, then that requires interpretation of the situation on the ground. (Environmental researcher, interview)

Data collection that required a higher-level of interpretation involved having a person with expert knowledge within the research field collect data based on visual observation and cognitive reasoning such as, identifying plant or bird species.

Another challenge with the data collection described by the environmental researchers was that the *data collection methods were not static, but continuously evolving*. The development of new and improved sensors, new procedures for manually collecting data, and new methods for analyzing the data have led to the adoption of new methods for data collection within different fields. One example of an evolving data collection method is as follows:

> [Researchers] extract DNA from the water samples, from the DNA they can estimate which species that are present in the lake. That is sort of the most automatic way of detecting species that is possible. [...] That is a fairly new technology

> that is getting a lot of traction. Before they used gillnets, where they put out nets in the water and caught the fish. (Environmental researcher, interview)

This excerpt illustrates how the detection of different species in a lake has evolved from a non-sensor-based data collection method to a more automated sensor-based one.

Although the informants described the use of IoT as uncommon or in an early adoption phase, a few described how *IoT has been utilized in specific research fields*. Three examples of how IoT has been used for environmental monitoring were presented during the interviews. IoT has been used for wildlife camera traps to transmit the pictures from the cameras to the researchers in real-time. Another field where IoT has been utilized was; monitoring of sea life using acoustic sensors, and the sensors transmit a continuous feed of data for the researchers to process and analyze. IoT has also been used for tracking animals: ”A GPS tracker can be used as an IoT device where the radio receiver also transmits the location of the animal to the researcher” (Environmental researcher, interview). These use cases for IoT were characterized by a long-term perspective for the data collection, resource-intensive installation and placement of sensors, and difficulties with collecting the sensors once installed.

5.2 Conflicting Perceptions of IoT Among Researchers

The participants provided different opinions about the utility, availability, and challenges with IoT technology. These conflicting perceptions were grouped into the following three categories: *IoT perceived as expensive infrastructure*, *IoT as a method to improve data collection*, and *IoT perceived as cheap and accessible devices*.

Some of the researchers described the increased cost of using IoT for data collection as a critical challenge with the technology. One issue related to increased cost was the price of the sensors. Sensors used for environmental monitoring risked getting damaged or lost when placed in nature over a long-term period, and ”it is not uncommon that sensors get lost” (Environmental researcher, interview). The price gap between a sensor capable of transmitting data and a sensor without this possibility impacts the potential loss for the researcher if the sensor is lost. The majority of the environmental researchers described IoT devices as expensive compared to conventional sensors.

Another issue related to increased cost was the process of transmitting data. ”It is also a lot of research economics in this as well. It is not for free to send this information ’home’” (Environmental researcher, interview). Transferring data between the sensors and researchers required technical infrastructures such as cables, GSM antennas, or other radio wave technologies, and installing new infrastructure or using existing infrastructure can be expensive.

The system engineers working with IoT and environmental researchers that have used IoT for data collections viewed IoT as a method to improve data collection. They focused on benefits with the technology, such as the continuous transfer of scientific data, preventing all the data from being lost in case the sensors are lost. ”It is a lot more affordable to install sensors that can collect and transmit data over multiple years, compared to using resources on manual observations” (Environmental researcher, interview). This excerpt illustrates how IoT can improve the efficiency of the data collection and save resources by providing a continuous feed of data over a long-term period.

The system engineers focused on IoT as cheap and accessible devices when they described the technology.

> Sensors are becoming more accessible and a lot more affordable and easier to mass-produce. [...] The coverage of different radio technologies are also improving. [...] and access points and antennas have also become cheaper. (System engineer, seminar notes)

According to the system engineers, the sensors and the technology to transfer the data have become a lot more available and affordable. It was a vital part of the motivation behind pushing for the adoption of IoT within environmental monitoring. This view of IoT as cheap and accessible devices conflicts with the perception of the environmental researchers described above.

5.3 Data Curation to Facilitate the Use of IoT

A recurring problem mentioned by environmental researchers when working with and collecting large datasets or a continuous stream of data; was the *need for a thoughtful strategy for the data curation process*. Storage space and processing power are limited resources. With increasing amounts of scientific data, optimizing the use of these limited resources has become an essential part of the work conducted by environmental researchers.

> [With] large amounts of data, storage is a challenge. So you have to build good routines to avoid storing data that you don't need. [...] Some people have routines where they fetch data from specific times they are interested in, process it, and deletes it afterward. Then they keep the results to minimize the storage requirements. If you have huge amounts of data, then you also need processing power for it. (Environmental researcher, interview)

This excerpt illustrates how environmental researchers have established routines for fetching, processing, and storing research data to overcome the challenges of limited processing power and storage space. The document analysis discovered that research organizations have guidelines for data stewardship. The informants confirmed this discovery but noted that there was limited enforcement of the guidelines. "[The research organization] got its own guidelines for handling the data, but it varies a lot between projects, and you decide how you want to do it for your project" (Environmental researcher, interview). The data curation process was described as fragmented, where individual researchers and research projects had different routines for handling the data.

Data analysis and data processing is another part of the data curation that has encountered challenges with the increased amount of data. Manual work methods and algorithms that require much oversight has become resource-intensive and inefficient when working with a live feed of data or large datasets. To handle these challenges, the researchers have adopted more sophisticated algorithms and machine learning to automate and optimize data analysis and data processing.

> Post-processing or processing of sensors data is a significantly important task. For example with the wildlife camera traps we have hired an IT-company that use AI algorithms to identify at least humans, since we are obliged to remove all images that show humans, [...] we started developing algorithms that detect humans, and they also have algorithms that can detect different species. (Environmental researcher, interview)

This excerpt describes how machine learning algorithms are used to automate the work-process of identifying species and removing unwanted pictures from wildlife camera traps. Environmental researchers described limited time and resources as critical challenges, therefore optimizing and automating the work has become increasingly important.

Another obstacle with the adoption of IoT related to the data curation process was the *need for a technical infrastructure to transmit, process, and store data*. IoT sensors require cables or radio wave technologies (e.g., DASH7, LoRA, and 4G) to transfer the data from the sensors in the field to the RIs, where it is processed and analyzed. The RIs also require servers that provide storage space and sufficient processing power to handle the continuous feed of data. The deficiency of infrastructure for radio wave technologies in remotes areas was a key obstacle addressed by the informants.

5.4 Data Curation as a Framework for Data Sharing

The software developers described how data curation could be used to enable the sharing of scientific data through a focus on *establishing standards for processing and storing research data*, *creating high-quality metadata*, and *developing a common platform for data sharing*. They proposed utilizing data curation as a framework to support and encourage the sharing and reuse of data.

> The current situation at [the research organization] is very fragmented, where individual researchers and research projects decide on how they want to store data, where they want to store the data, and how it is shared. [...] An essential part of our project is dealing with these challenges to try and improve the situation. (Software developer, interview)

To address the challenges caused by the fragmented environment that has been present at RIs, the software developers focused on establishing standardized methods for storing and processing data as a crucial part of the data curation process. By establishing standards for processing and storing scientific data, the goal was to make the data more trustworthy and easier to understand for other researchers.

The importance of metadata was another recurring topic among the informants. Metadata has allowed researchers to capture the context of how the data was collected, processed, and stored. Both the environmental researchers and software developers described how metadata is used by researchers to understand the scientific data and assert its trustworthiness. Research organizations have established policies that require researchers to make metadata, but the quantity and content of the metadata are decided by the researchers collecting or processing the data. A challenge with metadata was that it could be time-consuming and resource-intensive to create.

> If you produce the data through scripts or if you do a lot of data analysis, then you create so much data that it is simply not possible to create metadata for all the data you use, [...] At this point, we must create metadata, but we are still lacking good solutions for doing so that are also cost-efficient. (Environmental researcher, interview)

This excerpt illustrates how environmental researchers had to balance the requirements and benefits of creating metadata against the cost of making it, and it is a pressing issue with increased amounts of data.

Another issue with distributed sharing of scientific data was the lack of a common platform where the data is shared. ”Some data sharing is currently happening, but a lot of the data is 'hidden,' you need to know about its existence to find it” (Software developer, seminar notes). The work to find scientific data was described as time-consuming, and a vital part of facilitating data sharing was to make data findable and interoperable through a common platform where scientific data is shared.

5.5 Data Sharing Conflicts With the Present Research Culture

One issue with the sharing of scientific data described by the informants was the *lack of incentive for the researchers* that collected and processed the data. ”Another key issue is that there is currently no proper system for crediting the source of the dataset” (Sofware developer, interview). There was no proper system to give credit to the researchers sharing the data, such as the system for citing research articles. The researcher that collected the data also had to put extra work into the data curation process and create high-quality metadata to make it understandable, trustworthy, and interoperable for other researchers that wanted to reuse the data. The lack of compensation for the extra work required to make data suitable for reuse and the lack of incentives for the researcher that is sharing the scientific were recurring problems with data sharing among the environmental researchers.

Sharing of scientific data has been *predominantly required by research councils* and other institutions that provide research funding as part of the requirement to receive financing. The participants offered different explanations for why data sharing has been uncommon, such as lack of incentive for the researcher sharing the data, lack of platforms for sharing of scientific data, and it required extra work. Some of the informants described it as a cultural issue: ”Data sharing requires a cultural and technical shift” (Sofware developer, interview). The cultural argument was used to describe how data sharing were atypical, and that the benefits of data sharing have been neglected. The software developers focused on embracing the benefits of data sharing, for example, save resources by preventing redundant data collection and make data more available.

Another obstacle mentioned with data sharing was the *conflicting interests between different actors*. Research councils and other governing institutions have encouraged the sharing of scientific data through guidelines and requirements to receive funding. Some informants viewed data sharing as a benefit to the research community, and as a step to make research data more accessible. Other informants expressed concerns about the negative consequences of sharing data.

> There is also a trade-off that if you share the data, others may take your data and publish the paper you want to publish based on that data. So there is some anxiety among some colleagues that they are afraid that they spend much time collecting the data, and they only want to share it with other people once they are done analyzing it. (Environmental researcher, interview)

The research environment is competitive, and data sharing enables other researchers to compete against and challenge the work of the person sharing the data. Therefore it can potentially harm the work of the researchers that have collected, processed, and shared it. The informants also expressed concerns about legal issues related to licensing of data when using data from multiple sources.

5.6 Driving Forces Behind the Adoption of IoT Technology and Policies

The different driving forces behind the adoption of IoT and policies described by the participants were grouped into two conceptual categories: *Adoption to receive funding* and *adoption to solve a problem*. *Adoption to receive funding* were motivated by requirements and policies by research councils and other governing bodies that provide research financing.

> You can say that the main driver is economics. For example, the research council says all the data you collect and use in projects for the research council has to be made public, this is a regulation from the research council, but you do not have to do projects for the research council, but that is a lot of money for research that is available there. So if you want to get the money, then you have to publish the data. (Environmental researcher, interview)

This excerpt illustrates how research councils have encouraged the adoption of data sharing through requirements for research funding. Strategy documents by ESFRI and The Research Council of Norway also showed an increased focus on technological adoptions and data sharing.

Adoption to solve a problem consisted of the adoption of technology to improve existing work methods or develop new and improved work methods. The environmental researchers described how scripts, sophisticated algorithms, and machine learning is used to optimize and automate the data processing and data analysis at RIs, and handle more massive datasets. Another example of adoption to solve a problem was the use of IoT for monitoring sea-life, wildlife camera traps, and GPS tracking of animals to save time and resources during the data collection, and allow the researchers to collect larger datasets.

6 Discussion

This chapter discusses the findings in the context of the overarching theoretical framework, socio-technical focus, and research questions. It also addresses other observations of interest from the case study. There is one subsection dedicated to each research question and sub-question.

6.1 Adoption of IoT for Environmental Monitoring

Previous research regarding the use of IoT for environmental monitoring is limited. A majority of the research focuses on the use of IoT from a technical perspective (Truong et al., 2017; Park and Seo, 2017), while articles with a socio-technical focus are scarce (Parmiggiani et al., 2015). For this purpose, in this study, the author has addressed this research question: *How is IoT used at different environmental research infrastructures?* The findings described how the use of IoT for environmental monitoring is uncommon or in an early adoption-phase. However, IoT has been used in specific fields such as monitoring marine life, animal tracking, and wildlife camera traps.

One challenge described by the environmental researchers was the cost and resources required to conduct physical observations, and install and collect conventional sensors at regular time intervals. For example, installing and collecting sensors at the seafloor, tracking animals manually, and capturing or sedating animals to place and retrieve sensors. IoT sensors required a more substantial resource investment to install, and the sensors were more costly than conventional sensors. However, once installed, the sensors required minimal maintenance during their lifespan, and it gave the researchers access to a continuous stream of data. This trade-off between a more considerable start-up investment in return for the potential of saving time and resources during the sensors' lifespan coincides with previous research into the use of IoT and sensors for environmental monitoring (Monteiro and Parmiggiani, 2019; Borgman et al., 2012; Parmiggiani et al., 2015). Parmiggiani et al. (2015) look at the work to operate and install a sensor network for marine monitoring by an oil company and addressed the extra work required to install the sensors and the advantage of easy access to scientific data once the sensors are installed. The extra worked required by the researchers and the benefits with IoT are reflected in the analytical framework presented in Table 2.

Another difficulty was the harsh environment where the sensors were placed over more extended periods (e.g., deep in the forest, attached to animals, and on the seafloor). The sensors risked getting damaged and lost, resulting in the loss of all the research data. ”You are not always able to recover the sensors that you put out. [...] If none of the data must get lost, then people would use techniques where they send it over” (Environmental researcher, interview). The risk of losing all the scientific data if the sensors were damaged or lost was another reason for why the researchers have started to use IoT.

These challenges describe the motivation behind the adoption of IoT within specific environmental research fields, but they do not explain why IoT has been atypical in other fields. Some scientific data is difficult to collect with sensors (e.g., soil samples, blood samples, and measuring the size of animals). Researchers working with more sensor-based data collection also described IoT as uncommon. One explanation for why IoT has been rarely used is the conflicting perceptions among the informants described in chapter 5.2. Conflicting Perceptions of IoT Among Researchers. The informants working with IoT described it as cheap and accessible and highlighted how it could be used to improve the data collection, while many of the environmental researchers perceived it as expensive and complicated to use.

These conflicting perceptions could be caused by the informants’ cognitive model and how they associate new technology with their existing knowledge (Orlikowski, 1992; Tyre et al., 2018; Gash and Orlikowski, 1991). Orlikowski (1992) looks at the implementation of a knowledge sharing software in a consulting firm. She observed that the consultants had no prior experience with the technology, and sharing knowledge was atypical among the consultants. Therefore they struggled with understanding the purpose and benefits of the new system. She argued that humans have a cognitive model of the world, and we try to relate new experiences to previous knowledge. Environmental researchers that had little to no experience with IoT technology might struggle to understand the benefits and purpose with the technology and instead prefer more traditional data collection methods. This research project did not conduct a cost-benefit analysis of using IoT compared to sensor-based or non-sensor-based data collection. Therefore, it is difficult to determine if the technology is expensive or affordable, but empirical data showed that IoT was utilized in some environmental research areas, and the informants described how it was used to save time and resources.

Governing institutions and their guidelines for RIs to receive funding were highlighted as the main drivers for technological adoption at RIs by the participants. Kaltenbrunner (2017), Ribes and Lee (2010), and Kee and Browning (2010) describe how these policies shape the design of the RIs and impact researchers' day-to-day work. Karasti et al. (2010, 2006) criticize the guidelines for being limited, and primarily focused on technical aspects of the RIs; meanwhile, the socio-technical relations have been neglected. The document analysis of strategy documents corroborated the technical focus of these guidelines and policies. The environmental researchers explained how an increased focus on sensor-based data collection and IoT by governing bodies would gradually change the data collection methods to become more sensor-based.

6.2 The Importance of Data Curation

The environmental researchers emphasized the importance of data curation when working with large datasets or a continuous stream of data from IoT sensors. They described the need for a thoughtful plan on how the research data is processed, stored, and analyzed, and how sophisticated algorithms and machine learning is used to automate and optimize these processes. Data curation describes how scientific data is handled through its life cycle at a RIs (Lord and Macdonald, 2003), and the researchers rely on knowledge about the data curation process to establish trust and understand the research data (Karasti et al., 2006; Millerand and Baker, 2010). Sensor-based data collection, IoT, and the use of advanced algorithms for handling data makes understanding how the data is transformed through its life cycle essential to assess its trustworthiness (Leonelli, 2016). Therefore, the author decided to focus on this sub-question: *How do researchers assess the trustworthiness of data?*

The RIs required processing power and storage capacity, capable of handling extensive datasets and resource-intensive algorithms. With the adoption of IoT sensors, the RIs also need an infrastructure to transmit and receive the research data from the sensors. The RIs need to be continuously evolving to meet the needs of new technologies for data collection and data processing and new requirements from governing institutions (Pollock and Williams, 2010; Mongili, 2014). The data curation process at a RI extends beyond the researchers' work to steward the data, and also include the technical infrastructure that facilitates it (Karasti, 2009; Leonelli, 2016).

Distributed sharing of scientific data was another area of interest that highlighted the importance of data curation. The software developers described how established standards for data curation could be used as a framework to support the sharing of scientific data by making it easier to understand and assess the data's trustworthiness. They focused on standards for processing the data after it was collected and standards for storing the research data to make it less complicated to understand the context of how the data was created. However, the informants did not elaborate on how to deal with the problems connected to the use of standards presented in the literature. Finding a balance between using standards and the need for flexibility is difficult (Hanseth et al., 1996; Monteiro et al., 2013). The environmental research field is diverse, and previous attempts at introducing standards for data curations have struggled to accommodate the flexibility required by the researchers (Karasti et al., 2010; Karasti, 2009). Even though the use of standards for data curation might make it easier for outside researchers to reuse the research data, it is necessary to find a balance between standards and flexibility to be successful.

The creation of metadata is another crucial part of the data curation process. The informants described how metadata allows the researchers to capture the socio-technical context of how the data was collected, processed, and stored. For example, calibration and types of sensors used, algorithms and methods used to process the data, and the format for storing the data. Knowledge about the socio-technical context of scientific data is essential for researchers that want to reuse the data (Zimmerman, 2008; Arzberger et al., 2004b), making it uncomplicated for researchers to understand the research data and establish trust (Baker et al., 2000; Hobbie et al., 2003; Karasti et al., 2006). Empirical data from the case study corroborated the significance of metadata to support data sharing.

As illustrated in Figure 2, the environmental researchers working with IoT, focused primarily on how data curation could be used to improve data collection with IoT sensors. Data collection and processing are becoming more automated, and datasets are increasing in size. Due to this, a good plan for the data curation process and an in-depth understanding of how the data was handled at the RIs, were essential for the researchers to understand the data and assess its trustworthiness. The informants working with data sharing depended on data curation to make it easier for other researchers to evaluate the data's integrity and understand it. Both perspectives on data curation emphasized how researchers can use data curation to assess the trustworthiness of research data.

6.3 Extra Work Required by the Researchers

A common misconception when looking at the adoption of new technologies and information systems is viewing the adoption process as a planned organically growing process where ”good” solutions are successful. This mindset neglects the micro-events and the work done by the users to ensure the implementation is successful (Monteiro and Hepsø, 1998; Tyre et al., 2018; Ciborra, 1997). The work to install and operate IoT sensors for data collection, and the work as part of the data curation process falls on the environmental researchers. For this purpose, in this study, the author has addressed this sub-question: *How does IoT affect the daily work of the researchers?*

The findings illustrated how the use of IoT for data collection could be more resource-intensive and time-consuming to install the sensors. Once the sensors were installed, the environmental researchers described it as more beneficial than conventional sensors and non-sensor-based data collection. Even though IoT sensors might save resources for the researchers during the lifespan of the RI, they require higher proficiency for the users to install and operate the sensors. Environmental researchers have to spend extra time learning how to use IoT. The data curation connected to the use of IoT also causes similar issues with extra work and expertise required to process and handle the continuous stream of data.

The creation of metadata was another source of tension between the researchers and the requirements for metadata. The informants focused on how it benefits other researchers that want to reuse the research data, but they pointed out that creating metadata can be time-consuming and often holds little personal benefit to the researchers that collected the data. This lack of incentive for the researchers has led to the value of metadata often being neglected by the researchers. Articles into the use of software for collaboration and information sharing have highlighted the importance of creating an incentive for all users to increase the likelihood of user approval for the system. (Grudin, 1989; Orlikowski, 1992). Grudin (1989) addresses how uneven balance between work and benefit caused much resistance from users.

Sharing of scientific data is another area where the balance of work and benefit was problematic. The informants described how creating high-quality metadata and sharing data require extra work by the researcher, but there are no established systems in place for crediting research data. The research environment is competitive, and some of the participants expressed concerns about providing other researchers with their research data, and enable them to compete with their work. Orlikowski (1992) encountered similar challenges in the competitive environment within a consulting firm.

Research councils and other governing institutions incentivize technological adoptions and data sharing through research funding. The case study's empirical data illustrated how researchers had to balance how much time and resources they dedicated to making the research data understandable and trustworthy for other researchers.

> I think it is very important that you document what your results represent, because if you get the data on its own, then it can be extremely troublesome to figure out what it represents by yourself. In some situations, it is easier; in other situations, it is impossible. (Environmental researcher, interview)

The extra work involved creating high-quality metadata and documenting the work. Previous research supports the notion that data sharing requires more than only sharing the data itself (Zimmerman, 2003; Helly et al., 2002; Borgman et al., 2012), and the quality of the shared data depends on the work conducted by the researchers (Newman et al., 2003; Zimmerman, 2008). Governing institutions (e.g., ESFRI) seemed inadequate when it came to incentivizing high-quality data sharing and compensating the extra work required by the researchers.

6.4 IoT's Impact on Data Sharing

Sharing of scientific data for distributed collaboration or reuse is a recurring theme in previous research that looks at RIs for environmental monitoring (Hobbie et al., 2003; Karasti et al., 2010), data curation at RIs (Karasti et al., 2006; Parmiggiani and Grisot, 2020; Leonelli, 2016), and different methods for data collection (Borgman et al., 2012). Therefore, the author decided to include the research question: *How is the increased focus on data sharing affected by IoT?* The results of the case study substantiate how data sharing affects most research projects, primarily through requirements to receive financing and organizational policies. This increased focus on distributed sharing of research data is also reflected in the analytical framework developed by the author, as presented in the themes and conceptual categories in Table 2.

The informants described the growing importance of data curation with the adoption of IoT and more sensor-based data collection, and how data curation could serve as a framework to empower data sharing, as presented in chapter 6.2. The Importance of Data Curation. The increased focus on data curation to support technological advancements such as IoT (Leonelli, 2016; Karasti, 2009), aligns with some of the sub-processes in data curation to support distributed sharing of research data, such as the creation of metadata, capturing the social context of the data, and increasing focus on established practices.

The issues related to the trustworthiness of the data when the data is shared is not a unique problem to the environmental research field. Researchers have studied data sharing within different research areas, and work environments and problems with establishing trust are recurring (Luhmann, 2000; Jirotka et al., 2005; Van House, 2002). Jirotka et al. (2005) look at the sharing of mammograms between breast cancer clinics. They describe how the workers often relied on information about the social context of how the pictures were captured to interpret them, and that the pictures held little value outside the social-context where it was captured.

Another issue with data sharing was the need for a cultural and technical change among researchers. It is difficult for IoT to accommodate the need for a cultural change. However, the participants described how IoT technology often relied on cloud storage to transmit the data between the sensors and the researchers. The platform for cloud storage made use of access control to restrict user access to the data. The system engineers presented how the platform for cloud storage could be used as a platform for data sharing with simple access control modifications. The lack of a common platform to facilitate the sharing of research data has been a crucial obstacle with the notion of data sharing (Zimmerman, 2008; Arzberger et al., 2004b).

7 Conclusion

This thesis provides insight into the convoluted socio-technical environment present at RIs for environmental monitoring, and how the adoption of IoT influences it. The work of the author is by no means exhaustive. Still, it provides new empirical knowledge about the role of data curation, distributed data sharing in practice, and different methods for data collection, while looking at the complex institutional, technological, and socio-technical challenges affecting these themes. The importance of these themes is also reflected in the analytical framework presented in Table 2, and the overarching theoretical framework.

The use of IoT technology for environmental monitoring was in an early adoption phase at the time of this case study. IoT was primarily used by researchers to save time and resources in environmental research fields, where physical observations and conventional sensors were inefficient (e.g., using acoustic sensors to monitor sea life). The findings describe conflicting perceptions of IoT among the participants. The informants working with IoT described it as cheap and beneficial to their work. Meanwhile, the environmental researchers with little to no experience with the technology described it as expensive. The author proposed that one reason for the limited adoption of IoT might be the researchers' inability to understand the benefits and use cases for IoT, due to lack of experience with the technology.

More sensor-based and automated data collection, and advanced algorithms and more automation for data processing have led to an increased focus on data curation. Good routines for data curation and understanding the data curation process is vital to understand the research data and assess its trustworthiness. Furthermore, established practices for data curation and high-quality metadata can make it easier for outside researchers to understand the research data and establish trust.

Governing bodies, such as The Research Council of Norway, focus on data sharing and the technical infrastructure present at RIs through their policies. The findings of this thesis, as illustrated in Figure 2, describe how data curation can improve the use of IoT and how distributed data sharing depends on it to ensure the quality of the shared scientific data. The data curation process was described as fragmented, and the research data's quality was affected by the time and resources the researchers spent on the data curation. Recognizing the importance of data curation for data sharing and technologies, such as IoT, implementing data curation into the governing bodies' policies for RIs, can have a positive impact on the policies mentioned above, and the fragmented situation for data curation present at RIs.

Utilizing new technologies for data collection, increased focus on data curation, and creating metadata can be beneficial to the work conducted by the researchers. However, it often requires extra work by the researchers. The environmental researchers had to balance personal interests against the policies by governing institutions and interests of other researchers. The lack of incentives, time, and resources were recurring issues in their day-to-day work. On the face of it, it seems like these problems can be solved by increased financing. However, these are complex problems that require technical solutions, a cultural change among researchers, and support by governing bodies through policies and funding.

7.1 Limitations and Future Work

The author recognizes that the thesis is affected by some limitations. One fundamental limitation is the time constraint; this case study was limited to one semester. Therefore, the scope of this project was limited geographically and only included researchers working in Norway. Hence, the findings may not be representative for RIs in other countries. The time constraints also affected the number of participants the author was able to recruit for the interviews, and more time would have allowed for a higher degree of saturation.

Another limitation on the empirical data collection was the problems caused by the Covid-19 virus. The data collection had to be limited to only include interviews and observations from seminars, due to restrictions on physical meetings. Field studies with observations of the daily work conducted by environmental researchers could have provided a better understanding of the complex environment present at RIs.

In closing, the work presented in this thesis leaves room for future research. Future potential research projects that the author thinks are the most interesting, are described below.

First, conducting field studies at RIs to observe current practices for data collection and the use of IoT as initially planned by this thesis. Making physical observations of the data collection, the data curation, and the researchers' daily work could provide a better understanding of the use of IoT at RIs.

Another path to pursue is looking at existing data curation practices, and the efforts to establish standards for data curation at RIs. Previous research often focuses on the tension between standards and flexibility, but there have been some technological developments that attempt to overcome this issue.

Finally, data sharing has been encouraged by governing institutions for a long time. However, empirical data illustrated data sharing as fragmented and uncommon. Analyzing existing solutions for data sharing to understand the problem could be another path to pursue.

References

Angelakis, V., Tragos, E., Pöhls, H., Kapovits, A., and Bassi, A. (2017). *Designing, Developing, and Facilitating Smart Cities: Urban Design to IoT Solutions*.

Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., Moorman, D., Uhlir, P., and Wouters, P. (2004a). An international framework to promote access to data. *Science*, 303(5665):1777–1778.

Arzberger, P. W., Schroeder, P., Beaulieu, A., Bowker, G. C., Casey, K., Laaksonen, L., Moorman, D., Uhlir, P., and Wouters, P. (2004b). Promoting access to public research data for scientific, economic, and social development. *Data Science Journal*, 3:135–152.

Atkins, D., Droegemeier, K., Feldman, S., Garcia-Molina, H., Klein, M., Messerschmitt, D., Messina, P., Ostriker, J., and Wright, M. (2003). Revolutionizing science and engineering through cyberinfrastructure. *Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure, January*.

Baker, K. S., Benson, B. J., Henshaw, D. L., Blodgett, D., Porter, J. H., and Stafford, S. G. (2000). Evolution of a multisite network information system: The lter information management paradigm. *BioScience*, 50(11):963–978.

Baxter, P. and Jack, S. (2008). Qualitative case study methodology: study design and implementation for novice researchers.(report). *The Qualitative Report*, 13(4):544.

Birnholtz, J. and Bietz, M. (2003). Data at work: supporting sharing in science and engineering. In *Proceedings of the 2003 international ACM SIGGROUP conference on supporting group work*, GROUP '03, pages 339–348. ACM.

Boos, D., Guenter, H., Grote, G., and Kinder, K. (2013). Controllable accountabilities: the internet of things and its challenges for organisations. *Behaviour & Information Technology*, 32(5):449–467.

Borgman, C., Wallis, J., and Mayernik, M. (2012). Who's got the data? interdependencies in science and technology collaborations. *Computer Supported Cooperative Work (CSCW)*, 21(6):485–523.

Bossen, C., Chen, Y., and Pine, K. H. (2019). The emergence of new data work occupations in healthcare: The case of medical scribes. *International Journal of Medical Informatics*, 123:76–83.

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2):27–40.

Brown, C. (2003). The changing face of scientific discourse: Analysis of genomic and proteomic database usage and acceptance. *Journal of the American Society for Information Science and Technology*, 54(10):926–938.

Băjenescu, T.-M. I. (2018). From the internet of things (iot) to smart cities (sc). *Journal of Engineering Science (Chişinău)*, XXV(3):85–94.

Chin, Jr, G. and Lansing, C. (2004). Capturing and supporting contexts for scientific data sharing via the biological sciences collaboratory. In *Proceedings of the 2004 ACM conference on computer supported cooperative work*, CSCW '04, pages 409–418. ACM.

Ciborra, C. U. (1997). De profundis? deconstructing the concept of strategic alignment. *Scand. J. Inf. Syst.*, 9:2.

Council, N. R. (1997). *Bits of Power: Issues in Global Access to Scientific Data*. The National Academies Press, Washington, DC.

Dourish, P. (2016). *The Internet of Urban Things*, chapter 3, pages 27–48. Routledge, New York, NY.

ESFRI (2018). Strategy report on research infrastructures. http://roadmap2018.esfri.eu/. (Visited 16. April 2020).

European Commission (2019). What are research infrastructures? https://ec.europa.eu/info/research-and-innovation/strategy/european-research-infrastructures_en. (Visited 26. May 2020).

Forskningsrådet (2018). Norsk veikart for forskningsinfrastruktur 2018. https://www.forskningsradet.no/siteassets/programmer/veikart/norskveikartforforskningsinfrastruktur2018-del1omradestrategier.pdf. (Visited 20. April 2020).

Gash, D. C. and Orlikowski, W. J. (1991). Changing frames: Towards an understanding of information technology and organizational change. In *Academy of Management Proceedings*, volume 1991, pages 189–193. Academy of Management Briarcliff Manor, NY 10510.

Govoni, M., Michaelis, J., Morelli, A., Suri, N., and Tortonesi, M. (2017). An information-centric platform for social- and location-aware iot applications in smart cities. *EAI Endorsed Transactions on Internet of Things*, 3(9).

Greif, I. and Sarin, S. (1987). Data sharing in group work. *ACM Transactions on Information Systems (TOIS)*, 5(2):187–211.

Grudin, J. (1989). Why groupware applications fail: problems in design and evaluation. *Office, Technology and People*, 4(3):245.

Hanseth, O., Monteiro, E., and Hatling, M. (1996). Developing information infrastructure: The tension between standardization and flexibility. *Science, Technology & Human Values*, 21(4):407–426.

Hedstrom, M. (2004). It's about time: research challenges in digital archiving and long term preservation, i. *Microform & imaging review*, 33(1):15–22.

Helly, J., Elvins, T., Sutton, D., Martinez, D., Miller, S., Pickett, S., and Ellison, A. (2002). Controlled publication of digital scientific data. *Communications of the ACM*, 45(5):97–101.

Hey, A. J. G. and Trefethen, A. E. (2003). The data deluge: An e-science perspective. In Berman, F., Fox, G. C., and Hey, A. J. G., editors, *Grid Computing - Making the Global Infrastructure a Reality*, pages 809–824. Wiley and Sons. Chapter: 36.

Hilgartner, S. and Brandt-Rauf, S. I. (1994). Data access, ownership, and control: Toward empirical studies of access practices. *Knowledge*, 15(4):355–372.

Hobbie, J. E., Carpenter, S. R., Grimm, N. B., Gosz, J. R., and Seastedt, T. R. (2003). The us long term ecological research program. *BioScience*, 53(1):21–32.

ISO/IEC JTC 1 (2014). Internet of Things (IoT) Preliminary Report 2014. `https://www.iso.org/files/live/sites/isoorg/files/developing_standards/docs/en/internet_of_things_report-jtc1.pdf`. (Visited 18. November 2019).

Jirotka, M., Procter, R., Hartswood, M., Slack, R., Simpson, A., Coopmans, C., Hinds, C., and Voss, A. (2005). Collaboration and trust in healthcareinnovation: The ediamond case study. *Computer Supported Cooperative Work (CSCW)*, 14(4):369–398.

Kaltenbrunner, W. (2017). Digital infrastructure for the humanities in europe and the us: Governing scholarship through coordinated tool development. *Computer Supported Cooperative Work (CSCW)*, 26(3):275–308.

Karasti, H. (2009). Finltser network formation and the state of information management. *Ecological Circuits - Eco-informatics in Action*, pages 15–18.

Karasti, H., Baker, K., 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 (CSCW)*, 15(4):321–358.

Karasti, H., Baker, K., and Millerand, F. (2010). Infrastructure time: Long-term matters in collaborative development. *Computer Supported Cooperative Work (CSCW)*, 19(3-4):377–415.

Kee, K. and Browning, L. (2010). The dialectical tensions in the funding infrastructure of cyberinfrastructure. *Computer Supported Cooperative Work (CSCW)*, 19(3-4):283–308.

Klein, H. K. and Myers, M. D. (1999). A set of principles for conducting and evaluating interpretive field studies in information systems. *MIS Quarterly*, 23(1):67–93.

Kollerud, S. (2019). Iot and data curation in long-term environmental monitoring. Project report in TDT4501, Department of Computer Science, NTNU – Norwegian University of Science and Technology.

Kwa, C. (2005). Local ecologies and global science: Discourses and strategies of the international geosphere-biosphere programme. *Social Studies Of Science*, 35(6):923–950.

Leonelli, S. (2013). Global data for local science: Assessing the scale of data infrastructures in biological and biomedical research. *BioSocieties*, 8(4):449.

Leonelli, S. (2016). *Data-centric biology: A philosophical study*. University of Chicago Press.

Lord, P. and Macdonald, A. (2003). e-science curation report. `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.96.5156&rep=rep1&type=pdf`. (Visited 26. November 2019).

LTER-Europe (2019). What are research infrastructures? `https://www.lter-europe.net/lter-europe`. (Visited 15. November 2019).

Luhmann, N. (2000). Familiarity, confidence, trust: Problems and alternatives. *Trust: Making and breaking cooperative relations*, 6(1):94–107.

Millerand, F. and Baker, K. S. (2010). Who are the users? who are the developers? webs of users and developers in the development process of a technical standard. *Information Systems Journal*, 20(2):137–161.

Mongili, A. (2014). *Information Infrastructure(s): Boundaries, Ecologies, Multiplicity.* 1st ed.. edition.

Monteiro, E. and Hepsø, V. (1998). Diffusion of information infrastructure: mobilization and improvization. *IFIP*, pages 255–273.

Monteiro, E. and Parmiggiani, E. (2019). Synthetic knowing: The politics of the internet of things. *arXiv.org*, 43(1).

Monteiro, E., Pollock, N., Hanseth, O., and Williams, R. (2013). From artefacts to infrastructures. *Computer Supported Cooperative Work (CSCW)*, 22(4-6):575–607.

National Research Council (1995). *Finding the Forest in the Trees: The Challenge of Combining Diverse Environmental Data*. The National Academies Press, Washington, DC.

Newman, H., Ellisman, M., and Orcutt, J. (2003). Data-intensive e-science frontier research. *Communications of the ACM*, 46(11):68–77.

Oates, B. J. (2006). *Researching Information Systems and Computing*. Sage Publications Ltd.

Orlikowski, W. J. (1992). Learning from notes: Organizational issues in groupware implementation. In *Proceedings of the Conference on Computer-Supported Cooperative Work*, pages 362–369. Publ by ACM.

Park, Y. and Seo, J. (2017). Development of an iot-assisted environmental data monitoring system for agricultural applications. *Information (Japan)*, 20(8):5679–5684.

Parmiggiani, E. (2015). *Integration by Infrastructuring: The Case of Subsea Environmental Monitoring in Oil and Gas Offshore Operations*. PhD thesis.

Parmiggiani, E. and Grisot, M. ((forthcoming) 2020). Data curation as governance practice. *Scandinavian Journal of Information Systems*.

Parmiggiani, E., Monteiro, E., and Hepso, V. (2015). The digital coral: Infrastructuring environmental monitoring.(report). *Computer Supported Cooperative Work (CSCW)*, 24(5):423.

Passi, S. and Jackson, S. (2020). Trust in data science: Collaboration, translation, and accountability in corporate data science projects. *arXiv.org*, 2(CSCW).

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 (CSCW)*, 19(6):521–556.

Ribes, D. and Lee, C. P. (2010). Sociotechnical studies of cyberinfrastructure and e-research: Current themes and future trajectories.(editorial). *Computer Supported Cooperative Work (CSCW)*, 19(3-4):231.

Ribes, D. and Polk, J. (2014). Flexibility relative to what? change to research infrastructure. *Journal of the Association for Information Systems*, 15(5):287–305.

Romero, C., Barriga, J., and Molano, J. (2016). Big data meaning in the architecture of iot for smart cities. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 9714, pages 457–465. Springer Verlag.

Stein, L. D. (2008). Towards a cyberinfrastructure for the biological sciences: progress, visions and challenges. *Nature Reviews Genetics*, 9(9):678.

Sundmaeker, H., Guillemin, P., and Friess, P. (2010). *Vision and challenges for realising the Internet of things.* Publications Office, Luxembourg.

Tjora, A. H. (2019). Qualitative research as stepwise-deductive induction.

Truong, T., Dinh, A., and Wahid, K. (2017). An iot environmental data collection system for fungal detection in crop fields. Institute of Electrical and Electronics Engineers Inc.

Tyre, M., for, S., and Orlikowski, W. (2018). *Windows of Opportunity–Creating Occasions for Technological Adaptation in Organizations*. Creative Media Partners, LLC.

UK Research Council e-Science definition (2001). e-science definition. https://webarchive.nationalarchives.gov.uk/20080906044009/http://www.rcuk.ac.uk/escience/default.htm. (Visited 25. May 2020).

Van House, N. A. (2002). Digital libraries and practices of trust: Networked biodiversity information. *Social Epistemology*, 16(1):99–114.

Walsham, G. (1995). Interpretive case studies in is research: nature and method. *European Journal of Information Systems*, 4(2):74–81.

Walsham, G. (2006). Doing interpretive research. *European Journal of Information Systems: Including a Special Section on Mobile User Behaviour*, 15(3):320–330.

Weiser, M. (1991). The computer for the 21 st century. *Scientific American*, 265(3):94–105.

Wen-Tsai, S. and Chia-Cheng, H. (2013). Iot system environmental monitoring using ipso weight factor estimation. *Sensor Review*, 33(3):246–256. Copyright - Copyright Emerald Group Publishing Limited 2013; Last updated - 2019-09-06; CODEN - SNRVDY.

Willig, M. R. and Walker, L. R. (2015). *LONG-TERM ECOLOGICAL RESEARCH CHANGING THE NATURE OF SCIENTISTS*. Oxford University Press, Oxford, U.K.

Yang, C., Puthal, D., Mohanty, S. P., and Kougianos, E. (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):48–56.

Zhou, J., Cao, Z., Dong, X., and Vasilakos, A. V. (2017). Security and privacy for cloud-based iot: Challenges. *IEEE Communications Magazine*, 55(1):26–33.

Zimmerman, A. S. (2003). Data sharing and secondary use of scientific data: Experiences of ecologists.

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):631–652.

Appendix

A Interview Guide

Interview Guide

About the project:

The project looks at the use of IoT-sensor networks for environmental monitoring. The project aims to contribute with empirical insight into how IoT is utilized at research infrastructures today. We are also interested in understanding how the researchers cooperating with each other, and with other institutions. We want to get a better insight into the actual work that researchers carry out to make use and adapt the technology.

In addition, we are also interested in studying how the data collected is assessed as trustworthy, while also looking and the benefits and challenges with sensor networks.

It is voluntarily to participate in the project. We are not interested in the actual name of people and institutions, but only in ways of working with environmental monitoring. All the data collected will be anonymized and stored in an NTNU-approved secure folder.

Practical information, consent

Work area of the participants:

Topic 0 – background

Can you tell us a bit about yourself, and your background?

What do you do as part of your (daily) job?

What type of sensors do you work with?

Topic 1 – data collection

Is data collection mostly sensor based?

To what degree is IoT-sensor networks utilized?

Why is the current method for data collection preferred?

Do you collect a lot of parameters and why do you collect them?

Topic 2 – Benefits and challenges with data collection

What are the challenges with the current method for data collection?

What are the benefits?

Topic 3 – Data curation

What measures are taken to ensure the data collected are understandable?

Is it understandable outside the social context?

What measures are taken to ensure the data is trustworthy?

Error detection? Error handling?

Is the data curation process time consuming?

What happens with larger amounts of data?

Topic 4 – data sharing

Do you share the data collected with other researcher institutions or outside researchers?

How is it shared?

Challenges with sharing the data?

Is it understandable to outside researchers?

Topic 5 – Driving forces behind the adoption of new technology

What do you think are the driving forces behind the adoption of new technology?

NTNU

Norwegian University of
Science and Technology