UNIVERSITY
OF TURKU

GOVERNANCE OF PLATFORM DATA

From Canonical Data Models to Federative Interoperability

Tiina Nokkala

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. E OSA – TOM. 63 | OECONOMICA | TURKU 2020

UNIVERSITY OF TURKU

GOVERNANCE OF PLATFORM DATA

From Canonical Data Models to Federative Interoperability

Tiina Nokkala

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. E OSA – TOM. 63 | OECONOMICA | TURKU 2020

University of Turku

Turku School of Economics
Management and Entrepreneurship
Information Systems Science
Doctoral programme of Turku School of Economics

Supervised by

Professor of Practice Tomi Dahlberg
Turku School of Economics
University of Turku
Finland

Professor Jukka Heikkilä
Turku School of Economics
University of Turku
Finland

Reviewed by

Professor Boris Otto
Technische Universität Dortmund
Germany

Professor Samuli Pekkola
Tampere University
Finland

Opponent

Professor Boris Otto
Technische Universität Dortmund
Germany

The originality of this publication has been checked in accordance with the University of Turku quality assurance system using the Turnitin OriginalityCheck service.

Cover Image: Tarja Kankaala

ISBN 978-951-29-8132-8 PRINT
ISBN 978-951-29-8133-5 PDF
ISSN 2343-3159 (Painettu/Print)
ISSN 2343-3167 (Verkkojulkaisu/Online)
Painosalama Oy, Turku, Finland 2020

Sievin mummulle

UNIVERSITY OF TURKU
Turku School of Economics
Management and Entrepreneurship
Information Systems Science
TIINA NOKKALA: Governance of Platform Data – From Canonical Data Models to Federative Interoperability
Doctoral Dissertation, 180 pp.
Doctoral Programme of Turku School of Economics
September 2020

ABSTRACT

As the volume of data generated every day is constantly increasing and, at the same time, ever more complex business networks are using this voluminous data, there is a clear need for better data governance. So far, the academic and practical literature has focused mainly on data governance for intra-organisational purposes. However, in the era of multifaceted business networks and with a rising number of data-driven platforms, the scope of data governance needs to be widened to address inter-organisational contexts. From a practical point of view, data on the same subjects are scattered across numerous information systems and attempts to integrate them are often unsuccessful. Thus, new approaches are needed.

The object of this dissertation is to study data governance in platform contexts. The goal is to use the existing data governance frameworks as a basis for creating a new framework that encompasses aspects of the networked business and network and platform business models, and the specific features related to data sharing on platforms.

Three independent qualitative case studies were conducted to collect empirical evidence for this dissertation. The first case was about data federation in breast cancer patient and treatment data. The other two cases were conducted in the maritime industry to find out how data should be governed on platforms and what aspects affect the willingness of participating organisations to share data on platforms. The qualitative data consists of interviews with 29 people, material from several workshops, and group discussions and interview journals kept during the data collection. The theory building is based on the existing literature on data governance, networks and platforms and on the results of the case studies conducted.

The key theoretical contributions of this dissertation are threefold. First, a federative approach to data interoperability is presented, with related tools. The federative approach enables preserving the original and other contexts of data and is based on the use of metadata to explain the various meanings of data. Second, a platform data governance model that includes business model aspects for networks is proposed. This model considers data federation as a means of joining the network and the platform, and has a special focus on data access and ownership. Third and most importantly, this dissertation joins the discussion on whether data is universally or contextually defined by presenting new views on the latter position. The idea is that instead of aiming for objective definitions for data, data should be seen as the representation of facts in their contexts that affect how the facts should be seen. In data-sharing situations, there should be agreements on what data means and how it should be understood. These agreements can be stored as metadata for the data entries. There should be three types of metadata, i.e. information system (IS) technical, information processing and socio-contextual metadata, in order to give a thorough explanation of the meaning of the data.

This dissertation provides new insights into how data sharing in networked business environments could be arranged, taking into account the ownership and access issues of data, especially in the platform context. This study also widens the understanding of the ontological nature of data in more complex IS environments. So far, the academic literature on platform data governance is sparse, and that on data governance in general has focused more on data management functions and considered data to have universal definitions. On the practical side, the framework presents methods for implementing data governance policies on platforms and offers federative tools that are feasible for actual data integration.

KEYWORDS: Data, Data governance, Platforms, Networks, Data federation, Data interoperability

TURUN YLIOPISTO
Turun kauppakorkeakoulu
Johtaminen ja yrittäjyys
Tietojärjestelmätiede
TIINA NOKKALA: Tiedonhallinta digitaalisilla alustoilla – Kanonisista datamalleista federatiiviseen yhteentoimivuuteen
Väitöskirja, 180 s.
Turun kauppakorkeakoulun tohtoriohjelma
Syyskuu 2020

TIIVISTELMÄ

Syntyvän, luotavan ja tallennettavan datan määrä kasvaa yhä kiihtyvällä tahdilla, ja samanaikaisesti yhä monimutkaisemmat liiketoimintaverkostot haluavat hyödyntää näitä valtavia datamassoja. Tämän takia tarvitaan kattavampaa ja kokonaisvaltaisempaa tiedonhallintaa. Aiempi kirjallisuus on käsitellyt tiedonhallintaa pääasiassa organisaation sisäisestä näkökulmasta. Kasvavien ja monitahoistuvien liiketoimintaverkostojen, sekä erilaisten datavetoisten alustojen aikakautena tiedonhallinta tulee ulottaa koskemaan myös organisaatioiden välisiä konteksteja. Käytännön näkökulmasta tarve tiedonhallinnalle nousee samoja objekteja koskevan tiedon hajaantumisesta useisiin järjestelmiin ja organisaatioihin. Tämän hajaantuneen tiedon integraatioyritykset ovat usein epäonnistuneita ja siksi uudenlaisia lähestymistapoja tarvitaan.

Tässä väitöskirjassa tiedonhallintaa tutkitaan alustakontekstissa. Tavoitteena on olemassa olevia tiedonhallinnan viitekehyksiä lähtökohtana käyttäen luoda uusi viitekehys, joka huomioi verkostomaisen liiketoiminnan, verkosto- ja alustaliiketoimintamallit sekä alustoilla tapahtuvaan tiedonjakoon liittyvät erityispiirteet.

Väitöskirjaan sisältyy kolme itsenäistä tapaustutkimusta, joissa on kerätty tutkielman empiirinen aineisto. Ensimmäinen tapaustutkimus koskee rintasyöpäpotilaita ja heidän hoitoaan koskevan datan federaatiota. Kaksi muuta tapaustutkimusta toteutettiin meriteollisuudessa. Näiden kahden tapaustutkimuksen tavoitteena oli selvittää miten dataa tulisi hallinnoida alustoilla ja mitkä tekijät vaikuttavat organisaatioiden halukkuuteen jakaa tietoa digitaalisilla alustoilla. Laadullinen aineisto koostuu yhteensä 29 henkilöhaastattelusta, useiden työpajojen materiaalista, ryhmäkeskusteluista sekä aineistonkeruun aikana pidetyistä haastattelupäiväkirjoista. Tutkimuksen teoreettisen kontribuution rakentuminen pohjautuu siis aiempaan tiedonhallintakirjallisuuteen, kirjallisuuteen verkostoista ja alustoista sekä kolmen toteutetun tapaustutkimuksen tuloksiin.

Väitöskirjan teoreettinen kontribuutio jakautuu kolmeen osaan. Ensiksi esitellään federatiivinen lähestymistapa tiedon yhteen toimivuuteen ja tähän liittyviä työkaluja. Federatiivinen lähestymistapa mahdollistaa datan alkuperäisen ja muiden kontekstien samanaikaisen säilyttämisen sekä datan eri merkitysten ymmärtämisen metadataan perustuen. Toiseksi esitellään ehdotus tiedonhallintamalliksi digitaalisille alustoille. Tämä malli huomioi verkostomaisen liiketoiminnan erityispiirteet. Ehdotetussa tiedonhallintamallissa datafederaatio nähdään tapana liittyä dataa jakavaan verkostoon ja alustalle. Erityistä huomiota kiinnitetään datan käyttöoikeuksiin ja omistajuuteen. Kolmanneksi väitöskirja osallistuu keskusteluun datan luonteen määrittelystä: onko data universaalisti vai kontekstuaalisesti määriteltyä. Väitöskirjassa esitellään jälkimmäistä näkökulmaa puoltavia uusia tekijöitä. Pyrkimyksenä on objektiivisen datan määrittelyn sijaan nähdä datan kuvaavan tosiasioita erinäisissä konteksteissa. Nämä kontekstit vaikuttavat siihen, miten kyseiset tosiasiat tulee ymmärtää. Dataa jaettaessa on sovittava, mitä data tarkoittaa nämä kontekstit huomioiden. Nämä sopimukset voidaan tallentaa dataan liitettävänä metatietona. Datan merkityksen selvittämiseen perinpohjaisesti metatietoa tulisi olla kolmea eri tyyppiä: tietojärjestelmien teknistä, tiedon käsittelyyn liittyvää sekä sosio-kontekstuaalista metatietoa.

Tutkimuksessa esitellään uusia näkökulmia datan jakamiseen verkottuneissa liiketoimintaympäristöissä datan omistajuuteen ja saatavuuteen liittyvät tekijät huomioiden. Erityinen painopiste on alustakontekstissa. Lisäksi tutkimus lisää ymmärrystä datan ontologisesta luonteesta monimutkaisemmissa tietojärjestelmäympäristöissä. Tähänastinen akateeminen kirjallisuus tiedonhallinnasta on ollut niukkaa, keskittyen pääasiassa tiedon johtamisen eri toimintoihin. Lisäksi taustaoletuksena on ollut, että data on universaalisti määriteltyä. Käytännölliseltä kannalta väitöskirjassa esiteltävä viitekehys tarjoaa menetelmiä tiedonhallinnan toimintaperiaatteiden implementointiin digitaalisilla alustoilla. Lisäksi esitellään federatiivisia työkaluja, jotka soveltuvat käytännön dataintegraatioihin.

ASIASANAT: Data, Tiedonhallinta, Alustat, Liiketoimintaverkostot, Datafederaatio, Datan yhteentoimivuus

Acknowledgements

"It is the time you have wasted for your rose that makes your rose so important."

"It is the time I have wasted for my rose—" said the little prince, so that he would be sure to remember.

The Little Prince, Antoine de Saint-Exupéry

I have spent quite some time with this dissertation and doing research – reading, writing, interviewing, analysing, procrastinating, thinking and once again writing. However, I could have spent more time or less time doing that. Still, it does not matter now that I am here, my dissertation almost ready. What is important are the people I got to spend that time with.

The whole process actually started already in the bachelor's thesis seminar in the autumn of 2012 when I first met my supervisor-to-be, Professor of practice Tomi Dahlberg. If someone had told me then, I would never have believed the amount of discussions we would have during the following years. During supervising my bachelor's and master's theses, Tomi subtly convinced me that I should start doing more research and complete a doctoral degree as well. And I am most grateful that he did, and also so thankful for all the support and advice I have gotten from Tomi. All the projects – e.g. MDSS – you have involved me with have given me the opportunity to learn new things. Thank you for all the thought-provoking discussions we have had! I am glad to have such a wonderful supervisor and mentor in my life.

I am also thankful for my second supervisor, Professor Jukka Heikkilä. His importance to me has grown during the process, and I appreciate the time and space he has given me to do my research, still supporting me whenever needed. Your words "It is your dissertation; you must decide yourself on that" were among the best and most important advice I have got during these years.

I am very grateful for the excellent pre-examiners I had for this dissertation: Professor Samuli Pekkola and Professor Boris Otto. Thank you for the well-grounded critique and also for the kind words. Based on your feedback I was able to improve my thinking and hopefully also to express my thoughts more clearly. Professor Boris Otto, I am pleased that you have promised to act as my esteemed opponent. It was in the very beginning of this journey when I first found your research, and since then I have wanted to have a discussion with you in person.

I have been lucky enough to get a change to work at the department of Management and Entrepreneurship almost all the time I did research for this dissertation. First as a research assistant, a part-time teacher and an assistant at some projects (thank you ViBu and Digital Services). Then I have had funding from the doctoral programmes of Turku School of Economics and Mathematics and Computer Sciences (MATTI). In addition to that, I am grateful for the one-year working grant from KAUTE foundation and for a grant awarded to me by TOP foundation.

During the research projects that I have been involved in, I have had a change to visit and get acquainted with several people and companies. Special thanks to the contact persons of these projects, Anna Hammais, Jari Yli-Tolppa and Kari Sillanpää.

Research wise I have got help from many colleagues at our department and elsewhere. Professor Hannu Salmela has influenced my research very much, through the many discussions we have had over lunches, with answers to the questions I have rushed to ask him and through inviting me to participate in a research project D4Value. Thank you Hannu for all the support. During that project, I was also lucky enough to work with Dr Jouko Toivonen, doing interviews, sharing experiences and co-authoring one of the articles in this dissertation. I learnt a lot from you. I want also thank Dr Marikka Heikkilä for all the encounters we have had during the research projects and Dr Timo Leino for his support with all kinds of work-related matters and Jenni Heervä for the guidance with the practical matters of completing a degree. Thank you Professor Reima Suomi and Assistant Professor Matti Mäntymäki for your feedback in the research seminars and Assistant Professor Hongxiu Li, Dr Ville Harkke and Dr Eija Koskivaara and other colleagues at the Turku School of Economics for the discussions and encounters we have had.

Then there are the friends and colleagues that I have spent the most working hours with. Dr Kai Kimppa – without you, coffee, (Brunberg's) kisses and occasional beers it would have been intolerable. Thank you for all the discussions, laughs and tears I have got to share with you. Dr Jonna Järveläinen, it has been a pleasure to have my office next to yours. I am very happy that you are not able to say NO most of the time, as I have made use of your knowledge and friendliness so many times. Thank you, all the fellow doctoral candidates, (some already graduated, some still in the process) I have had the pleasure to study with, it has been much nicer in good company. Special thanks to Dr Altti Lagstedt, Dr Sanna Ilonen and Ari Helin for providing excellent peer support during the last steps of this journey.

Still, most importantly I want to thank my family and friends outside the academia. Thank you for bearing with me in my best and worst moments, for listening to my joys and frustrations and taking my attention elsewhere when necessary. Noora, Lotta, Marjukka and Dr Anu, thank you for always being there. My parents in law Timo and Tuula Nokkala, thank you for keeping my feet on the ground and for your help in combining career and family by always being there to take care of the little princess and arranging some extra working hours for me. I want to thank Timo also for the excellent work with the figures of this dissertation. My parents Kari Kuusela and Tarja Kankaala, thank you for supporting me with all the ideas I have ever got, making me believe that everything is possible. Mom, thank you for setting an example of life-long learning and studying new things, and for the cover image of this book. Thank you, dad, for eagerly asking (sometimes too often) when "the book" is ready, and sorry for the sometimes grumpy answers. My siblings Sanna, Mikko, Jenni and Anni, thank you for being the best sisters and brother that I could ever have.

The most beloved ones – Tomi and Inna – I do not have words to thank you enough. There have been obstacles on the way that only you Tomi have been able to help me with, as well as moments of success only you have really understood. Thank you for all the love, support and understanding. And Inna, my little sunshine – I love you.

Turku, 7.8.2020
Tiina Nokkala

> "And now here is my secret, a very simple secret: It is only with the heart that one can see rightly; what is essential is invisible to the eye."
>
> "What is essential is invisible to the eye," the little prince repeated, so that he would be sure to remember.
>
> The Little Prince, Antoine de Saint-Exupéry

Contents

Tables

Figures

List of Original Publications

This dissertation is based on the following original publications, which are referred to in the text by their Roman numerals:

I Dahlberg, T. – Nokkala, T. A Framework for the Corporate Governance of Data – Theoretical Background and Empirical Evidence. *Business Management and Education*, 2015;Vol 13, No 1. pp. 25-45.

II Nokkala, T. – Dahlberg, T. Empowering citizens through data interoperability - data federation applied to consumer-centric healthcare. *Finnish Journal of eHealth and eWelfare*, 2019; Vol 11, No4.

III Nokkala T. – Salmela H. – Toivonen, J. Data Governance in Digital Platforms. *AMCIS 2019 Proceedings*, 2019.

IV Dahlberg, T. – Nokkala, T. Willingness to Share Supply Chain Data in an Ecosystem Governed Platform – An Interview Study. *Proceedings of the 32nd BLED Conference*, 2019.

The original publications have been reproduced with the permission of the copyright holders.

1 Introduction

1.1 Motivation and background

The importance of well-managed data has grown in the era of data as the "new oil",[1] with organisations eagerly making use of all the data at their disposal. On the other hand, now that the value of data has been recognised, data has become a business in its own right. However, that value is not yet clear to all actors, as they often find it difficult to define and measure the value of their data (Otto 2015). Just as electricity was once seen as merely useful but is now essential, data is quickly becoming crucial in almost any field of business.

This dissertation focuses on the governance of data on platforms. The initial motivation for this research was aroused during the writing of my bachelor's and master's theses on data management and governance. I found the various uses of data to help both individuals and organisations very interesting and researching this subject has opened my eyes both to the enormous possibilities opened up by data and the issues hindering their realisation. An increasing number of operations, business models and supply chains are being integrated and automated, forcing organisations to face issues of ownership of data, business models regarding data use, data quality and data definitions. The problem remains only partly solved for single organisations, and for inter-organisational contexts, the search for a solution has only started. The reason behind this is the move from closed to open systems environments.

"Data as an asset" is an issue that has been gaining more and more attention in both the academic literature and among practitioners (Khatri & Brown 2010), data governance aiming at strategic focus on data (Abraham, Schneider, & vom Brocke 2019). Considering data as a valuable asset makes it easier for the top management to understand data's importance (Horne 1998). Still, it seems that many organisations lack a high-level understanding of their own data assets, or of the level of risk related to those assets. In many cases, the leaders of organisations are not even aware of all the data the organisation they lead owns and handles every day. In Finland, some

1 "Data is the new oil", original quote from Clive Humby (2006), see also *The Economist* vol. 423, issue 9039 (2017).

pioneering organisations have presented a statement of their data assets,[2] together with mandatory financial statements. This shows that at least some organisations have awakened to the real value of data and are openly stating that they consider data important. In terms of research, there have been proposals for data governance models and frameworks on different levels (e.g. Abraham et al. 2019; Alhassan, Sammon, & Daly 2019; Khatri & Brown 2010; Otto 2011a, 2011b; Rosenbaum 2010; Weber, Otto, & Österle 2009; Wende 2007), which have guided approaches to the subject.

However, as Panian (2010) notes, organisations must share their data across the whole business instead of keeping it only for independent applications and departments. Research on that topic is still scarce, and in practice, the issue has become much more complicated due to factors such as ownership issues and business models related to data that is to be shared. Data sharing can happen in contexts such as digital data-driven platforms.

In the IT governance cube by Tiwana, Konsynski and Venkatraman (2013), the research gap in this area is shown as the absence of IT governance research on the ecosystem level in general and on the content of IT artefacts (e.g., data and information) dimension specifically. In my research, the aim is to take the firm level governance of data and extend it through three case studies to ecosystem-level data governance to provide a conceptual framework for data governance on inter-organisational platforms. From the cube's "how" dimension of Tiwana et al. (2013), my research involves all three governance mechanisms: decision rights, control and architecture.

As mentioned above, there have been proposals of data governance frameworks as well as other research on the topic. On the specific research gap of ecosystem-level data governance, Lee, Zhu & Jeffery (2018b) have started data governance research on platforms and ecosystems with their proposed data governance framework for process management. Digital platforms as enablers of value-creating interactions via digital tools (Constantinides, Henfridsson, & Parker 2018) in general have been a central focus of research lately (de Reuver, Sørensen, & Basole 2018; Tiwana 2013; Tiwana, Konsynski, & Bush 2010). That highlights the need for data governance research on digital platforms.

Furthermore, all the existing data governance models and frameworks nevertheless focus on accountability and decision domains of data, which are unquestionably important issues, but at the same time, they lack a notion of the meaning of the data. The meaning and the value of data can differ from one use context to another, meaning that a deeper understanding of the data's value is

[2] For example, OP Group has presented its first statement of data assets as a part of its financial statements (Eskola 2019).

necessary (Otto 2015). During the last few decades, the number of information systems storing information has grown due to the growing number of contexts and use situations in which data is created, modified and used. This leads to the need for a data governance framework that considers the different meanings and values of data attributes. All this is part of the shift from silos (Tallon, Ramirez, & Short 2014) to networks.

1.2 From canonical data models to federative data interoperability

A canonical data model (also known as a common or abstract data model), is a simplified, mutually accepted data model that is used to integrate information from different sources, often via translators (Garcia-Molina et al. 1997; Hohpe & Woolf 2004), that can be described as "The definition of a standard organization view of a particular information subject" (Schmidt & Lyle 2010 p. 401). A canonical data model is practical when there are more than just a couple, but still a controllable number, of information systems or data sources to be integrated.

In the past, information systems were mostly owned and developed in-house by organisations themselves, or owned after purchasing a customised solution from a software vendor. Use of canonical data models was suited to these situations. Data models in systems were known and, as differences between the models of different systems were understood, it was relatively easy to combine data from different systems. The number of systems was controllable, and the external environment did not have much influence on information systems.

More recently, organisations act more openly and in various networks. Information systems are acquired from outside as standard packages and data models are unknown. Some systems have much more data than others; some data are much more detailed and technical than others, and informational definitions for the very same data items vary. Thus, combining data from a number of systems is challenging. In addition, new data types, such as temporal and spatial data, audio and video data, social media data and internet of things (IoT) data are increasingly used.

Organisations are connected with each other in and between networks, creating various layers and linkages,[3] which leads to the need for connecting and linking information systems as well. With the growing size of these networks (or in larger organisations, the number of independent departments) and the increasing number of use contexts for data, a canonical data model seldom functions. To be able to integrate or federate the data from these numerous systems and storage methods and

[3] Formation of the networks and connections (in Life Sciences as an example context) is illustrated e.g. by Powell et al. (2005).

still understand the original meaning and value of these data, new methods are needed. Data federation can be defined as the integration of heterogeneous databases that act independently of each other, making the data in them interoperable (Heimbigner & McLeod 1985; Sheth & Larson 1990). To be integrated, data from different sources should be made interoperable (Janssen, Estevez, & Janowski 2014). Janssen et al. (2014) state that "At the data level, where interoperability is so critical for information-based application creation, interoperability is the ability of two or more datasets to be linked, combined and processed." In my research, tools for data federation are proposed based on the ideas of networked organisations and contextually defined data. This proposed approach is called federative data interoperability, referring to data being interoperable via federated metadata.

1.3 Research objectives and the structure of the thesis

Governance of data is the highest possible structure for managing an organisation's data and that data's relation to business objectives, and is connected to data management through data governance. Figure 1 illustrates the relationships between the concepts of governance of data, data governance and data management, and also provides definitions for the first two. Data management is a collection of activities contributing to "planning for, controlling and delivering data and information assets" (Mosley, Mosley, Brackett, Earley, & Henderson 2010). This dissertation aims to contribute to the field of governance of data, i.e. to enhance knowledge of the strategic level of data collection, storing and usage. Governance bodies for data must understand what data is used and needed in an enterprise, and must be capable of evaluating, directing and monitoring the data cycle. On the other hand, strategic objectives regarding data must be translated from the governance of data to daily management operations. That is done through data governance systems. Premises for the strategic level and governance of data are given both from corporate governance (Shleifer & Vishny 1997) and IT governance (Weill & Ross 2004) perspectives.

According to Abraham et al. (2019), data governance is a framework formalising data management as a cross-functional effort, prescribing decision rights and domains, and affecting and shaping standards and processes and monitoring their realisation. Thus, the approach to data must be selected as part of data governance. Strategic-level bodies need to be aware of the data-related issues and should make decisions based on them. In inter-organisational platform situations, the issue is more complicated, as there are several ways to arrange the general governance of the platform (Otto & Jarke 2019) and following data governance issues.

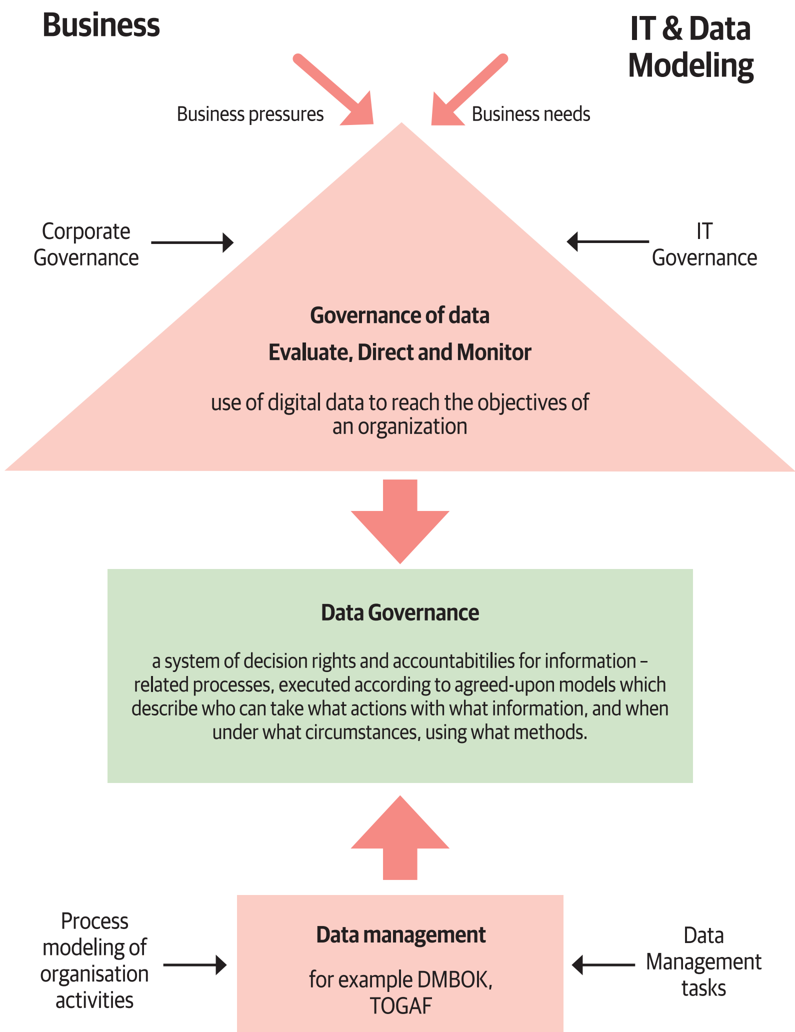

Figure 1. Governance of data, data governance and data management.

In the governance of data field the existing approaches are canonical (Berson & Dubov 2007; Loshin 2010; Otto 2012; Weber et al. 2009) and federative (Dahlberg, Nokkala, Heikkilä, & Heikkilä 2017). In this dissertation, one aim is to describe the features of both and to suggest their suitability for different purposes, as illustrated in Figure 2. In this dissertation, the focus is on inter-organisational data governance, i.e. data governance on digital platforms. Digital, inter-organisational platforms are a kind of technological industry platform (Gawer 2014) that are used as a means for data sharing. The aim of platforms is to enable transactions and connections between two or more parties that otherwise would not have a "conduit" between them (Gawer

2014). In the context of this dissertation, the types of platforms discussed are mainly multi-sided or networked platforms. In two-sided platforms the sides are heterogeneous, e.g. buyers and sellers (Eisenmann, Parker, & Van Alstyne 2009; Rochet & Tirole 2003), whereas multi-sided platforms directly join together two or more heterogeneous sides, which can be called consumers and "complementors" affiliated with the platform (Boudreau & Hagiu 2009; Hagiu & Wright 2015). Digital, inter-organisational platforms are data-sharing environments in which governance of data is needed, and in which the premises for governance are different from those in a single-organisation case. Bringing in more actors to govern data between them challenges traditional data governance thinking and requires theoretical reflection about the nature of data. Within the networks, ecosystems and on platforms the actors need to be aware of their stance towards data and the consequences of that stance.

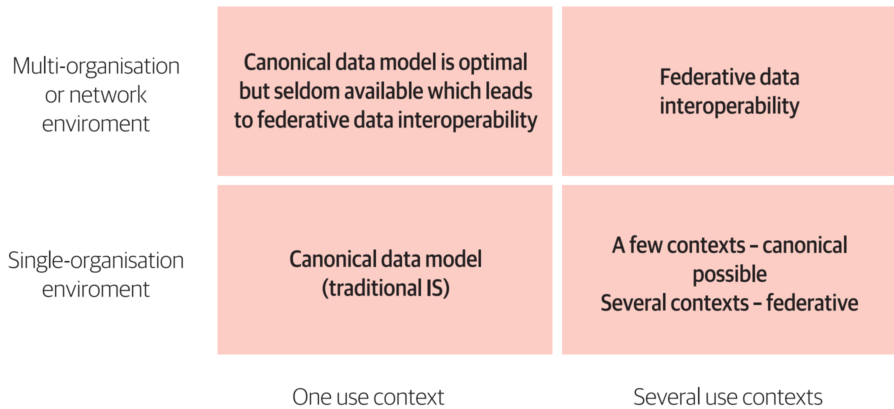

Figure 2. Different data governance approaches are needed in different data usage environments.

In this light, the research questions of this dissertation are the following:

Main research question (MRQ): *How does governance of data facilitate data sharing on inter-organisational platforms?*

RQ1: *Why are governance of data models needed for inter-organisational data sharing platforms?*

RQ2: *How can data from distinct organisations be made interoperable when the original data varies in terms of critical attributes?*

The main research question focuses on the governance of data on inter-organisational platforms and how data sharing on these platforms can be facilitated by better governing the data. The question is answered by using the two maritime industry cases, both of which are about platforms for sharing data. The question arises from digital transformation on a fundamental level, and more specifically from the growing importance of both data and data sharing business networks. Generally, digital platforms have been of interest in the academic community as of late (de Reuver et al. 2018), and data on them have been considered a boundary resource (Otto & Jarke 2019). The existing governance of data frameworks focus strongly on intra-organisational data, and for inter-organisational data sharing, they focus mainly on organising the sharing and other operations rather than on governing the actual data (e.g. van den Broek & van Veenstra 2015). A well-designed governance model for data and data sharing on platforms allows for data sovereignty and privacy, thus affecting the willingness to share data and enhancing business opportunities of the platforms. In this dissertation, I aim to answer the main research question by justifying the general need for the governance of data models (RQ1) and by proposing practical, albeit theory-based tools for enabling data interoperability between the data sources (RQ2). The interrelationships of the research questions and connections of the articles and cases included in this dissertation are shown in Figure 3.

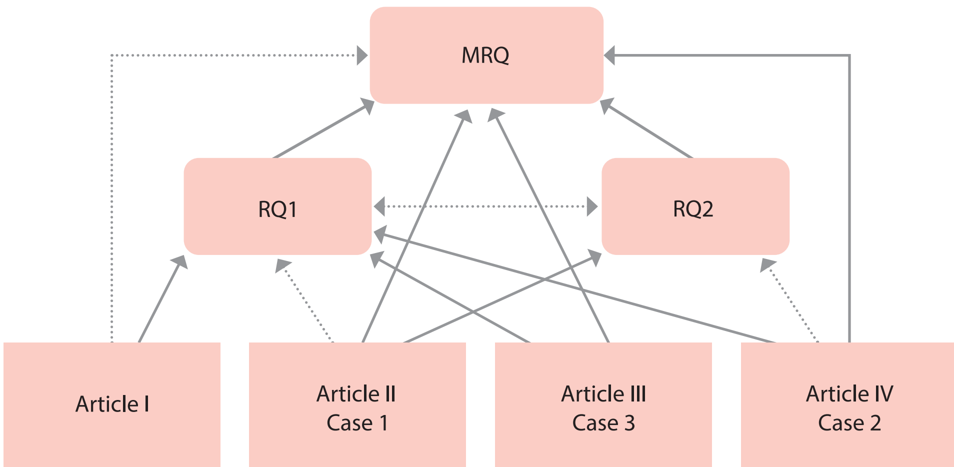

Figure 3. Interrelationships of research questions and articles.

RQ1 is about justifying the need for governance of data models. Data governance for intra-organisational contexts has already been justified in both practice-oriented (e.g. Earley & Henderson 2017) and academic literature (e.g. Abraham et al. 2019; Alhassan, Sammon, & Daly 2016; Brous, Janssen, & Vilminko-Heikkinen 2016). However, as inter-organisational platforms become more widespread for data sharing, there are changes in power relationships, data ownership definitions, and overall governance mechanisms (Otto & Jarke 2019), and data privacy and data sovereignty are affected. Having a sound governance of data model in place when designing and establishing a platform will help future platform owners. To answer this question, all three case studies (the Breast Cancer Case and the two Maritime Industry Cases) are used to explain the current situation and the problems with it.

RQ2 aims to identify and describe the two different approaches to data interoperability and to list the various attributes (e.g. content, format and volume) that must be understood in data governance functions. This more practice-oriented question is linked to the shift towards open and linked information systems that are required to share data. These information systems are both intra-organisational and inter-organisational. With the explosion of digital data and the rise of new digital data formats and sources, this question is relevant and topical. Data are becoming increasingly unstructured, and they flow more rapidly, e.g. from social media and various types of sensors. Governance over all that data is needed, and the interoperability of data is part of that governance. The question is answered through presenting a breast cancer case study in which a federative tool is used, and by evaluating the tool. By answering this question, this dissertation contributes to the governance and data management literature through presenting and justifying data interoperability tools that are based on the federative interoperability of data. The answer also complements the research on linked data and data spaces (see e.g. Bizer 2009; Heath & Bizer 2011), and linked data can be seen as a federative interoperability solution. Although the Breast Cancer Case is not about the inter-organisational platform, it has been selected as a means to answer this research question due to its complex nature and the difficulty of data integration, as later explained in the case selection criteria in chapter 3.3.

The research context is that organisations have ever-growing amounts of data, as described above. Networked, competing and cooperating organisations govern data both inside a single organisation and between multiple organisations. Thus, organisations practice data governance by themselves and together, e.g. on digital platforms. To implement digital platforms, data integration and the tools to achieve this are needed. To integrate the data between and within organisations, it must be made interoperable first (Janssen et al. 2014). The context is shown as the grey square in

Figure 4, in which the research domains of data governance, platforms (and ecosystems) and data interoperability are situated at the top. The research questions cover different sections of the domains, as shown in the figure.

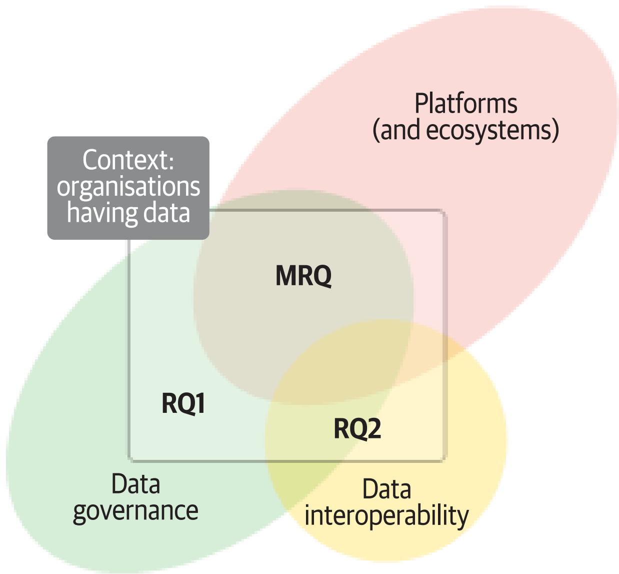

Figure 4. Research domains and research questions.

The research methodology selected for this study is qualitative case study research, and the methods used include interviews, focus groups and interview journals. I investigated three independent cases, one in healthcare and two in the maritime industry. All three cases contributed to one or two research domains, as described in Table **1**. The unique contributions of the case studies are presented in Articles II, III and IV, while Article I is more theoretical and explains the premises and the theoretical starting point for my research.

Table 1. Domains and cases connected.

Cases / Domains	Governance of data	Platforms	Data interoperability
Breast Cancer Case	X		X
Maritime Case 1		X	
Maritime Case 2	X	X	

My research makes several theoretical contributions to the field of information systems science. On the deepest level, the theoretical contribution made by this dissertation complements the thinking of Wand and Weber (1988, 1990, 1993, 1995). Their ideas focused on constructs and mappings between real-world and information systems, that is between ontological constructs and design constructs, and the different states and properties these constructs can have. The idea behind my research is that while Wand and Weber have justified their thoughts for single-organisational environments that have only a few contexts (see section 1.2), their model needs updating to be applicable to the multi-organisational, multi-context era. The more specific contribution of this dissertation is its examination of how ontological constructs and mappings relate to inter-organisational data sharing on digital platforms in ecosystems and of how this knowledge improves data governance and increases willingness to share data. As a result, a conceptual framework for platform data governance is proposed.

For the integration and interoperability of shared data, I present a federative approach and data federation matrixes that enable data integration via data interoperability. I also present findings regarding other factors promoting and preventing data sharing in ecosystems, and a governance of data framework to be used in inter-organisational contexts. The practical contribution is based on the notion that to share data inter-organisationally, partners in different ecosystems need to have adopted a sophisticated data governance framework and defined fine-grained business models to enable the trust and foundation required for data sharing.

The structure of the rest of this dissertation is the following. In chapter 2, the theoretical background for the governance of data, business ecosystems, networks and platforms is explored through a literature review. In chapter 3, the research design, underlying philosophical assumptions, and selected methodology and methods are described, and these selections are justified. In chapter 4, an overview of the independent articles is given. In chapter 5, the findings are presented, and chapter 6 includes the conclusions and discussion. Thereafter, the original articles are reprinted.

2 Theoretical background

In this chapter, I review the existing literature related to data in its various forms; the importance of data and its governance; data management; data governance; platforms and data governance on them; ecosystem data governance, and data ontologies.

Literature reviews of the field of data governance have already been made by several authors (Abraham et al. 2019; Al-Ruithe, Benkhelifa, & Hameed 2018; Alhassan et al. 2016; Brous et al. 2016). However, the number of concepts used is enormous (Brous et al. 2016), and most generally lack agreed-upon definitions. Many of the concepts are transferred from the IT governance literature (Weber et al. 2009). In addition, the current data governance frameworks are quite general and focus on middle-level issues, not strategic issues (see Figure 1). Alhassan et al. (2016) divide the literature into scientific and practice-oriented approaches, comparing the concepts and actions that are present in both streams. The most important actions in data governance, according to the review by Alhassan et al. (2016) are to define, implement and monitor, which in turn is related to Khatri and Brown's (2010) study of decision domains. Great emphasis is placed on defining roles and responsibilities, and on data quality. Brous et al. (2016) gather together the key concepts and principles for data governance, sorting them out under organisation, alignment, compliance and common understanding. Abraham et al. (2019) include the antecedents and consequences of data governance under the data governance umbrella. In my opinion, the existing frameworks, both scientific and practice-oriented, provide a good starting point for further development. Thus, in this chapter I review the literature not only on data governance but also on premises for data governance, data ontologies and IS ontologies. I also include platform and ecosystem literature in this chapter, as the rise of platform ecosystems brings new challenges to data governance as the number of concepts and linking points grows. My aim is to show how data governance can be enhanced by taking into account the properties and ontologies of data and to then use the enhanced data governance framework in platform and ecosystem environments.

The most important elements of an information system are processes, people, information or data and technologies (Alter 2013). The more formal (implementation-oriented) definition is: “An information system is a representation that is implemented using data and transformations on data (processes)” (Wand & Weber 1988). Owing to that, research on information systems phenomena can be on any of these elements individually or in combination. In this dissertation, information systems are understood to combine all these elements, and the scope of the research is data. Also, information systems phenomena can be viewed from an external viewpoint, as a planning and process management “black box” producing services, information and different impacts, or from an internal viewpoint considering how information systems—and data and other elements in them—should be organised and what are the best ways to provide the system to its users, who are also an element of the system (Wand & Weber 1995).

2.1 Data as an asset

Organisations hold various kinds of assets as the basis of their businesses. These assets can be tangible, like real estate, machinery and inventory, or intangible like intellectual property, special systems and data. However, unlike financial and physical assets, data are rarely valued as an item on an organisation’s balance sheet. In fact, although the value of data as an asset is recognised, its exact value is rarely known by an organisation’s decision-makers. Khatri and Brown (2010) offered the following definition: “Information assets (or data) are defined as facts having value or potential value that are documented.” They note that data and information are not differentiated in their definition. Weill and Ross (2004 pp. 5–7) mentioned “Information and IT assets” as one of the six key assets of organisations; these assets include “digitized data, information and knowledge about customers, processes performance, finances, information systems and so on”. According to the same authors, the maturity of the governance model of a key asset varies, and typically information assets are among the worst governed. Thus, regardless of the potential value of data assets, the appreciation of data does not manifest itself in governance practices. From the data governance point of view, Abraham et al. (2019) noted that “data governance focuses on data as a strategic enterprise asset”.

In Khatri and Brown’s (2010) data governance framework, one of the five domains is data principles, i.e. “clarifying the role of data as an asset”. This means that organisations need to define what the value of data is for them, and to decide how this evaluation should be materialised in organisational activities. Otto (2015) presented a study on how large organisations manage both the quality and value of their data resources over the data’s life cycle. The focus of these studies is on master data management, but in general, they prove that determining data’s value based on the quality of data is challenging. However, they also note that organisations recognise the strategic value of data (Otto 2015). As a valuable asset, data must be managed like any other valuable asset (McGilvray 2006), and the quality of data should be managed just like the quality of any other asset (Wang 1998; Wang & Strong 1996). However, the silo effect of data is present in both intra-organisational and inter-organisational data environments, meaning that data is created, stored, managed and used in silos. Optimising data-related processes and streamlining both intra-organisational and inter-organisational processes can be difficult when the whole picture of data assets is missing (Vayghan, Garfinkle, Walenta, Healy, & Valentin 2007).

When an organisation officially knows what data it has and uses and how and when it uses the data, as well as knows the risks related to contingencies such as losing all that data, data can become a competitive differentiator and provide an enormous competitive advantage. Different data governance models provide a tool to underline the importance of data (Khatri & Brown 2010; Rosenbaum 2010; Weber et al. 2009). Stating the stewardship responsibilities for data (Rosenbaum 2010) promotes data awareness, and further providing final accounts of data on a yearly basis also announces data's worth to the stakeholders. By data stewardship, Rosenbaum (2010) refers to "a collection of data management methods covering acquisition, storage, aggregation, and de-identification, and procedures for data release and use".

2.2 Data as a concept

A definition of data is rarely given without defining information and knowledge at the same time. All three—data, information and knowledge—are important and fundamental when we talk about knowing something about anything. Information technology processes only data, not information or knowledge (Galliers & Newell 2003), as data are facts representing the real world, and act as the raw material of information (Earley & Henderson 2017).

Davenport and Prusak (1998) gave the widely used definitions for data, information and knowledge. They stated that data is objective and tells discrete facts about things and events. When these things and events are situated in a context, data becomes information that is relevant and has a purpose. Information can be communicated as a message that informs its sender and receiver. Davenport and Prusak (1998) considered the methods that are used to shape data to become information to be contextualisation, categorisation, calculation, correction and condensation. For knowledge to be derived from information, methods like comparison, consequences, connections and conversation, all of which only humans are capable of using, must be employed (Davenport & Prusak 1998). The DAMA International Guide to Data Management Body of Knowledge DMBOK (Earley & Henderson 2017) and Zins (2007) followed quite similar routes, stating that data is the base layer or raw material, on which information is built by adding the context, and that knowledge is on top of information, using information as raw material. In addition, the DMBOK talked about wisdom as an even higher-level concept, but, as Ackoff (1999) pointed out, wisdom is not a self-evident consequence of data transformed into information and knowledge. Boisot and Canals (2004) in turn defined the data – information – knowledge chain as follows: "information is an extraction from data that, by modifying the relevant probability distributions, has a capacity to perform useful work on an agent's knowledge base". They also concluded that only data can be commonly shared between agents, while information and knowledge are commonly shared on only a very limited basis.

The idea of data's objectivity has been challenged by scholars such as Tuomi (1999), who claimed that in order for data to emerge, information is needed, and for information to form, knowledge is presumed. His justification for this is that someone's knowledge is required before isolated facts can be created and collected. Meaning structures and semantics are needed to store information in a database as raw data.

In information systems, data is the representation of real-life things, their properties and states. States can be lawful or unlawful, depending on natural and/or artificial rules restricting the possible states of things. Things, properties and states are ontological constructs that are represented by the design constructs of information systems (Wand & Weber 1990, 1993, 1995), as shown in Figure 5. Ontological constructs aim to be rigorous definitions for conceptual modelling constructs, making the use of the latter more effective (Wand, Storey, & Weber 1999). Wand et al. (1999) also provided mappings between the ontological and conceptual modelling constructs and rules for the use of constructs. In Figure 5, different contexts through which real world is represented as data in information systems, and contexts through which data in information systems are interpreted, are shown as lenses that I have added to the original figure by Wand and Weber (1993). As stated by Wand and Weber (1988), "In an information system implementation, data represent states and events; processes are the implementation of laws." Following from this, data in an information system representing a real-world system must be based on "someone's or some group's perception of the real-world system" (Wand & Weber 1995); that perception is located in a specific concept and can also vary when the context changes. Natural or artificial laws are limiting the possible values and states implemented in processes. In the era of open systems, the number of these perceptions presented as information systems is enormous and colliding, and data derived from the same real-life situations can have varying meanings and even errors due to varying artificial laws.

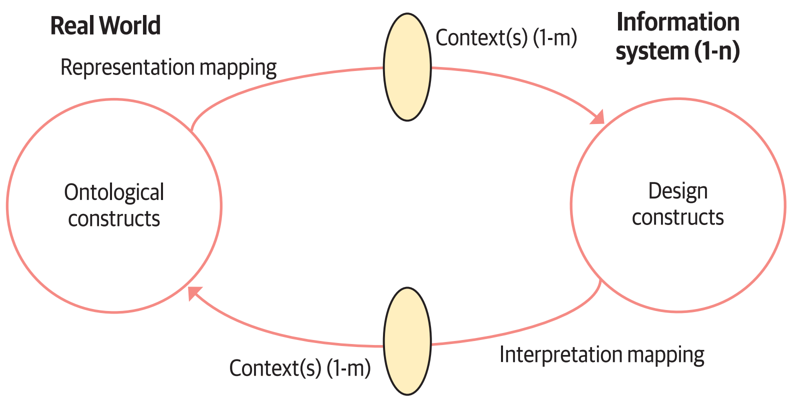

Figure 5. Modelling the real world and interpreting the model – contexts as lenses (modified from Wand and Weber (1993)).

Information systems as presenting the meaning of real-world systems should be able to represent things, their properties and states, and to track their states. As for state tracking, an information system, with the data included, has the following conditions that it must satisfy:

1. Each real-world system state must map to at least one information system state.
2. When the real-world system changes states, the information system must be able to change from a state that corresponds to the initial real-world system state to a state that corresponds to the subsequent real-world system state.
3. If an external (input) event occurs in the real-world system, an event that is a faithful representation of the real-world external event must occur in the information system.
4. Events in the information system that represent external events in the real-world system must be ordered the same way as the real-world system external events they represent. (Wand & Weber 1995)

These conditions can be satisfied quite easily in a one-context case, such as one in which the state changes are unambiguous. An example of this is a situation in which an information system is tracking an organisation's stock and an order is dispatched from the stock. However, when multiple contexts are tracking the same changes, e.g. the transport company delivering the order, the customer who has ordered the items and the bank financing the order follow the changes in their own systems, the situation may be ambiguous. Condition 4 on the order of events can be particularly challenging to agree on, as in the example case of order delivery, state changes may happen not simultaneously as seen in the context of different organisations. Clear agreements on when the data should represent a state change must be made, or the result will be an imbalance in information systems. Still, while organisations only track these changes internally in their own systems, problems are usually solvable. When the same organisations try to share these data on a platform, issues arise more often and are more difficult to solve. Once the state tracking becomes imbalanced, organisations will not able to agree on what has happened and when. In a situation in which the parties sharing that information are legally separate, the solution cannot be that they use the same information system to track states.

Wand and Wang (1996) concentrated on data quality issues and pointed out the reasons for deficiencies in data. As stated above, an information system is someone's or some group's perception of a real-world system represented as a constructed system, and that construction is further interpreted by its users and then compared to same user's interpretation of the real world. This is often the cause of a data quality deficiency (Wand & Wang 1996). Figure 6 is our view of the deficiency creation mechanism in a multi-context environment, e.g. in the case of a platform with many creators, sources and users of data. However, there are several data quality dimensions that affect the perceived quality of information systems (Wang & Strong 1996).

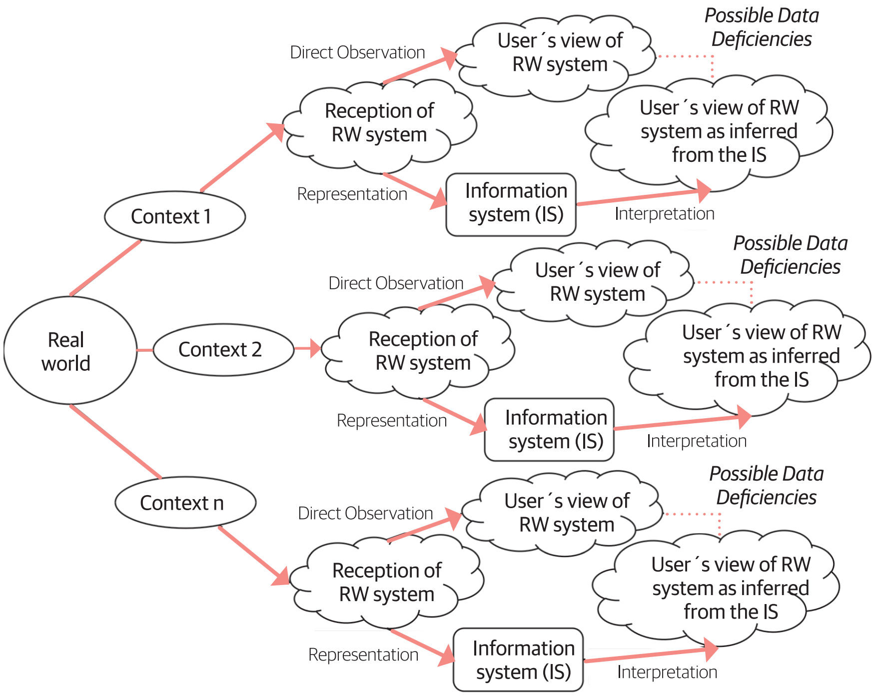

Figure 6. Possible data deficiencies from the user's viewpoint in multi-context-environment (modified from Wand & Wang (1996)).

The literature presented here on information systems and data ontology provides a thorough view of information systems and data as organisations' internal matters. However, as stated already, in the era of networked organisations and multiple contexts, organisations interact with each other and even want to integrate data into their systems. In the absence of other means of doing so, integration is often achieved by selecting a so-called "golden record" and then forcing other systems to use that golden data and to remove their own, possibly context-specific, data entries. I aim to extend the information systems ontologies presented here so that they can be used in the open systems era and on digital platforms where organisations try to integrate their data into their information systems. That new approach is called a federative approach. However, as my focus is on the business side of data governance, I have decided to exclude more detailed technical discussion, e.g. on conceptual data modelling. Next, I will present some specific features related to data ontology, i.e. different data types, and discuss the importance of metadata and digital data.

2.2.1 Data types

Dreibelbis et al. (2008 pp. 32–36) defined metadata, master data, reference data, transactional data and historical data as the most important data types. Cleven and Wortmann (2010) made a similar listing, although they omitted historical data. Historical data was added to the list by Vayghan et al. (2007). In a data taxonomy, master and transactional data are classified as domain data; metadata has two sub-types, informational and operational metadata, and reference data are used for all other data types (Cleven & Wortmann 2010). The listing of data types can be complemented by adding reporting/analysis data. Having so many different data types with distinctive features and management requirements is a challenge for data management and governance (Earley & Henderson 2017). I have collected different data types with their respective characteristics in Table 2. Metadata is further described in section 2.2.2, as metadata is the most important data type for the governance of data.

Master data is the core data of organisations, and any errors in them cause significant financial losses (Haug & Stentoft Arlbjørn 2011). Being *core data* means that master data includes business-critical data on *parties, things and locations* (Cleven & Wortmann 2010). Master data is typically created only once and used for purposes such as transactions over a long period without changing (Berson & Dubov 2007; Dreibelbis et al. 2008; Loshin 2010). Transactions use master data as the basis, but, on the other hand, all data, including master data, are created in transactions (Solomon 2005). The problem with master data is that, in spite of the fact that it is quite often created in different contexts (e.g. customer service, sales, production, procurement, accounting), and its use contexts vary, most organisations are willing to integrate all master data and make a "system of record" in which all parties use and store master data (Berson & Dubov 2007). Centralised master data records often wipe out the various context-specific data features and are an example of a golden record or canonical approach to data.

Master data, like other types of data, has its life cycle, but due to its "referential" nature, it is used in different ways in different situations (Otto 2015). The main life cycle phases for master data (management) are "data procurement, data storage and maintenance, data use for information production and data disposal" (Otto 2015). Different approaches are needed to govern the master data in each phase.

Table 2. Data types and their characteristics.

Data type	Characteristics
Metadata	Data telling the meaning of master or other data in the use context for each data user (Khatri & Brown 2010). Technical, operational and socio-contextual metadata each tell different facts about the data entry.
Master data	Stable, non-transactional basic data on the organisation's core entities, e.g. customers, products, suppliers, accounts, contracts and employees that are used organisation-wide and by different departments (Berson & Dubov 2007; Dreibelbis et al. 2008; Loshin 2010). Master data is of high importance for an organisation and organisations aim to keep it clean and unambiguous. Master data is independent of any other objects (Cleven & Wortmann 2010) but offers context for transactional and historical data (Mosley et al. 2010).
Reference data	Standard and regularised data coming from outside the organisation that are used to check the validity of some data entries (Dreibelbis et al. 2008 p. 34).
Transactional data	Data originating in all business transactions, such as communications with client/supplier, RFQs, offers, orders, order confirmations, changes to orders, invoices, payments, lists of collection, delivery instructions/documents, acceptance of delivery. Master data is used in these transactions. (Dreibelbis et al. 2008 p. 35)
Historical data	Master and transactional data become historical data over time, as these data stop being used anymore or become replaced by newer entries. Historical data is collected both for analytic purposes and for legal and regulatory use. Historical data can be used as a basis for decision making. (Dreibelbis et al. 2008 pp. 35–36; Vayghan et al. 2007 p. 671)
Reporting/analysis data	New data derived from master, transactional and historical data and their combinations, e.g. segmentation, productivity and classification reports per customer or product. Some reports are also created for legal or regulatory purposes. Metadata has a very important role in analysis, as it gives the timeframe and meaning to analysed data.

Reference data often include external sources like registries or internal and external standards that are used to check data's validity. Different codes, abbreviations, official names and allowed values are examples of data that use reference data (Dreibelbis et al. 2008 p. 34).

Historical data include the accumulated master and transactional data and are mostly used for analytical purposes to create an outlook on the current situation for decision-making purposes (Dreibelbis et al. 2008 pp. 35–36).

The various data types and their relationships are illustrated below in Figure 7. Master data and transactional data are the biggest ones, although they are very different in nature. All data types use metadata, either directly or indirectly. It is also possible to generate analysis and reports using only master data, e.g. it is possible to analyse the customer base of an organization using only customer addresses.

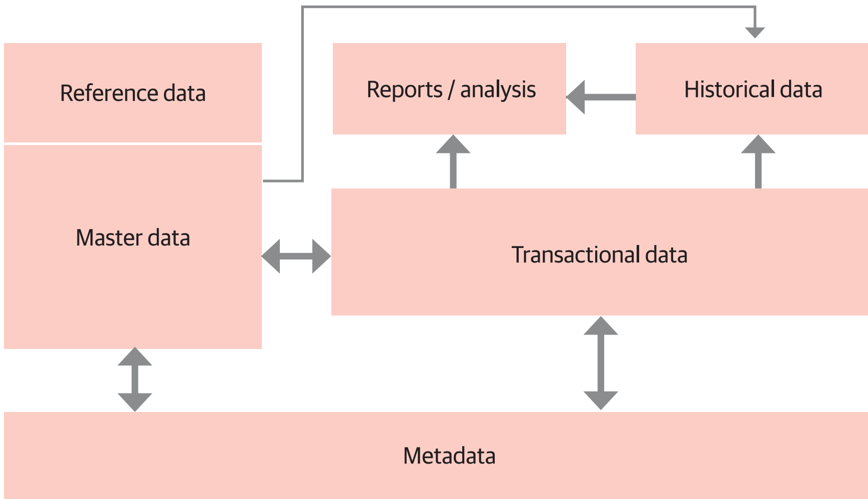

Figure 7. Data types and their relationships.

2.2.2 Metadata

Metadata is "data about data" and explains the meaning of data stored in an information system, helping to interpret that data (Khatri & Brown 2010). Metadata enable understanding of the context and circumstances of the data creation, which is essential for the proper use and interpretation of that data (Janssen et al. 2014). Metadata is created during the various processes related to the data, e.g. its creation, use and processing, and it also exists in the design and architecture of the information systems handling the data. Even knowing the name and type of system that stores specific data is metadata about the data. Metadata's importance has been neglected in the history of digital data (Sen 2004). It has been classified and categorised in many ways. Khatri and Brown (2010) listed physical, domain-independent, domain-specific and user metadata. Cleven and Wortmann (2010) divided metadata into informational and operational metadata. Another approach was offered by Berson and Dubov (2007 pp. 128–129), who divided metadata into business, technical and operational metadata. Inmon et al. (2008) made a distinction between structured and unstructured metadata. Singh, Bharathi et al. (2003) classified metadata under physical metadata, replication metadata and content metadata. All these classifications are important and useful.

In this dissertation and in the research articles and in the Breast Cancer Case, the metadata types used are IS technical metadata, information processing metadata and socio-contextual metadata, as described in Table 3. This classification was selected for the purpose of making data interoperable based on metadata. The selected metadata types describe the data entries' meaning and context for federation purposes. Other metadata classifications are useful for different purposes, and also partly included in this classification.

Table 3. Metadata types used in the federative approach and their descriptions.

Metadata type	Description
IS technical metadata	Technical, physical and operational characteristics of the data. Information about the size, format and type of the data. Enables designing and operating on data on the technical level, e.g. fitting data fields from several sources together.
Information processing metadata	Data type of the attribute (Table 2), to what function data is related, what is the source of the data and where it is modified and used. Structure of the data and information about internal/external sources.
Socio-contextual metadata	Meaning of the data in the given context, semantics of the data in its creation and many use contexts. Information on data's life cycle. Allows data to be interpreted. Value and importance of a data attribute. User metadata annotations.

IS technical metadata include the description of the data item; the medium on which it is stored; the name, value, length and type of the data field; information on allowed and prohibited items and characters, and field and data structure. IS technical metadata can be seen as the most basic form of metadata, but as was also noted in a case study by Otto and Jarke (2019), that syntactical metadata is not sufficient alone, but metadata on ownership, use conditions and access rights must also be included. Information processing metadata gives information on where and by whom data has been created, checked and modified, who is responsible for its storing and against what reference data it is possibly checked. Socio-contextual metadata tells about the context and meaning of different phases of data's life cycle, like what data meant when it was created, what its meaning is in different uses and how it becomes outdated.

Metadata is also part of Khatri and Brown's (2010) data governance framework, where it appears as one of the five decision domains. To make the best out of data, metadata must be systematically documented, and any changes in it must be managed. Management and governance of metadata are also discussed in both editions of the DMBOK (Earley & Henderson 2017; Mosley et al. 2010), in which the relationship between data and all other assets is discussed, with a recommendation that we need data about assets of all kinds in order to manage them; in data's case, that data is metadata.

The three metadata types presented here are characteristic of the federative approach to data. Information processing and socio-contextual metadata both give a great deal of information about the context and environment in which data are created and used, and therefore they are not highlighted in the canonical approach. The canonical approach sees data as something that can just be integrated, meaning that all the silos or parts of the organisation can use the same master data without needing to understand its original context. Data integrations can easily be conducted without any deeper understanding of metadata, but for more sustainable integration, an understanding of data semantics is required and data needs to be made interoperable. Thus, the canonical approach is useful and cost-efficient in smaller-scale data integrations in which there are only a few contexts.

2.2.3 Digital data

During recent decades, the data that mankind stores and communicates have been mainly digital, and the amount of data has been growing enormously (Hilbert & López 2011). In spite of the parallel growth in storage space and the development of analytical and communication tools, this has led to a situation in which we are no longer able to store, manage and use all the data we generate. At the same time, new ways of using data are being developed with the assumption that data is digital.

The digitisation of data has affected underlying business models in many industries and changed related aspects of the social infrastructure, such as content providers and content platforms (Tilson, Lyytinen, & Sørensen 2010). Another interesting notion about data in the context of digital innovation—especially in the open systems era—is the idea of the homogenisation of data (Yoo, Henfridsson, & Lyytinen 2010). Compared to analogue data, which is tightly coupled with the devices that use it, digital data can be accessed, stored, used and moved by almost any digital device. In the open systems era, when almost all data are digital, originating from uncountable sources and often combined together from a number of sources, this homogenisation is present. For the organisations storing and using data, this means that data is not medium dependent and that metadata and content are important.

In the case of analogue data, metadata in the most old-fashioned sense was stored as notes on the other side of a piece of paper, or taped on the top of a VHS cassette. For an instance of analogue data, IS technical, information processing and socio-contextual metadata may consist only of what has been interpreted from the data itself. For instance, in the case of a physical, printed photograph, the metadata indicate that it is black and white and tell how old it is, as well as recording where it was taken and who took it. As for a digital photo, some metadata often follows with the file; in some cases, metadata can even be very broad and rich and include information such as the details of the camera with which the photo was taken, the location and date of the occasion and many technical details about the format and settings of the camera. It also applies to simpler and more complex data types, such as text messages, video and audio material, sensor and social media data. Digitalisation has also enabled open data to be shared by governmental and other organisations to be used by individuals or organisations.

The DMBOK (Earley & Henderson 2017) pointed out that due to the digital storage of facts—information and data—many things that in the era of analogue data would not have been called data, are nowadays also considered data. Digitalism also allows us to measure and store data about things that were not measurable before, like a continuous heartbeat that is monitored with a sports watch, or real-time updates on events happening in space.

2.3 Governance of data in the single-organisational context

The need for governance of data arises from digitalisation in particular and the data generation that follows from this. As previously mentioned, nowadays, many things are data that were not considered data before, and the amount of data in general is growing rapidly. The functional roots of data governance lie in IT governance (De Haes & Van Grembergen 2004; Van Grembergen & De Haes 2008; Weill 2004, 2005). In relation to corporate governance, both IT and data governance form a part of it and need to be in dialogue with each other (Cheong & Chang 2007). However, as Henderson and Venkatraman (1999) have shown with their business/IT alignment model, business and IT strategies must interact with each other. The same goes for data, and all four alignment perspectives (Henderson & Venkatraman 1999) can also be used with data; data strategy and infrastructure can act as a source of competitive advantage, as a transformative force or as supporting business strategy. That is to say, data must be a concern at the highest level of corporate governance, not only part of the daily management routine. The current approach seems to be that the concern about data is to make it fulfil the needs of a particular application, not to act as a strategic asset for the whole business (Panian 2010).

Governance in general can be relational or based on formal contracts (Poppo & Zenger 2002). Formal structures or contracts are based on binding agreements between transacting partners (Hambrick, Werder, & Zajac 2008), where the strength of the “legal bonds” is used to describe the level of detail and specific obligations of the contracts made (Cannon, Achrol, & Gundlach 2000). Relational governance is more flexible and is based on social relationships (Poppo & Zenger 2002) and behavioural structures and processes (Hambrick et al. 2008). Usually, governance involves both mechanisms—formal contracts and relational governance—depending on governance needs, uncertainty and the environment (Poppo & Zenger 2002). IT governance has been divided into structural, procedural and relational mechanisms (De Haes & Van Grembergen 2004; Van Grembergen & De Haes 2008). In these mechanisms, structures guide roles and responsibilities, e.g. the positions of strategic key holders related to IT; processes are different measurements and agreements related to IT decision making and monitoring, and relational governance includes participation and communication between related parties. Also, data governance can be based on formal contracts or relational agreements; it manifests itself through mechanisms of structural, procedural and relational governance, and governance type can even change from domain to domain in different combinations.

Data governance is significant in both scientific and practice-oriented literature (Abraham et al. 2019; Alhassan et al. 2016). However, most of the data governance literature is focused on governing data management functions such as appointing rights and responsibilities, and defining the roles and different activities related to data governance (Alhassan et al. 2016; Mosley et al. 2010). Among other things, data governance should develop “data policies, standards and procedures” (Abraham et al. 2019). What seems to be lacking is a comprehensive (theoretical) model on which organisations and their top-level management could base their conception of data owned by the organisation, as well as the value of and level of risk associated with the data. With such a model, it would be possible for the strategic management team to have a common understanding of data as an asset for the organisation and to discuss them using the same language. One consequence of the establishment of such models could be that data asset statements (as part of mandatory financial statements) would gain popularity, as data is currently lacking from the balance sheets (Panian 2010). After all, as a part of corporate governance, data governance should direct the actions of the managers in all possible situations to maximise the profits of financiers and other stakeholders (Shleifer & Vishny 1997).

The importance of data governance for master data management and its success and implementation were noted by Vilminko-Heikkinen (2017), who stated that data governance is linked to the whole process of developing master data management. However, Vilminko-Heikkinen also referred to data governance as mainly a system through which responsibilities and roles are distributed. In their study of roles, responsibilities and ownership regarding data, Vilminko-Heikkinen and Pekkola (2019) also noted that for data that are used and created within a single organisational function, ownership was more clearly defined, as the owner is commonly within that specific function.

In the literature, the most widely used frameworks and models for data governance are those in the DMBOK (Earley & Henderson 2017; Mosley et al. 2010), which offers a more practice-oriented framework for data management; Khatri and Brown’s (2010) framework with the five decision domains for data governance; Wende’s (2007) model for data governance in relation to data quality management, and the contingency approach to data governance (Weber et al. 2009), which also concentrates on different roles and responsibilities. Rosenbaum (2010) also discusses the topic, proposing that data governance is about managing the stewardship of different data-related functions. In this dissertation, many of those frameworks are used as a basis for creating the governance of data framework, as opposed to the above-mentioned data governance frameworks.

Based on the data and data governance literature for single-organisation contexts, Figure 8 has been designed to illustrate how the concepts are connected. Governance of data, which is guided by strategy, mission and vision, together with the values of the organisation, is the highest-level approach an organisation takes regarding its data assets. It encompasses understanding the value of data as an asset and the risks related to the data. When executed carefully, governance of data can manifest itself in contexts such as official value statements of data.

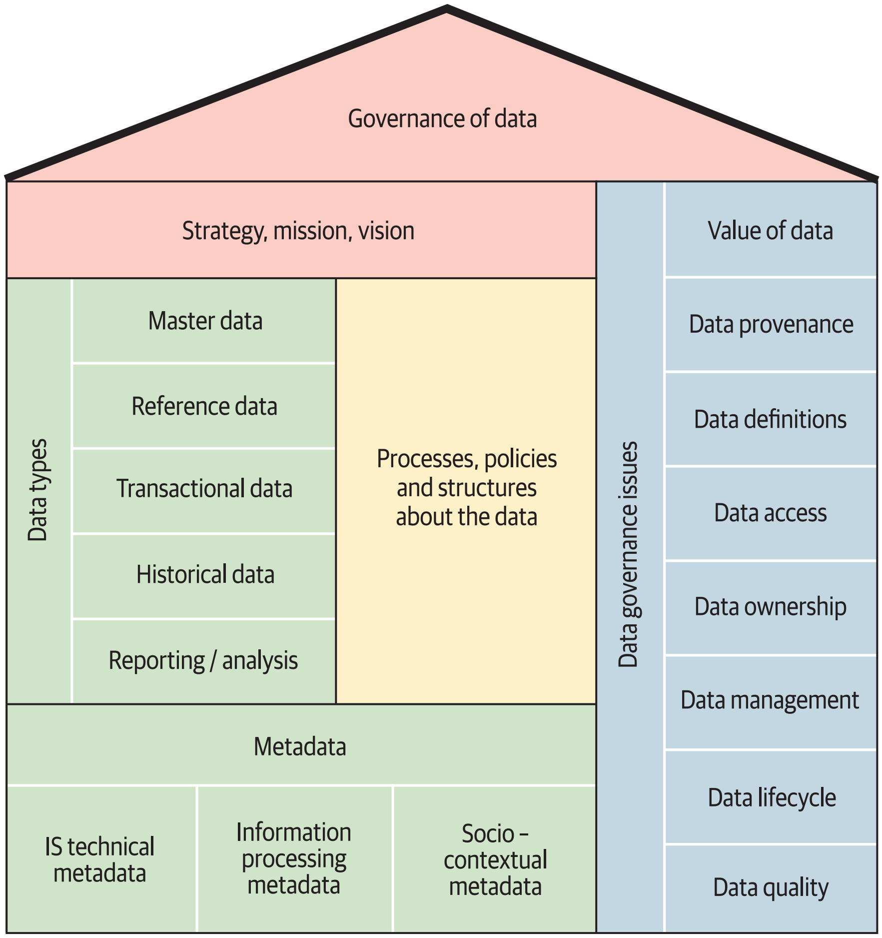

Figure 8. Data and data governance concepts combined.

The combined conceptual framework presented in Figure 8 includes the domains and issues found in the (intra-organisational) data governance literature. Based on the literature, these domains are concepts that are important issues in data governance and that should act as a basis for its further development. The idea of the framework is to show "what is" in the data governance field and what should be governed there, with no causal or predictive explanations (Gregor 2006). Table 4 presents the data governance issues one by one in more detail. A similar kind of framework is presented in Article I of this dissertation, as it was the starting point for this research and initiated the framework-building process.

Table 4. Data governance issues.

	Definition	References
Value of data	Data's value needs to be understood and used as strategic leverage.	(Abraham et al. 2019; Khatri & Brown 2010)
Data provenance	Data's history, including actions of creation, use, changes and deletion.	(Khatri & Brown 2010)
Data definitions	Data need to be defined with regard to aspects such as data type, structure, meaning, access/ownership and processes, and these definitions need to be maintained. Closely related to metadata.	(Khatri & Brown 2010)
Data access	Who, when and on what conditions can access the data objects.	(Rosenbaum 2010)
Data ownership	Ownership of data objects must be clearly identified and assigned together with responsibility for data. In intra-organisational environments, there can still be externally owned data involved.	(Abraham et al. 2019; Weber et al. 2009)
Data management	Data management processes and tasks need to be linked to data governance objectives. Data's value is actualised through data management.	(Abraham et al. 2019; Earley & Henderson 2017; Mosley et al. 2010; Weber et al. 2009)
Data life cycle	Data changes in different stages over their life cycle and the criticality and strategic value changes as well. Different meanings and values for different stages must be understood in data governance.	(Khatri & Brown 2010)
Data quality	Data quality affects organisations on many levels. Important data quality dimensions include completeness, timeliness, accuracy and credibility. A framework for decision rights and responsibilities for data quality and its management should be included in data governance.	(Khatri & Brown 2010; Weber et al. 2009; Wende 2007)

In Article I, the framework is referred to as a *"governance of data framework"* instead of a *"data governance framework"*, indicating that the proposed framework takes one step further and aims to be more comprehensive, not only about controlling the data but also about ensuring the achievement of a set of strategic goals, as indicated in Rau's (2004) definition of the term "governance". However, the naming of the framework is not a central issue, but the ultimate aim is to take the single-organisational governance of data framework and reshape it to answer the needs of digital platform data.

2.4 Digital platforms, business networks and ecosystems

Ecosystems have received a great deal of attention in the academic literature lately, and the concept is often used to describe various kinds of networked organisations and the technology, processes and services related to them (Aarikka-Stenroos & Ritala 2017; de Reuver et al. 2018). Digital transformation and digitalisation, on the other hand, have generated capabilities for innovative platforms (Yoo, Boland, Lyytinen, & Majchrzak 2012). In this section, I will provide definitions for the central (digital) ecosystem, network and platform concepts I use in this dissertation, as well as the distinctions between the concepts. I will also briefly go through the dynamics that affect the interplay between the members of (digital business) ecosystems and users of platforms. Data governance on platforms forms part of the contribution of this dissertation, and the existing literature on the topic is presented in section 2.5.

2.4.1 Networks and digital business ecosystems

The ongoing change in the business environment—the replacement of traditional markets by business networks—requires that organisations look further outside their own boundaries (Möller & Halinen 1999). In networks, organisations are forming patterns and structures for exchange and communication (Powell 1990), i.e. networks are a way to orchestrate business. Rapidly expanding networks with new ties created with increasing pace have been illustrated in studies such as that by Powell et al. (2005). In Figure 9, I have collected their illustrations of network expansion (showing the new ties created with the main organisation each year) from research in life sciences networks over the years 1989–1997. These new ties being added to the existing network structures, and different layers being formed on other layers make the situation increasingly complex. One-to-one connections and structures become too heavy to maintain with even the key partners. Following these changes, new governance models are already needed to ensure value creation and capture within the networks.

Continuing on the same evolutionary path, Aarikka-Stenroos and Ritala (2017) propose that the shift towards wider environments—ecosystems—is happening. Increased dynamism, connectivity and variability are characteristic of these new networks and ecosystems. The boundaries of networks and ecosystems are blurry and changing, while ecosystems are perceived to be compositions of networks and their actors affiliated with the ecosystem (Aarikka-Stenroos & Ritala 2017). Such blurriness and dynamism have implications for the governance (of data) structures.

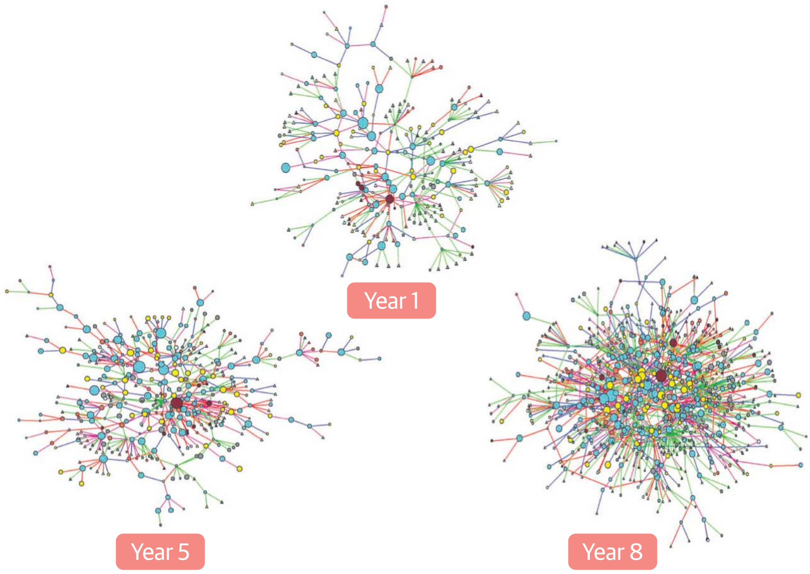

Figure 9. Expansion of a network – number of new ties formed (original figures from Powell et al. (2005).

Another way to look at business ecosystems is to consider the term's roots in biology: just as in natural ecosystems organisms are interdependent, in a business ecosystem there is an interdependency between the actors doing business together and competing with each other (Moore 1993, 2016). In the digital world, the term is digital business ecosystem (DBE). Compared to a natural ecosystem, a DBE is a digital environment inhabited by digital creatures, using digital and technical infrastructure and digital representations such as software, agreed-upon documents and standards (Nachira 2002; Nachira, Dini, & Nicolai 2007). A business ecosystem consists of networked actors often using shared technologies and institutions in their value creation and exchange (Adner 2017; Iansiti & Levien 2004a; Moore 1993). For research purposes, the business ecosystem has been a widely used concept to describe different sets of systems that co-evolve (Aarikka-Stenroos & Ritala 2017). However, ecosystems do not just emerge but are formed as the result of a process in which organisations try to optimise their businesses' value creation and extraction (Jacobides, Cennamo, & Gawer 2018).

According to Nachira (2002), a DBE is the last phase in ICT adoption evolution, following the path from e-mail to websites, e-commerce, e-business and networked organisations. The preceding digital phases have led the way to organisations becoming gradually more dependent on digital solutions and becoming more networked through ICT services. Nachira (2002) described the DBE evolution as follows:

> As a consequence of this evolution, the dynamic networking of the organisations, drives the dynamic cooperation of the players on the territory and the connection of the resources in a system, building a community that shares business, knowledge, and infrastructures. This will dramatically affect the ways enterprises are constructed and business is conducted in the future, and the actual slowly changing organisations will be replaced by more, fluid, amorphous and, often, transitory structures based on alliances, partnerships and collaboration.

Nachira (2002) also pointed out that these DBEs can be sector specific, meaning that some business areas or industries start to adopt certain software components and other features related to a DBE, gradually starting a new DBE for the sector. These sector-specific DBEs, in turn, can be either local or global.

When using a DBE as a metaphor, the evolution follows similar paths to natural evolution in that organisms that are more complex are formed and gradually gain space in the ecosystem, and that, to survive, an ecosystem needs a sufficient amount of inhabitants (Nachira 2002). However, using the (digital) business ecosystem metaphor to analyse real networked organisations is not in itself sufficient, as the flows and exchange of resources are not as simple as in biological ecosystems (Corallo 2007). Still, DBE as a concept allows organisations to be viewed as members of industrial ecosystems, in which the respective members use the same resources, platforms and services. The use of the same resources allows competition and cooperation of the members in an ecosystem (Iansiti & Levien 2004a, 2004b).

The holistic view of DBE dynamics shows how societal trends, enterprise practices and capabilities, and digital platforms are intertwined as the core elements of a digital business ecosystem (El Sawy & Pereira 2013). According to El Sawy and Pereira (2013), these elements cannot be separated, as they dynamically interact with each other. This explains how digital platforms are an inseparable part of digital business ecosystems.

For this dissertation, networks and ecosystems play an important role. The need for more sophisticated data governance follows from the appearance of more complicated networks (as illustrated in Figure 9) and the more dynamic and ever-changing structures of those networks. Peltoniemi and Vuori (2004) defined a business ecosystem as “a dynamic structure which consists of an interconnected population of organizations. These organizations can be small firms, large corporations, universities, research centers, public sector organizations, and other parties which influence the system.” Given the additional complexity caused by digitalisation and digital business, it is understandable that tools—such as digital platforms—are needed to make sense of all the data these networks and ecosystems are sharing within them.

2.4.2 Digital platforms

Tilson et al. (2010) proposed that the new capabilities and possibilities created by digital ecosystems lead to new kinds of systems and services and new ways to integrate them, also bringing new social and technical dimensions on board. Digital platforms are at the centre of such new applications, bringing in several technological and managerial challenges. Following from the expanding networks discussion in the previous section, digital platforms can be seen as tools to give networks and ecosystems a proper form and to handle the chaotic network of one-to-one, one-to-many and many-to-many relationships and the data exchanges involved. Networks are thus platform-mediated when access to resources—such as data—is provided by the platform and participation is encouraged at different possible levels of openness (Eisenmann et al. 2009). As solutions, digital platforms can be defined from two different viewpoints (de Reuver et al. 2018):

- Technical view: “An extensible codebase to which complementary third-party modules can be added.”
- Sociotechnical view: “Technical elements (of software and hardware) and associated organisational processes and standards.”

In this dissertation, digital platforms are seen as sociotechnical concepts. For instance, Constantinides, Henfridsson and Parker (2018) used a technological-sounding definition: “a set of digital resources—including services and content—that enable value-creating interactions between external producers and consumers”. That definition, however, implies the interaction between different parties and exploitation of common resources. In the context of my research, it is important to take into account the sociality of the concept and include processes, standards and rules in the definition of digital platforms; a point that illustrates this is that willingness to share data on a platform is presumably not solely based on the platform’s technical properties. As Iivari (2017) pointed out, an IS artefact (such as a digital platform) consists of IS applications, together with social artefacts, IT artefacts and information artefacts.

Digital platforms, consisting of items from various levels and contexts, can also be analysed on various levels (de Reuver et al. 2018). Therefore a researcher must be very clear with the definitions of the phenomena studied. For instance, de Reuver et al. (2018) noted that the domain of mobile payments includes platform components scattered on various levels, and that often these scattered components need to be considered in very different contexts, regardless of the fact that they belong to the same domain. A very similar issue is present in the research articles of this dissertation, as the platforms that are investigated are located in domains, and the platform is not just a single element in that domain, but rather consists of several levels and layers combining technical and organisational elements.

Digital platforms can be either internal (for a product or an organisation) or external, industry-wide platforms. Organisations develop, own and further develop sets of assets (such as technologies, services, applications and products) for the former, while the latter comprises similar assets that are, however, external and developed by one or more actors and can act as a basis for organisations' own creations. Both forms are based on a structure that is common and mutually agreed on (Gawer & Cusumano 2014).

Digital platforms need to be divided into single-sided, two-sided and multi-sided platforms, referring to the number of actors operating on them and providing resources for them (Helfat & Raubitschek 2018). In this dissertation, I focus on multi-sided platforms (Boudreau & Hagiu 2009; Hagiu & Wright 2015; Otto & Jarke 2019; Tan, Pan, & Zuo 2018), and data governance is seen as part of governing the whole platform infrastructure. However, in the literature, platforms are generally looked at as having a clear platform leader or owner (Cusumano & Gawer 2002; Gawer & Cusumano 2008; Ghazawneh & Henfridsson 2013; Helfat & Raubitschek 2018; Huber, Kude, & Dibbern 2017; Ozalp, Cennamo, & Gawer 2018; Tiwana 2013), or a keystone firm (de Reuver et al. 2018) who owns the platform and allows others to use it.

Against the traditional view of keystone-centred or single-organisational platforms, Otto and Jarke (2019) have presented a case study on a multi-sided platform that is alliance-driven. This means that a platform is owned by—and its architecture and design are defined by—multiple owners, who have to integrate their shared interests into the platform (Otto & Jarke 2019). Thus, the ultimate need for governance of data on multi-sided platforms seems to be a consequence of sharing data within complex networks where one-to-one data-sharing models are not sufficient, due to the enormous number of actors sharing and making use of the data. In the keystone-driven platforms architecture, design and other issues are decided mainly by one keystone firm (de Reuver et al. 2018), and the coordination of value creation is mainly conducted via pricing (Gawer 2014), whereas in alliance-driven platforms the starting point is already complex and dynamic, requiring clear structures and decision-making mechanisms (Otto & Jarke 2019).

2.4.3 Inter-organisational data sharing

Digital ecosystems and platforms are means of sharing data between organisations, particularly for managing and integrating supply chain activities. Firm performance is enhanced when supply chains are integrated, as data consistency is an important factor in firm performance (Rai, Patnayakuni, & Seth 2006). This indicates that sharing consistent data between supply chain members (possibly on a digital platform) can increase revenues.

High technological connectivity and a high degree of willingness to share data indicate strategic, trust-based relationships between data-sharing organisations, also involving technological integration and sharing relevant information on decision making. On the other hand, willingness to share data and connectivity are low in situations in which relationships are reserved, resources are scarce and the mutual history between the data-sharing parties may include bad experiences, decreasing trust. In the latter situation, only minimal information is shared, which hinders the possibilities for performance enhancement (Fawcett, Osterhaus, Magnan, Brau, & McCarter 2007). Both the willingness and the technical capabilities are needed to make data sharing work meaningfully (Prajogo & Olhager 2012). Willingness and trust-based relationships indicate that for data governance, the relational mechanisms are at stake (De Haes & Van Grembergen 2004; Van Grembergen & De Haes 2008).

Environmental uncertainty, intra-organisational facilitators and inter-organisational relationships all have an impact on information and data sharing and also on information quality in supply chain management. To share data, organisations need managerial support and enabling technologies, and the relationships with other partners need to be good enough. Mutual vision and commitment from all the partners are needed. External elements promoting or preventing information sharing are related to uncertainties; organisations may wish to cope with the uncertainties by sharing their data (Li & Lin 2006). Uncertainty, inter-organisational relationships and other such elements also affect the type of governance needed (Poppo & Zenger 2002), both on data-sharing arrangements and data governance. Allen et al. (2014) have described data-sharing agreements (DSA) as a means for governing the data and meeting the legal and trust-based requirements for inter-organisational data sharing in healthcare. DSAs are from the contractual governance end of the governance type continuum (Hambrick et al. 2008; Poppo & Zenger 2002).

Van den Broek and Van Veenstra (2015) presented four modes of inter-organisational data governance in different types of data collaborations, i.e. market, bazaar, hierarchy and network modes. They stated that the selection of the governance mode for inter-organisational data collaboration depends on the type of data sharing. In the context of this dissertation, data sharing is supposed to happen on digital platforms on which data is shared and exchanged laterally and used to complement the sharing partners' often mutual products, so the most suitable governance mode is the network mode.

Lee and Whang (2000) listed types of information that are shared on supply chains. The list includes inventory levels, sales data, order statuses for tracking/tracing, sales forecasts, production/delivery schedules and other information such as performance metrics and capacity. In Maritime Industry Case 1, organisations' willingness to share these and some other types of data on a supply chain digital platform are investigated. Lee and Whang (2000) have already noted that an effective system or platform, such as the internet, is crucial for data sharing. Models they propose for sharing are the information transfer model, third-party model and information hub model (H. L. Lee et al. 2000).

On the more ontological level of data, data sharing is also about matching master data, as sharing data for orders, for example, requires that the basic customer and product data is unambiguous for all the organisations sharing that data. For intra-organisational data sharing, Otto (2015) noted that there is a difference between global and local master data. Global master data is shared, e.g. a certain material has a unique ID number, whereas locally there are master data on that material's specific features in the local context (Otto 2015). For the inter-organisational context, this means that master data should be shared partially, possibly excluding the local master data, and if local master data is shared, their context needs to be described in metadata.

2.5 Data governance on platforms

Clear data governance and processes for data management are needed for a multi-sided platform to be successful in data sharing (Otto & Jarke 2019). The whole general governance structure for the ecosystem and multi-sided platform, including the access rules of the platform (Helfat & Raubitschek 2018) and architecture and design of the platform (Otto & Jarke 2019) must be defined and agreed on, and data governance must be included in this process. The actual structure for defining governance depends on the stage of evolution and the developer of the platform and on whether the platform is driven by a single organisation or by an alliance, i.e. it depends on whose interests matter the most (Otto & Jarke 2019).

Considering the concept proposed by de Reuver et al. (2018) of multiple levels of platforms, governance is not easily designed. On the general governance level, governance means who decides what. For a platform with a clear owner—e.g. a keystone-driven platform (Otto & Jarke 2019)—the first governance issue is how decision rights are divided between platform owners and platform contributors. The second issue is the control mechanisms, i.e. how the platform is governed formally and informally. The third issue is pricing and, more generally, the business model (Tiwana 2013). For the case of data governance, similar issues exist. If platform users are to share data on a platform, access and decision rights must be agreed on, as well as control and stewardship of data. These issues refer to who can add and access what data on a platform, and how the quality of data is ensured. In a case study by Otto and Jarke (2019), access to and ownership of data were noted as especially important aspects of data governance: all the industry partners in the case study insisted that data access must be delicately controlled, and that metadata should contain information about access rights to data items.

Like any other inter-organisational platform, on a data-sharing platform a business model is needed; data is valuable (Khatri & Brown 2010), meaning that it is important to define who can extract value from the platform data, and how the costs are divided. And, if someone wishes to join the platform and start sharing data, that party should have something valuable to be integrated into the platform.

A couple of articles have been written on data governance on platforms (ecosystems) that address this topic or even propose a framework for data governance on platforms or platform ecosystems (S. U. Lee, Zhu, & Jeffery 2017, 2018c, 2018a; Otto & Jarke 2019), and some note data as a boundary resource for platform governance (Otto & Jarke 2019; Schreieck, Wiesche, & Krcmar 2016). Lee et al. (2018b) proposed a data governance approach for platform ecosystems that seems to be most comprehensive in combining data governance and platform ecosystems. They proposed four data governance principles and seven decision domains in their framework. The principles are (S. U. Lee et al. 2018b):

- Align with platform governance concepts
- Meet the needs of all participating groups
- Address all types of data
- Consider the platform context

The decision domains they propose are data ownership and access definitions, the regulatory environment, contribution estimation, data use cases, conformance, monitoring and data provenance. The framework and its components were based on a literature review and a review of 19 existing governance frameworks (S. U. Lee et al. 2017, 2018b). Together with the proposed framework, Lee et al. (2018b) presented processes (data collection; data management; surveying, research and productisation; data consumption and data termination) that brilliantly clarify the governance framework and how the decision domains presented in it are present in day-to-day processes for governing data on platforms.

Another model is that proposed in Lee et al. (2018c), in which the authors note that data governance on platforms can be centralised or decentralised, both options having their pros and cons. They then present a decentralised model in which data governance is partitioned between three main types of platform users (data providers, data users and user groups) and both closed data governance and open data governance are used. These entities interact with each other. In this model, closed data governance is performed by platform owners to set up models for data access and data ownership and to classify the data on the platform, as well as to align the strategy with business objectives. Open data governance is for the platform users to get access to data on the platform and to see all activity happening on the platform. This model (S. U. Lee et al. 2018c) is quite practice-oriented and enables an understanding of the different roles and tasks in platform data governance.

In another article, Lee et al. (2018a) proposed a contingency model for data governance on platform ecosystems, that, for example, considers internal and external contingencies in relation to the characteristics and design type of data governance.

Lee et al. (2017, 2018a, 2018c, 2018b) clearly indicated that traditional data governance is focused on in-house control of data, and prior research on platform governance is still in its infancy. Despite the existence of platform data governance models and frameworks, a similar kind of research gap exists as in intra-organisational data governance (section 2.3), i.e. the existing frameworks focus only on the middle level (data governance), instead of the strategic, governance of data level (see Figure 1). However, Lee et al.'s (2018b) framework and Otto and Jarke's (2019) case study provide an excellent starting point for creating a more strategic and multi-level framework for data governance on platforms. Therefore, the concepts and principles proposed by them are present in the framework presented in this dissertation.

2.6 Summary and framework

Data governance is a topic that has mainly been studied from an intra-organisational perspective (S. U. Lee et al. 2018a, 2018b). The intra-organisational data governance framework in Figure 8 was drafted based on the existing literature. In the few articles touching on data governance in platforms (S. U. Lee et al. 2017, 2018c, 2018a; Schreieck et al. 2016), the focus is on the owner of the platform; however, Otto and Jarke (2019) presented two different kinds of ownership, i.e. keystone-driven and alliance-driven. On the other hand, data governance even in an intra-organisational perspective strongly focuses on defining roles and responsibilities with respect to data and governing data management functions (Alhassan et al. 2016; Mosley et al. 2010). Strategic understanding of the data types, sources, formats and other data characteristics, as well as an understanding of the value of the data and the risks related to them, seems to be lacking. Lee et al.'s (2018b) framework is, as stated above, the most comprehensive framework in the field that I have seen, and thus it has been relied on heavily in creating the framework proposed in this dissertation.

Executing the governance of data requires understanding the characteristics of all possible data types and understanding how data types are manifested in an organisation's data assets. In particular, metadata with its sub-types should be considered crucial. When the characteristics and processes related to each data type are already considered in strategy building and the design of organisational processes, data can advance the business strategy substantially.

Data governance issues, on the other hand, are different types of decision domains organisations and their strategic management need to consider part of strategic and operational processes. Some of these issues can be dealt with by using RACI matrixes (e.g. Wende 2007) and other types of responsibility definition and allocation tools.

For data governance on platforms and platform ecosystems, different kinds of issues need to be considered (S. U. Lee et al. 2018b). On the platforms, data-related processes are different from those in single-organisation cases, and the whole governance structure is more complex. Even the regulatory environment affects data sharing on platforms, as data ownership must be defined, and several kinds of privacy and security concerns arise in a multi-organisational context. The governance mechanisms used need to be considered as well, in order to determine whether they should they be structural, procedural or relational (De Haes & Van Grembergen 2004; Van Grembergen & De Haes 2008), or some combination of these (Abraham et al. 2019). In chapter 5 I will present a platform data governance framework that is based on the data governance frameworks for single-organisation contexts (starting with the one presented in Figure 8), the literature reviewed in this chapter and the empirical material collected during the process of writing my dissertation.

3 Research design and methodology

In this chapter, I will discuss the scientific approach and the methods selected for this dissertation and the ontological and epistemological stances leading to these choices. Then I will present the selected methods in more detail, describe the three cases and describe the data collection and analysis processes and methods. The research design consists of the components described in this chapter, which acted as a leading guideline from the initial interest to actually conducting the research (Eriksson & Kovalainen 2008; Flick 2007; Yin 2014)

3.1 Scientific approach and methodological choices

The way a researcher chooses to conduct research depends on the way they see the world. Knowledge is produced by humans and also used by humans. We can only know about humans and their environments—including other living creatures, nature, space, etc.—both physical and social. Research methods include certain implicit and explicit assumptions about the nature of the world and how knowledge is constructed. Epistemological and methodological assumptions are beliefs about knowledge and its justification, and ontological assumptions are those related to the (objective or subjective) existence of the research object (Chua 1986). In addition to those beliefs, assumptions are made about human intention and rationality, as well as about the relationships between people and between people and society. Other assumptions have to do with the relationship between theory and practice (Chua 1986). Different combinations of these assumptions lead to different research paradigms (Mingers 2001), which in information systems science are mainly positivist, interpretive and critical (Orlikowski & Baroudi 1991). In some cases, only positivist and interpretive approaches are mentioned (e.g. A. S. Lee 1991).

Usually, positivist and interpretative approaches are considered to be opposed to each other, as the world views on which they are based are so different. This confrontation of approaches has many streams, e.g. objective versus subjective (Burrell & Morgan 1985; Orlikowski & Baroudi 1991) and quantitative versus qualitative (Van Maanen 1979). I find that confrontation very difficult because my personal view of the world combines the two. In some cases I think that it is possible to measure anything, and that reality is objective; in other situations I believe in the importance of the context and socially constructed reality. Neither research nor reality can be black and white. Luckily, Lee (1991) has shown that it is possible to integrate the two approaches in organisational research. Also, Mingers (2001) stated that using only one paradigm shows only one aspect of the world and keeps the researcher blind to the others.

Another issue related to ontological and methodological assumptions is the suitability of different types of research methods. In extracting knowledge about reality, epistemology, ontology and methodology are connected (Eriksson & Kovalainen 2008). In the positivist approach, quantitative methods are preferred, as reality is seen as objective and quantifiable, and in the interpretative approach, qualitative methods are more often used (Kaplan & Duchon 1988; Myers 1997; Orlikowski & Baroudi 1991). In business research, qualitative methods such as case studies; ethnographies; grounded theory; focus groups; action, narrative, discursive, critical and feminist research are used (Eriksson & Kovalainen 2008). Quantitative methods include surveys, statistical testing and controlled experiments to test the hypotheses created. Due to the nature of the methods, they are suitable for different situations and phases of the research, both potentially contributing to different stages (Kaplan & Duchon 1988).

In information systems science, the positivist stream was the mainstream for a long period, and it still has strong supporters; however, the interpretive paradigm has been rising steadily (Hirschheim & Klein 2012; Orlikowski & Baroudi 1991). Thus, both are accepted approaches for research conducted in information systems science.

In my research, I use qualitative case studies that include interviews, focus groups and other material. However, I do not go to the deepest level of interpretation, but rather attempt to remain more objective. If forced to define my research philosophical stance, I would say that it lies somewhere between the positivist and the interpretive, perhaps slightly tilted towards the latter. The approach I use is more ideographic than nomothetic, as the ultimate aim is to understand the people that are part of the information systems (Burrell & Morgan 1985). My assumption is that there are different levels of reality, e.g. the framework proposed is “hard”, and can be evaluated and discussed more objectively, whereas the willingness to share data and factors related to it are more subjective, socially constructed and require more interpretation and co-creating of reality. Also, nominating data as an asset is a result of social construction, and the importance of data is understood only within a context created by humans. When looking at data federation from a purely technical point of view, it could be researched objectively and measured, but when the interpretation of data and their representativeness—instead of the objective facts—are at stake, social constructions and human interpretation come into play. This indicates that the framework cannot be built without subjective meanings.

The ultimate aim of research—in the information systems field, as in others—is to build *good* theories and to make sense of the underlying world and the phenomena in it (Gregor 2006). For theory building, the theoretical contribution I am aiming at is explaining "how" and "why" something happens, i.e. it is a "theory for understanding" (Gregor 2002, 2006). The research questions posed aim at explaining the need for governance of data, describing the interoperability of data from distinct sources and understanding the importance of governance of data for inter-organisational data sharing on platforms. However, the conceptual frameworks presented in this dissertation belong to the "theory for analysing" category (Gregor 2006), which means that the frameworks gather and classify the concepts related to the governance of (platform) data. This approach was selected for the framework because, currently, there is no framework for the inter-organisational governance of platform data.

3.2 Case study research

In information systems science research and also more generally in social science research, the case study method is among the most popular qualitative research methods (Dubé & Paré 2003). Case studies can be either positivist (Yin 2018) or interpretive (Klein & Myers 1999; Walsham 1995) in nature. With "how" and "why" research questions, when the researcher has little control over the events investigated, and in the real-life context of a contemporary phenomenon, the case study method is preferred (Yin 2014). All the three cases investigated in this dissertation are real-life cases in natural settings over which the researcher had very little control, indicating the suitability of the case study method. The main "how" research question followed by "why" and "how" sub-questions also paved the way for the case study method. The case study method has also allowed theory generation from practice, including in areas where little research has been done before (Benbasat, Goldstein, & Mead 1987).

The philosophical basis of a case study investigation in information systems can be positivist, interpretive or anything between the two; it can also involve one or multiple cases and use qualitative or quantitative methods (Cavaye 1996). Thus, it allows plenty of possibilities and alternatives, but also requires clearly stating and defending the choices that have been made.

For theory building from single cases, the cases need to be carefully selected. The case should be especially revealing; it should either be an extreme case or one in which the access provided to the research subject was unusual (Yin 2014). The rationale for selecting a single case can be that it is "critical, unusual, common, revelatory or longitudinal" (Yin 2018 p. 49). The aim of my research was to develop theory instead of testing it, and thus the case selection does not need to be especially representative. However, I have selected cases that provide great insights, represent interesting new approaches and openings and in which I have had, if not unusual, still very good access to the research subject. Thus, these single case studies provide good grounds for theory development, while more cases are still needed to generalise the tentative theories (Eisenhardt & Graebner 2007).

For the three case studies presented in this dissertation, the guidelines of Yin (2011, 2014, 2018) have been followed, indicating a more positivist than interpretive approach. For theory building based on the case studies, I have followed Eisenhardt (1989), Eisenhardt and Graebner (2007) and Yin (2018). Multiple means of data collection were used, no experimental controls or manipulation were involved, and, as no variables had been defined beforehand, the results were based on the researcher's own integrations of the material (Benbasat et al. 1987). The methods for data analysis are described below in section 3.5. Eisenhardt (1989) advised comparing the emergent concepts from the case studies to the wide range of extant literature, including writings that conflict with the concepts, in an effort to find rival explanations.

For the strength of theory building from case studies, Eisenhardt (1989) mentioned the likelihood of generating novel theories, the likelihood of the creation of testable constructs and hypotheses and the likelihood of empirically valid the resulting theory. These all are linked to the closeness between the empirical evidence and the theory. The weaknesses of case study theory building include the possibility of ending up with very complex theories, as the rich case material allows one to capture "everything", and the risk of narrow and idiosyncratic theory. In the selection of the case study research methodology, I have also selected the risk of difficult generalisation of the theory.

However, it is possible to make generalisations from even single case studies and to go even further than just making hypotheses in the first stage of the research process (Flyvbjerg 2011; Ruddin 2006; Stake 2008). In my dissertation, the three case studies are instrumental (Stake 1995 p. 3, 2008 p. 123) by nature, serving as sources for insight into the issue of data interoperability and governance on digital platforms. Combining the three case studies in this dissertation also falls into the category of multiple or collective case studies (Stake 2008 pp. 123–124), even though the cases are dissimilar to each other. However, it was not planned as a collective case study, but the independent cases were selected one after another and planned separately, with the interest being on data governance on digital platforms.

In the case study research, the type of generalisation can move from description to theory (A. S. Lee & Baskerville 2003), or as Yin (2018 p. 37) calls it, analytic generalisation. In contrast to statistical generalisation in which findings are generalised from a sample to a larger population, analytic generalisation is based on case study results in order to either advance initial theoretical propositions or propose new concepts (Yin 2018 p. 38). In my research, the case study results do both: they advance initial theoretical propositions (intra-organisational governance of data models and frameworks are further developed to suit inter-organisational governance of data situations), and new concepts are proposed (new features that arise in my case studies are added to the framework). Implications of the contextual nature of data are discussed in light of the multi-sided data platforms and the network business models.

The three independent case studies presented in this dissertation serve different functions. Their role in and relationship to the whole is explained in the case descriptions in the next section. There I also justify the selection of the cases in light of the research questions of this dissertation.

Benbasat et al. (1987) recommend determining the unit of analysis based on its appropriateness for the research project. That has been done for all three case studies. In the breast cancer project, that unit was the data related to breast cancer diagnosis and treatment. In the Maritime Cases, the units were the respective platforms or ecosystems.

3.3 Case descriptions and selection criteria

The empirical material for this dissertation has been collected in three individual cases. Figure 10 illustrates how the cases are located on the research domain map. In this section, I will give a brief description of the cases and their backgrounds, and more detailed descriptions can be found in the research articles included in this dissertation.

In relation to each other, the cases are complementary. As mentioned before, they are instrumental case studies (Stake 2008) that act as sources of information and insight into various issues (Stake 2008 p. 126) related to data interoperability and data governance on platforms. The rationale for selecting each case was different, as explained in the following paragraph.

The breast cancer case (3.3.1) was selected because it is an extreme case, i.e. especially problematic (Flyvbjerg 2011 pp. 306–307), and it provided great learning opportunities (Stake 2008 pp. 130–131). For data interoperability, the breast cancer case provided evidence about an extremely complicated case—maybe even a critical one (Ruddin 2006)— in which the data needed to be made interoperable from a large number of information systems owned and governed by independent departments, as well as from external parties such as regional healthcare centres and occupational healthcare providers. As seen in Figure 10, the breast cancer case is not about platforms. However, it contributes to the RQ2 by presenting tools to make data interoperable as a part of data governance. In the data governance on digital platforms, the situation is similar in regard to data interoperability. Different sources need to be linked and data should be shared between them.

The first maritime case (3.3.2) was also selected based on its criticality and its many learning opportunities (Ruddin 2006; Stake 2008). In addition to these characteristics, selection was also based on the unusually good access (Stake 2008 p. 130) to the case context due to the research project. As the project was about creating and implementing a platform within an ecosystem to enable the automation of data sharing in a supply chain, this case provided excellent insights about inter-organisational data sharing on platforms. The case is connected to the research questions as it is about an ecosystem that aims at mutually governed data sharing via data interoperability on a digital platform.

The second maritime case (3.3.3) was selected for its context and prevailing circumstances: the digital platform was in the planning phase, and the participants were free of the limitations of the existing platform structures. Ownership of the planned platform was not yet decided; thus, the governance structure was fully open to discussion. This case also offered great learning opportunities (Stake 2008) and provided an excellent instrument (Stake 1995 p. 4) to discuss platform data governance. Regarding the research questions in this dissertation, this case contributed to RQ1 and the main research question, as it illustrated the need for a model to govern data on the inter-organisational settings and on a platform, and it also provided insights about inter-organisational data sharing.

Overall, the selected cases—presenting, e.g. different domains and industries—may not be the most representative, but they can be described as extreme (Flyvbjerg 2011) or critical (Ruddin 2006), and they provide wide opportunities for learning (Stake 2008). Considering the cases together, generalisations can be made regarding the phenomena of data governance on digital platforms.

In addition to the context in Figure 10, the cases are connected to each other through data interoperability as a leading theme. In case 1, data interoperability is clearly present in the tools proposed. In the maritime cases, cases 2 and 3, data interoperability is intrinsic in how data are planned to be shared and in the data governance.

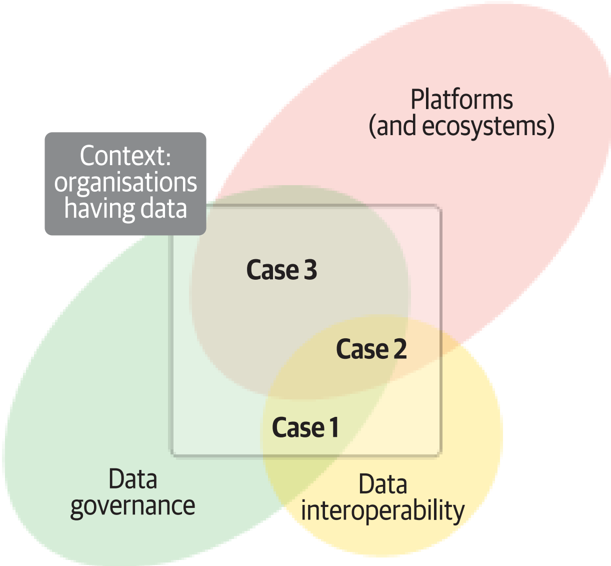

Figure 10 Research domain and the cases

The timeline of the case studies data collection is illustrated in Figure 11. The Breast Cancer Case, conducted in 2016, was the first one, and the two Maritime Cases were conducted partly overlapping in 2017-2018.

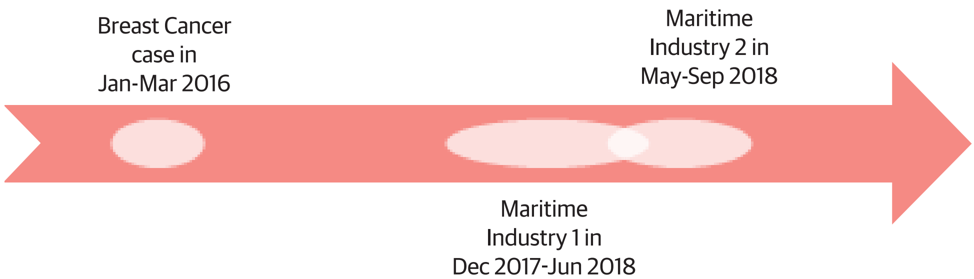

Figure 11 Timeline of the case data collection

3.3.1 Breast cancer case

The first case study, the Breast Cancer Case, was conducted in a Finnish hospital district at a university hospital. The case was selected for this dissertation primarily for its contribution to RQ2, on how data can be made interoperable when it varies in terms of critical attributes. The actual case was a real-world situation in which the data needed to be integrated, as the data specialists at the hospital were trying to compile patient data on breast cancer patients from various sources in order to be able to use the data for analyses to detect breast cancer earlier, to make treatments more effective and to improve the survival rates for widely spread breast cancers. The data was in many parts overlapping in the various information systems, and thus this case provided a good basis for building the federation matrixes and testing the federative approach. Also, the access and data ownership issues were well represented in this case, as the information systems at the hospital are domain-specific, and usually used for one context only. Some other important aspects regarding data governance were clearer than usual, e.g. the data's value was quite clearly understood, and the risks were quite obvious due to the sensitive nature of healthcare data. One more reason for selecting this case was the unusual level of access, as I had the opportunity to discuss the case with the data specialists, who in turn were able to connect me and the other researchers with doctors and other staff for the workshops. The workshop participants had a wide variety of backgrounds and thus provided a comprehensive view of the matter. In Article II, in which this case study is used, it is to illustrate another context—empowerment of citizens through data interoperability—which brings in one more group of individuals who are related to the data. This illustrates how data federation and federative tools can provide benefits to many user groups.

University hospitals and hospital districts have dozens of information systems actively in use, and through these systems, they have access to enormous amounts of data. However, as we found out during the research project, they have difficulties in using that data, as it differs from system to system in its format, definitions, access, use and modifying rights, and there can be overlapping data for the same events in different information systems, while some data or systems are not properly connected and linked to the other systems. Thus, we needed to find a way to recognise the data and information systems that are relevant for the task and to describe and link these data to make them interoperable and to enable data integration.

Breast cancer is one of the common causes of early death for women. In 2017, 4,947 new breast cancer cases were diagnosed in Finland and the survival rate of patients five years after the diagnosis in follow-up period in 2014-2016 was 91 % ('Cancer in Finland - Syöpärekisteri' 2018). That indicates that despite its high incidence, breast cancer can be effectively treated and, in many cases, cured. Still, cancer is a growing problem in Finland, as the number of people diagnosed with cancer is forecasted to rise, and the risk of breast cancer is significant for the female population. Women's risk of developing breast cancer is 2.22 %; the risk of developing any cancer is 5.63 % for women over 50, and for women over 60, the risk is 5.03 % for breast cancer and 11.28 % for any cancer, with the risk increasing steadily with age ('Cancer Registry Statistics' 2016).

Breast cancer can be detected during screening studies, at an appointment with a doctor at a public or private hospital, during occupational healthcare or on some other occasion. To confirm the diagnosis, X-rays, laboratory and other tests and appointments of various sorts are needed; data is also accumulating from medical prescriptions, cover letters, pathology reports and patient records. Treatments include radiotherapy, surgical treatments, cytostatic treatments and other cancer-specific treatments. Patients usually spend some time at the hospital ward and visit policlinics regularly. Follow-ups and regular examinations are needed, even after the breast cancer has been cured. It is not uncommon for breast cancer tumours to develop metastases that also need treatment.

To gain a basic understanding of breast cancer and its treatments, we had a working group meeting to discuss the issues. Specialists explained the diagnosis and treatment processes and described the data that is used and created and the parties involved in those processes on a general level. Then I had one meeting with the data specialist, who showed me the information systems used at the university hospital that included data related to breast cancer. Together we arranged workshops (focus groups) with the specialists and doctors. We had altogether three workshop sessions with 4–7 participants, including 1–3 researchers, in which we discussed breast cancer-related information systems at the hospital and how data is defined in them. The duration of the workshops was 1.5–3 hours.

Based on the interviews and workshop sessions, we gradually developed the data federation matrixes presented in Article II. We were able to find four key data attributes for the interoperability: social security number, appointment time/date, tumour node metastasis (TNM) code and diagnosis code. We then went through the breast cancer-related patient information systems, starting from the main patient information system and proceeding to the laboratory, radiotherapy and pathology information systems, describing the four key attributes and their metadata in each system.

We were not able to conduct the actual data integration, as patient data is very sensitive, and we did not have access or usage rights to the information systems and data. However, we presented the tools and solutions that we created to the data specialists and discussed their feasibility and usefulness. The specialists found our ideas and tools useful. The results have been presented in several academic forums in addition to Article II (Dahlberg, Lagstedt, & Nokkala 2018; Dahlberg et al. 2017; Nokkala & Dahlberg 2018).

3.3.2 Maritime Industry 1

The first Maritime Industry Case is a part of a group of research projects under the title DBE Core. The research projects under that title are currently the digital supply chain (DSC) project in the bio-refinery industry ecosystem, maritime digital supply space (MDSS) project and SmartLog project in container traffic ecosystem. In this dissertation I mainly report results from the MDSS project, although in article II results from the DSC project are also presented.

The MDSS case was selected for the DBE project as it is critical for the field and for the automation of data exchange in the supply chain. Data governance is crucial in such an attempt, and with this case, I had a chance to witness the creation of a data-sharing platform from the very beginning. I have also had unusual research access to participating organisations, as the project has been strategic for them as well. The project offers the possibility of a longitudinal approach after the completion of this dissertation.

In the DBE Core projects, the aim is to automate and integrate the supply chains in the industry ecosystems using a shared digital platform for business data and technical data exchange. On the digital platform, blockchain technology will also be used. Automation and integration of the supply chains are achieved through agreeing on standardised messages that are used to exchange data and messages on the platforms. These standards are based on the OASIS UBL 2.2 standard and, if needed, on other general standards. Member organisations of an industrial ecosystem can send and receive standardised messages to the platforms through API (application programming interface) programmes ('DBE CORE – Digital Business Ecosystem' 2019). The potential cost savings in the bio-refinery industry alone if 100 % automation of 8.5 million business documents and 2.5 million annual invoices is achieved is 580 million euros per year (Korpela 2014).

The MDSS project targets the automation and integration of the shipbuilding supply chain in the maritime industry ecosystem in Southwest Finland. Enhanced value-creation mechanisms are sought for through automation and integration. In the maritime ecosystem, there are various kinds of organisations, mainly the shipyards (two shipyards participated in the project) and their suppliers and sub-contractors. The suppliers and sub-contractors included design companies, IT firms, engineering companies, parts suppliers, hull parts suppliers, cabin factories, etc. Some banks were also involved in the project.

I interviewed 17 representatives of project partner organisations between December 2017 and June 2018. The interviewees were recruited by asking for recommendations from the project's contact persons; some interviewees were contact persons themselves, and some were other suitable specialists. We used the snowball method to find more interviewees, asking after each interview if there were some other persons in the respective organisation that could have something to add to the topics that were discussed. The aim of the interviews was to find out what factors promote or prevent willingness to share data in the ecosystem, and what data partner organisations are willing or unwilling to share on the DBE platform (Article

IV). In the interviews, I also investigated other issues not yet reported in scientific articles belonging to this dissertation. Those other issues are willingness to share transactional data, use of electronic invoices in the organisations and willingness to share technical product data as a part of standardised messages on the platform. The findings on these other topics are briefly presented in section 5.4.

3.3.3 Maritime Industry 2

The second Maritime Industry Case was selected because it offers remarkable insights into platform data governance in the pre-platform phase. The project partners were only just planning on establishing a platform for data sharing, and that allowed the research to be conducted without the limitations of existing platform structures. The interviewees were free to ponder various viewpoints and opportunities, as no execution of the actual platform had been decided on yet. I also had access to multiple sides of the shipbuilding process, from the shipowner to the shipyard's sub-contractors and IT service providers.

The case was also part of a larger research project, i.e. the D4Value project, which ran from 2016 until spring 2019. This project consisted of many parts and topics, ranging from autonomous ships and ports to the regulation and insurance environment in the maritime sector and to the transmission and use of sensor data collected from the ships during the operation ('D4V – Design for Value Program Report' 2019). I was involved in the project from February 2018 until spring 2019 and participated in research on creating a data-driven digital platform for data created on cruise ships during their operations.

The business environment of the cruise business is currently undergoing many changes. On the one hand, cruise ships are becoming large entertainment centres involving spas, amusement parks, theatres, multiple restaurants, etc. On the other hand, climate issues, global warming, emission control and security issues are restricting the business, both through official regulation and consumer demand. Shipping companies must respond to these changes and even try to transform themselves in advance to avoid a bad reputation. Automatisation and the use of data and the internet of things are seen as possibilities to tackle these issues.

With another researcher, I conducted 10 semi-structured interviews with 12 persons at the shipping company, shipyard, sub-contractors and IT service providers who were all involved in cruise ships' operations on the sea. The interviewees were selected based on our contact person's recommendations, and we used the snowball method to acquire more interviewees. The interviews were conducted in 2018. The topics discussed during the interview varied slightly between the interviews because the backgrounds of the persons interviewed were very different from each other and the organisations they represented had different roles in the ecosystem.

3.4 Data collection

In the case study method, there is no formal data collection protocol to follow, but instead, the relevant data to be collected can emerge during the process (Yin 2018 p. 83). Thus, we had an initial data collection plan and case study protocol (Yin 2018 pp. 93–94) for all three cases but allowed other data to accumulate as well. This approach yielded some interesting additional findings. The interview journal was used as one way of collecting such findings (see section 3.4.4).

Yin (2018 p. 111) lists six potential sources of case study evidence: documentation, archival records, interviews, direct observation, participant observation and physical artefacts. I have used documentation (e-mails, agendas, reports, administrative documents, and news clippings), archival records (public files, organisational records, charts, drawings and mappings), interviews (shorter and longer interviews, semi-structured and other methods, both individual and group interviews), a small amount of participant observation and physical artefacts (information systems and databases). Yin (2018 pp. 126–130) recommends using multiple sources of evidence for triangulation purposes.

During the data collection, we followed ethical guidelines ('TENK' 2019) to ensure that we did not harm the participants in our study. All the participants participated voluntarily and gave their consent after receiving sufficient information about the research; we also guaranteed the anonymity and confidentiality of the interviews (Eriksson & Kovalainen 2008). In the Breast Cancer Case, we did not use actual patient data in any phase, but data attributes were used at the generic level. If needed, made-up data items were used for illustrative purposes. In the Maritime Cases, the organisations were often in seller-buyer relations with each other, or direct or indirect competitors, and thus we needed to be cautious when discussing and reporting the interview results not to harm their businesses.

The main data collection method was, however, the interviews. In the following sections, I will describe the interview methods used and briefly discuss how I conducted the interviews.

3.4.1 Semi-structured interviews

In information systems science, the semi-structured or unstructured interview is the interview type most often used in qualitative research (Myers & Newman 2007). Such an interview follows an incomplete script, having some questions prepared beforehand, but allowing the researcher to improvise if and when needed. A semi-structured interview can also be a group interview, as in some of the interviews I conducted.

Qualitative interviews are most often interpretive in their philosophical stance (Rubin & Rubin 2005). This means that the approach is deeper and adaptive, allowing different contexts and interpretations of events. The aim is not to find out pure facts, such as how many companies are absolutely willing to share their data, but instead to study the factors behind the phenomenon.

Purely semi-structured interviews were used in Maritime Case 2, in which I conducted interviews together with another researcher. All interviews were conducted face-to-face, thus allowing more personal discussion. Both researchers made the improvised additional questions on the topics the most important, and the initial interview script was modified at times, depending on the interviewee's background.

Semi-structured interviews, or, indeed, any type of qualitative interview, have many potential pitfalls, as the situation in the interview can be complex. Lack of trust and time and level of entry can vary, and interviewees can have their own agendas for the interview (Myers & Newman 2007). We confronted some of these problems during some of the interviews. However, the advantages of the qualitative interviews were also present in many cases, as we had some excellent conversations and gained some particularly interesting insights.

3.4.2 Enhanced interactive (multi-stage) interview method

In the MDSS project interviews, I used a new interview method called the enhanced interactive multi-stage interview method, which has also been used in a study by Dahlberg, Hokkanen and Newman (2016). The aim of this method is to conduct the interviews efficiently, saving both interviewees' and interviewer's time.

Before the interviews, the questions are drafted carefully and tested in trial interviews. Clear interview protocols and scripts are drafted. Questions are sent to the interviewee before the interview. In the actual interview setting, the questions are shown one by one and answers are typed and displayed on the screen or wall at the same time as the discussion. That helps both the interviewer and interviewee to focus on the discussion, and less side-talk is involved. Answers are not written down word-by-word, but the first round of interpretation is done immediately. As the interviewee can see all the time what is being written down, (s)he can check the interpretation immediately and make corrections if needed. However, the interview script allows the posing of additional questions when needed, e.g. if some new and interesting concept is mentioned. Interviews are also recorded with permission from the interviewee, which facilitates the recall of any additional discussions. These recordings are then listened to, and the notes of the answers completed. The complete notes are then sent to the interviewee for approval. As the interviews are not transcribed verbatim, the notes are usually completed within a week from the interview, and thus the interviewee still has the interview fresh in their memory when checking the notes.

In terms of the scientific approach, on the continuum from positivist to interpretive, the enhanced interactive multi-stage interview method is located in the middle. It is still interpretive, as the answers to the questions are interpreted and the reality is “co-created” between the interviewee and the interviewer while the discussion and interpretation take place. However, even though conducting the interviews with this method can be considered to be interpretive, the analysis of the interview material can be rather positivist, as described later in section 3.5.

3.4.3 Focus groups

Focus groups as a research method differ from group interviews in the sense that in focus groups participants are facilitated in interacting and working with each other, not just responding to the researcher’s questions. The research interest in a focus group is the collective outcome of the discussion and how different actors construct a shared understanding (Eriksson & Kovalainen 2008).

Focus groups were used as a data collection method in the Breast Cancer Case. We also called the focus groups “workshops” to underline the mutual goal we had for the sessions. We collected specialists from the hospital units to discuss the data different units have on breast cancer patients and treatments. During the sessions, participants expressed their opinions and beliefs and described their ways of acting on the data. Some ideas for improvement were also expressed.

The role of the researcher(s) in the sessions was to facilitate discussion and, if needed, to pose some additional questions that guided the discussion. In all the sessions, both subject specialists and data specialists were present, and they were able to learn from each other during the discussions.

3.4.4 Interview journal

While conducting the interviews, an interview journal was kept to make notes of any special events or anomalies during the interviews. Also, any deviant behaviour on the part of any of the interviewees or any differences in the interview circumstances were noted down. An interview journal was kept during all cases, but especially meticulous notes were taken when the enhanced interactive interview method was used, as that method allows fewer opportunities for interpretation interview notes.

An interview journal was also used in the data analysis. Some of the events mentioned in the journal supplemented the results of the data analysis.

3.5 Data analysis

Data analysis of the three case studies was done following the guidelines provided in the methodology literature (e.g. Eisenhardt 1989; Miles & Huberman 1984; Rubin & Rubin 2005; Yin 2018). Due to the slightly different natures of each case, and the differences between the data collection methods, there are also some differences in the analyses. However, the rich case study material provided excellent opportunities for analysis and interpretation.

In all three cases, the data analysis was done in cooperation with other researchers. Benbasat et al. (1987) noted that this can improve the capturing of the richness of data collected and allow more accurate inferences to be based on them.

3.5.1 Breast Cancer Case

Data analysis in the Breast Cancer Case was slightly different from the Maritime Cases, as the main contribution was created during the workshops. We did not use any software to analyse the interviews, but the case and related information systems were illustrated as case write-ups. The illustrations served as the starting point for the workshops in focus groups, together with the theoretical tools created beforehand.

To sum up, data analysis in the Breast Cancer Case was conducted when the tools were used during and after the workshops, based on the discussions. The results were then sent to the participants, who had the possibility to comment, request changes and to accept the results.

3.5.2 Maritime Cases

In the two Maritime Cases, the data analysis was conducted following the case study data analysis guidelines (Eisenhardt 1989; Eriksson & Kovalainen 2008; Yin 2018). Different interview methods were used for these cases, which caused some differences in the analysis. To start with the case study data analysis, detailed case study write-ups (in the form of notes and figures) were written (Eisenhardt 1989) and then discussed with the other researchers involved in the cases. This helped to create an initial understanding of the cases. That understanding was then deepened by going through the interview data and starting to build nodes (in the nVivo software package) and classifying the statements from the interviews into these nodes (or codes (Miles & Huberman 1984)). Based on the initial theoretical understanding and the analysis of interview material, more detailed and sharper constructs were then developed. The whole process was iterative, making use of several types of material in addition to the interviews, e.g. brochures, web-pages, newspapers, interview journals.

When the enhanced interactive (multi-stage) interview method was used, the data analysis with nVivo was more straightforward, as part of the interpretation had been done already while conducting the interviews. The transcripts were also different and easier to analyse, as the answers to the interview questions were already classified under the questions in the interview situations. There are both strengths and weaknesses to this method with regard to the data analysis. On the one hand, the analysis is faster, and to some extent easier, as part of the analysis has already been done during the interviews, together with the interviewee. Verification of the initial interpretation is requested from the interviewees after the interview session, which strengthens its reliability. On the other hand, there is a high risk of disregarding some important points the interviewee has said, as when typing the responses immediately on the screen, the interviewer is to some extent leading the discussion. This makes the interview more focused, but may also limit the statements of the interviewee. We had tried to compensate for that limitation by encouraging the interviews to add anything they see as important both during the interviews and afterwards when they were sent the initial scripts for approval.

For the more traditional semi-structured interviews, the analysis process was conducted as described at the beginning of this section. The interview material was more extensive, as the interviews were transcribed verbatim. The analysis was more complex and required more mixing of the themes that were discussed and picking up points from here and there. For the semi-structured interview method, the analysis problems presented by Miles and Huberman (1984) became more prominent; there were multiple forms and sources of data, and everything seemed important and interesting. In analysing the Maritime Case 2 data, initially there were huge amounts of nVivo nodes or codes that needed to be classified again. Codes were sorted, weighted and compared, and in some cases integrated, modified or even deleted (Rubin & Rubin 2005). Some of the data and the respective codes fell outside the scope of the research for this dissertation and were thus left out of any further analysis.

3.6 Summary

The data collection and analysis methods used in the research articles and cases are summarised in Table 5. Article I is not part of any of the empirical cases presented in this dissertation, but rather a conceptual paper that describes the governance of data phenomenon and proposes a framework. However, it includes results from a quantitative survey that was conducted by the first author. This survey is not part of this dissertation, and thus the method is not covered in this chapter.

The other three articles involve qualitative data collected for the purposes of this dissertation. For Article II (the Breast Cancer Case), I did the data collection partly by myself (an initial interview and a focus group discussion), and partly together with the co-author of the article. A master's thesis student was present in one of the focus group discussions. Data analysis was done together with the co-author.

For Article III (Maritime Industry Case 2), qualitative interview data was collected by me together with the third co-author, and two of the interviews were done by the third co-author alone. I did the data analysis for the article by myself. The quantitative interview data in Article IV consists of interviews I conducted in the maritime industry (17 persons interviewed; both authors were present for one of the interviews) and interviews done by a master's thesis student under the supervision of the co-authors in the bio-refinery industry. I did the initial data analysis by myself and discussed the results with the co-author.

Table 5. Summary of methods in articles and cases.

	Article I	**Article II**	**Article III**	**Article IV**
Case		Breast Cancer Case	Maritime Industry 2	Maritime Industry 1
Title	A Framework for the Corporate Governance of Data – Theoretical Background and Empirical Evidence	Empowering citizens through data interoperability - data federation applied to consumer-centric healthcare	Data governance in digital platforms	Willingness to share supply chain data through platforms in digital business ecosystems – an interview study in two ecosystems
Authors	Tomi Dahlberg Tiina Nokkala	Tiina Nokkala Tomi Dahlberg	Tiina Nokkala Hannu Salmela Jouko Toivonen	Tomi Dahlberg Tiina Nokkala
Forum	Business, Management and Education journal	Finnish Journal of eHealth and eWelfare	AMCIS 2019 conference proceedings	Bled conference proceedings
Year	2015	2019	2019	2019
Research aims	To propose a framework and discuss its theoretical basis.	To illustrate the data federation tools and show how they can be used to empower citizens.	To extend the single-organisation data governance models to the platform environments.	To find out the factors promoting and preventing data sharing and what data organisations are willing to share.
Data collection	Conceptual paper	Qualitative: specialist interviews and focus group discussions	Qualitative: semi-structured interviews (10 interviews, 12 persons interviewed)	Qualitative: enhanced interactive (multi-stage) interviews (17 interviews in maritime, 8 in bio-refinery)
Analysis method		Content analysis	Content analysis	Content analysis

4 Articles

This chapter presents the main findings of the research articles included in this dissertation and their relations to the whole. Each article is from a different research project, even though all of them belong to this research stream. The selection of these articles was based on their centrality to this dissertation topic and on the share of contribution in the writing of the articles. In section 5.4, other results from the case studies are presented.

4.1 Article I. A Framework for the Corporate Governance of Data – Theoretical Background and Empirical Evidence

This article is the first article we published on our research on governance of data. The objective of the article is to justify the need for governance of data and to explain the theoretical idea of how different ontological stances affect the way data is governed and managed. The context in which governance of data is discussed in the article is ageing societies and elderly citizens. However, it is noted that governance of data issues are fundamentally similar in almost any other context.

4.1.1 Findings

In the article, we propose a framework for the corporate governance of digital data. The need for such a framework is argued for based on the idea that there does not seem to be any data governance framework that addresses data governance as a corporate governance-related issue. It is also noted that in our view data governance is a managerial issue rather than a technical one.

With respect to the corporate governance side of data governance, we argue that, as investments in an asset such as data must be assured to produce returns, our framework is needed (Shleifer & Vishny 1997). Taking the governance of data as the responsibility of the corporate management level and giving clear guidelines for treating the data as a corporate asset in any situation in which this is reasonably possible helps managers to act on data in an agreed way.

In the article, we also consider the distinct of defining data, i.e. contextual and universal. The framework we propose assumes contextually defined data. We use the thinking of Wand and Weber (1993, 2002) to justify the different data ontologies and increasing need of more data storages.

The proposed framework, given in Figure 12, consists of three building blocks that illustrate different domains for the governance of data. One block describes the data types to be considered; one presents the data sources, and the third illustrates data's internality-externality and structured-unstructured dimensions.

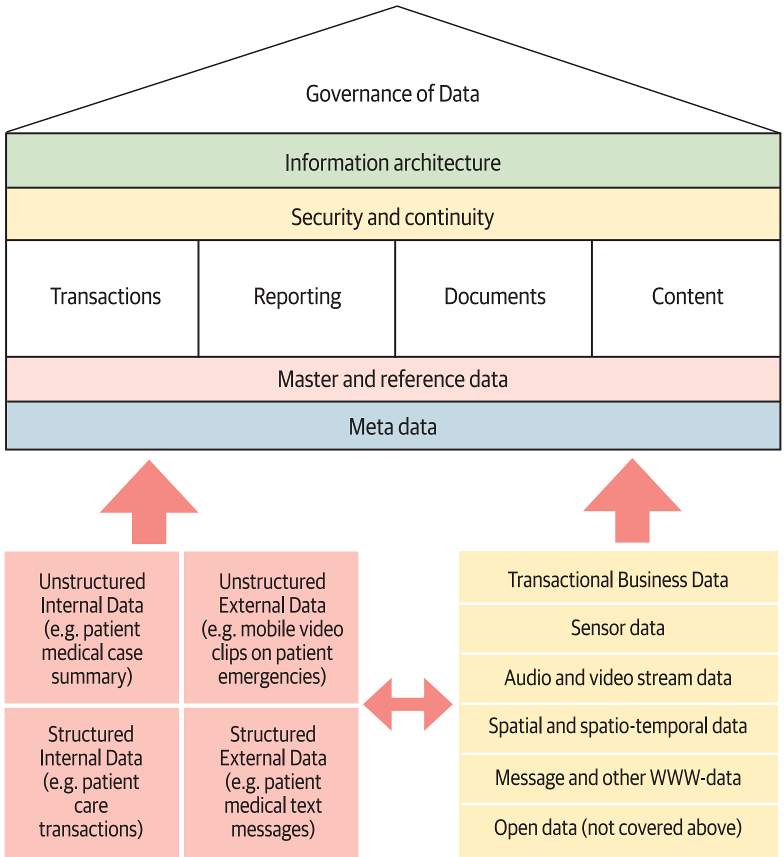

Figure 12. Framework proposed in the article.

4.1.2 Relation to the whole

The article answers RQ1 by justifying the need for data governance frameworks. It also explains the main assumption behind this dissertation: that data is contextually, not universally defined. This assumption is in the backdrop of the whole dissertation and affects all the thinking presented in my work.

4.2 Article II. Empowering citizens through data interoperability - data federation applied to consumer-centric healthcare

In Article II, we present a practical tool for data federation based on the theoretical framework presented in Article I. The aim of the article is to address three research questions on the federative approach, i.e. does the federative approach support the interoperability of citizens' healthcare data? (RQ1), is the federative approach able to support the empowerment of citizens in accessing and using healthcare data? (RQ2), and what benefits does the federative approach offer to citizens in the use of healthcare data? (RQ3).

The research context for this article is citizens' healthcare data, namely diagnostic and treatment data on patients who have breast cancer that has spread widely. The federative tool is used to make interoperable that data to forecast the effectiveness of treatments and survival rates. In the article, data federation is discussed in the context of empowering citizens with respect to their own healthcare in general, and the Breast Cancer Case is used as an example of data federation. The same case study has been used in other articles as well, from different perspectives (Dahlberg et al. 2018, 2017).

4.2.1 Findings

The use of data federation matrixes (Figure 13 and Figure 14) is explained in the article, and the matrixes are found to be suitable for data federation purposes. The ontological stance behind the use of the federation tool is described and justified. The theoretical background regarding citizen empowerment in healthcare and health literacy is reviewed and used to justify the suitability of the federative approach for this purpose. The research question is answered on the basis of the conceptual study, and the answer is that the federative approach supports the interoperability of healthcare data. We also conclude that by using the federative tool, citizen empowerment can be supported, and the use and understanding of healthcare data by citizens can be increased (RQ2). The benefits of the federative approach to the citizens are the potential convenience of combining healthcare data and using it to communicate with citizens; the increased possibilities of understanding the different views of the real-world situations from the perspectives of the healthcare professionals and citizens themselves and the potential for supporting citizens better with interoperable data from various sources (RQ3).

We also note that making data interoperable is needed to integrate data from the growing amount of data sources (e.g. in healthcare), and that it is not enough to pour all the data into a "data lake" or common data repository. Data interoperability requires linking the data from their original sources and using (IS technical, information processing and socio-contextual) metadata to explain the meaning of the original data in its various use contexts.

	Patient IS	Laboratory IS	Surgical IS	Radiotherapy IS	Pathplogy IS	Information System N
Social security identification	X	X	X	X	X	X
(Cancer) diagnosis code	X	X	X	X	X	X
Tumor node metastasis (TNM) code	X	X	X	X	X	X
Date of events	X	X	X	X	X	X

Figure 13. Matrix tool to identify the shared data attributes in information systems.

Personal ID / Diagnosis / TNM code	Patient IS	Laboratory IS	Surgical IS	Radiotherapy IS	Pathplogy IS	Information System N
IS Technical metadata						
– Description						
– Field name/ lenght etc.						
Information processing metadata						
– Place of initial data entry creation						
– Creator of data entry						
– Responsible for data storage						
Sociocontextual metadata						
– Meaning when saving patient data						
– Meaning for treatments						
– Meaning in diagnosis						
– Meaning for medication						

Figure 14. Matrix tool to collect together the metadata properties of the shared attributes in each information system.

4.2.2 Relation to the whole

The article provides an answer to Research Question 2 regarding how data can be made interoperable when it has varying characteristics. The article also contributes to answering Research Question 1, as it shows the importance of data federation, which is a consequence of governance of data.

4.3 Article III. Data governance in digital platforms

This article takes a stand on how data should be governed on platforms. Based on data governance frameworks for single-organisation purposes, we first propose an analytical data governance framework for digital platforms. Then we present results from 10 interviews conducted in a shipbuilding network and at their major customers. The research question for this article is: how should data governance models for a single-organisation context be extended to fit them into inter-organisational digital platforms?

The context for the study is a research project conducted in a shipbuilding network consisting of shipbuilding partners, their sub-contractors and their customers. In the research project as a whole, the aim was to investigate various aspects of automated ship operations. In our part of the research, we investigated the possibility of integrating data that are created and used during the ship's operations using a shared digital platform.

4.3.1 Findings

Based on our analytical framework, the analysis of the interview material supported many of the domains we included in our framework. The most significant differences in data governance in single-organisational contexts and on platforms were in domains of ownership and data access, usage and value and data stewardship. This illustrates the differences between the contexts.

On the practical side, this article reveals that there is a need for a data governance model in establishing a data-sharing digital platform. A clear business model is needed when a platform is implemented. Organisations understand the value of their own data quite well, as well as the added value that is available when the data are integrated from different partners and used to create new services for the customer.

It is also noted that in this specific case, all parties highlighted the fact that the owner of the data is the main customer, the shipping company. This is a fact that can either complicate or simplify the situation in data sharing.

4.3.2 Relation to the whole

This article answers to the main research question regarding how a data governance framework can facilitate inter-organisational data sharing on a shared digital platform. A supporting answer to Research Question 1 is also provided, as it is noted that there is a need for the data governance model.

4.4 Article IV. Willingness to share supply chain data through platforms in digital business ecosystems – an interview study in two ecosystems

In this article, we present factors that promote or prevent data sharing on a digital platform within a maritime ecosystem and within the bio-refinery industry ecosystem. The theoretical background of the factors impacting willingness to share data is reviewed, and results from the interviews are compared with the theoretical background. Another theme handled in the article is what data organisations are willing and unwilling to share on a platform.

Altogether, 25 sourcing and accounting specialists were interviewed, 17 from the maritime industry and 8 from the bio-refinery industry. Material was collected in two different contexts, but we claim that these two industries can be treated as one case study, as the subject of investigation—the supply chain and its automation—is the same, as the platform is developed for the use of both industries.

4.4.1 Findings

We found 11 factors promoting data sharing and 9 factors preventing willingness to share data on a digital platform. Control of (supply chain) processes, further development of the processes, resource savings, data quality and external pressures were found to be the most important promoting factors. Also, an interesting concept, situational opportunity, came up in our interviews and interview journal. This refers to the global situation that unexpectedly allows the development of the inter-organisational supply chain processes within the ecosystem. To our surprise, factual cost savings were not mentioned as being among the most important factors. Internal factors, risks, trust, situational factors and costs were found to be the most important factors preventing data sharing on digital platforms. Concrete benefits and clear measurements for them were insisted on, as the data-sharing platform needs to be well-reasoned in its presentation to the governance bodies of organisations. On the other hand, some organisations were worried about their own, old-fashioned information systems and poor internal data quality.

Detailed product data was seen as very sensitive, and organisations expressed the need to keep their know-how and competitive advantage-related data to themselves. Moreover, sensitive internal data and business-sensitive drawings were also considered data not to be shared. Still, planning materials, project schedules and instructions were seen as sharable. Organisations were also interested in sharing invoices and payment data through the platform.

In the article, we describe the creation and division of systematic business value on an industry level digital platform. Collaboration and mutual coordination and governance of data are needed to achieve the expected benefits from the implementation of a supply chain platform.

4.4.2 Relation to the whole

The article contributes to the main research question by uncovering the factors that affect data sharing on digital platforms. From the contributions of this article, we note that value creation and division on digital platforms is possible, but it requires shared governance, also with respect to governance of the data that is shared on the platform.

5 Findings

Each publication included in this dissertation makes a unique contribution. In this chapter, I will connect those conclusions and answer the three research questions (i.e. one main question and two sub-questions). The ultimate finding is the proposed model for governance of data on platforms presented in Figure 15. That model includes the main concepts to be considered in governing the platform data. The research projects and the empirical studies I have conducted have yielded some other results as well. Those results will be briefly introduced in section 5.4

5.1 Research question 1

The first sub-question is *"Why are governance of data models needed for inter-organisational data sharing platforms?"* The need for governance of data models has been argued for in Article I with respect to the single-organisation context and in Article III for the inter-organisational context.

On a theoretical basis, the reasoning for the need for data governance models is based on the fact that data governance is needed to assure that the governance body of an organisation is aware of, understands and properly exploits and protects the data assets that the organisation holds. The stakeholders who have invested in the organisation expect returns on their investments. In the case studies, especially in the maritime cases, the varying attitudes towards understanding the value of data were visible. Many of the interviewees stated that most actors collect data on their operations and on the operations of the devices they produce. However, they acknowledged that not all parties understand the value of that data, and the collection is done on a nice-to-have basis. In the Breast Cancer Case the context was different, and the value of patient data was more self-explanatory. Data were collected and often used in planning and research. However, the breast cancer case also yielded the finding that a governance of data model is needed if data is to be shared among different stakeholders.

In the platform context, data governance models have a slightly different bearing than intra-organisationally. Supposedly, each organisation participating in a platform already has some data governance model for internal use. When data are then shared, integrated and managed inter-organisationally, an agreement on the governance model is needed. In Maritime Case 2, interview analysis yielded nodes such as "data ownership", "data sharing" and "integrator role" (some quotes from the interviews are presented in Article III). These nodes were characterised by the question of governance of data: an agreement was needed, e.g. on data ownership and data sovereignty in order for data to be shared. Many of the organisations were willing to act as integrators or platform providers but were unwilling to let others take that role, which led to the conclusion that a governance of data model is needed, no matter who the integrating partner is. Whereas intra-organisationally at least some authority over all the data can be presupposed, on inter-organisational premises the general authority must be stated. Data stewardship, ownership, access, and definitions must be agreed on, considering the prevailing regulatory and business environments. Also, an agreement on a business model, as well as the cost and benefits allocation, is needed. In addition to Maritime Case 2, these issues were also present in Maritime Case 1 as a reluctance to rush headlong into the automation of data sharing. A business model for the automation platform was requested in many of the interviews (some quotes from the interviews are presented in Article IV). Interviewees were also worried about limiting access rights so that only the partners who were really in need of some data could access it. These kinds of issues can be agreed upon in the governance of data model.

When integrating data into a digital platform, new needs for metadata emerge. Supposedly, many organisations have overlapping data on the same subjects, and when sharing those data, explanations of the meanings are needed. These metadata constructs can be included in the data governance structures. When asked about this, many interviewees—especially in Maritime Case 1 but also in Maritime Case 2—described the contextual differences of many of their data domains. For instance, the prices of product parts differ even between the internal units of the organisations (e.g., some departments use the price of raw material, whereas others also include other costs in the price of certain product parts, and the prices also vary depending on their arrival times and batches). For inter-organisational purposes, many felt that it was impossible to share their price data at all, expect their selling prices. These statements have informed the need forof metadata to be included in the governance of data model for platforms.

However, each platform has its unique context, and thus it is not possible to define an all-inclusive data governance model. A generic model considering the main aspects, such as data types and metadata and the main decision domains is proposed in this dissertation.

This quote from Tan, Pan and Zuo (2018) illustrates how important it is to agree on effective management structures in order to gain a business advantage from the use of a multi-sided platform:

> Our premise is that systems integration itself may not always be sufficient for business integration, especially if the systems integration was overinvested or wrong. The results of this study confirm that the effective management of interdependencies enables firms to increase their capacity to respond quickly and effectively to market forces, to improve the quality of conformance to customer requirements, and to reduce costs.

With the existence of such general models, it is easier for both individual organisations and ecosystems governing platforms to make the agreements.

5.2 Research question 2

For the second sub-question: "*How can data from distinct organisations be made interoperable, when the original data varies in terms of critical attributes?*" the answer was presented in detail in Article II, where we presented the tools that we used to make interoperable the breast cancer data. The idea behind these tools (both matrix tables) is in original practical tools used by a local company providing digitalisation services. Even though the Breast Cancer Case is just a small example of such data integration, one can speculate that tools and ideas similar to those used in this case be used in other, larger contexts as well. This is also backed up by the fact that the tools are originally SAP-based. In the use of the tools, information systems can be added one by one to the data federation, adding similar key attributes and definitions and linking them to the earlier ones. Also, on the digital inter-organisational platforms, where the original data comes from legally separate entities, it is possible to use these tools.

Table 6 illustrates a tool that can be used for identifying the shared data attributes in information systems for the federation. The number of shared attributes needed for data federation depends on the case and context, as the need to make data interoperable is different in every case. In our Breast Cancer Case, we had first three shared attributes but added one during the process to make a distinction between different periods of events. First, three shared attributes (social security identification, cancer diagnosis code and tumour node metastasis (TNM) code) were quite easily found in our first workshop with the specialists as we introduced and used the tools.

In the Breast Cancer Case, the question to be solved guided our use of the tools quite clearly. To find individual patients with breast cancer and to make the data interoperable between information systems, the patient's social security identification was needed to define which patient we were collecting data about. Furthermore, we needed information on whether the patient had breast cancer (diagnosis code) and whether the cancer had created metastases (TNM code). The two latter codes were used in recognising the data entries related to cancer diagnoses and treatments in the patient information systems, as the same patients could have been in the same systems for many different reasons; e.g., in the laboratory IS, there could be data not only on a patient's breast cancer-related blood and other tests but also on their other diseases as well. However, the date of the events was added to the shared attributes list, as some of the, e.g., laboratory tests could be similar for many diseases, and with the date, we were able to confirm that test results were related to breast cancer events.

Table 6. Tool for identifying the shared attributes in information systems for data federation.

	Information system 1	**Information system 2**	**Information system n**
Shared attribute 1			
Shared attribute 2			
Shared attribute n			

After identifying the shared attributes using the Table 6 tool, the metadata of these attributes need to be described.

Table 7 illustrates the matrix tool that can be used for that. The metadata items in Table 7 are presented on a general level, and in the use of the matrix tool, the amount of metadata items can be increased or decreased. Our aim in developing the tool was to create general-level tools that can be modified for other use contexts.

The aim of the matrix tool is to describe the meaning of each shared attribute in each federated system in such detail that when the federation is used to gather together data from the various information systems for analytical, reporting or any purposes, the meaning of the data can be understood in each system, and the possible different entries for the same object can be explained. For example, if the name of the customer is entered differently in different information systems, the metadata explains the process that is used in each system, and the name of the client can be verified.

The process of identifying and describing the shared attributes can happen in various phases. It is possible to start with just two systems and add more in later phases. The same goes for adding the shared attributes; however, adding new attributes requires revisiting the information systems that have already been included in the federation.

A specific metadata repository is needed for storing the information collected with the matrix tools. This repository can also be used for storing the created linkages of data attributes and information systems. The technical execution of such a repository is beyond the scope of this dissertation.

Table 7. Tool for describing the metadata of each shared attribute in each information system for data federation.

Shared attribute 1/ shared attribute 2/ shared attribute n	**Information system 1**	**Information system 2**	**Information system n**
Technical metadata			
description			
platform			
field name			
field value			
field type			
field length			
indexed			
obligatoriness			
standard used			
min. length			
max. length			
duplicates allowed			
structure			
date style			
number of decimals			
Information processing metadata			
place of creation			
reference used			
maintenance			
storage			
creator of entry			
editing rights			
responsibility			
Socio-contextual metadata			
meaning in context 1			
meaning in context 2			
meaning in context n			

The most important aspect of these tools is the use of three types of metadata: IS technical, information processing and socio-contextual metadata. The use of metadata enables one to overcome the issues of different formats, contexts and definitions. When the meaning of the data attribute is explained in the metadata on three or more levels, and the meaning and links are stored in the metadata repository, it is possible to follow the data to its original sources and understand its meanings and contextual nuances.

5.3 Main research question

The main research question of my dissertation is "*How does governance of data facilitate data sharing on inter-organisational platforms?*"

To answer that question, a framework combining different data governance aspects has been crafted (Figure 15), illustrating what governance of data is in inter-organisational platforms. The parts—or building blocks—of the framework originate from the existing literature, the case studies I have conducted, articles belonging to this dissertation and the answers to the two sub-questions.

Platform context, data types, metadata (of the original data), platform business models, platform governance models and data governance-related issues form the main parts of the framework. Data types and metadata are important concepts in the single-organisational contexts, but are perhaps even more important in the platform context, in which data need to be governed inter-organisationally. Thus, the **data types** and **metadata** blocks originate from Articles I and II, the Breast Cancer Case, as well as from the literature review. The data types were also discussed during the Breast Cancer Case workshops with data specialists in order to confirm that our classification was seen as reasonable. Consequently, the data and metadata types in the framework blocks originate from the Breast Cancer Case,, which is why they were included in the framework in Maritime Case 2.

The platform context is a vital issue for data governance on a platform. In addition to its discussion in the context of the literature review, the **platform context** was also discussed in the two Maritime Case studies. In the Maritime Case 2 interviews, the very specific context of the case arose in every interview. The customer who owns the cruise ship is the owner of all the data. In addition, the maritime industry is quite regulated, e.g. concerning emissions and routes. Many interviewees talked about how the specific features of the industry affect the prospective data sharing platform. However, the maritime industry is by no means unique in its special context. On the contrary, most platform environments are affected by specific requirements, and thus, I have included the platform context in the framework.

For the **platform (data) business and governance model**, the building blocks originate from various sources. A quote from one interview in Maritime Case 2 summarises the importance of a clear business model: "Whose business is it, who benefits and in what way?" Also, in Maritime Case 1, the business model and "the winner" was often discussed, often in relation to the price of platform participation. Having a clear business model and trusted partners to manage the platform actions is crucial. In the maritime cases, different business models were proposed, e.g. the alliance-driven platform and the keystone-driven platform. However, various issues affect the way the governance and business model can be arranged, and the most important issues are included in the framework under the platform (data) business and governance model block. As Lee et al. (2018b) noted as their fourth principle, different levels of governance control are needed based on the **regulatory environment**, data strategy quality, platform strategy and other configurations. In the case studies, the regulatory environment of the Breast Cancer Case was totally different to the maritime cases. However, the maritime cases were different from each other as well. Maritime Case 1 was about shipbuilding, where most important regulations in relation to the platform were about commerce and competition. In Maritime Case 2, the regulated side was for reporting many operational aspects. It is important to consider these issues in the business model to make sure that the requirements are met. The level of trust between the members and the openness of the platform affect the control sharing. The role of trust, however, can be diminished by using blockchain or other distributed ledger technologies on the platform. **Contribution measurement**, **cost allocation** and **benefits division** were discussed in the literature review (see section 2.5), and in more detail in Maritime Case 2. These factors contribute to the plausibility of the platform and facilitate sharing data when the agreements over them have been made. In the Maritime Case 1 interviews, the cost allocation factor was most often mentioned, as the plan was to establish an alliance-driven platform which would be owned by all the partners. In Maritime Case 2, the contribution measurement was discussed regarding how the benefits are shared equally if all participants benefit from the shared data. The interviewees shared many views on how that could be arranged, and thus, this is an important aspect in the framework.

Platform alignment was an important aspect of Maritime Case 1, in which willingness to participate in platform activities was explained by the interviewees to be partly dependent on a sufficient amount of business being conducted on the platform. In both maritime cases, the interviewees representing the shipyard's subcontractors stated that, because they have other buyers as well, the platform should be well aligned with their business functions. The bigger the customer the shipyard was for the subcontractor, the more interested the interviewees were to adopt the platform. The smaller companies were different, as they felt they had no other choice than to adopt the platform if the shipyard decided to do so. In the framework, platform alignment also refers to the role of the platform in business, whether it is just a tool or whether it has a more significant role as a product itself. In the networks and ecosystems, the role must be mutually agreed upon. **Data stewardship** on platforms and as a part of platform data governance was discussed in Maritime Case 2. No matter the general governance model of the platform, there should be a party taking care of the data's lifecycle and sufficient metadata descriptions. In Article I and in the Breast Cancer Case, the metadata's importance was shown and underlined, but the creation of metadata connections and repositories can be laborious; thus, it is important to include the decision on data stewardship for a platform as part of the governance of data framework. In all three cases, the interviews showed how scattered and heterogeneous the data are in the information systems, and thus, the data steward role is an important one.

The blue part of the framework—**data governance issues on platforms**—consists of data related issues that need to be agreed on as part of data governance for a platform. They could also be named decision domains, following Khatri and Brown (2010). Some of the blocks are similar to those in the single-organisational data governance framework (see Figure 8), as they originated from the same sources and have also proved to be important in the platform context. However, in the platform context—as discussed in both Maritime Cases—the data management and data life cycle issues remain more the responsibility of the original data sources, i.e. the organisations participating in the platform activities. The same was actually noted for the **data quality** domain as well, but I decided to keep the data quality domain in the platform framework. The reason for this is the concern that emerged in the interviews, especially in both maritime cases, that other partners might not understand or might misunderstand the data due to their low quality. However, this may not be an issue of low quality, but rather a context issue that can be solved using metadata.

Value of data was a prevailing theme in all the case studies. Especially in the Maritime Cases, interviewees talked about the value of data and making a business out of data. However, as many interviewees especially in all three cases stated, the same data can have different kinds of value for different parties, and in a platform context where a common business model is needed, the value of data must be understood in the context of sharing and integrating them with others. It is not enough to give the data to others, as the value is only created when the data is used for something that creates value. This contributes to the business model, contribution measurement and benefits allocation in many ways. The value of data is also dependent on the metadata, as they explain the meaning of the data entries and can make them more or less valuable to the user depending on the context. For example, in the Breast Cancer Case, the value of a certain data entry can depend on who created it and under which circumstances. Data value can also be dependent on the other data available, as in Maritime Case 2, in which some interviewees stated that certain data are only useful, e.g. if the weather data are available for the same period of events.

In the review of the literature on data governance, for both intra-organisational and inter-organisational settings, **data provenance** was seen as important by many authors. Data provenance is essential in the platform context, and it is also related to the access rights of data. It was also discussed in all three cases, as the history of creation, usage, changes and deletion for each data entry is important in understanding the value and credibility of data. In Maritime Case 1, this was discussed quite a lot in the context of the blockchain options available for the case platform.

Shared data definitions are needed when data is shared and federated over the platform to prevent misunderstandings. However, as shown in the Breast Cancer Case and in Article II, that does not necessarily mean that all participants are forced to use same definitions and formats of data, but rather that integration can be facilitated using federative tools and making data interoperable. In any case, the data definitions must be agreed on as a part of platform data governance, and some (stewarding) party must be made responsible for them. **Data structure** is not only related to the data definitions but also to the openness of the platform. A clear and dynamic structure from the beginning helps new parties in joining the platform and starting to contribute. Structure also refers to the storage and infrastructure of a platform’s data assets, as well as to the policies and guidelines for data creation. Shared data definitions and data structure are more theory-based blocks in the framework. However, in the Maritime Case 1 interviews, these were discussed in relation to the automated data exchange in the order-delivery chain, where the messages were meant to be shared in a standardised format. Also, in Maritime Case 2, the standards were discussed, as some of the data, e.g. those produced by the sensors, must be reported in a standardised format.

Cybersecurity, hacking and continuity were seen as important issues with respect to **data-related risks,** especially in Maritime Case 1. In that case, the risks were often mentioned as prohibitive to data sharing. A platform’s security must be taken care of, and the security measures need to be communicated to the participants. Another data risk was seen in direct competitors having access to each other’s data. In Maritime Case 1, the interviewees were worried about giving up their competitive advantage if data were shared too openly.

That risk of losing competitive advantage is highly related to the **data ownership and access** block of the framework. Ownership and access are already supremely important factors in the single-organisation context, as they promote responsibility for data. Vilminko-Heikkinen and Pekkola (2019) and Silvola et al. (2011) related undefined data ownership and structural elements to non-existing policies and processes for data creation. A lack of clear ownership causes a lack of motivation to manage the data and related processes (Vilminko-Heikkinen & Pekkola 2017). The issue of the ownership of data is multifaceted, as was noted in the case studies. In Maritime Case 2, all the parties clearly announced the client as the owner of the data. However, in that case even the owner rarely had access to all the data. Thus, the ownership issue is not simple in any platform context, and must always be clarified with agreements. The access issue was similarly complicated in the pre-platform phase in Maritime Case 2, as access to data was granted and executed in various ways. The ownership and access block should also take into account the copyright and usage right issues of data (e.g. in the music or movie industries these are particularly important aspects), and the information about these rights should be included in the metadata.

Data governance over the issues and domains included in the framework can be executed in several ways. As discussed in section 2.3, governance mechanisms can include structural, procedural and relational—or contractual and relational—mechanisms. Otto and Jarke (2019) have also studied governance mechanisms in the early phase of platform design. Similarly to their study, I noted that it is already important to have different types of agreements and regulatory elements in the early stages and over the platform's life cycle. In the interviews I conducted, many partners expressed their insistence on having clear agreements on several issues on the platforms, referring to the structural or contractual type of governance. However, they also see relational mechanisms as important in communicating the acceptance terms for new joiners to the platforms, for example. In Maritime Cases 1 and 2, the creation of the platforms was in the very early stages and thus it was not possible to, for example, clearly state the ultimate ownership of the platform (but in Maritime Case 1, the plan was for it to be owned by an alliance company owned by the members of the ecosystem). Thus, it was understandable that, much as Otto and Jarke (2019) argued, that several types of regulatory instruments are needed, together with relationship-based controls. It can be concluded that it is not possible to agree on all the domains included in the framework (Figure 12, Figure 15) in one contractual agreement, but that different kinds of governance mechanisms are needed, based on the environment—both business and regulatory—and the platform's context, alignment and other issues. The governance mechanisms need to be dynamic and negotiable, as in both Maritime Cases it was clear that the ecosystem or network is not closed and that its members will change over time.

For the recognised promoting and preventing factors for data sharing (Article IV), governance of data offers solutions. Having a clear governance structure helps to reduce risks, strengthens trust and helps to control costs. Governance of data on a platform promotes controlling the whole (e.g. supply chain) process and streamlines it. Resources are saved when data is clearly linked between organisations. Even internal data quality can be improved, as data's *fitness for use* (Otto, Hüner, & Österle 2012) is ensured in communication with others. That, of course, might require some additional resources for stewarding the data quality.

The value creation and the value of using the data become more prominent when data is governed based on a clear structure. How a platform's business model treats data, e.g. how the contribution is measured and how the benefits are distributed, can facilitate data sharing. In the case of a supply chain data-sharing platform, consistent data and clear business cases of expected benefits are promoting factors for data sharing (Fawcett et al. 2007; Rai et al. 2006).

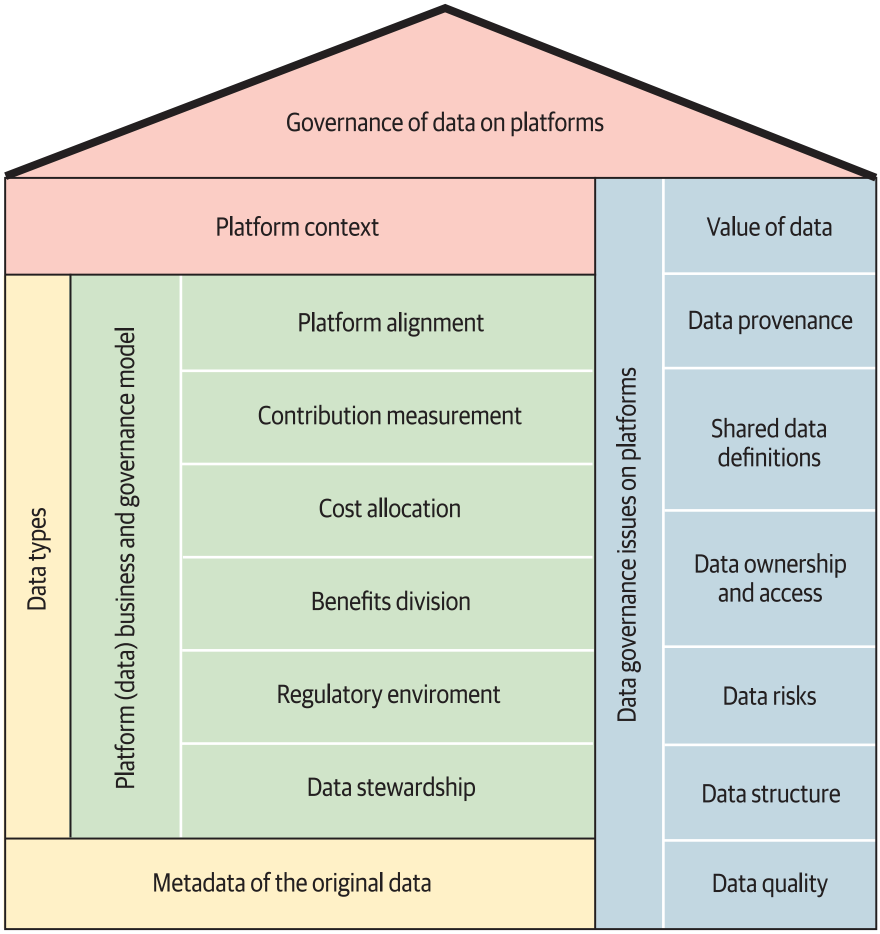

Figure 15. Proposed model for governance of data on platforms.

Many actors acknowledge the value of the data itself and speak about this, but yet, this often leads to no action whatsoever. As a motivation for this study, this notion has become clearer during my research. Thus, the value of data is included in the framework, as it should be clarified in the first place when implementing data governance in a platform context. Clarifying the value of data also facilitates sharing the data on the agreed-on basis.

Data types relate strongly to the data structure. When the original sources of data are not within a single entity but scattered within an ecosystem, the structure of platform data is very complex. Metadata must be used to make the data on the platform interoperable (see Article II). Metadata is also used for shared data definitions that allow usability of data for the purposes of various stakeholders.

Even though this is mainly an issue of data governance, the solution—the framework—also offers answers to other governance issues, such as the free-rider problem on platforms (Eisenmann 2008), as the framework takes into account the division of costs and benefits, and contribution measurement.

5.4 Other findings

In Maritime Case 1, the interviews generated interesting findings that have not yet been reported, as Article IV focused mainly on the willingness to share supply chain data. In addition to factors affecting data-sharing willingness and data that organisations will and will not share, issues like commercial transactions data, e-invoices and technical product data were discussed with the interviewees. Here I will present a brief overview of those results.

Interviewees were asked about their organisations' willingness to automate data sharing on commercial transactions. The willingness to automate the sharing of individual transactions is illustrated in Figure 16. It should be noted that in general interviewees were willing to automate these transactions, and they saw many benefits in automation. In particular, invoicing and ordering-related transactions were seen as supply chain stages to be automated.

If the transactions discussed were automated, that would reduce the need for manual work and decrease the risk of typos and other errors. Information security and employees' resistance to automating their jobs were seen as negative sides of automation.

In order to automate these transactions in the supply chain, most of the interviewees thought that all the partners in the ecosystem should use the new standards and automated processes and that the common good must be put before individual benefit. However, the price of such a new system, the need to maintain personal relationships, the cost of cybersecurity measures and the number of transactions executed by each company (as the received benefits depend on the amount of trade) were seen as possible preventing factors for implementation of automation.

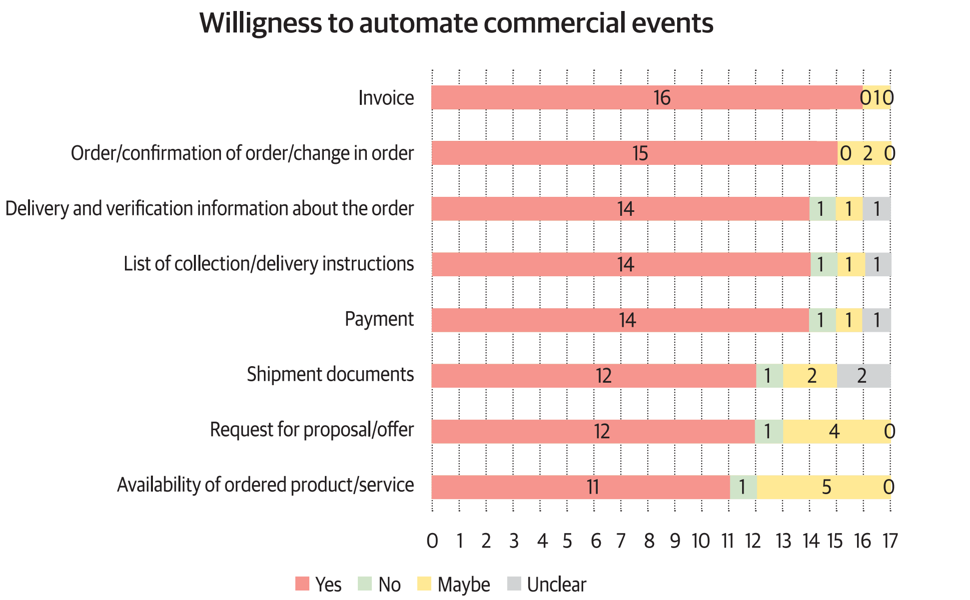

Figure 16. Willingness to automate commercial transactions.

Regarding electronic invoices, interviewees were asked about the current use rate (interviewees' estimate), factors that could increase the use of e-invoices and reasons for not using them in all situations. Also, the effects of e-invoices were discussed. The estimated use of e-invoices was 90–100 % (only one interviewee said that about 20 % of invoices are electronic). Supply chain processes were seen as having become faster and smoother due to the use of e-invoices, as less manual work is needed, leading to cost and time savings. In many cases, the use of e-invoices was demanded by business partners. However, the needs of business partners—especially those outside Finland—were seen as a reason for not always using e-invoices, as some small companies (and in international business, the partners) are not able to handle e-invoices. In some cases, the reason for not using e-invoices was that the systems did not support all required attachments. Some interviewees also reported that some of their partners did not understand what e-invoices really were, as they thought that scanned pdf-documents were e-invoices.

The exchange of technical product data as part of transactions within the supply chain was discussed in the interviews. Many interviewees reported that they already exchanged such information via e-mail or product catalogues. Changing the format and medium for doing that was not seen as a problem. The benefits of having such data in digital format for later usage and archiving purposes were seen clearly. Attaching technical product data to business transactions could also speed up the data-sharing process. Factors preventing such data sharing included the fear of providing access to the technical product data to competitors and harming the organisation's competitive advantage.

6 Discussion and conclusions

In this chapter, I will summarise and discuss the results and contributions of my research. The limitations of this study and suggestions for further research are also presented. There are both theoretical contributions to the scientific community, especially the information systems science field, and more practical contributions to be used in practice communities.

The research has concentrated on several aspects of data governance and data sharing. To start with, the basic idea of data being contextually or universally defined was gone through, and I have declared that the contribution offered here is based on contextually defined data thinking. Also, the value of data in general seems to be well understood on the speech-level, but that valuation is not translated into action often enough.

Next, I have discussed the digital platforms and ecosystems using and running the platforms. Referring to de Reuver et al. (2018), platforms have both theoretical and practical importance in contemporary business. However, the governance of the platforms and the data on them has hardly remained defined (Tiwana et al. 2010). Thus, I have taken data governance in the single-organisational context to the platform level and figured out what are the special issues in platform-level data governance. As a result, a framework is proposed.

6.1 Theoretical contributions

The results of this dissertation provide a theoretical contribution that is divided into three parts. First, I describe the federative approach and data federation tools involved; second I present the conceptual framework for platform data governance and discuss the implications of data governance for the platform business model; and the third contribution joins the discussion started by Wand and Weber (1988, 1990, 1993, 1995) by updating their thinking for the era of networks and numerous contexts for data interpretation.

The theoretical contributions of my dissertation consist of both supporting and advancing theory. The case study analyses and results from the three case studies support the data governance literature and data interoperability research. New tools for data interoperability are proposed. Furthermore, the case studies shed light on digital platforms and the ecosystems operating them, as well as on data governance related to the platforms. The results from both maritime cases propose new concepts and features for the data governance theories, and a framework for the governance of platform data is compiled. The discussion of the ontological nature of data is resurrected in light of the networked business and platform business models. Finally, my research lays the foundation for future platform data governance research.

6.1.1 The federative approach

The first part of the theoretical contribution is the demonstration of how the federative approach tools created enable data interoperability from scattered information systems. The requirement for this is that the meaning of the original data is known and expounded in three types of metadata. Data federation is then executed by recognising the shared attributes and making linkages using them. These shared attributes and their linkages act as an ontology for the shared and federated data by giving meaning to real-world things and events (Wand & Weber 1990).

The federative approach is opposed—and complementary—to the prevailing data management thinking (e.g. Dyché & Levy 2011; Mosley et al. 2010), where data is seen as universally defined, and integrations are often done by simply deciding the “one version of the truth” data entries and linking the rest of the data to them, in some cases even by replacing the rest of the entries with the “one version of the truth” values. This new federative approach allows different contexts of data to remain and to be understood even after the federation. That helps and simplifies data federation that draws on multiple information systems that are meant for very different purposes and contexts. In the literature review and findings sections, I have presented the theoretical basis for this thinking and argued for the use of data federation matrixes. The federative approach also answers to Abraham et al.’s (2019) call for metadata’s role in enabling data interoperability between organisations should be investigated.

The federative approach participates in the information systems science discussion on data interoperability and integrations in both intra-organisational and inter-organisational settings. The metadata types included in the tools complement those classified, e.g., by Khatri and Brown (2010), Cleven and Wortmann (2010) and Berson and Dubov (2007).

The data federation matrixes presented (Table 6 and Table 7) are dynamic and suitable for multiple situations, from intra-organisational data federations to larger-scale inter-organisational settings with some modifications. They can be used as a basis for creating a shared data ontology for a data-sharing platform. Having the growing networks and the change in the way the business is done in mind, this contributes to the very contemporary issue of using data to support the business, as well as running a data-driven business. The researchers in the data governance and platform research fields could find this approach interesting in creating and testing platform-based solutions for inter-organisational data sharing (e.g. S. U. Lee 2017; S. U. Lee et al. 2018b; Otto & Jarke 2019).

The presented data federation matrixes are based on practical tools provided by a company producing digitalisation services, and they are originally SAP-based. In our research, the theoretical foundations of these tools have been justified as they are suited, e.g. to mapping from the ontological to the design constructs through different lenses presented by Wand and Weber (1993). Federative interoperability also promotes semantic and pragmatic interoperability and enables technical and syntactic interoperability, as discussed by Janssen et al. (2014). Regarding Linked Data (Bizer 2009; Heath & Bizer 2011), the concepts seem to be fairly similar; Linked Data is a practical application of the federative approach to Web-based data sharing and reusing. Furthermore, the federative approach to data interoperability could benefit the data space architecture research (Otto & Jarke 2019) by providing practical tools for connecting data between data providers. The metadata repository proposed as a solution to store contextual information about the interoperable data could be used in Linked Data solutions and in data spaces.

The federative tools presented in this dissertation are aimed at very general use. However, in the Breast Cancer Case the case study context was health care, and thus, the usability of the matrixes there has been justified. In other contexts, the attributes and metadata needed may be different, but the basic idea remains the same. The case study has shed light on the different metadata types that are needed in making data interoperable.

In summary, my research provides a contrasting approach—the federative approach—to the prevailing "one truth" way of data integration. Still, I do not claim that the "one version of the truth" thinking can be fully replaced by the federative approach, but I think they can both co-exist and serve different purposes.

6.1.2 Platform data governance and data business model in business networks

The second part of the contribution relates to the first part, as data federation is one practical way to implement data governance principles. The actual platform data governance framework proposed in my work (Figure 15) combines the intra-organisational data governance frameworks (e.g. Khatri & Brown 2010; Rosenbaum 2010; Weber et al. 2009), some of the first platform data governance frameworks presented in the literature and the findings from my case studies. The field of data governance on platforms is so new that there are only a few earlier theoretical frameworks (e.g. S. U. Lee et al. 2018b) available for comparison. The Abraham et al. (2019) framework includes both intra- and inter-organisational scopes, thereby being possibly the closest relative to my framework in Figure 15. My results do not negate these earlier frameworks but are complementary in the sense that I bring a new point of view to the governance of data. My case study results clearly show the need for such a viewpoint and for a platform data governance model, and they propose domains to be included.

Besides the data itself, platform data governance includes issues regarding the business model of the data platform, as shown in my case study results. The value of data and how it is measured; contribution measurement on data contribution; cost allocation; ownership of the data and the distribution of the benefits are among these important topics. Contribution to the platform cannot be measured only as the amount of data shared to the platform, but the business model should be based on the added value of the data. This indicates that platform contributions should generate income when data is used for business that adds value.

In that sense, the parties on the platform can have different entitlements to access, use, copy and distribute the data, as in all three case studies in this dissertation. However, owning certain data or having access to them does not ensure any profits will be realised, as many of the interviewees stated. Parties collect data "just in case" and perhaps integrate it internally on intra-organisation platforms and information systems (Maritime Industry Case 2, Article III). Still, greater value is often achieved through cooperation within the network, where others own data as well as sharing them to add value together. Who gets the benefits from the added value, and how the income is divided, are crucial issues.

On the other hand, data is seen as a competitive advantage, but without a clear business case and model, these trials do not add value. Usually external data are needed as well, leading to different entitlements over data. This shows how data are not valuable in themselves, but that their value lies in being combined with others. However, as Abraham et al. (2019) pointed out, we do not know much about how control over data and the ownership of data are ensured in inter-organisational relationships. If control and ownership are not managed carefully, it is difficult to show who should get benefits from the shared data. That point is thus taken into account in the framework I am proposing.

Following from the intertwining of models for networked business and data governance, my framework presented in Figure 15 consists of domains of platform context, data types, metadata, the platform (data) business and governance model and data governance issues on platforms. Each domain includes several issues that have resulted from my case study analyses. Here, the theoretical contribution follows from the use of initial theoretical propositions fortified and advanced through the findings of the case studies. The contribution is made to both the literature on data governance (Abraham et al. 2019; Khatri & Brown 2010; S. U. Lee 2017; S. U. Lee et al. 2018b; Otto & Jarke 2019) and the literature on digital platforms as enablers of networked business (de Reuver et al. 2018; Tiwana 2013; Tiwana et al. 2010). Understanding the interrelatedness of data governance and networked business is crucial for further developing these business models. Having set clear metadata structures for the platform data, the ownership and access issues can be solved by integrating information on them into the metadata.

For the networked business, the value of my findings is on the platform (data) business and governance model part of the framework. The governance model that involves both the business model and the data governance can enable and enhance data sovereignty that has recently been called for by both research and practice (Jarke, Otto, & Ram 2019; Otto & Jarke 2019).

The framework could be termed a middle-level theory (Rubin & Rubin 2005). The empirical case study material and relevant prior academic literature have provided concepts and processes that seem to be important in light of my research. To broaden the theory, more empirical cases are needed.

Another contribution to this matter is showing the need for governance of data frameworks. That has been done by explaining how these frameworks can be used to facilitate data sharing and also to ensure that organisations know what data they have and understand how valuable they are. It is also noted that no comprehensive framework for platform data governance currently exists, as suggested by Lee et al. (2018b).

These results are important due to growing and more complex business networks in which data play a very important role. For researchers, this amalgamation of data governance and business network thinking should offer a new line of reasoning with regard to the importance of data and where that importance comes from.

6.1.3 Contextual representation of data

The third, and, in my opinion, the most important contribution is the idea of the contextual representation of data. Data, as spoken of in this dissertation, are never objective, but a representation of someone's perception of the real world (Wand & Weber 1995), and thus, when interpreted, data need to have specification on that real-world context. In Article I, this is exemplified with the discussion of universal and contextual approaches and how they differ when data attributes in two or more information systems have the same meaning, or the same data attributes can have two or more meanings. A universal approach that is based on objective "one truths" replaces all these attributes with the "golden" values, whereas the contextual approach aims at keeping the original data attributes, explaining their meanings and linking them for use.

In a sense, this thinking is nothing new, as the same ideas were presented by Wand and Weber (1990, 1993, 1995) and Wand and Wang (1996) and are still valid. In my research, however, I have updated their thinking for an era in which there are a larger number of contexts for data's life cycle due to business networks and data platforms. Whereas in the past, the number of contexts and actors sharing data was controllable, today multi-sided and multi-level networks all have data on the same subjects with slightly different definitions.

This part of the contribution is quite theoretical and builds on the aforementioned sources. However, the analysis of my case studies supports my presumption that data are contextually defined, which has consequences on their governance. These consequences are on a larger scale in the platform contexts, where data are shared between organisations which all have their own contexts and which follow definitions for their data. In the search for more efficient ways to integrate the data, the need to understand the meaning of the context may have been forgotten. Thus, my research also contributes by broadening the levels of interoperability (technical, syntactic, semantic and pragmatic) presented by Janssen et al. (2014).

The ontological clarity and expressiveness of data in information systems are crucial when decision making is based on that data. The meaning of the data in the original contexts must remain when data are integrated from various systems and in various formats. By understanding this, both the researchers and practitioners can handle data governance issues—both intra-organisationally and inter-organisationally—in a different way.

6.2 Practical contributions

For the practical community, this research makes a contribution through establishing clear areas that must be governed as part of data governance on platforms. When establishing a new digital data-driven platform, this list or framework can be used for checking that all required data management, governance and other needed orchestrations are considered during the implementation and later in maintaining the platform. In particular, the notion of clear business models and income sharing principles based on not the amount of data contributed, but on usage of shared data is important. On the other hand, the practical community can adapt the idea of contextually defined data when making integrations and thus use federative matrixes to create linkages between the data storages.

The contribution the framework offers is in its comprehensiveness. The framework includes aspects of platform context, data type, metadata (of the original data), platform business model, platform governance model and data governance-related issues. Under each of these topics, there are more detailed issues to be clarified with respect to platform data governance. The proposed framework offers a comprehensive basis for starting to govern data on a platform. The framework is mainly intended for platforms used inter-organisationally to share data on mutual end-products within a supply chain or, for example, internet of things related data.

The results of my case studies also show that the general governance of a platform, including data governance, might be best organised through having a neutral, external organisation, e.g. a mutually owned company, to own the platform instead of one of the ecosystem partners owning and managing it. That result was present in both the maritime industry cases, as the interviewed persons indicated that they would find it hard to have enough trust in many of their competitors to provide a shared platform. Similar results have been presented in studies such as those by Otto and Jarke (2019) and Tura et al. (2018). Neutrality and freedom from any business or political interests of the platform owner would provide a stronger capability to expand the platform and attract more partners to join.

Another practical contribution is the matrix tools for data federation. Having been tested in the Breast Cancer Case with patient data, these tools have proven their usability in a complex environment in which there is a good deal of overlapping data. The tools can also be used in digital platform contexts, where the participating organisations need to define their own data attributes that are to be integrated on the platform.

In the practice-oriented literature, the approach presented in my research contradicts the view of DAMA's DMBOK first edition (Mosley et al. 2010), but the second edition (Earley & Henderson 2017) already discussed data federation as it is understood in my research. This can be seen as a sign that data federation and contextually defining data are issues that need more attention in the future.

6.3 Limitations

My study, as any piece of research, has its limitations. Consisting of three single case studies, each conducted in only one industry and within one country, the generalisability cannot be based on statistical sampling, but instead, the generalisation is done to a valid theory. Lee (1989) pointed out several problems with single case studies, including controlled observations and deductions and allowing for replicability and generalisability. These issues are present in my research as well. To eliminate these problems, I have followed the case study protocols, used multiple sources of evidence and a case study database (Yin 2018) and followed guidance provided in (information systems science) the literature. That allows generalisation to theoretical propositions (A. S. Lee & Baskerville 2003). However, the complexity of the theory resulting from case studies is an issue. The framework resulting from the three case studies consists of several pieces and thus is difficult to test and evaluate in a single study. It is also noteworthy that even though there are three single cases, the possibility of making comparisons across the cases are extremely limited, given the differences in the case study structures from one another.

Specific features of the case environments and industries may have affected the results, as institutional conditions (Orlikowski 1992) can still affect the use of data and information systems. Thus, having only three cases which are all drawn from just two contexts is not enough to build a comprehensive theory for all possible platform situations. However, e.g. the federative tools—the interoperability matrixes—are meant to be general enough to be used after required modifications in several contexts. In addition, the conceptual governance of the platform data framework may need modification (and validation of the practical applicability); its concepts may be of different levels of practical importance, some of them describing decision domains rather than being actual activities. However, the framework presented in this dissertation is one step ahead of the framework of Abraham et al. (2019) with regard to empirical observations.

My results also lack the more practical approach concerning data modelling approaches. I have focused on the production of a more general-level approach to data interoperability, neglecting the actual data storage and processing architectures and possibilities and limitations within them. For instance, Linked Data and data space architectures may have provided valuable insights into the interoperability part of my research.

6.4 Future research

As mentioned previously, the field of data governance on platforms is quite new, even though data governance in single-organisation contexts has been studied for some time already. Some of the results from the single-organisation context can clearly be extended to inter-organisational platforms, but some additions must be made, as shown in my case study findings. Empirical testing of frameworks is definitely needed, as well as further studies on the factors promoting and preventing data sharing on platforms. An interesting aspect is also the effect that blockchain technology can have on the need for trust in data sharing.

A more technical approach to data interoperability and a federative approach should be taken, and the impacts of data modelling and related features should be tested and included in the practical versions of the conceptual framework. Different data architectures and their relationships to data interoperability provide an interesting research stream as well.

The building blocks of the data governance framework need to be evaluated and tested within a real-world case in which data is integrated on a platform from multiple organisations. By comparing the building blocks of the framework with the dimensions, issues and data types within a real case, it would be possible to either confirm the framework or make some adjustments to it. For the further development of the data governance framework, I also suggest considering copyright information and various types of data security and protection issues. Whether these dimensions should also be included in the metadata and as parts of the framework is worth investigating.

One interesting subject to study in this field could be taking the platforms' boundary resource thinking (Ghazawneh & Henfridsson 2013) and using that to consider data as a resource on the platform. Seeing data as an important resource puts it in a position in which it has a defined value and governing it formally makes even more sense.

List of References

Aarikka-Stenroos, L., & Ritala, P. (2017). Network management in the era of ecosystems: Systematic review and management framework. *Industrial Marketing Management*, *67*, 23–36.

Abraham, R., Schneider, J., & vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management*, *49*, 424–438. https://doi.org/10.1016/J.IJINFOMGT.2019.07.008

Ackoff, R. (1999). *Ackoff's best*. New York: John Wiley & Sons.

Adner, R. (2017). Ecosystem as Structure. *Journal of Management*, *43*(1), 39–58. https://doi.org/10.1177/0149206316678451

Al-Ruithe, M., Benkhelifa, E., & Hameed, K. (2018). A systematic literature review of data governance and cloud data governance. *Personal and Ubiquitous Computing*. https://doi.org/10.1007 /s00779-017-1104-3

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

Alhassan, I., Sammon, D., & Daly, M. (2019). Critical Success Factors for Data Governance: A Theory Building Approach. *Information Systems Management*, *36*(2), 98–110. https://doi.org/10.1080/10580530.2019.1589670

Allen, C., Des Jardins, T. R., Heider, A., Lyman, K. A., McWilliams, L., Rein, A. L., … Turske, S. A. (2014). Data governance and data sharing agreements for community-wide health information exchange: lessons from the beacon communities. *EGEMS (Washington, DC)*, *2*(1), 1057. https://doi.org/10.13063/2327-9214.1057

Alter, S. (2013). Work System Theory: Overview of Core Concepts, Extensions, and Challenges for the Future. *Journal of the Association for Information Systems*.

Benbasat, I., Goldstein, D. K., & Mead, M. (1987). The Case Research Strategy in Studies of Information Systems. *MIS Quarterly*, *11*(3), 369. https://doi.org/10.2307/248684

Berson, A., & Dubov, L. (2007). *Master data management and customer data integration for a global enterprise*. McGraw-Hill, Inc.

Bizer, C. (2009). The emerging web of linked data. *IEEE Intelligent Systems*, *24*(5), 87–92. https://doi.org/10.1109/MIS.2009.102

Boisot, M., & Canals, A. (2004). Data, information and knowledge: have we got it right? *Journal of Evolutionary Economics*, *14*(1), 43–67. https://doi.org/10.1007/s00191-003-0181-9

Boudreau, K. J., & Hagiu, A. (2009). Platform rules: Multi-sided platforms as regulators. In *Platforms, markets and innovation* (pp. 163–191). Cheltenham, UK: Edgar Elgar Publishing.

Brous, P., Janssen, M., & Vilminko-Heikkinen, R. (2016). Coordinating Decision-Making in Data Management Activities: A Systematic Review of Data Governance Principles. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (pp. 115–125). https://doi.org/10.1007/978-3-319-44421-5_9

Burrell, G., & Morgan, G. (1985). *Sociological paradigms and organisational analysis : elements of the sociology of corporate life* (Repr). Gower.

Cancer in Finland - Syöpärekisteri. (2018). Retrieved May 28, 2019, from https://cancerregistry.fi/ statistics/cancer-in-finland/

Cancer Registry Statistics. (2016). Retrieved May 28, 2019, from https://tilastot.syoparekisteri.fi/syovat/

Cannon, J. P., Achrol, R. S., & Gundlach, G. T. (2000). Contracts, Norms, and Plural Form Governance. *Journal of the Academy of Marketing Science*, *28*(2), 180–194. https://doi.org/10.1177/0092070300282001

Cavaye, A. L. M. (1996). Case study research: a multi-faceted research approach for IS. *Information Systems Journal*, *6*(3), 227–242. https://doi.org/10.1111/j.1365-2575.1996.tb00015.x

Cheong, L. K., & Chang, V. (2007). The need for data governance: a case study. *Proceedings of 18th Australasian Conference on Information Systems ACIS*, 100. Toowomba: AIS.

Chua, W. F. (1986). Radical Developments in Accounting Thought. *The Accounting Review*, *61*(4), 601–632. Retrieved from https://uomustansiriyah.edu.iq/media/lectures/10/10_2019_02_17!07_44_28_PM.pdf

Cleven, A., & Wortmann, F. (2010). Uncovering Four Strategies to Approach Master Data Management. *2010 43rd Hawaii International Conference on System Sciences*, 1–10. https://doi.org/10.1109/HICSS.2010.488

Constantinides, P., Henfridsson, O., & Parker, G. G. (2018). Introduction—Platforms and Infrastructures in the Digital Age. *Information Systems Research*, *29*(2), 381–400. https://doi.org/10.1287/isre.2018.0794

Corallo, A. (2007). The business ecosystem as a multiple dynamic network. In A. Corallo, G. Passiante, & A. Prencipe (Eds.), *The digital business ecosystem* (pp. 11–32). Edward Elgar Publishing.

Cusumano, M. A., & Gawer, A. (2002). The elements of platform leadership. *MIT Sloan Management Review*, *43*(3), 51.

D4V – Design for Value Program Report. (2019). Retrieved May 28, 2019, from http://hightech.dimecc.com/results/d4v-design-for-value-program-report

Dahlberg, T., Hokkanen, P., & Newman, M. (2016). How Business Strategy and Changes to Business Strategy Impact the Role and the Tasks of CIOs: An Evolutionary Model. *Proceedings of the 49th Hawaii International Conference on System Sciences (HICSS), 2016*, 4910–4919. IEEE.

Dahlberg, T., Lagstedt, A., & Nokkala, T. (2018). How To Address Master Data Complexity In Information Systems Development – A Federative Approach. *Proceedings of the 26th European Conference on Information Systems*. Portsmouth, UK.

Dahlberg, T., Nokkala, T., Heikkilä, J., & Heikkilä, M. (2017). Data federation with a governance of data framework artifact as the federation tool: Case on clinical breast cancer treatment data. *Frontiers in Artificial Intelligence and Applications*, Vol. 292, pp. 31–42. https://doi.org/10.3233/978-1-61499-720-7-31

Davenport, T. H., & Prusak, L. (1998). *Working knowledge : how organizations manage what they know*. Harvard Business School Press.

DBE CORE – Digital Business Ecosystem. (2019). Retrieved May 28, 2019, from https://dbecore.com/

De Haes, S., & Van Grembergen, W. (2004). IT governance and its mechanisms. *Information Systems Control Journal*, *1*, 27–33.

de Reuver, M., Sørensen, C., & Basole, R. C. (2018). The digital platform: a research agenda. *Journal of Information Technology*, *33*(2), 124–135.

Dreibelbis, A., Hechler, E., Milman, I., Oberhofer, M., van Run, P., & Wolfson, D. (2008). *Enterprise master data management: an SOA approach to managing core information*. Pearson Education.

Dubé, L., & Paré, G. (2003). Rigor in Information Systems Positivist Case Research: Current Practices, Trends, and Recommendations. *MIS Quarterly*, *27*(4), 597. https://doi.org/10.2307/30036550

Dyché, J., & Levy, E. (2011). *Customer data integration: Reaching a single version of the truth*.

Earley, S., & Henderson, D. (2017). *DAMA-DMBOK : data management body of knowledge* (2nd ed.). Data Management Association.

Economist, T. (2017). The world's most valuable resource; regulating the data economy. *The Economist, 423, 9*. Retrieved from https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, *14*(4), 532–550.

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory Building From Cases: Opportunities And Challenges. *Academy of Management Journal*, *50*(1), 25–32. https://doi.org/10.5465/amj.2007.24160888

Eisenmann, T. R. (2008). Managing Proprietary and Shared Platforms. *California Management Review*, *50*(4), 31–53. https://doi.org/10.2307/41166455

Eisenmann, T. R., Parker, G., & Van Alstyne, M. (2009). Opening platforms: how, when and why? In A. Gawer (Ed.), *Platforms, Markets and Innovation* (pp. 131–162). Chaltenham, UK: Edward Elgar.

El Sawy, O. A., & Pereira, F. (2013). *Business modelling in the dynamic digital space: An ecosystem approach*. Springer.

Eriksson, P., & Kovalainen, A. (2008). *Qualitative methods in business research*. SAGE.

Eskola, H. (2019). OP julkaisi ensimmäisenä tietotilinpäätöksen – tuo esiin vastuullista tiedon käsittelyä ja hyödyntämistä. *Kauppalehti*. Retrieved from https://www.kauppalehti.fi/uutiset/op-julkaisi-ensimmaisena-tietotilinpaatoksen-tuo-esiin-vastuullista-tiedon-kasittelya-ja-hyodyntamista/

Fawcett, S. E., Osterhaus, P., Magnan, G. M., Brau, J. C., & McCarter, M. W. (2007). Information sharing and supply chain performance: the role of connectivity and willingness. *Supply Chain Management: An International Journal*, *12*(5), 358–368.

Flick, U. (2007). *Designing Qualitative Research*. https://doi.org/10.4135/9781849208826

Flyvbjerg, B. (2011). Case Study. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (4th ed., pp. 301–316). Thousand Oaks (CA): SAGE.

Galliers, R. D., & Newell, S. (2003). Back to the future: from knowledge management to the management of information and data. *Information Systems and E-Business Management*, *1*(1), 5–13. https://doi.org/10.1007/BF02683507

Garcia-Molina, H., Papakonstantinou, Y., Quass, D., Rajaraman, A., Sagiv, Y., Ullman, J., … Widom, J. (1997). The TSIMMIS Approach to Mediation: Data Models and Languages. *Journal of Intelligent Information Systems*, *8*(2), 117–132. https://doi.org/10.1023/A:1008683107812

Gawer, A. (2014). Bridging differing perspectives on technological platforms: Toward an integrative framework. *Research Policy*, *43*(7), 1239–1249. https://doi.org/10.1016/j.respol.2014.03.006

Gawer, A., & Cusumano, M. A. (2008). Platform Leaders. *MIT Sloan Management Review*.

Gawer, A., & Cusumano, M. A. (2014). Industry platforms and ecosystem innovation. *Journal of Product Innovation Management*, *31*(3), 417–433.

Ghazawneh, A., & Henfridsson, O. (2013). Balancing platform control and external contribution in third-party development: the boundary resources model. *Information Systems Journal*, *23*(2), 173–192. https://doi.org/10.1111/j.1365-2575.2012.00406.x

Gregor, S. (2002). A Theory of Theories in Information Systems. In S. Gregor & D. Hart (Eds.), *Information Systems Foundations: Building the Theoretical Base* (pp. 1–20). Canberra: Australian National University.

Gregor, S. (2006). The Nature of Theory in Information Systems. *MIS Quarterly*, *30*(3), 611. https://doi.org/10.2307/25148742

Hagiu, A., & Wright, J. (2015). Multi-sided platforms. *International Journal of Industrial Organization*, *43*(C), 162–174.

Hambrick, D. C., Werder, A. v., & Zajac, E. J. (2008). New Directions in Corporate Governance Research. *Organization Science*, *19*(3), 381–385. https://doi.org/10.1287/orsc.1080.0361

Haug, A., & Stentoft Arlbjørn, J. (2011). Barriers to master data quality. *Journal of Enterprise Information Management*, *24*(3), 288–303. https://doi.org/10.1108/17410391111122862

Heath, T., & Bizer, C. (2011). Linked Data: Evolving the Web into a Global Data Space. *Synthesis Lectures on the Semantic Web: Theory and Technology*, *1*(1), 1–1. https://doi.org/10.2200/S00334ED1V01Y201102WBE001

Heimbigner, D., & McLeod, D. (1985). A federated architecture for information management. *ACM Transactions on Information Systems (TOIS)*, *3*(3), 253–278.

Helfat, C. E., & Raubitschek, R. S. (2018). Dynamic and integrative capabilities for profiting from innovation in digital platform-based ecosystems. *Research Policy*, *47*(8), 1391–1399. https://doi.org/10.1016/J.RESPOL.2018.01.019

Henderson, J. C., & Venkatraman, H. (1999). Strategic alignment: Leveraging information technology for transforming organizations. *IBM Systems Journal*, *38*(2.3), 472–484. https://doi.org/10.1147/SJ.1999.5387096

Hilbert, M., & López, P. (2011). The world's technological capacity to store, communicate, and compute information. *Science*, *332*(6025), 60–65.

Hirschheim, R., & Klein, H. K. (2012). A glorious and not-so-short history of the information systems field. *Journal of the Association for Information Systems*, *12*(4), 188–235.

Hohpe, G., & Woolf, B. (2004). *Enterprise integration patterns : designing, building, and deploying messaging solutions*. Addison-Wesley.

Horne, N. W. (1998). Putting information assets on the board agenda. *Long Range Planning*, *31*(1), 10–17. https://doi.org/10.1016/S0024-6301(97)00086-1

Huber, T. L., Kude, T., & Dibbern, J. (2017). Governance Practices in Platform Ecosystems: Navigating Tensions Between Cocreated Value and Governance Costs. *Information Systems Research*, *28*(3), 563–584. https://doi.org/10.1287/isre.2017.0701

Humby, C. (2006). Data is the new Oil. *ANA Sr. Marketer's Summit*. Retrieved from https://ana.blogs.com/maestros/2006/11/data_is_the_new.html

Iansiti, M., & Levien, R. (2004a). Strategy as ecology. *Harvard Business Review*, *82*(3), 68–81.

Iansiti, M., & Levien, R. (2004b). *The keystone advantage: what the new dynamics of business ecosystems mean for strategy, innovation, and sustainability*. Harvard Business Press.

Iivari, J. (2017). Information system artefact or information system application: that is the question. *Information Systems Journal*, *27*(6), 753–774.

Inmon, W. H., O'Neil, B. K., & Fryman, L. (2008). *Business metadata : capturing enterprise knowledge*. Elsevier/Morgan Kaufmann.

Jacobides, M. G., Cennamo, C., & Gawer, A. (2018). Towards a theory of ecosystems. *Strategic Management Journal*, *39*(8), 2255–2276. https://doi.org/10.1002/smj.2904

Janssen, M., Estevez, E., & Janowski, T. (2014). Interoperability in big, open, and linked data-organizational maturity, capabilities, and data portfolios. *Computer*, *47*(10), 44–49. https://doi.org/10.1109/MC.2014.290

Jarke, M., Otto, B., & Ram, S. (2019, October 1). Data Sovereignty and Data Space Ecosystems. *Business and Information Systems Engineering*, Vol. 61, pp. 549–550. https://doi.org/10.1007/s12599-019-00614-2

Kaplan, B., & Duchon, D. (1988). Combining Qualitative and Quantitative Methods in Information Systems Research: A Case Study. *MIS Quarterly*, *12*(4), 571. https://doi.org/10.2307/249133

Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, *53*(1), 148–152.

Klein, H. K., & Myers, M. D. (1999). A set of principles for conducting and evaluating interpretive field studies in information systems. *MIS Quarterly*, *23*(1), 67–94.

Korpela, K. (2014). *Value of Information Logistics Integration in Digital Business Ecosystem* (Lappeenranta University of Technology). Retrieved from http://lutpub.lut.fi/handle/10024/102292

Lee, A. S. (1989). A Scientific Methodology for MIS Case Studies. *MIS Quarterly*, *13*(1), 33. https://doi.org/10.2307/248698

Lee, A. S. (1991). Integrating Positivist and Interpretive Approaches to Organizational Research. *Organization Science*, *2*(4), 342–365. https://doi.org/10.1287/orsc.2.4.342

Lee, A. S., & Baskerville, R. L. (2003). Generalizing Generalizability in Information Systems Research. *Information Systems Research*, *14*(3), 221–243. https://doi.org/10.1287/isre.14.3.221.16560

Lee, H. L., Lee, H. L., & Whang, S. (2000). Information Sharing in a Supply Chain. *International Journal of Technology Management*, *20*, 373--387. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.90.6864

Lee, S. U. (2017). *Design Choices for Data Governance in Platform Ecosystems: A Contingency Model*.

Lee, S. U., Zhu, L., & Jeffery, R. (2017). Data Governance for Platform Ecosystems: Critical Factors and the State of Practice. *Proceedings of the 21st Pacific Asia Conference on Information Systems (PACIS 2017)*, 1–12. Langkawi, Malaysia.

Lee, S. U., Zhu, L., & Jeffery, R. (2018a). A Contingency-Based Approach to Data Governance Design for Platform Ecosystems. *Proceedings of the 22st Pacific Asia Conference on Information Systems (PACIS 2018)*.

Lee, S. U., Zhu, L., & Jeffery, R. (2018b). A Data Governance Framework for Platform Ecosystem Process Management. In M. Weske, M. Montali, I. Weber, & J. vom Brocke (Eds.), *Business Process Management Forum. BPM 2018. Lecture Notes in Business Information Processing, vol 329* (pp. 211–227). https://doi.org/10.1007/978-3-319-98651-7_13

Lee, S. U., Zhu, L., & Jeffery, R. (2018c). Designing Data Governance in Platform Ecosystems. *Proceedings of the 51st Hawaii International Conference on System Sciences*. https://doi.org/https://doi.org/10.24251/hicss.2018.626

Li, S., & Lin, B. (2006). Accessing information sharing and information quality in supply chain management. *Decision Support Systems*, *42*(3), 1641–1656.

Loshin, D. (2010). *Master data management*. Morgan Kaufmann.

McGilvray, D. (2006). Data Governance: A Necessity in an Integrated Information World. *DM Review*, *16*(12), 24–26.

Miles, M. B., & Huberman, A. M. (1984). Qualitative Data Analysis: A Sourcebook of New Methods. In *London Sage Publications*. https://doi.org/10.1016/0149-7189(96)88232-2

Mingers, J. (2001). Combining IS Research Methods: Towards a Pluralist Methodology. *Information Systems Research*, *12*(3), 240–259. https://doi.org/10.1287/isre.12.3.240.9709

Möller, K. K., & Halinen, A. (1999). Business Relationships and Networks:: Managerial Challenge of Network Era. *Industrial Marketing Management*, *28*(5), 413–427. https://doi.org/10.1016/ S0019-8501(99)00086-3

Moore, J. F. (1993). Predators and prey: a new ecology of competition. *Harvard Business Review*, *71*(3), 75–83.

Moore, J. F. (2016). *The death of competition: leadership and strategy in the age of business ecosystems*. Retrieved from https://scholar.google.fi/scholar?hl=fi&as_sdt=0%2C5&q=The+Death+of+Competition%3A+Leadership+and+Strategy+in+the+Age+of+Business+Ecosystems&btnG=#d=gs_cit&u=%2Fscholar%3Fq%3Dinfo%3Ayaat3Spocn0J%3Ascholar.google.com%2F%26output%3Dcite%26scirp%3D0%26hl%3Df

Mosley, M., Mosley, M., Brackett, M., Earley, S., & Henderson, D. (2010). *The DAMA guide to the data mangement body of knowledge : DAMA-DMBOK guide* (First edit). Bradley Beach, NJ: Technics Publications, LLC.

Myers, M. D. (1997). Interpretive Research in Information Systems. In J. Mingers & F. A. Stowell (Eds.), *Information systems : an emerging discipline?* (pp. 239–266). Berkshire, England: McGraw-Hill.

Myers, M. D., & Newman, M. (2007). The qualitative interview in IS research: Examining the craft. *Information and Organization*, *17*(1), 2–26.

Nachira, F. (2002). Towards a network of digital business ecosystems fostering the local development. In *Digital Business Ecosystems*. Retrieved from http://www.digital-ecosystems.org/doc/discussionpaper.pdf

Nachira, F., Dini, P., & Nicolai, A. (2007). A network of digital business ecosystems for Europe: roots, processes and perspectives. *European Commission, Bruxelles, Introductory Paper*.

Nokkala, T., & Dahlberg, T. (2018). *Data Federation in the Era of Digital, Consumer-Centric Cares and Empowered Citizens*. https://doi.org/10.1007/978-3-319-97931-1_11

Orlikowski, W. J. (1992). The Duality of Technology: Rethinking the Concept of Technology in Organizations. *Organization Science*, *3*(3), 398–427. https://doi.org/10.1287/orsc.3.3.398

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

Otto, B. (2011a). A morphology of the organisation of data governance. *Proceedings of 19th European Conference on Information Systems, ECIS 2011*.

Otto, B. (2011b). Organizing Data Governance: Findings from the telecommunications industry and consequences for large service providers. *Communications of the Association for Information Systems*, *29*(1), 45–66.

Otto, B. (2012). How to design the master data architecture: Findings from a case study at Bosch. *International Journal of Information Management*, *32*(4), 337–346. https://doi.org/10.1016/J.IJINFOMGT.2011.11.018

Otto, B. (2015). Quality and Value of the Data Resource in Large Enterprises. *Information Systems Management*, *32*(3), 234–251. https://doi.org/10.1080/10580530.2015.1044344

Otto, B., Hüner, K. M., & Österle, H. (2012). Toward a functional reference model for master data quality management. *Information Systems and E-Business Management*, *10*(3), 395–425. https://doi.org/10.1007/s10257-011-0178-0

Otto, B., & Jarke, M. (2019). Designing a multi-sided data platform: findings from the International Data Spaces case. *Electronic Markets*, 1–20. https://doi.org/10.1007/s12525-019-00362-x

Ozalp, H., Cennamo, C., & Gawer, A. (2018). Disruption in Platform-Based Ecosystems. *Journal of Management Studies*, *55*(7), 1203–1241. https://doi.org/10.1111/joms.12351

Panian, Z. (2010). Some practical experiences in data governance. *World of Academy of Science, Engineering and Technology*, *62*(1), 939–946.

Peltoniemi, M., & Vuori, E. (2004). Business ecosystem as the new approach to complex adaptive business environments. *Proceedings of EBusiness Research Forum*, 267–281.

Poppo, L., & Zenger, T. (2002). Do formal contracts and relational governance function as substitutes or complements? *Strategic Management Journal*, *23*(8), 707–725. https://doi.org/10.1002/smj.249

Powell, W. W. (1990). Neither market nor hierarchy. *Research In Organizational Behavior*, *12*, 295–336.

Powell, W. W., White, D. R., Koput, K. W., & Owen-Smith, J. (2005). Network Dynamics and Field Evolution: The Growth of Interorganizational Collaboration in the Life Sciences. *American Journal of Sociology*, *110*(4), 1132–1205. https://doi.org/10.1086/421508

Prajogo, D., & Olhager, J. (2012). Supply chain integration and performance: The effects of long-term relationships, information technology and sharing, and logistics integration. *International Journal of Production Economics*, *135*(1), 514–522.

Rai, A., Patnayakuni, R., & Seth, N. (2006). Firm performance impacts of digitally enabled supply chain integration capabilities. *MIS Quarterly*, 225–246.

Rau, K. G. (2004). Effective Governance of It: Design Objectives, Roles, and Relationships. *Information Systems Management*, *21*(4), 35–42. https://doi.org/10.1201/1078/44705.21.4.20040901/84185.4

Rochet, J.-C., & Tirole, J. (2003). Platform Competition in Two-Sided Markets. *Journal of the European Economic Association*, *1*(4), 990–1029. https://doi.org/10.1162/154247603322493212

Rosenbaum, S. (2010). Data Governance and Stewardship: Designing Data Stewardship Entities and Advancing Data Access. *Health Services Research*, *45*(5p2), 1442–1455.

Rubin, H., & Rubin, I. (2005). *Qualitative Interviewing (2nd ed.): The Art of Hearing Data*. https://doi.org/10.4135/9781452226651

Ruddin, L. P. (2006). You Can Generalize Stupid! Social Scientists, Bent Flyvbjerg, and Case Study Methodology. *Journals.Sagepub.Com*, *12*(4), 797–812. https://doi.org/10.1177/1077800406288622

Schmidt, J. G., & Lyle, D. (2010). *Lean integration - an integration factory approach to business agility*. Addison-Wesley.

Schreieck, M., Wiesche, M., & Krcmar, H. (2016). Design and Governance of Platform Ecosystems-Key Concepts and Issues for Future Research. *ECIS 2016 Proceedings*.

Sen, A. (2004). Metadata management: past, present and future. *Decision Support Systems*, *37*(1), 151–173. https://doi.org/10.1016/S0167-9236(02)00208-7

Sheth, A., & Larson, J. (1990). Federated database systems for managing distributed, heterogeneous, and autonomous databases. *ACM Computing Surveys (CSUR)*, *22*(3), 183–236. https://doi.org/10.1145/96602.96604

Shleifer, A., & Vishny, R. W. (1997). A survey of corporate governance. *The Journal of Finance*, *52*(2), 737–783.

Silvola, R., Jaaskelainen, O., Kropsu-Vehkapera, H., & Haapasalo, H. (2011). Managing one master data – challenges and preconditions. *Industrial Management & Data Systems*, *111*(1), 146–162. https://doi.org/10.1108/02635571111099776

Singh, G., Bharathi, S., Chervenak, A., Deelman, E., Kesselman, C., Manohar, M., … Pearlman, L. (2003). A Metadata Catalog Service for Data Intensive Applications. *Proceedings of the 2003 ACM/IEEE Conference on Supercomputing - SC '03*, 33. https://doi.org/10.1145/1048935.1050184

Solomon, M. D. (2005). It's All About the Data. *Information Systems Management*, *22*(3), 75–80. https://doi.org/10.1201/1078/45317.22.3.20050601/88747.9

Stake, R. E. (1995). *The art of case study research*. Retrieved from https://www.google.com/books?hl=fi&lr=&id=ApGdBx76b9kC&oi=fnd&pg=PA7&dq=+The+art+of+case+study+research+Stake,+Robert+E.&ots=KvOHl4Jjaq&sig=oJ7LcIkXn8UbIt9Hx3yf8OY8zUM

Stake, R. E. (2008). Qualitative Case Studies. In N. Denzin & Y. Lincoln (Eds.), *Strategies of Qualitative Inquiry* (3rd ed., pp. 119–150). Thousand Oaks (CA): Sage Publications, Inc.

Tallon, P. P., Ramirez, R. V, & Short, J. E. (2014). Toward a Theory of Information Governance. *Journal of Management Information Systems*, *30*(3), 141–178. https://doi.org/10.2753/ MIS0742-1222300306

Tan, F. T. C., Pan, S. L., & Zuo, M. (2018). Realising platform operational agility through information technology - enabled capabilities: A resource - interdependence perspective. *Information Systems Journal*, *29*(3), 582–608.

TENK. (2019). Retrieved June 5, 2019, from https://www.tenk.fi/en

Tilson, D., Lyytinen, K., & Sørensen, C. (2010). Research Commentary — Digital Infrastructures: The Missing IS Research Agenda. *Information Systems Research*, *21*(4), 748–759. https://doi.org/10.1287/isre.1100.0318

Tiwana, A. (2013). *Platform ecosystems : aligning architecture, governance, and strategy*. San Francisco: Morgan Kaufmann.

Tiwana, A., Konsynski, B., & Bush, A. A. (2010). Research commentary—Platform evolution: Coevolution of platform architecture, governance, and environmental dynamics. *Information Systems Research*, *21*(4), 675–687.

Tiwana, A., Konsynski, B., & Venkatraman, N. (2013). Special Issue: Information Technology and Organizational Governance: The IT Governance Cube. *Journal of Management Information Systems*, *30*(3), 7–12. https://doi.org/10.2753/MIS0742-1222300301

Tuomi, I. (1999). Data is more than knowledge: implications of the reversed knowledge hierarchy for knowledge management and organizational memory. *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences HICSS-32*. https://doi.org/10.1109/HICSS.1999.772795

Tura, N., Kutvonen, A., & Ritala, P. (2018). Platform design framework: conceptualisation and application. *Technology Analysis & Strategic Management*, *30*(8), 881–894. https://doi.org/10.1080/09537325.2017.1390220

van den Broek, T. A., & van Veenstra, A. F. (2015). Modes of Governance in Inter-Organizational Data Collaborations. *ECIS*.

Van Grembergen, W., & De Haes, S. (2008). *Implementing information technology governance : models, practices, and cases*. IGI Pub.

Van Maanen, J. (1979). Reclaiming Qualitative Methods for Organizational Research: A Preface. *Administrative Science Quarterly*, *24*(4), 520. https://doi.org/10.2307/2392358

Vayghan, J. A., Garfinkle, S. M., Walenta, C., Healy, D. C., & Valentin, Z. (2007). The internal information transformation of IBM. *IBM Systems Journal*, *46*(4), 669–683. https://doi.org/10.1147/sj.464.0669

Vilminko-Heikkinen, R. (2017). *Data, Technology, and People: Demystifying Master Data Management*. Tampere University of Technology.

Vilminko-Heikkinen, R., & Pekkola, S. (2017). Master data management and its organizational implementation. *Journal of Enterprise Information Management*, *30*(3), 454–475. https://doi.org/10.1108/JEIM-07-2015-0070

Vilminko-Heikkinen, R., & Pekkola, S. (2019). Changes in roles, responsibilities and ownership in organizing master data management. *International Journal of Information Management*, *47*, 76–87. https://doi.org/10.1016/J.IJINFOMGT.2018.12.017

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

Wand, Y., Storey, V. C., & Weber, R. (1999). An ontological analysis of the relationship construct in conceptual modeling. *ACM Transactions on Database Systems*, *24*(4), 494–528. https://doi.org/10.1145/331983.331989

Wand, Y., & Wang, R. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, *39*(11), 86–95. https://doi.org/10.1145/240455.240479

Wand, Y., & Weber, R. (1988). An Ontological Analysis of some Fundamental Information Systems Concepts. *ICIS 1988 Proceedings*. Retrieved from https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1008&context=icis1988

Wand, Y., & Weber, R. (1990). An ontological model of an information system. *IEEE Transactions on Software Engineering*, *16*(11), 1282–1292. https://doi.org/10.1109/32.60316

Wand, Y., & Weber, R. (1993). On the ontological expressiveness of information systems analysis and design grammars. *Information Systems Journal*, *3*(4), 217–237. https://doi.org/10.1111/j.1365-2575.1993.tb00127.x

Wand, Y., & Weber, R. (1995). On the deep structure of information systems. *Information Systems Journal*, *5*(3), 203–223. https://doi.org/10.1111/j.1365-2575.1995.tb00108.x

Wand, Y., & Weber, R. (2002). Research Commentary: Information Systems and Conceptual Modeling—A Research Agenda. *Information Systems Research*, *13*(4), 363–376. https://doi.org/10.1287/isre.13.4.363.69

Wang, R. Y. (1998). A product perspective on total data quality management. *Communications of the ACM*, *41*(2), 58–66.

Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, *12*(4), 5–33. https://doi.org/10.1080/07421222.1996.11518099

Weber, K., Otto, B., & Österle, H. (2009). One size does not fit all---a contingency approach to data governance. *Journal of Data and Information Quality (JDIQ)*, *1*(1), 4.

Weill, P. (2004). Don’t just lead, govern: How top-performing firms govern IT. *MIS Quarterly Executive*, *3*(1), 1–17.

Weill, P. (2005). A Matrixed Approach to Designing IT Governance. *MIT Sloan Management Review*, *46*(2), 26.

Weill, P., & Ross, J. W. (2004). *IT governance : how top performers manage IT decision rights for superior results*. Harvard Business School Press.

Wende, K. (2007). A model for data governance-Organising accountabilities for data quality management. *ACIS 2007 Proceedings*, 80.

Yin, R. K. (2011). *Applications of case study research*. Sage.

Yin, R. K. (2014). *Case study research : design and methods* (5th ed). Los Angeles: SAGE.

Yin, R. K. (2018). *Case Study Research and Applications: Design and Methods* (6th editio). Los Angeles: SAGE Publications.

Yoo, Y., Boland, R. J. J., Lyytinen, K., & Majchrzak, A. (2012). Organizing for Innovation in the Digitized World. *Organization Science*, *23*(5), 1398–1408.

Yoo, Y., Henfridsson, O., & Lyytinen, K. (2010). Research Commentary — The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research. *Information Systems Research*, *21*(4), 724–735. https://doi.org/10.1287/isre.1100.0322

Zins, C. (2007). Conceptual approaches for defining data, information, and knowledge. *Journal of the American Society for Information Science and Technology*, *58*(4), 479–493. https://doi.org/10.1002/asi.20508

UNIVERSITY
OF TURKU

Painosalama Oy, Turku, Finland 2020

ISBN 978-951-29-8132-8 PRINT
ISBN 978-951-29-8133-5 PDF
ISSN 2343-3159 (Painettu/Print)
ISSN 2343-3167 (Verkkojulkaisu/Online)