Data spaces and the (trans)formations of data innovation and governance

Dragana Paparova

UiA University of Agder

Doctoral Dissertations at the University of Agder 455

Data spaces and the (trans)formations of data innovation and governance

Dragana Paparova

Data spaces and the (trans)formations of data innovation and governance

Dissertation for the degree philosophiae doctor (PhD)

University of Agder
Faculty of Social Sciences

Department of Information Systems

2024

Doctoral dissertations at the University of Agder 455

ISSN: 1504-9272
ISBN: 978-82-8427-172-9

Print: Make!Graphics

Kristiansand

Preface

This thesis consists of two parts. Part I extends the six papers (two of which are single-authored, four of which are co-authored) by conducting a meta-analysis that binds them together. The six papers constitute Part II, as an appendix.

The papers included are as follows:

1. "Data governance spaces: The case of a national digital service for personal health data". Authors: Dragana Paparova, Margunn Aanestad, Polyxeni Vassilakopoulou, Marianne Klungland Bahus. Publication outlet: *Information & Organization*. Volume: 33. Issue: 1. Paper no: 100451. Year: 2023.
2. "Exploring the ontological status of data: A process-oriented approach". Author: Dragana Paparova. Publication outlet: *Thirty-first European Conference on Information Systems (ECIS)*. Year: 2023.
3. "Data hierarchies: The emergence of an industrial data ecosystem". Authors: Daniel Stedjan Svendsrud, Dragana Paparova. Published in edited version: *Fourty-fourth International Conference on Information Systems (ICIS)*. Year: 2023.
4. "Beyond organizational boundaries: The role of techno-legal configurations". Authors: Dragana Paparova, Margunn Aanestad, Ela Klecun. Published in edited version: *Fourty-fourth International Conference on Information Systems (ICIS)*. Year: 2023.
5. "Opening-up digital platforms to accommodate patient-generate health data". Authors: Dragana Paparova. Publication outlet: *8th International Conference on Infrastructures in Healthcare*, *InfraHEALTH*. Year: 2021.
6. "Governing innovation in e-health platforms: Key concepts and future directions". Authors: Dragana Paparova, Margunn Aanestad. Publication outlet: Selected Papers from the Information Systems Research Seminar (IRIS). Issue: 11. Paper no. 4. Year: 2020.

Acknowledgments

This journey was as intellectually stimulating, as it was daunting. The main outcome of it – this thesis – would not have been possible without the guidance and support of several people who have left their footprint on my professional and personal development.

First, my main supervisor, Margunn Aanestad. You have been a mentor, a role model, and above all – humane. I am filled with pride and honor that I got to be taught, influenced, and challenged by you. Your wisdom, advice, and critical thinking have sharpened a young researcher's brain. If I could choose anyone, all over again, to guide me throughout this journey, I would always choose you.

Second, my co-authors, co-supervisors, colleagues, and friends at the University of Agder, the London School of Economics and Political Sciences, and the University of Oslo. Your involvement has made this academic endeavor invaluable. Polyxeni Vassilakopoulou, thank you for supporting me during a steep learning curve. Alexander Kempton, thank you for being receptive and encouraging towards different and/or controversial ideas. Ela Klecun, thank you for being a door opener to new academic impulses at a critical time in my career. Carl Erik Moe, thank you for your leadership in co-teaching, and beyond. My co-supervisors, Sara Hofmann, and Marianne Klungland Bahus, thank you for sticking with me until the very end. And this list is not exhaustive.

Third, my wonderful family; my mom Stanka, who has taught me hard work and sacrifice; my father Metodi, who has shown me integrity and being true to your words; and my brother Ljubomir, whose intellectual curiosity has set a great example. Thank you for being an infinite support. Everything I am, or ever aspire to be, I owe it to you.

Lastly, my partner, co-author, colleague, and best friend – Daniel. Thank you for believing in me when I struggled to do so myself, and for unconditionally standing by my side at my highest, and lowest. This journey would be meaningless and impossible without you.

Summary

In this thesis, I theorize data innovation and governance as simultaneous processes and account for the distinctive nature of data. Utilizing the concept of space, I show how data innovation and governance in multi-actor environments unfold across certain structures of possible forms, and how the realities data refer to condition the forms innovation and governance can take.

The uniqueness of data entities has been of interest to information systems scholars, imparting distinct value-creation possibilities and dedicated governance approaches. In the literature on digital innovation, data have been referred to as semantic entities whose value can be open-endedly explored once assigned meaning by actors to fulfill various goals and purposes. Across the literature on data governance, data have been referred to as strategic assets that are governed by organizations. This duality of data – as valuable resources that at the same time require proper governance – has also been central in practical debates, such as the European Union's aspirations for developing data spaces as shared infrastructures for innovating with data, while preserving European values, laws and regulations.

Data innovation commonly requires recombining data that are produced, copied, shared, and used across multiple actors, requiring forms of governance extending beyond the boundaries of single organizations. In this thesis, I build on the process-oriented, realist ontology of assemblage theory to account for data's distinctive nature and utilize the concept of space to theorize processes of innovation and governance in multi-actor environments. Data spaces, as argued in this thesis, are neither solely geometrical, nor networked; instead, provide forms across which processes of data innovation and governance can change their spatial configurations.

Empirically, I study data spaces through an embedded case study in the highly regulated Norwegian healthcare sector dealing with personal and sensitive health data. The cases take an information infrastructure perspective on studying how health data (including electronic patient record data and patient-generated health data) were innovated with and governed across multiple public and private actors. Overall, the meta-analysis shows how innovation and governance with health data

took on different forms as data were processed for various purposes across multiple intertwined data spaces.

This thesis is aimed at theory-building and its contribution is two-fold. First, it shows how the concept of data spaces can be used to study processes of data innovation and governance as unfolding across various organizations, digital technologies, legal basis, and data sources, by changing their spatial configurations as certain thresholds are reached. Second, it shows how data do not simply decouple from the realities they refer to, rather, these realities condition the forms data innovation and governance can take and are shaped by these processes in return.

Oppsummering

I denne avhandlingen teoretiserer jeg datainnovasjon og datastyring som samtidige prosesser og redegjør for den særegne karakteren til data. Ved å utnytte begrepet om rom, viser jeg hvordan datainnovasjon og datastyring i multiaktørmiljøer utspiller seg på tvers av visse strukturer av mulige former, og hvordan virkeligheten dataene refererer til betinger formene datainnovasjon og datastyring kan ha.

Det unike med dataenheter har vært av interesse for informasjonssystemforskere, og har gitt distinkte verdiskapingsmuligheter og dedikerte styringstilnærminger. I litteraturen om digital innovasjon, har data blitt referert til som semantiske enheter med åpen verdi når aktørene har tildelt dem mening for å oppfylle ulike mål og formål. På tvers av litteraturen om datastyring, har data blitt referert til som strategiske eiendeler som er styrt av organisasjoner. Denne dualiteten av data – som verdifulle ressurser som samtidig krever passende styring – har også vært sentralt i praktiske debatter, som for eksempel den Europeiske Unionens ambisjoner om å utvikle datarom som delte infrastrukturer for innovasjon med data, mens europeiske verdier, lover og forskrifter bevares.

Datainnovasjon krever ofte å rekombinere data som produseres, kopieres, deles og brukes på tvers av flere aktører og innføring av styringsformer som strekker seg utover grensene til enkeltorganisasjoner. I denne oppgaven bygger jeg på den prosessorienterte, realistiske ontologien til assemblage teori for å redegjøre for datas særegne natur, og utnytte rombegrepet til å teoretisere prosesser for datainnovasjon og -styring i fleraktørmiljøer. Datarom, som argumentert i denne avhandlingen, er verken geometriske eller nettverksbaserte; i stedet, gir datarom former for datainnovasjon og -styring prosesser, som derav kan endre deres romlige konfigurasjoner.

Jeg studerer datarom gjennom en empirisk casestudie i den sterkt regulerte norske helsesektoren som omhandler personlige og sensitive helsedata. Casene tar et informasjonsinfrastrukturperspektiv på å studere hvordan helsedata (inkludert elektroniske pasientjournaldata og pasientgenererte helsedata) ble innovert med og styrt på tvers av flere offentlige og private aktører. Samlet sett viser metaanalysen

hvordan innovasjon og styring med helsedata tok ulike former ettersom data ble behandlet til ulike formål på tvers av flere sammenvevde datarom.

Denne avhandlingen er rettet mot teori bygging og dens bidrag er todelt. For det første, viser den hvordan begrepet datarom kan brukes til å studere prosesser for datainnovasjon og -styring som utspiller seg på tvers av ulike organisasjoner, digitale teknologier, juridisk grunnlag, datakilder, ved å endre deres romlige konfigurasjoner etter visse terskler nås. For det andre, viser den at data ikke bare kobles fra virkeligheten de refererer til; istedenfor, virkeligheten betinger formene datainnovasjon og datastyring kan ta, og til gjengjeld blir formet av disse prosessene.

Contents

Figures

Tables

Abbreviations

AT	Assemblage theory
GDPR	General data protection regulation
EPR	Electronic patient records
PGHD	Patient-generated health data
RCM	Remote care monitoring
RHT	Regional Health Trust
API	Application programming interfaces
II	Information infrastructures

PART I: THE KAPPA

1 Introduction

1.1 Background

Data innovation – the recombination of data into larger objects that yield organizational value – has been increasingly at the center of information systems (IS) research. Within organizations, data have been studied around their open-ended value potential once assigned meaning and purposes when used by actors (Aaltonen and Penttinen 2021; Aaltonen and Tempini 2014; Alaimo et al. 2020). However, advanced technological developments, such as social media platforms, cloud computing, and internet-of-things infrastructures brought in vast amounts and varieties of data, which are seldom produced, shared, and used within single organizations (Haugjord & Kempton, 2022; Mikalsen & Monteiro, 2021; Østerlie & Monteiro, 2020). Rather, a large part of data's technological setup can lie outside of organizational boundaries, beyond the control of specific organizations, and evolve at the will of multiple actors with autonomous, but interconnected value-creation goals and interests (Alaimo, Kallinikos, & Valderrama, 2020; Kazemargi et al., 2023; Monteiro & Parmiggiani, 2019).

Several research streams in IS have been covering topics related to data innovation in multi-actor environments. In the information infrastructures literature, scholars have shown how data shared across multiple actors can yield different forms of value (Barrett et al., 2016), where the value-creation of one actor can result in value disruption for another (Tempini, 2017). In the digital ecosystems literature, data have been commonly referred to as complementarities, or resources that yield larger value when combined and aligned across shared structures, rather than when used separately by actors (Alaimo, Kallinikos, & Valderrama, 2020; Kazemargi et al., 2023). Overall, data have been recognized as resources of distinctive nature and semantic entities whose value stems from meaning-making, rather than fitting together a set of technical components (Alaimo, Kallinikos, & Aaltonen, 2020).

Data took on significance in the IS discourse with the advent of big data stemming from more pervasive digitalization and datafication (Constantiou & Kallinikos, 2015; Günther et al., 2017; Kallinikos & Constantiou, 2015; Lycett, 2013). The empirical impetus for these studies came not the least from organizations starting to use social media data. Such data represented massive, heterogeneous, and dispersed user activity – a very different information resource from the more well-structured data that were traditionally collected and used within a centrally controlled scheme. Constantiou and Kallinikos (2015, p. 54) describe the challenges of organizations using "the heterogeneous, unstructured, agnostic, trans-semiotic nature of big data". These conceptual studies were followed, as per the request of Jones (2019), with more detailed empirical investigations that conceptualize data collection and use in work practices. The empirical studies (Aaltonen et al., 2021; Aaltonen & Tempini, 2014; Alaimo, Kallinikos, & Valderrama, 2020; Barrett et al., 2016; Tempini, 2017) emphasize how data's value stems from recontextualizing and reinterpreting data based on actors' goals and purposes, and not solely from recombining data into larger objects.

Scholars have argued that data are distinct entities from IT components, as data put together do not become programs or software, and do not embody functions (Alaimo, Kallinikos, & Aaltonen, 2020). Therefore, data do not follow the recombination logic of modular architectures but are sign tokens referring to objects, events, people in the real-world (ibid.). In other words, due to their semantic nature, data need to be worked on, produced, transformed, and interpreted to create organizational value. Moreover, data can be edited, ported, recontextualized, aggregated into larger objects, and assigned meaning, based on actors' goals, needs, and purposes, instead of being read as pre-defined metrics (Aaltonen & Penttinen, 2021)

Data can also be shared without depleting (Vassilakopoulou et al., 2019), copied and stored at different places, bringing distinctive issues related to data governance. Traditionally, the data governance literature has been treating data as inherent to IT, as noted by Benfeldt (2017), with few exceptions (Rosenbaum, 2010; Tallon et al., 2013). Within organizations, data have been treated as strategic assets (Zhang et al., 2022), or as economic goods that can be owned (Zhang et al., 2022). In inter-organizational settings, scholars have discussed the different governance mechanisms to allocate roles and responsibilities around data sharing (Abraham et al., 2019), the governance structures e.g., hierarchies or networks (Jagals & Karger, 2021; Van den Broek & Van Veenstra, 2015) and stakeholders' conflicting and complementary interests (Benfeldt et al., 2020). However, the specifics of data governance imparted by the distinctive nature of data in multi-actor environments remains an underexplored topic, with few exceptions (Gegenhuber et al., 2023; Jarvenpaa & Essén, 2023).

The duality of data in multi-actor environments – as entities used for innovation and requiring proper governance approaches – has also been central in practical debates. The European Commission (2020) has recently introduced the necessity of developing data spaces as shared, domain-specific infrastructures for producing, sharing, and using data while preserving rules, laws, and regulations. As a rising practitioner's challenge, data's duality has also been discussed by IS scholars (Kazemargi et al., 2023; Vial, 2019), and data spaces have become a topic of interest (Geiss et al., 2023; Hutterer et al., 2023). However, in the IS field, beyond metaphorical usage, space as a concept has not received sufficient attention for studying data innovation and governance in multi-actor environments.

This thesis aims to theorize data innovation and governance as simultaneous processes and account for the distinctive nature of data by utilizing the concept of space. Building on the realist, process-oriented ontology of assemblage theory (DeLanda, 2000, 2006, 2013, 2016), I argue how processes of data innovation and governance change their spatial configurations as they unfold across various data spaces; how the realities data refer to, condition the forms innovation and governance can take and are shaped by these processes in return. By taking this stance, I treat space and time, structure and process, stability and fluidity not as dualisms – or opposing, but as mutually-enabling dualities (Farjoun, 2010).

This thesis consists of two parts. In Part I – *The Kappa* – I bind the individual papers together by investigating data governance and innovation as simultaneous processes utilizing the concept of space based on the ontology of assemblage theory. In Part II – *The Individual Papers* – I present empirical studies conceptualizing data in multi-actor environments and argue for understanding data's ontological status as a duality of structure and process.

Empirically, this thesis investigates innovation and governance of personal health data in the multi-actor, highly regulated Norwegian healthcare context. By conducting an embedded study consisting of two cases, I show how data's value potential was not open-endedly explored, but required constant reconfigurations of actor relationships, digital technologies, organizational means, legal basis, and purposes for data processing, as data were produced, used and shared across multiple public and private actors.

The thesis contributes to IS research on data, and more specifically to data-driven value creation (originating from the literature on digital innovation) (Aaltonen et al., 2021; Alaimo, Kallinikos, & Aaltonen, 2020; Alaimo, Kallinikos, & Valderrama, 2020; Alaimo & Kallinikos, 2022; Kazemargi et al., 2023; Mikalsen & Monteiro, 2021; Østerlie & Monteiro, 2020) and data governance (Davidson et al., 2023; Gegenhuber et al., 2023; Jarvenpaa & Essén, 2023; Vial, 2023), by theorizing data innovation and governance utilizing the concept of space. I identify the following reasons for engaging with these debates in IS.

First, the importance of protecting data while maximizing their value potential has been raised by IS scholars (Vial, 2023), however, the literature streams of data innovation and data governance have been commonly developing separately. Both literature streams have identified different dynamics in creating value from data (Barrett et al., 2016; Tempini, 2017), and governing data (Abraham et al., 2019; Jagals & Karger, 2021; Van den Broek & Van Veenstra, 2015) within organizations and across inter-organizational environments. However, assuming a clear separation between organizations' intra- and inter-organizational environments is challenging due to the involvement of data. Data are seldom produced within single organizations and do not simply travel by transferring responsibilities from one organization to another. Instead, as the empirical cases in this thesis show, access is not exclusive, and data are not simply "owned" like other organizational assets. Data can be stored at one place, accessed from another, and simultaneously exist "here" and "there" while being prone to different rules, copied, aggregated, and repurposed across multiple actors. As argued in this thesis, the concept of space can accommodate both, the emergence of various forms of data innovation and governance encompassing multiple actors beyond the micro-macro divide, as well as the changing configurations from one form to another.

Second, IS scholars have recognized how data as distinctive entities, can decouple from the realities they refer to, as data have a semantic nature and engage in processes of meaning-making and knowledge production (Alaimo et al. 2020). It is argued how data's attributes of editability, portability, and recontextualizability allow data to form larger objects (Alaimo, 2021; Alaimo & Kallinikos, 2022), and move from being tokens (such as call detail records) to becoming commodities, as aggregates of data aligned with business objectives (Aaltonen et al., 2021). These "[d]ata objects are made by aggregating data and metadata under a given structure.

This is a key passage. Once embedded in a structure, data become less dependent on external referents and able to produce new insights by relating to each other and, as objects, to other objects" (Alaimo 2021, p. 05).

However, little attention has been paid to the heterogeneous realities data refer to, and the conditions these realities impart to the processes of innovating with and governing data. Data can be non-personal and personal, open and sensitive, anonymized and non-anonymized, non-regulated and regulated. The initial claims for data decoupling from the realities, objects, and events, they refer to, come from empirical settings around "digital born" companies, such as telecommunication networks, or social media platforms. However, in other empirical settings, such decoupling can be a matter of degree. For instance, Østerlie and Monteiro (2020) show how, despite their ability to decouple, the digital representations of physical objects – in their case sensor data about sand in oil and gas production – never fully detached from the realities they refer to.

I argue how the realities data refer to are heterogeneous, and this heterogeneity affects the degree to which data can decouple from the real-world objects, events, or people data are about. Moreover, using the ontology of assemblage theory, I argue for zooming out of data's realities beyond the real-world objects, events, and people they refer to, and acknowledge the underlying industrial, technological, legal, organizational structures that, as part of that reality, affect how data can be innovated with, and governed.

1.2 Motivation

My motivation behind this work comes from my engagement with the IS literature on data and personal health data as an empirical domain covered in this study. The research question and theoretical matters investigated in this thesis are also motivated by engaging with the case study of interest, which began as an exploratory case study (Yin, 2017). In the preliminary interviews, I expected to learn more about innovating with person-generated health data in the multi-actor healthcare context. However, the initial insights exposed how innovating with personal health data is a challenging process that requires navigating through a complex techno-legal landscape incorporating various public and private actors. Aiming to understand data's innovation potential and the necessary governance approaches to realize that potential has thus been driving this project from the start.

My practical motivation comes from recent developments in regulating and innovating with (personal) data, including the European General Data Protection Regulation (GDPR) (Befring, 2021) and the European Data Spaces (European Commission, 2020). The data spaces are aimed at working as domain-specific and shared infrastructures that harmonize rules and overcome legal barriers in access, sharing, processing, and usage of data for innovation across various industries and sectors; as well as ensure semantic, technical, and legal interoperability across various sectors, e.g., healthcare, or finance. As of now, the term “data spaces” is used as a metaphor in practice and research; the conceptual development of data spaces requires further investigation, although initial work on conceptual clarity is already present (Hutterer et al., 2023).

The choice of topic is also motivated by my educational, professional, and personal background. As a master’s student in e-business management, I researched the governance of the national e-health platform in North Macedonia; sparking my interest in researching healthcare as a context, and health data’s innovative potential. During my work experience in the industry, among other tasks, I also “exploited” data’s potential for personalized marketing value using search engines and social media platforms. Last, but not least, my personal beliefs have been shaping this project since the start. Having experienced the instability of institutions and the discrepancy between defined rules, laws, regulations, and their actual implementation, I resisted to regard institutions as stable structures existing “out there” bringing absolute order to the world we live in. Instead, I aspired to regard them as relatively stable entities whose function is dependent on the socio-political entities that give rise to them. This has been a central thought I have been carrying throughout this project.

1.3 The Empirical Case and its Role in The Research

My thesis aims to contribute with theory building (Alvesson & Kärreman, 2007; Corley, 2022). The role of the empirical material is not to fit, or justify theory, but to serve as a resource for inspiring the development of theory by problematizing the existing understanding of data innovation and governance. I provide a set of concepts that are abstracted from a case study, iterated with theory, and generalizable across contexts (Lee & Baskerville, 2003). The set of theoretical concepts provided posits the existence of relationships between entities which cannot be directly observed in the empirical material but can only be theorized (ibid).

Using the term "empirical material" is not an accidental, but a deliberate choice. As per the work of Alvesson and Sköldberg (2010), I consider empirical "data" (hereafter referred to as empirical material) to be constructions that emerge through my interactions with the informants and the empirical field. Therefore, I do not solely encounter the empirical material and let it lead me; rather, I am actively framing and constructing it. With this, I acknowledge that my personal understanding and engagement with healthcare as an empirical field, the larger socio-institutional context, and the knowledge shared within and beyond information systems (IS) as a research community, have shaped my interpretations of the empirical material. I would characterize my methodology as a reflexive and abductive (Alvesson & Sköldberg, 2010), as I treat theory and the empirical material as mutually co-evolving, while also allowing myself to be surprised and let the insights shape my own theoretical understanding of data in multi-actor environments.

My theory building falls under the paradigm of process theorizing (Cloutier & Langley, 2020; Langley & Tsoukas, 2022). By adopting a process view, I construct theoretical concepts which focus on the relations between entities, and not on independent entities as the primary objects of theorizing. This helps me acknowledge the complexity of data and their ontological status, the multi-actor environments they are produced, shared, and used across, and healthcare as an extreme context. I treat processes as the dynamic unfolding of spatio-temporal relations over time; I also acknowledge the existence of outcomes and relatively stable structures across the process over time.

My empirical study encompasses two embedded empirical cases; the choice of two cases helped me bring more robustness to the theoretical generalizations regarding data (Yin, 2017). The first case provides a retrospective study of the 10-year evolution of a national citizen-facing digital health service, HealthNorway. The study follows how data were governed across multiple actors as HealthNorway provided functionalities related to producing, sharing, and using personal health data for citizens and various healthcare services. The user interface for citizens at

HealthNorway and some of the functionalities offered are presented in Figure 1.

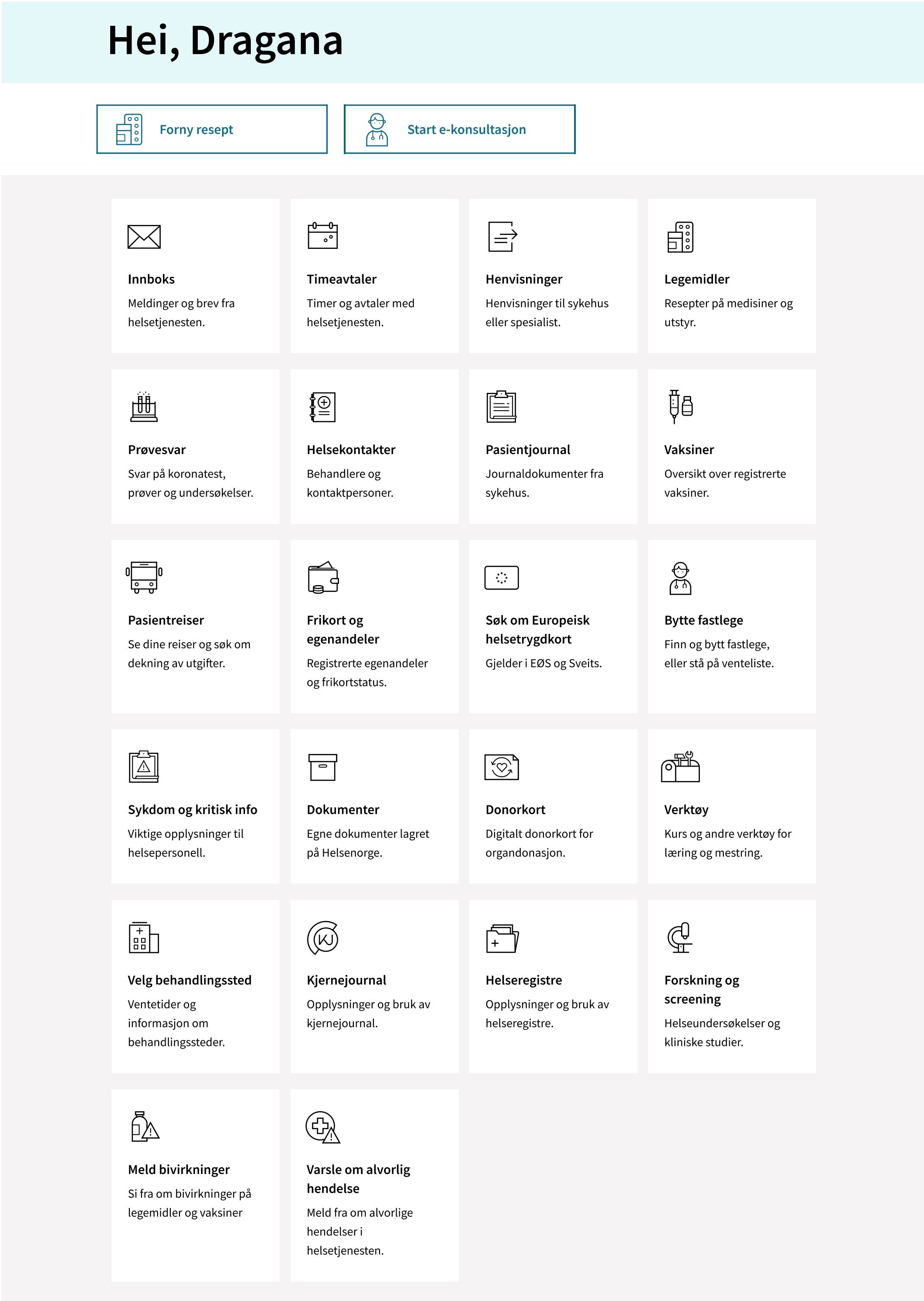


Figure 1: Citizen user interface on HealthNorway

The second case follows a real-time process platform initiative in the southeast region of Norway, aiming to provide a shared infrastructure for innovating with work processes around remote-care monitoring (RCM). RCM services include sensor-based technologies, digital consultations, and structured data forms (see Figure 2), provided by private vendors, and used by patients at home. The services are catered for well-defined patient groups, e.g., diabetes or epilepsy, offered by individual hospital departments and operated through local procurement agreements with different commercial vendors. Re-use of RCM solutions and service models, or sharing and secondary use of the data, are generally not considered. I conduct a retrospective study of the 11-year unfolding of techno-legal arrangements for sharing personal health data across specialist healthcare services in the southeast region of Norway and end the account with the current process platform initiative. The study shows how the interplay of technology and law defined territories across which patient data from EPR systems and patient-generated health data (PGHD) could be produced, shared, and used; these territories were not defined by organizational boundaries, but by techno-legal configurations.

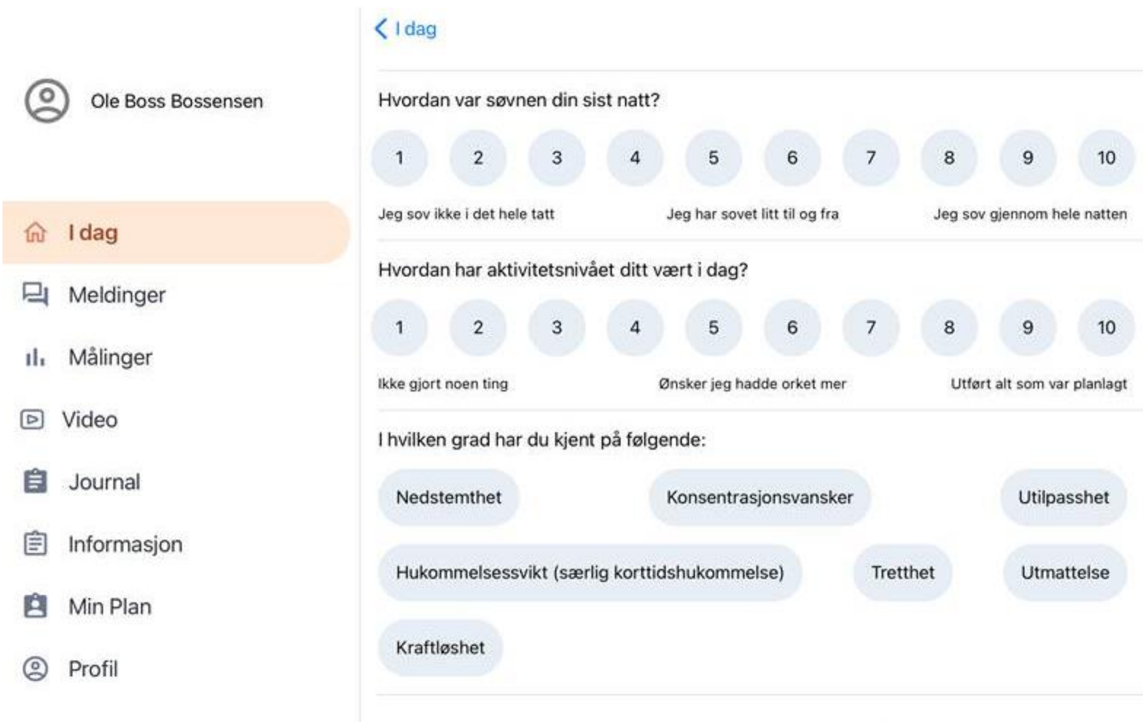


Figure 2: Sample form for remote care monitoring of patients

1.4 Research Questions

Data's dual role in engaging in processes of value creation and governance is an emergent topic of interest to IS scholars (Davidson et al., 2023; Gegenhuber et al., 2023; Jarvenpaa & Essén, 2023; Vial, 2023). However, data innovation and governance have been developing as separate literature streams in IS, commonly distinguishing between governing and innovating with data in intra- and inter-organizational environments. More recently, scholars have also called for "the need to move beyond organization-level IS research to explore more fully how data governance is nested across micro-, meso-, and macro-levels" (Davidson et al., 2023, p. 06). Moreover, it was emphasized how "data have no [such] fixed boundaries. Once they are obtained, they can be easily copied, altered, falsified, and used for a purpose that is vastly different from their original intent." (Vial, 2023, p. 10). Therefore, data can be produced, stored, copied, shared, and used across various actors, imparting distinct innovation and governance dynamics, unfolding beyond organizational boundaries. To revisit the conceptualization of data innovation and governance as simultaneous processes beyond the intra–, and inter-organizational divide, I pose the following research question:

> *RQ1: how can processes of data innovation and governance in multi-actor environments be theoretically accounted for, utilizing the concept of space?*

This thesis utilizes the concept of space to theorize data innovation and governance in multi-actor environments as processes that unfold across certain structures of possible forms. The main focus is not on mapping out the outside boundaries of space, but to show how data governance and innovation form and transform across various "nested" spaces, which do not contain each other. The empirical studies, as presented in the individual papers, show how multiple data spaces were formed around processes of data innovation and governance with personal health data (based on my empirical work), and data about physical assets in the oil and gas industry (based on a co-author's empirical work).

Data's uniqueness as entities has also been discussed in the IS around data's properties of being editable, portable, and recontextualizable (Alaimo, Kallinikos, and Aaltonen 2020). Scholars have argued how these properties allow data to decouple from the real-world events, objects, people they refer to (ibid.), although empirical studies from traditional industries also show how the digital and physical do not simply mirror one another, nor do they completely decouple (Østerlie and Monteiro, 2020). Therefore, data can refer to various heterogeneous realities, e.g., physical assets on oil and gas platforms (Østerlie and Monteiro, 2020), people's health (Grisot et al., 2019; Tempini, 2017), people's online behavior (Alaimo, Kallinikos, & Valderrama, 2020). To revisit IS scholars' theorizing of data, and accommodate the heterogeneous realities data refer to and their implications for data innovation and governance, I pose the following research question:

> *RQ2: how are processes of data innovation and governance conditioned by data's unique nature?*

The empirical studies presented in this thesis show how, beyond the people, objects, or events data refer to, data's realities also encompass larger structures, such as actor structures (industrial/sector actor relations, e.g., collaboration, cooperation, competition), organizational structures (functions, contracts), or legal structures (laws, regulations). I argue how these larger structures allow certain forms of innovation and governance to take place, and constrain others.

1.5 Expected Contributions

This thesis aims to contribute to the IS literature on data (Aaltonen et al., 2021; Alaimo, Kallinikos, & Aaltonen, 2020a; Mikalsen & Monteiro, 2021), data innovation (Aaltonen et al., 2021; Haugjord & Kempton, 2022; Tempini, 2017), and data governance (Davidson et al., 2023; Jarvenpaa & Essén, 2023; Vial, 2023) by conceptualizing data innovation and governance as simultaneous processes in multi-actor environments. This thesis contributes to these literature streams in two ways.

First, it shows how the concept of space can be utilized to conceptualize data innovation and governance in multi-actor environments beyond intra- and inter-organizational boundaries. Data spaces, as conceptualized in this thesis, are neither solely Euclidean or "containers", or open-ended networks; instead, data spaces are structures across which processes of data innovation and governance can form and change their forms. Second, this thesis argues for zooming out of data's reality beyond the objects, people, and events they refer to, and accommodate the larger technical, legal, and actor structures. I state how these structures condition the forms data innovation and governance can take and are shaped by these processes in return.

This thesis also has a practical contribution, as the conceptualization of data spaces can inform practical debates on the European data spaces and their innovation and governance aims across sectors, industries, and countries' borders. Moreover, it also provides insights into how organizations can understand the role of the law, as both, an actor that can delegate roles and responsibilities, and as a structure that conditions how organizations can produce, share, and use data.

1.6 Structure of the Thesis

The thesis consists of this introductory chapter as first, and six other chapters.

In Chapter **Two**, I review the IS literature on data as organized around the two research questions. First, I provide an overview of the literature streams on data innovation and governance. Then, I move into exploring IS scholars' conceptualization and ontological assumptions around data.

In Chapter **Three**, I provide an overview of debates on space in the philosophy of science and the IS literature. Then, I show how space is theorized in the realist, process-oriented ontology of assemblage theory, based on which I conceptualize data spaces.

In Chapter **Four**, I present my research approach. First, I describe the embedded case study. Then, I present the research design, including my ontological and epistemological stance, the methodology adopted, and the role of theory. Lastly, I elaborate on the data collection and analysis process.

In Chapter **Five**, I provide a meta-analysis of the individual papers, as included in this thesis, answering the two research questions. The meta-analysis shows how processes of innovation and governance unfold across various data spaces by changing their spatial configurations. It also shows how, in the empirical cases followed in this thesis, data did not simply decouple from the realities they refer to, but were conditioned by, and shaping these realities in return.

In Chapter **Six**, I discuss the meta-analysis and present the overall contribution of the thesis, which is two-fold. First, I show how the concept of data spaces can be utilized to study data innovation and governance beyond single organizations' boundaries. Second, I argue for zooming out of data's realities beyond the objects, events, and people data refer to, and accommodate the larger organizational, technical, and legal structures.

In Chapter **Seven**, I conclude the thesis, discuss the thesis' limitations, and suggest further directions for research.

2 IS Research on Data

In this chapter, I provide an overview of the IS literature on data, organized around the two research questions. First, data innovation and governance with a particular focus on multi-actor environments. Second, IS scholars' theoretical assumptions around data as distinct entities.

2.1 Data Innovation

The literature stream on data innovation (Aaltonen & Penttinen, 2021) – data's recombination potential to be aggregated into larger objects (Aaltonen et al., 2021; Alaimo, Kallinikos, & Aaltonen, 2020a) – originates from the literature on digital innovation. However, within organizations, data and value have been explored earlier, as part of the literature stream on big data.

Big data, referring to the large volumes, diversity, frequency, and speed of growth of data (Lycett, 2013), were conceptualized by IS scholars as different than the data traditionally collected across well-structured schemes in organizations (Constantiou & Kallinikos, 2015). These messy, heterogeneous, unstructured data sources, e.g., social media data, were captured without a pre-determined purpose "used, combined and interpreted to become relevant to strategic pursuits" (Constantiou and Kallinikos 2015, p. 47), i.e., to yield organizational value (Woerner & Wixom, 2015). Within organizations, the big data literature was conceptually focused on two aspects, as defined by Günther et al. (2017): 1) data work practices related to data gathering, which can be inductive or without a pre-defined purpose, e.g. social media data, or deductive and with a pre-defined purpose, e.g. healthcare; and 2) organizational capabilities to create value from big data, including innovating with business models and defining centralized or decentralized governance structures.

As empirical investigations followed, data's attribute of *big* was slowly abandoned; instead, scholars started focusing on the uniqueness of data as a resource for value creation. One of the early empirical works is Aaltonen and Tempini´s (2014) study on a telecommunications operator aiming to create advertising audiences from a pool of data. The authors show how data are not valuable in themselves, instead, data are useful or meaningful only if organizations set up mechanisms to actualize their value potential. Data's value potential to achieve organizations' strategic aims remains of interest in recent studies (Aaltonen et al., 2021). However, beyond single organizations, empirical works have also been increasingly concerned with the dynamics of value creation (Barrett et al., 2016) and disruption (Tempini, 2017) in multi-actor environments. For instance, Barrett et al.'s (2016) and Tempini's (2017) studies on online communities where various stakeholders interacted around shared infrastructures, show how different forms of value can be created, changed, and disrupted – the value creation dynamics are not necessarily additive but represent potentially conflicting interests (see also Vassilakopoulou et al. 2019, Monteiro and Parmiggiani 2019). More recently, in the context of digital ecosystems, data have been understood as carriers of value that create complementarities across ecosystem actors (Alaimo, Kallinikos, & Valderrama, 2020; Kazemargi et al., 2023). Data are seen as "an essential and specific type of resource whose value is contingent on its constant updatability, portability, and sharing" and "data are also a key medium by which business relationships and connections are forged in the digital economy" (Alaimo et al. 2020, p. 26).

Therefore, within organizations, scholars' focus was predominantly on creating value from data for achieving strategic aims. In inter-organizational environments, scholars' focus was predominantly on actors' interdependencies and complementarities as different forms of value from data were created and disrupted.

2.2 Data Governance

Traditionally, works on data governance have been building on IT governance as a conceptual framework, treating data as inherent to IT, as noted by Benfeldt (2017), and recognized the distinctiveness of data only to a limited degree. Like the literature on data innovation, within organizations, works on data governance have been characterizing data as strategic assets (Black et al., 2023; Fadler et al., 2021; Fadler & Legner, 2020; Khatri & Brown, 2010) aimed at achieving organizational goals. The governance of data assets gave rise to new organizational roles, such as data managers, data stewards (Rosenbaum, 2010), data owners (Fadler & Legner, 2020), but also new challenges related to the specific nature of data, such as issues related to data quality (Fadler et al., 2021), integrity (Winter

& Davidson, 2019), privacy and security (Abraham et al., 2019; Janssen et al., 2020; Rosenbaum, 2010).

However, as Janssen et al. (2020) point out, a common challenge "is that the data flow and logic may not follow the structure of an organization. The mismatch between organizational structure and data usage can easily result in data silos, duplications, unclear responsibilities, and missing control of data over its entire life-cycle." (p. 03). Therefore, the involvement of data commonly implies involving digital technologies and stakeholders beyond organizational boundaries. Works on data governance in inter-organizational environments are also present (Jagals & Karger, 2021; Van den Broek & Van Veenstra, 2015; Winter & Davidson, 2020), however, similar to the literature on data innovation, these works predominantly highlight stakeholders' relationships, and pay less attention to the distinctive nature of data as a resource to be governed, with some exceptions.

For instance, Van den Broek and Van Veenstra (2015) acknowledged how the governance of personal health data should be hierarchical due to the sensitivity of data, while the governance of data e.g., about energy, or municipal activities can be open, or organized in networks, such as around digital platforms. In their framework of data governance, Abraham et al. (2019) also differentiated between intra- and inter-organizational governance, and between traditional and big data, but no category has been devoted to data being personal or non-personal, used for primary or secondary usage, open data or regulated data. Scholars have also argued how, in the case of personal data, such as personal health data, laws play a key role in determining the governance approaches (Rosenbaum, 2010; Winter & Davidson, 2019, 2020).

Overall, data governance has been commonly defined as a framework for organizational decision-making rights, roles, and responsibilities (Abraham et al., 2019; Benfeldt, 2017). However, more recent works have been claiming how the nature of data brings in specific challenges for both, governing and innovating with data (Davidson et al., 2023; Jarvenpaa & Essén, 2023; Vial, 2023), which can be summarized around two main points.

First, these works (Davidson et al., 2023; Jarvenpaa & Essén, 2023; Vial, 2023) argue how data are not fixed but move across organizational boundaries and beyond organizations' unilateral control – such as across networks, ecosystems, information infrastructures. For instance, Jarvenpaa and Essén (2023) argue how, data do not solely transcend organizational boundaries, but could span across technological and human generations. As Davidson et al. (2023) raise "many data resources are situated outside a single organization's boundaries and beyond its unilateral control" (p. 03) as data are commonly nested across societal levels, and need to incorporate governance approaches that account for the rights of individuals the data are about. This, "entails new, distributed organizational forms enacted by individuals, technology vendors, data-holding (or using) organizations, and regulatory agencies (ibid, p. 04).

Second, data's future use is unpredictable and can have unprecedented societal consequences. For instance, Vial (2023) claims that "data governance mechanisms in place at time t may not anticipate potential use cases for data at time t + n. As a result, data governance can be perceived as series of measures that hinder digital innovation because it constrains the ability to find innovative uses for data" (p. 05). Therefore, favoring one element (e.g., protecting data) at the expense of the other (e.g., digital innovation). These insights have implications for how we theorize data (Vial, 2023), as "once they [data] are obtained, they can be easily copied, altered, falsified, and used for a purpose that is vastly different from their original intent." (ibid., p. 10). Jarvenpaa and Essén (2023) also argue how data's use in future socio-technical regimes, and by heterogeneous actors, requires "to learn to go back to data resources that were previously pushed aside, forgotten, or viewed as inferior as we struggle not only to imagine alternative futures, but also to understand, solve, or even prevent problems in those futures." (p. 10).

These insights require rethinking if studying data governance as a framework distinguishing between intra- and inter-organizational environments is enough to account for the nature of data, and for how data are innovated with, and governed in multi-actor environments.

2.3 The Distinctiveness of Data Entities

In the late 1990s, data were regarded as facts which are combined into meaningful structures to become information, and put into a context to become knowledge; "data are a prerequisite for information, and information is a prerequisite for knowledge" (Tuomi, 1999, p. 104). Tuomi (1999), challenged this hierarchy by arguing how data emerge last, only after structure and semantics are fixed to represent information. This view on data as not just existing "out there" waiting to be used, but instead, being produced and worked on to yield organizational value has been dominant in IS. IS scholars have advocated how "data do not 'have' a structure but are made by a structure that confers data their capacity to represent contextual facts" (Aaltonen and Penttinen 2021, p. 5922). By referring to structure, Aaltonen and Penttinen (2021) did not only regard structure as a technical matter but as entangling social practices, institutional settings, and organizational processes to create new types of value. Therefore, data's structure is relational; data are structured to fulfill specific purposes and are embedded into a structure to be contextualized as per their semantic context.

In the early big data works, beyond their volume, variety, veracity, and value, data were also conceptualized around their ability to *dematerialize* – separate from their context, *liquify* – be ported once unbundled or dematerialized, and *density* to be recombined and mobilized into other contexts (Lycett, 2013). Data could also be updated over time (Constantiou & Kallinikos, 2015), and interconnected across various data sources (Günther et al., 2017). Similar views were adopted in the literature on data innovation (or data-driven value creation). As individual tokens, data were defined as *granular*, as they can be aligned, aggregated, and juxtaposed with other data into larger objects (Aaltonen & Tempini, 2014), *editable* (can be continuously revised, renewed and expanded), *portable* (can be shared across various digital technologies) and *re-contextualizable* (can be distanced from their origin and re-assigned meaning) (Alaimo, Kallinikos, & Aaltonen, 2020).

Similar to the big data literature (Constantiou & Kallinikos, 2015; Günther et al., 2017; Kallinikos & Constantiou, 2015; Lycett, 2013), a common understanding in the literature stream on data innovation was how, once data were structured into larger objects, their value-creation potential was *open-ended*, as data are semantic objects related to meaning-making (Aaltonen & Tempini, 2014). However, the main focus in the literature on digital innovation was the uniqueness of data entities, instead of their property of being *big*.

Alaimo, Kallinikos, and Aaltonen (2020) characterized data as distinct from IT components; data do not embody functions, but are carriers of signs, meaning, and knowledge about the realities that they refer to. They argued how data do not follow the recombination logic of modular architectures. Instead, once data are aggregated into larger objects under a given structure (Aaltonen et al., 2021;

Alaimo, 2021; Alaimo & Kallinikos, 2022) they "become less dependent on external referents and able to produce new insights by relating to each other and, as objects, to other objects. Rather than representing existing entities, data objects construct new entities out of data produced in dispersed digital environments. Data and data objects are mutually co-constitutive" (Alaimo, 2021, p. 05).

IS scholars' assumptions about data and the realities they refer to can also be traced across works on digital representations. For instance, by following the mechanisms for monitoring sensor data about sand in oil and gas production, Østerlie and Monteiro (2020) argued: "[D]igital representations, despite their theoretical capacity, were never dichotomously separated from their physical origin. Rather, the mechanisms through which digital representations come to be implicated in organizational action transcend characteristics of the referent/reference relationship" (p. 12). Therefore, digital representations do not decouple from the physical domain but resemble it to a degree.

In summary, IS scholars have acknowledged how data are distinct from digital technologies, as they have a semantic nature related to meaning-making once used by actors for value-creation purposes. This uniqueness of data as entities requires further consideration in empirical and conceptual works studying data's dual role in innovation and governance processes.

2.4 A Synopsis

The review of IS research on data can be summarized in the following ways. First, data are often produced beyond single organizations' boundaries, stored at multiple locations, copied across various digital technologies, and used for different purposes – this brings distinctive challenges to studying data innovation and governance (Davidson et al., 2023; Vial, 2023). Second, data are unique entities, with distinctive properties (Aaltonen & Tempini, 2014; Alaimo, Kallinikos, & Aaltonen, 2020a) which have the capacity to decouple from the digital technologies carrying them, and the realities they refer to (Alaimo, Kallinikos, & Aaltonen, 2020a; Østerlie & Monteiro, 2020). For instance, Alaimo (2021) argues how data objects, despite their resemblance with the reality from which they originate, are simply digitized versions of such realities; therefore, once structured into digital data, they decouple from the realities they refer to (Alaimo, 2021; Alaimo, Kallinikos, & Aaltonen, 2020b).

In this thesis, I complement these works in the following ways. First, I argue how the involvement of data challenges the distinction between intra- and inter-organizational environments. In the chapter that follows, I show how the concept of space can serve as a useful tool for conceptualizing data innovation and governance as simultaneous processes in multi-actor environments. Second, I argue how the decoupling of data from the realities they refer to should be more nuanced to account for the heterogeneous realities data refer to. While such decoupling might be more emphasized in big data environments (Alaimo, Kallinikos, & Aaltonen, 2020) where data can be anonymized and pooled together to yield e.g., population-level insights, in other settings, such as personal health data (Günther et al., 2017), the production, sharing and usage of data is highly regulated. Despite the digitized version of personal health data acquiring the properties of editability, portability, and recontextualizability, the laws, institutional, organizational, and social norms do not allow for a complete decoupling of data from the persons, objects, or events they refer to.

3 Philosophical and Theoretical Grounding

In this chapter, I provide an overview of the philosophical debates on space and time and their substantialist or relationist worldview. Then, I present an overview of core IS works conceptualizing (or metaphorically building on) space. Lastly, I show how the concept of space can be understood as both, Euclidean and networked, bounded and unbounded, by building on assemblage theory (DeLanda 2006, 2013, 2016).

3.1 Space and Time in The Philosophy of Science

Scientific ontology deals with foundational beliefs of what the world we are researching is comprised of. To be more concrete: what kinds of entities, relations, processes, and structures, exist in such a world. Time and space, as top-level ontological concepts, have been the main focus of such philosophical debates throughout centuries (Barbour, 1982; Dainton, 2014; Massey, 2005). Space is regarded as creating the conditions for certain opportunities, and not others, distinguishing between "here" and "there". Time is regarded as a dimension of change, referring to what is "now" and "then".

One crucial question in these debates is whether space and time exist as entities in their own right, leading to two opposing views, *substantialism* and *relationism* (Ballard, 1960; Barbour, 1982; Dainton, 2014). As Dainton (2014) elaborates, "Substantivalists maintain that a complete inventory of the universe would mention every material particle and also mention two additional entities: space and time. The relationist denies the existence of these entities. For them, the world consists of material objects, spatiotemporal relations, and nothing else." (ibid, p. 02). Accordingly, both stances bring in certain assumptions about space and time; substantialists commonly regard space and time as *static (containers)*; relationists commonly regard space and time as *dynamic*. In this subchapter, I review some of the static and dynamic views on space and time as per Dainton's (2014) work.

	Substantialist	Relationist
Space	Space as an entity in its own right: External, absolute	Space is nothing but spatial relations between entities
Time	Time as an entity in its own right: Objective, clock-time	Time is temporal relations between events: Subjective, process time
Structure	Container, geometrical, block	Relational, open
Space and objects	Space contains material objects	Space is an ordering of objects
Time and events	Time contains events	Time is an ordering of events
Change	Movement in relation to absolute time and space	Movement in relation to other objects and events
Whole/totality	Totality: space is a sum of all coexisting material objects; time is a sum of all coexisting events	Whole: space emerges through relations between objects; time emerges through relations between events

Table 1: Comparison of substantialist and relationist conceptualizations of space and time

3.1.1 Substantialist Views on Space and Time

In a *substantialist* view – commonly regarded as a block-view – the world is a *container* in which everything else exists and occurs. The ocean contains water, fish, algae, microorganisms, and other sea life. These entities exist independently, as fixed and finished forms, and can be clearly separated from their environment. For substantialists, a complete worldview of the universe would contain every material particle and two additional entities: space and time. Space and time are the biggest things there are and everything else exists and occurs *in* them. They are, therefore, finite and absolute. Conceptualizing space as absolute means that space exists independently of the objects that occupy it. Therefore, if we were to remove all material objects, space would still be "out there". Similarly, conceptualizing time as absolute, indicates that time is objective and flows independently of the events that are taking place; if we were to remove all events, time would still be "out there".

The container is commonly regarded as geometrical (Euclidean) and a four-dimensional block (see Figure 1); space is a three-dimensional arrangement in which places coexist; time is a linear one-dimensional series of coexisting events. The coexistence does not indicate that the places and events exist all at once; rather the block is an ensemble of places and events coexisting in different locations. Therefore, space can be reduced to the sum of its material parts, and time can be reduced to the sum of all events that have taken place. Since space and time make up the four-dimensional shape of the block, the block does not exist in space and time, but space and time exist *in* the block.

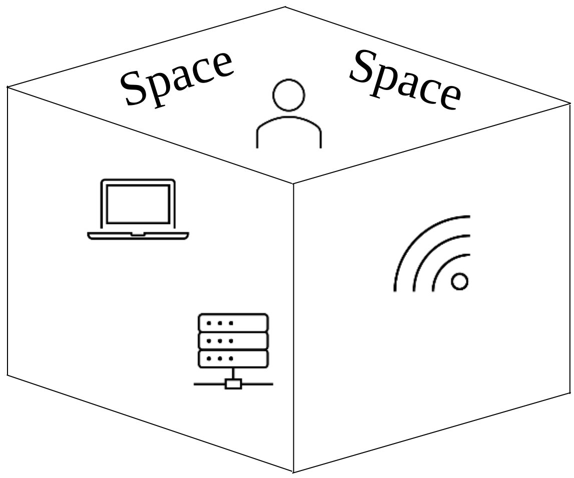


Figure 3: Substantialist views on space as a four-dimensional container, recreated from Dainton (2014)

The main characteristic of substantial space is the relation between objects located in a particular space, and the space itself. Although substantialists do not deny that objects can change their place relative to other objects, *what distinguishes substantialism from relationism is the ability to move relative to absolute space*. Therefore, objects move by occupying different places in space; and objects occupy a pre-existing place which is theirs in that space. For instance, Aristotle believed that the four elements: earth, water, fire, and air had natural places and natural motions in space; if removed from this place, they would aim to return there. According to him, the natural motion of heavy things – earth and water – is downwards; the natural motion of light things – air and fire – is upwards, away from the center of the Earth. Therefore, if objects are to be moved from this natural place, a force is applied, and they are moved back to *their* place.

Overall, adopting a substantialist view would imply how data innovation and governance are fixed and finished entities, contained by higher-level entities, such as organizations, or societies. Moreover, it would also imply how organizations are clearly demarcated from their environment, and although relating to such environment, these relations would be treated as secondary. Therefore, structural unity would be the primary focus of analysis (Cooper, 2005), and entities such as structure and process would be characterized as dualisms, or independent substances, instead of interdependent and mutually enabling and constraining. As stated by Farjoun (2010), "[d]uality resembles dualism in that it retains the idea of two essential elements, but it views them as interdependent, rather than separate and opposed" (p. 203).

3.1.2 Relationist Views on Space and Time

The substantialist view of space and time has been challenged by relationists. Relationists conceive ontology based on events and processes, instead of substances and things. For relationists, it is the relations between things, rather than the things themselves that matter. Relations can be defined as connections, interactions, sequences, causes and effects, or as spatio-temporal. Entities such as individuals, groups, and society are not independent, yet always in relation, characterized by changing relationships. Therefore, for relationists, space and time do not exist as independent entities; instead, space and time are constantly produced and altered, as objects and events change their relations relative to other objects and events.

This is the main difference between a substantialist and a relationist view on space and time. According to relationists, there is no such thing as absolute space or absolute time; objects and events do not move in relation to a four-dimensional block, but *move by changing their spatio-temporal relations relative to other objects and events*. Therefore, space and time are irreducible to the material objects that occupy them or the events taking place; space and time are more than the sum of the parts. Space does not exist "out there" independently from material bodies but is a network of relations between material bodies; time does not exist "out there" independently from events, but time is branching as events relate to each other. Therefore, a relational understanding of space does not focus on object-space relations, but on object-object relations; a relational understanding of time does not focus on event–time relations, but on event-event relations.

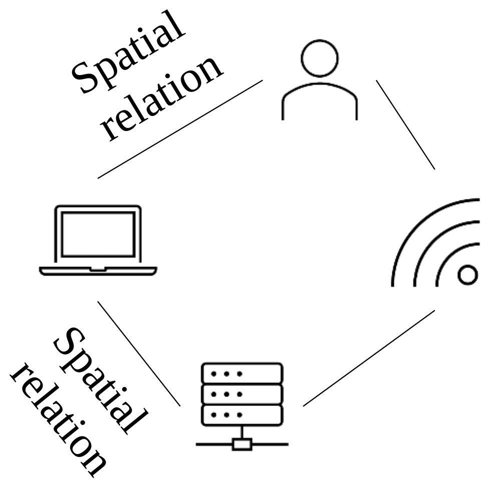


Figure 4: Relationist view on space as a network of spatial relations, recreated from Dainton (2014)

According to relationists, objects and events can have various motions and places across space and time. One of the most influential philosophers of relationism, Leibniz, claimed that space has an *order of coexistence*, the same way that time has an *order of successions*, and the places that material bodies occupy across space do not differ from one another (Ballard, 1960; Northrop, 1946). Therefore, the ordering of objects and events is significant; space is not a mere aggregate of coexisting objects, and time is not a mere ensemble of coexisting events. This is another crucial argument that differentiates relationism from substantialism; time and space are not the sum of objects and events, but the focus is on *temporal and spatial ordering*.

In IS, data are commonly conceptualized as relational entities; (Aaltonen & Penttinen, 2021; Alaimo, Kallinikos, & Aaltonen, 2020a), although authors do not clearly state if their view is based on a relational ontology. Overall, a relational ontology would indicate how data innovation and governance are formed across a network of relationships, where forms of innovation or governance are never complete in themselves, but always subjected to changing relationships (Cooper, 2005). Moreover, it could also move the focus toward conceptualizing data innovation and governance as dualities that are mutually enabling and constraining, instead of opposing. I now turn to showing how the concept of space has been utilized in IS.

3.2 The Concept of Space in The Information Systems Field

Developing ontological assumptions is core to any scientific field, including IS. As noted by Little (2016), it is not possible to research a domain well if we do not know what things or processes it consists of. Time and space, as high-level ontological concepts, have been at the center of various philosophical and scientific fields; however, their explicit conceptualization in IS has been scarce, with few exceptions (Haj-Bolouri et al., 2023; Mousavi Baygi et al., 2021). Although a comprehensive review of IS conceptualizations of space is beyond the scope of this thesis, in this chapter I build on core literature, e.g., scholars who have specifically promoted the usefulness of the concept of space in the IS (Haj-Bolouri et al., 2023; Sahay, 1997); and scholars who have used the concept of space metaphorically (Henfridsson et al., 2018; Winter et al., 2014). Therefore, the review of the literature is not representative of IS fields' understanding of space; instead, it brings significant insights that can be utilized when studying data innovation and governance.

The early studies of information systems, from the 1970s to the 1980s, were marked with substantialist, space-centric views, as Cecez-Kecmanovic (2016) notes. "IS researchers tended to adopt a positivist research paradigm, assuming that the world consists of discrete entities – human beings, organizations, technologies, processes, products, accounts, and others – that exist independently of observers" and "[o]f particular interest to positivist research is how information systems and information technologies (IS/IT) as an autonomous, exogenous force impact on organizational processes and structures." (ibid, p. 11). Sahay (1997) also argued how the IS field was studying IT implementation in line with Newtonian substantialism, identifying factors and forces that impede and enable IT implementation processes. Beyond paradigmatic distinctions of scholars' philosophical choices, the concept of space has also been utilized to study virtual worlds (e.g., Goel et al. 2011) or geographical information systems (e.g., D'Mello and Sahay 2007; O'Leary and Cummings 2007).

In the literature on digital innovation, space has been used metaphorically to conceptualize the open-ended value recombination of digital resources (Henfridsson et al., 2018). Henfridsson et al. (2018) define *value spaces*, as "an evolving network of digital resources interlinked through connections established and dissolved by actors seeking to generate and appropriate value." (p. 92). As the authors argue, innovating with digital resources unfolds across multiple spaces which hold possibilities for value-creation and capture. A digital resource belongs to a particular value space but can be part of multiple value paths at once. Therefore, in this work, although used metaphorically, space can be understood as relational, enabling the recombination of digital resources to create value.

A metaphorical usage of the concept of space can also be noted in the work of Winter et al. (2014). The authors claim that in IS "[e]ven in cases where research went beyond the organizational container, such as studies of inter-organizational systems, organizational boundaries were essentially treated as given" (ibid. p. 258). The authors argue how, in contrast to treating work systems as being contained by organizations, the focus should be on how the goals, values, and meaning of work systems are renegotiated and can change over time. "Work systems can derive purpose, meaning, and structure from the multiple contexts in which elements are embedded and they may pass on purpose, meaning, and structure to the sociotechnical systems that emerge around them." (ibid, p. 260).

Beyond the metaphorical usage of space, more recently Haj-Bolouri et al. (2023) conducted an extensive literature review on the concept of space in IS and other related disciplines. They define four spatial themes that can be of value to IS researchers: 1) representing space (substantial) working as a container of objects, people, and events, clearly separating what is within space's boundaries, and what are the outside areas; 2) differentiating space (relational), which is socially constructed, produced and reproduced through social practices and socio-political hierarchies in society; 3) disclosing space, which is multi-dimensional, enabling possibilities for action and interaction by connecting objects, events, places, people, or separating them; and 4) intuitive space, which is fluid, with melted boundaries due to the mixed realities e.g., physical and hybrid.

Works in related disciplines also provide useful insights into the utility of the concept of space. For instance, Pentland et al. (2020) conceptualize how the spaces of possible paths expand, shift, or contract as processes unfold. The authors suggest how the spaces of possible paths can be estimated; if the process has low visibility (cannot be observed), high granularity (variety when zoomed-in), or if there are multiple possibilities for aggregating or changing the paths, then the spaces of possibilities will be larger.

Overall, space can be claimed to be an understudied concept in IS, as noted by Haj-Bolouri et al. (2023). In what follows, I show how the concept of space is theorized in assemblage theory, an approach lifting the concept of space beyond the substantialist-relationist divide and the dualism of Euclidean (geometrical) or networked spaces.

3.3 Assemblage Theory

Assemblage theory (AT) (DeLanda, 2006, 2013, 2016; Deleuze & Guattari, 1987) has been rarely used by scholars in information systems, with few exceptions (Hanseth & Rodon Modol, 2021; Patel et al., 2022; Tarafdar & Kajal Ray, 2021). A distinctive feature of AT is its realist, process-oriented ontology. Traditionally, realism in the social sciences has been based on structure-oriented ontologies (also referred to as object-oriented) demanding form, order, clarity, and simplicity. The realism in AT, instead, takes heterogeneity as a starting point, where the complex, dynamic, and open world is not settled enough to be reducible to independent entities, such as things and categories. Instead, this is a realism that can accommodate processes and structures, arguing how there are different degrees of chaos and order, openness, and closure of systems (Rutzou, 2017). Overall, AT's "ontology is a complex interplay between heterogeneity and homogeneity, dynamism, and recurrence, but heterogeneity and dynamism always seem to have the upper hand. "(Rutzou and Elder-Vass 2019, p. 406). As argued in Paper #2 in this thesis, this realism can accommodate structures, processes, relations, instead of treating structures and processes as dualities (Farjoun, 2010)

3.3.1 Assemblages and Processes

Heterogeneity is at the core of AT's central concept of assemblages. Its original meaning in French stands for "*agencement*", referring to the process of fitting together a set of heterogeneous components; therefore, putting the focus on the process of assembling, and not the final outcome.

At the core of an assemblage's heterogeneity are the *relations* between its parts characterized as *relations of exteriority*, where "a component part of an assemblage may be detached from it and plugged into a different assemblage in which its interactions are different" (DeLanda, 2006, p. 11). DeLanda (2006) opposes relations of exteriority to totalities where "the components parts are constituted by the very relations they have to other parts of the whole. A part detached from such a whole ceases to be what it is since being this particular part is one of its constitutive properties" (p. 10). The latter are referred to as relations of *interiority*.

Although DeLanda (2006) makes it unclear whether an assemblage can be characterized by relations of exteriority *only*, others have argued how an assemblage can also incorporate relations of interiority but gives primacy to relations that do not constitute the parts (Rutzou & Elder-Vass, 2019). By focusing on relations of exteriority, DeLanda (2006) aims to indicate that the properties of the component parts cannot explain the relations that constitute a whole, as the relations do not have the properties as their *causes*, but depend on the exercise of capacities. The properties of the assemblage are emergent, and irreducible to those of the parts, as the assemblage is not an aggregate of the parts' properties, but emerges out of parts exercising their capacities.

Rutzou and Elder-Vass (2019) stated how "[f]or D&G [Deleuze and Guattari], assemblages and other systems should not be thought of as unities but rather as compositions, defined by difference. Assemblages are not structures but rather "living" arrangements, unsettled and mobile by nature, rather than fixed or hierarchical. Instead of having a stable form or an essence that indicates an underlying unity or homogeneity, an assemblage is characterized by an unstable set of interior and exterior relations between parts and wholes. Assemblages are open and heterogeneous systems; they are diffuse networks that connect different components into complex ensembles that resemble "rhizomes" (p. 405, 406)".

For instance, in the cases introduced in this thesis, the forms data innovation and governance take can be conceptualized as assemblages of heterogeneous and interacting components. Such components include data (electronic patient records, patient-generated healthcare data, wellness data, structured data, unstructured data); digital technologies (electronic patient records systems, laboratory systems, mobile apps, sensor devices); public and private actors (Directorate of Health, Directorate of eHealth, national registries, national portals, hospitals, municipalities), laws (Health Record Act, Health Register Act, Personal Data Act). The processes of innovating with and governing data, therefore, get formed by relating these heterogeneous components.

Taking heterogeneity as a starting point does not mean that the relations between components are chaotic; instead, in AT these relations have elements of both, chance and determinism. This brings the need to discuss the assemblages' structure.

3.3.2 Assemblages and Structure

As a realist ontology, AT accounts not solely for how entities change, but also how they keep their identity over time. The structure of assemblages – what gives order and identity to the heterogeneous whole – is a distribution of multiplicities. Multiplicities are a core concept in the ontology of AT, as they replace what is essentialism in other realist ontologies. According to essentialism, fully formed entities behave in a particular way due to their essences, and their structure is a copy of the essence. For instance, in critical realism, mechanisms as independent entities that are distinguishable from one another, causally connect real structures to actual structures.

While essences possess a clear and distinctive nature, "multiplicities are, by design, obscure and distinct: the singularities which define a multiplicity come in sets, and these sets are not given all at once, but are structured in such a way that they progressively specify the nature of a multiplicity as they unfold following recurrent sequences" (DeLanda, 2013, p. 8). As per the original text of Deleuze and Guattari (1987) multiplicities resemble rhizomes, and not trees; they do not have points of departure like roots, nor do they have an end, instead look more like a map that is open and connectable on all of its dimensions.

In AT, multiplicities define the *degree* to which assemblage can be formed and change, i.e., *they structure the possibility space*. "Multiplicities specify the structure of spaces of possibilities, spaces which, in turn, explain the regularities exhibited by morphogenetic processes" (p. 3). However, the multiplicities are not a blueprint that defines the final product of processes across which assemblages get formed; instead, "[m]ultiplicities give form to processes, not to the final product, so that the end result of processes realizing the same multiplicity may be highly dissimilar from each-other" (DeLanda, 2013, p. 14). Due to this, the form assemblages acquire is not a copy of the structure, but corresponds to it *only to a degree*; the structure gives direction but is not a prediction of the assemblages that will be formed. As Rutzou (2017) notes, AT focuses on structured processes of production, where process and structure are inextricably related. Therefore, in contrast to mechanisms, multiplicities do not only connect structures to structures, but continuously entangle a variety of entities, structures, processes, and forces, which are not produced by causes, but become contingent through historical evolution.

Overall, in AT, the world is not portrayed around persistent forms, definite boundaries, essences, and causal capacities, but the focus is on "how things come to be the way they are", i.e., how assemblages get formed and change (Rutzou and Elder-Vass 2019, p. 402).

For instance, as shown in the empirical cases of this thesis, the spaces of possibilities for innovating with and governing personal health data were structured by meshworks of data, digital technologies, organizations, laws, allowing some forms of innovation and constraining others. Personal health data could be shared for the purposes of treatment and diagnosis, but could not be sold for commercial purposes. This did not imply that these structures, e.g., the laws, worked as predictions for the forms of data innovation and governance would take. Instead, they worked as structures across which various forms of innovation and governance can unfold.

3.3.3 Assemblage Theory and its Flat Ontology

A distinctive feature of assemblage theory is its flat ontology which accommodates both structures and processes, instead of arguing for a hierarchical, stratified conception of structures, as object-oriented ontologies would do. The flat ontology of AT consists of *the actual, the virtual, and the real*. The *actual* encompasses the assemblages that exist here and now, e.g., all the ways in which data *are* innovated with and governed. However, assemblages also possess *virtual* structures across which they can be formed and change, e.g., all the possible forms across data innovation and governance *can* change, out of which some will actualize, and others will not. The actual and the virtual define the assemblage's *reality*.

The virtual is a novel ontological category whose reality is a structure defined by multiplicities. As per the words of Delanda (2014) ”the virtual must be defined as strictly a part of the real object – as though the object had one part of itself in the virtual into which it is plunged as though into an objective dimension” (p. 272). The virtual provides limits for the degrees to which assemblages can be formed and change, and regularities across which processes can acquire certain forms. However, the actualized assemblages are not a simple copy of the virtual, as the virtual does not predict the forms that will take place but provides probabilities.

Concept	**Description**
Realism	Heterogeneity, differentiation, accommodating structures, processes, relations, entities
Assemblage	The fitting of a set of components into a coherent, relative stable whole.
Properties	Emergent, arising from parts exercising their capacities to form wholes
Capacities	Not caused by properties, but dependent on other parts's capacities
Multiplicities	Obscure and distinct singularities whose distribution defines the structure of assemblages
Virtuality	The spaces of possibilities; the degrees to which assemblages can be formed and changed

Table 2: Summary of concepts from assemblage theory

Characterizing the ontology of AT as flat indicates that higher-level entities do not impose a form on lower-level entities, i.e., the real does not contain the virtual, which contains the actual. The structure of assemblages in the virtual does not contain the processes across which assemblages will get formed; instead, the structure and processes unfold progressively and mutually shape each other. For instance, this would indicate how societies, do not contain organizations, which further on contain data or digital technologies. Instead, various forms of producing, sharing, and using data can occur across various digital technologies, organizations, and societal structures which are inextricably meshed together, allowing some forms of innovation or governance, and constraining others.

3.3.4 Assemblages and Progressive Spatialization

The ontology of AT also brings in distinctive assumptions around the concept of space. Assemblages get formed through processes of double articulation, hereby referred to as *double spatialization.* The first spatialization fits heterogeneous components into a *spatial structure* by homogenizing the components into a specific form, and increasing the sharpness of the assemblage's boundaries. DeLanda (2006, 2016) refers to this initial spatialization as *territorialization* – the more territorialized an assemblage is, the higher the degree of its stability. The second spatialization gives the formed assemblages an identity by further differentiating them from other entities through a *functional structure*. DeLanda (2006, 2016) refers to this spatialization as *coding*, where the heterogeneous components get interlocked into a specific pattern of behavior further materializing or expressing the identity of an assemblage. DeLanda at times refers to territorialization and coding as processes (DeLanda, 2006), other times as parameters (DeLanda, 2016); in this thesis, I side with the latter view.

Assemblages have a progressive spatialization. In the actual, assemblages acquire a specific spatial structure that differentiates them from other entities. In the virtual, assemblages can be formed across multiple spatial structures. However, as assemblages are never settled, they keep on forming and transforming, i.e., c*hanging their spatial configurations*. What is significant is not to map out the relatively bounded spatial structures assemblages acquire, but the *thresholds* at which they change from one spatial structure to another. For instance, in Paper #1, data governance transforms from horizontal to vertical by reaching a threshold connected to the purposes for which data are being processed. In this case, data innovation and governance exist in several different forms and switch from one spatial structure to another, where such forms do not exclude each other but coexist. The progressive spatialization of assemblages across the virtual, the actual, and the real are illustrated in Figure 5.

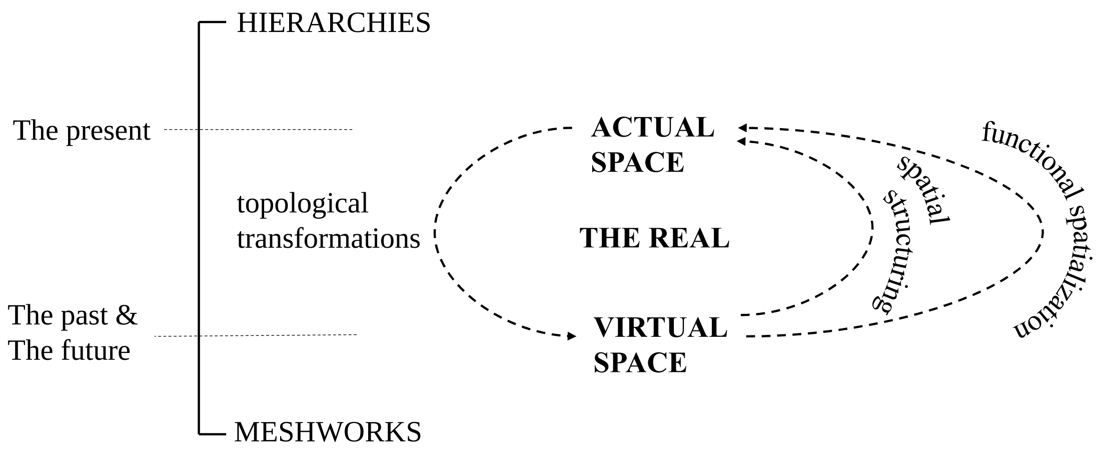


Figure 5: Illustration of progressive spatialization of assemblages across the virtual, the actual, and the real.

3.3.5 Assemblages and Space: A Synopsis

To summarize, in the actual, assemblages acquire a spatial structure that differentiates them from other entities and can be more, or less bounded. In the virtual, space is a structure that allows multiple forms to unfold – space is unbounded over time. DeLanda (2013) regards the virtual space as a manifold that provides degrees of freedom across which assemblages can be formed and change. Although the term manifold commonly designates a geometrical space, in assemblage theory, the term refers to *topological space* where various geometrical shapes can transform into one another as they reach certain thresholds. DeLanda commonly illustrates this through the transition of water into ice or gas. At higher than 90 degrees, water transforms into gas, at below 0 degrees, it transforms into ice; therefore, by reaching certain thresholds, water acquires different forms.

In AT, "space, [is] a notion which must not be purely geometrical but also capable of being linked to questions of process" (DeLanda 2013, p. 3). In the actual, processes can acquire many different (geometrical) forms. What is significant is not to look for a set of properties that are common to all forms, but instead to see how these forms transform into one another by reaching certain thresholds. Therefore, in AT, Euclidean spaces (metric) can be the product of progressive spatialization of networked spaces (non-metric). The Euclidean spaces, as outcomes observed in the actual, would not be embedded into networked spaces as higher-level n+1 dimensions; instead, in the virtual, space is a surface across which an object can have various possible states. Rutzou and Elder-Vass (2019) write how, in AT a space can be regarded as “a field of possible assemblages with related structures (…) each point in the possibility space is a different possible structure for an assemblage.” (p. 412).

Building on assemblage theory, in this thesis I define *data spaces as structures across which processes of data innovation and governance can form and transform as they change their spatial configurations by reaching certain thresholds.*

4 Research Approach

To explore data spaces empirically, I adopted a qualitative research methodology (Sarker et al., 2018b, 2018a). Concerning ontology, I underline my study as a *realist* ontology which acknowledges the existence of both, material and mental entities, independently of my constructions of them. The ontology in this thesis is *process-oriented*; it is concerned with a reality in which entities are always in relations, instead of self-sustaining. Epistemologically, I adopt an "interpretation-centric approach" (Sarker et al. 2018b, p. 761); the empirical material is a flexible text whose meaning is influenced by my interpretations as a researcher; it is not an objective text that represents reality. Therefore, these cases were not "given", but constructed through a careful dialogue between the informants, my role as a researcher, and the guidance of my main supervisor. The role of theory in conducting this empirical research was to, both, enlighten the gathering and analysis of the empirical material, and to guide the iteration between collection and analysis. This resulted in theoretical claims about data's ontological status and the processes of innovating with and governing data in multi-actor environments.

This chapter proceeds as follows. First, I describe the case study. Then, I provide an overview of the research design and the case study as a method of choice. Then, I explain the methodology around the two-embedded case studies. Further on, I elaborate on the role of theory and the abductive process of engaging with the collection and reasoning of the empirical material. Lastly, I provide a detailed description of the collection of empirical material and its meta-analysis.

4.1 Description of The Case Study

Norway offers free public healthcare, organized around national, regional, and local (municipal) healthcare service delivery involving a variety of public bodies and private actors. The national strategies for health and e-health are defined by the Directorate of Health and Directorate of eHealth, respectively; the regional agendas for specialist healthcare services are governed by five Regional Health Trusts which own the regional hospitals; the municipalities deliver primary healthcare services.

Utilizing the potential of digital technologies and health data has been at the focus of various national and political agendas. Back in 2012, the Norwegian government released a white paper “One citizen – one record” aimed at making necessary, relevant, and correct information available to healthcare personnel quickly and efficiently when needed, regardless of where the patient receives, or has received healthcare before. Sharing, and providing access to health data for healthcare personnel is expected to provide improved, more effective health and care diagnosis, treatment, and follow-up of diseases and yield societal value related to population-level analysis used for predicting, or preventing diseases. However, realizing data’s innovation potential remains challenged by technical, organizational, and legal means enabling and constraining data’s value potential.

As of recently, a new type of data, generated by patients, using remote care devices, such as wearables and smartphone apps started being adopted in the healthcare delivery, hereafter referred to as patient-generated health data (PGHD). Moving services outside hospitals, and in patients’ homes has been advocated by the Directorate of Health as a strategic aim; these services (uniformly referred to as digital home follow-up in Norwegian policy documents), are based on PGHD. This includes patient-reported outcome measures, as structured data forms that measure how patients experience conditions related to health and illness and treatment effects, and patient-reported experience measures. PGHD includes novel data types (e.g., lifestyle-related, behavioral, and activity data), more continuous and complete measurements than the episodic data healthcare providers have access to currently, and can facilitate personalized and proactive services, including early-stage warning, prevention, and detection of diseases. However, PGHD currently resides with the vendor, and citizens have little control over them beyond what the front end of the app allows. The various health and fitness apps are typically not interconnected, and data are not being integrated and used to their potential, with some exceptions.

The strategic goals of sharing health data are anchored in Norwegian national strategies for healthcare, aimed at improving the data flow across organizational work processes, reusing health data across service levels, but also providing citizens with access to their own health information so that they can be active participants in decision about their health. However, as of now, the IT landscape of healthcare is siloed, fragmented, and not rigged for external communications. To share health data, the existing public health infrastructures need to be able to receive data (after authentication and access management), process it (assure data provenance), and store data (with technical and semiotic interoperability). Both the vendor side (currently monopolizing data) and the healthcare provider side (not prioritizing integration and utilization), need to change for the innovation potential to become realized.

Beyond the technical heterogeneity, the sharing of patient data is also regulated by a complex legal landscape. Some important regulations include:

- The *Personal Data Act* consists of national rules and the General Data Protection Regulation (GDPR) and deals with the processing – i.e., the collection and use – of personal data.
- The *Health Record Act* deals with the processing of health information when providing healthcare. The purpose of the Act is to make relevant and necessary information available to health personnel, protecting information against unauthorized use and protecting patients' privacy, safety, and their right to participate and be informed. This law encompasses processing data for primary use – providing healthcare treatment – e.g., storing data in electronic patient health records (EPRs) in hospitals, GPs, and municipalities.
- The *Health Register Act* regulates the processing of health information in health registers intended for secondary use, such as statistics, health analyses, research, quality improvement, planning, management, and preparedness. Examples of such registers are the Cancer Register, the Prescriptions Register, and the Norwegian Stroke Register. The information in these registers is often collected in connection with the provision of health care to individual patients and forwarded to the registers. These registers have a different character from electronic patient records and other treatment-oriented registers.

In practice, the implementation of these Acts covers separate, but overlapping areas related to storing, sharing, or using personal health data.

4.2 Description of The Two Embedded Cases

The case study is designed around two embedded cases (Yin, 2017), which are complementary to each other. The study was not initially intended to be an embedded case study. However, the limitations of the empirical material on patient-generated health data's value potential brought in the need to seek additional empirical insights. Choosing two embedded cases was expected to provide more breadth around data innovation and governance in multi-actor environments, as each case emphasized limited insights in specific areas. The embedded case study helped me use the preliminary findings from the first case, as guidance in approaching the second case, and use the insights from both cases to redirect and refine my theoretical conceptualizations. The second case was chosen a year into researching the first one, but ever since, until the rest of the time conducting research for this thesis, the cases were explored simultaneously (Thomas, 2015).

The intention was not to compare these cases but to analyze them in their shared context – health information systems in Norway, and specifically personal health data, including patient-generated health data in the area of remote care monitoring. Each case provided insights into the processes of innovation and governance of personal health data, and contributed to the whole case of data in multi-actor environments; the cases were subunits connected to a larger whole (Thomas, 2015). The analytical conclusions arose independently from the two cases but were integrated to understand the governance and innovation of personal health data in multi-actor environments. Therefore, I conducted a multi-level analysis – from the parts (the embedded cases) to the whole (the case study), and the other way around; each case gained its wholeness from the wider case study.

The first case is a national digital health service, HealthNorway, which allows residents and citizens to store, access, and share some of their data with healthcare services. HealthNorway was launched in 2011 as a single point of access to trustworthy, quality-assured health-related information for citizens. Subsequently, additional citizen-centric services were added, including personalized access to information and interactive services (see Figure 6). HealthNorway is intended to be the major access point for citizen's digital health information and as such, the potential to integrate with HealthNorway is of great interest and relevance for various technology vendors and service providers. The service is used by 80% of citizens and residents by October 2021. It is integrated with 55% of GP EPR systems, all Regional Health Trusts have made available at least one portal functionality to their citizens, and the same holds for one out of every fourth municipality. Providing citizen-centric functionalities through HealthNorway requires coordination across multiple national, regional, and municipal actors on how to govern citizen data on top of these services. This process of governing data across multiple actors is the main empirical focus of this case.

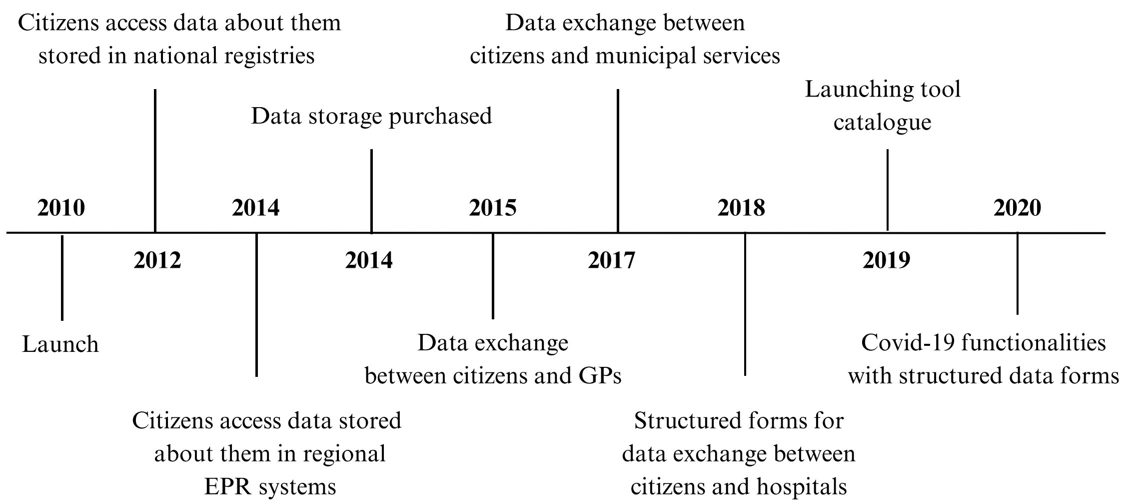


Figure 6: Timeline of citizen-centric functionalities provided on HealthNorway

The second embedded case is a study conducted in the southeast health region in Norway which offers specialist health services to 57% of the total population in Norway. The Regional Health Trust (RHT) is the administrative body overseeing 11 public hospital trusts, 5 private, non-commercial hospital trusts, and its own IT company (HospitalPartner) that works together with IT vendors and hospitals to implement the necessary digital technologies. The RHT's Strategic Development Plan towards 2035 emphasized the importance of moving services outside of the hospital. This comprises temporary home-based cases using connected medical equipment (so-called home hospital services), long-term monitoring with sensor technologies (called digital home follow-up), and more episodic communication services such as video and chat. The strategic emphasis on moving services to the home aligns with national policy as well as general trends.

The region has a complex, fragmented IT portfolio of applications and data silos, and it has struggled to achieve the required responsiveness to ongoing innovations. Going back to 2011, the region had multiple initiatives aimed at developing and implementing a regional IT architecture and standardizing patient data across a variety of IT systems. Such initiatives have only shown limited success. As of the spring of 2017, the regional authority also started working on a regional strategy related to remote care technologies. In the meantime, many of the hospitals already started implementing sensor devices, patient-reporting of data, and digital consultations e.g., to support early discharge of newborns, allow patients with cancer or on long-term antibiotics treatment to stay at home. Several of the hospital trusts in the southeast health region had already initiated various home hospital projects, where some were in the pilot phase and others in routine service. There is, however, no dedicated digital infrastructure in place that could support the deployment of remote care monitoring (RCM) at scale. In the autumn of 2020, the regional authorities started work to consolidate the fragmented portfolio of RCM services. This was connected to a larger initiative that aimed to provide a shared infrastructure that would enable the regional authority to scale up RCM beyond the stand-alone projects, through implementing a new process platform.

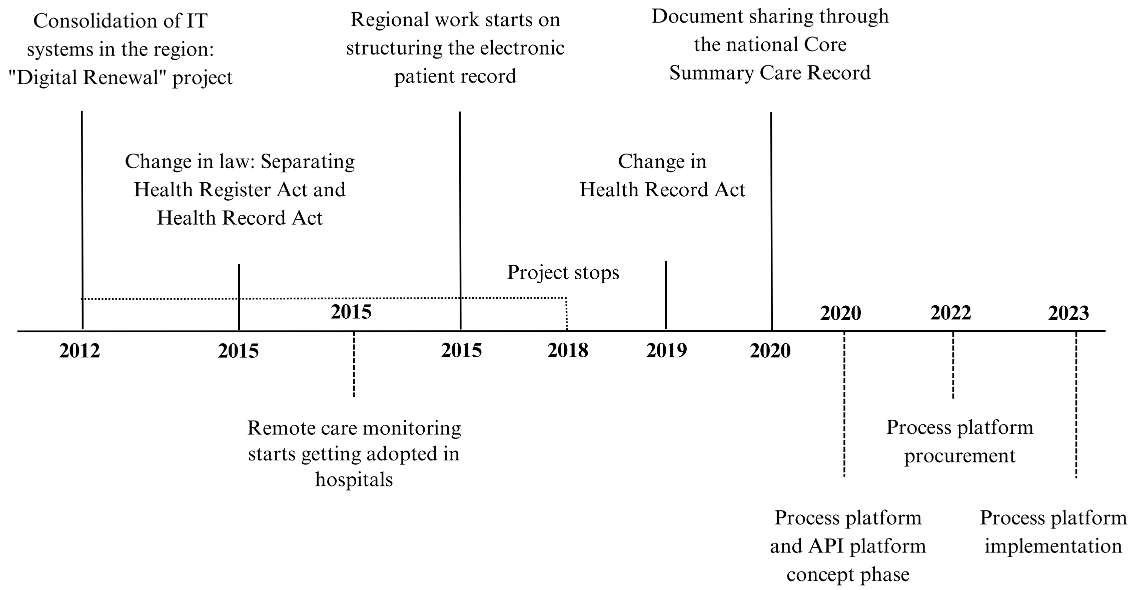


Figure 7: Significant events in the evolution of the regional information infrastructure

The latter initiative aimed to take a different approach and install a process platform above the existing siloed systems. The area of RCM was early seen to fit well as a use case supporting the argumentation for the process platform. It was argued that the event-driven architecture of the platform will help the healthcare providers shift from today's model of follow-up and care which is calendar-governed, to becoming needs- and events-driven. At the time, hospitals were entering individual contracts with RCM vendors impeding larger-scale innovation with work processes around RCM in the region.

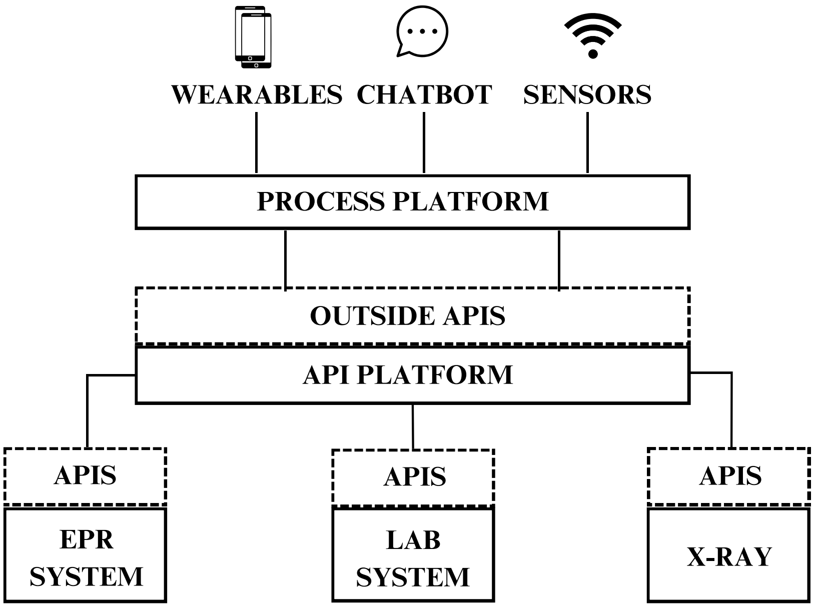


Figure 8: Proposed regional process platform architecture

Realizing the regional vision required standardizing the existing portfolio of IT systems provided by private vendors. The introduction of the process platform would require that RCM vendors develop the necessary integrations to the regional APIs, instead of integrating with EPR systems in the hospitals. EPR vendors would also have to change their existing architectures to be able to receive PGHD. The process platform architecture is presented in Figure 8, as adopted from Paper #4.

This study aligns with the process platform, starting in October 2020 when the infrastructure project was in its initial concept phase.

4.3 The Case Study and its Evolving Research Design

To study data spaces in their naturalistic context, I chose case studies as a method for data collection and analysis (Piekkari & Welch, 2022). As defined by (Yin, 2017), a case study "investigates a contemporary phenomenon (the 'case') in depth within its real-world context, especially when the boundaries between phenomenon and context may not be clearly evident" (p. 15). The case study, as a flexible analytical approach which I amended and modified, was thus suitable for theory development (Ragin, 2014). As noted by Dubois and Gadde (2002), "the main arguments against *it* has been that case studies provide little basis for scientific generalizations" (p. 554), as they are too context-specific and generate findings that are unstable over time. I position the genre of my qualitative study as an "interpretive case study" (Klein & Meyers, 2022; Sarker et al., 2018a; Walsham, 1995, 2006). However, instead of solely describing the case in-depth, I used the case study to provide generalizable theoretical concepts (Yin, 2017). For that purpose, I relied on pre-defined concepts that guided the collection of empirical material and helped me explore a variety of meanings (Dubois & Gadde, 2002). Both the design of the case study, as well as its product – theory development – were evolving over time. I use the word theory development, instead of theory generation (ibid.), to indicate that an existing theory was refined, instead of generated by my research.

The case study was designed progressively (Dubois & Gadde, 2002), as I redefined the research design, the research questions, and the theoretical framings progressively. For instance, exploring the ontological status of data was not the initial intention with this thesis. When engaging with the first case, I expected to explore data's potential for value creation, particularly around patient-generated health data. However, the initial insights from the case uncovered technical and legal challenges in sharing such data across multiple actors; despite recognizing data's vast potential for value creation. In the attempt to study personal health data's value potential, I engaged in a second case, expecting to follow the set-up of inter-organizational, technical, and legal structures for larger-scale production, sharing, and usage of patient-generated health data. At the end of my first year of conducting research for this thesis, the research design was adjusted as an embedded case study.

4.4 The Role of Theory and The Abductive Approach

In this thesis, I adopt an abductive way of engaging with and analyzing my empirical material (Alvesson & Sköldberg, 2010; Dubois & Gadde, 2002). I was iterating between induction and deduction where theory was used "as a lens to interpret or unfold complicated social processes" (Sarker et al. 2018b, p. 759) and supported the iterative process between gathering the empirical material and its analysis. In contrast to the recommendations of Yin (2017) who suggests defining a sharp research design and not changing it, I engaged with my case study iteratively.

Initially, I used multiple theoretical resources as a starting point, such as digital platforms, digital ecosystems, digital infrastructures, and tried not to limit myself to one theoretical explanation, while having some guidance in making sense of the empirical material. However, the preliminary insights of my empirical material brought in the need to look back at the literature and consider alternative theoretical assumptions. The chosen concepts illuminated how an organizational understanding of the phenomenon of data is limiting, as data rarely originate from, or stay within the boundaries of single organizations and their IT systems. The initial findings from my cases showed how data do not solely move in digital platforms, digital infrastructures, or ecosystems, but acquire a life on their own and actively produce and are produced across processes of governance and innovation.

As it became evident how personal health data's value potential cannot be separated from the governance structures necessary to produce, share, and use these data across multiple actors – I started exploring the role of the law in enabling and constraining data's relations. In my empirical context, the law was not simply "out there" as an antecedent or the outcome, but was another actor that could delegate roles and responsibilities. This surprising observation brought the need to engage with alternative theoretical assumptions, as the observations and current theories did not fit. I used this insight to articulate new theoretical concepts which could help me make sense of my empirical material; and started exploring the concept of space.

With these preliminary findings, I went back to the IS research on data and decided to move away from particular concepts, such as digital platforms, digital ecosystems, or information infrastructures, and engage in theorizing data according to my empirical insights. By going back and forth between the empirical phenomenon and theory, I managed to expand my understanding of both, my empirical cases and data innovation and governance in multi-actor environments. The process resembles Dubois and Gadde's (2002) "systematic combining", as I was simultaneously directing and redirecting the study, instead of forcing the empirical material to fit specific categories.

4.5 Collecting The Empirical Material

The gathering of empirical material is divided into three phases. Phase one was exploratory and took place during summer 2020 – spring 2021. This phase was focused on gathering empirical material about the national citizen portal and the private vendors, from which I derived an initial set of theoretical themes and research questions to be further investigated in phase two. Phase two was conducted during spring 2021 – spring 2022 and was influenced by the initial analysis of the empirical material in phase one. This phase included follow-up interviews with participants from the national citizen portal. However, a central focus in this phase was the regional process platform initiative which helped me acquire novel insights. Phase three was concluded during spring 2022 – spring 2023 and included more gathering of empirical material about the regional initiative. This phase ended with an overall meta-analysis of the empirical material.

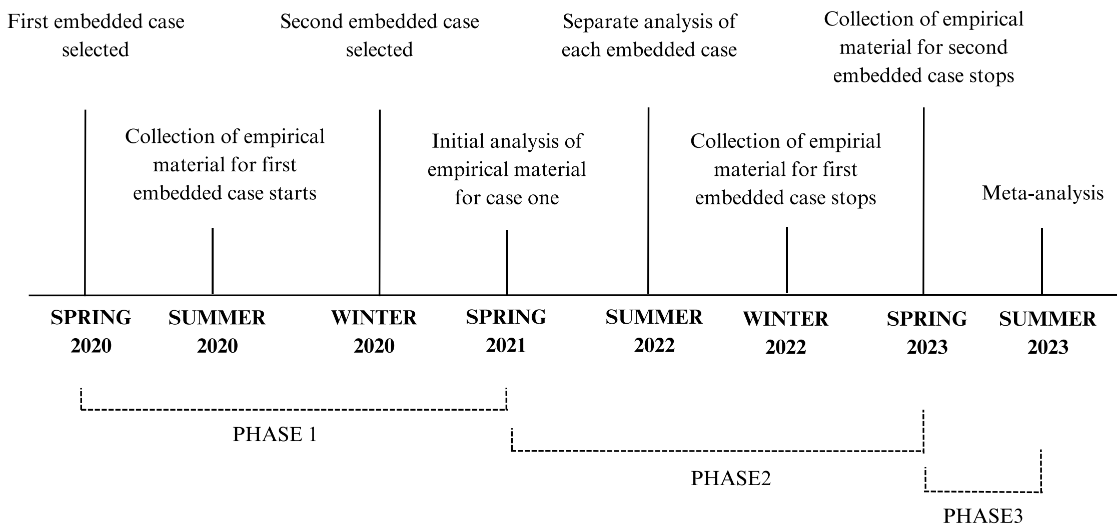


Figure 9: Timeline with phases of collecting empirical material and analysis

In collecting the empirical material, I took the role of an "outside observer" (Walsham, 1995) as I was not part of the organizations I was studying, but an outsider. The initial data gathering was "researcher provoked" (Sarker et al., 2018a) and relying on interviews, but along the way I also saw the need for complementing these interviews with "naturally occurring" materials such as strategy documents, videos, presentations (Sarker et al., 2018a). I did not approach the interviews as statements or representations revealing the truth about the empirical “data”, but as interpretations of the informants, upon which I managed to build my own interpretations by fusing the empirical material with theory. I, therefore, treated the interviews as “social events” (Alvesson 2010, p. 05), which call for the need for theory to be understood. The interview types were both “creative” and “active” (Sarker et al., 2018a), where I was trying to go beyond what the informants were saying, but also construct reality jointly with them. The latter was particularly present in the second embedded case which I followed in real-time; both myself and the informants were uncertain about how the case will unfold in the future.

Interviews were the primary method for collecting the empirical material, which helped me acquire rich accounts of the empirical phenomenon. I conducted semi-structured interviews and adopted a certain degree of liberty in adjusting the interview guides and themes of interest while speaking to informants. This was challenging at times, as I had to navigate between what was significant and what was not in the conversation with the informants. When conducting the interviews, I aimed to acquire both a “subjective understanding” and “negotiated meanings” in the interaction between the subjects and me, as a researcher (Sarker et al., 2018a). The formal instrument for seeking control over the empirical material and my informants as subjects was the interview guide and the set of theoretical assumptions I entered the fieldwork with (Alvesson, 2003). While the initial interviews were looser and sought to grasp an overview of a wider variety of themes, over time, I managed to tighten up the themes discussed. I aimed to do this by analyzing the empirical material as I was gathering it, as well as seeking alternative explanations from additional sources, such as documents, presentations, and online videos (Yin, 2017).

As per the terminology of Alvesson (2003), I took a “reflexive pragmatism” approach to conducting interviews. This included “working with alternative lines of interpretation and vocabularies and reinterpreting the favored line(s) of understanding through the systematic involvement of alternative points of departure” (p. 14). This helped me treat the interviews as not simply “data which reveals reality”, but also use my creative abilities to build upon the richness of the empirical material.

The interview guides were evolving as my understanding of the phenomenon of interest was changing. I prepared separate interview guides for the two embedded cases and adjusted them according to the background of the informants, their job position, their role in the case of interest, and my understanding of the empirical material at that point in time. The interview guides for the first embedded case revolved around the rationale for HealthNorway's functionalities, the ecosystem strategy, the legal landscape in which the service operates, and the challenges in sharing patient data. The interview guides about the second embedded case revolved around the rationale for the process platform initiative, the concept phase and procurement process, and the different views/needs of the various stakeholders included.

Case study focus	**Interviews**	**Description**	**Other empirical material**
First embedded case: National digital service (HealthNorway)	Conducted: 17	One-on-one interviews Approximate duration: 11 interviews of 1 hour, 1 written answer, 1 interview of 1h 50min	Public documents: 48 Internal documents: 5 Short videos online:15 Video presentations: 5 Presentations (docs):8
Second embedded case: Regional process platform initiative (Health South-East)	Conducted: 13	Group interviews with project leaders, hospital managers, and private vendors (10 participants in total) Duration: 1-2 hours	Documents: 21 Presentations: 5 Meeting observations: 2 Videos: 3 Podcasts: 8 Press releases: 7

Table 3: Summary of the collection of empirical material

In terms of size, the interviews varied depending on the embedded case, and the topic discussed. In the first case, I predominantly did one-on-one interviews, as the informants were employed by the same organizations, currently or previously. The case was studied retrospectively, and the informants provided their own opinions, perceptions, and reflections on the matter of interest. The second embedded case, on the other hand, was predominantly conducted as group interviews. This approach was chosen as the case was followed in real-time and multiple stakeholders within one organization were commonly chosen to engage in committee meetings, evaluate the initiative, or were to be affected by its implementation in the future.

The informants were well-educated and experienced professionals, with different backgrounds, and professions, such as software engineers, hospital managers, consultants, lawyers, project leaders, and software architects; their expertise on the subject matter was varied. In the interview settings, I was usually the youngest, least experienced, and at the time, the (only) non-Norwegian speaking person. This, sometimes, made me feel insecure about my local knowledge of the Norwegian healthcare sector and the topics I was inquiring about. I managed to compensate for such shortcomings by having my main supervisor, as a senior researcher, join some of the interviews.

The interviews were conducted in English, and on Zoom (with one exception of a face-to-face interview) and were later transcribed. English was not the first language for either me as a researcher or the informants. For each interview, the participants were asked to accept a consent form with a detailed explanation of how the interview material will be processed and used.

4.6 The Triangulation Approach

As per the suggestion of Yin (2017), I also relied on additional sources of empirical material – triangulation. This helped me gain a more in-depth understanding of my phenomenon and guide myself in redirecting the study and the chosen theoretical concepts (Yin, 2017). Beyond interviews, I included official documentation, strategy documents, presentations, and online videos, in the empirical material. This helped me rely not only on the opinions of informants but also on official communication which could reflect the strategies, desires, and directions at the time, provide a clearer overview of events, and put separate stories together. Including these additional sources helped me gain more details into the empirical case, revealed aspects of it that were previously obscure to me, and provided not only a more comprehensive understanding but also the ground for redirecting my study.

Using triangulation also brought more rigor to the findings generated from the study, and as Yin (2017) suggests, helped me make such findings generalizable across theoretical concepts. I, therefore, relied on both active and passive empirical material (Dubois and Gadde, 2002), as some material was acquired by my purposeful search, such as interviews and conversations with informants, other parts appeared with discovery, by observing meetings, reviewing presentations, listening to online videos, events, and strategic documents. During the collection of empirical material, I continued reorienting my analytical lens, where the theoretical assumptions, empirical fieldwork, and case analysis evolved simultaneously, which shows to be particularly useful for the development of new theories (Dubois & Gadde, 2002).

Documents played a crucial role in the collection and analysis of empirical material in this thesis. I treated documents as material objects created for a reason and connected to a world outside themselves, i.e., I treated documents as relational material objects (Asdal & Reinertsen, 2020). These documents uncovered factual past events – what has happened; factual present events – the state of projects at that specific point in time; and possible events – what was expected to happen. Some documents were attachments to larger cases – such as opinion documents on legislation changes, others were written to contribute to the cases being handled – they were tools with a specific function (e.g., process platform initiative in Health South-East). The documents were varied and included, e.g., discussions around the introduction of new legislation, concept documents regarding upcoming initiatives, national or regional strategy documents, and reference architectures.

The documents collected and used for the analysis were created by official institutions, such as the Directorate of e-health, the Ministry of Health and care, Regional Health authorities, Norwegian Health Network; therefore, they were created by legitimate authorities. All documents were written in Norwegian, except research papers; and were translated into English for data collection and analysis.

The collection of documents was at times with a specific purpose, e.g., to understand how a functionality works; at other times I was reading documents to understand more about the overall healthcare context around the case, with no pre-defined intention in mind. After reading some documents, I would organize the important text in a separate work document; first around significant events and categories, and then I would order them chronologically to follow the unfolding of a process over time.

Initially, I used these documents to understand the context; as I realized the value, breadth, and depth of matters discussed, I decided to use them as part of the analysis. This helped me fill in gaps in my understanding of the cases, as the interviews were conducted between 2020 and 2023; however, both cases were retrospective and covered an extended period of time going back to 2010. Using documents created at different times, places, for different events, and by different actors helped me understand the discussions taking place at a specific point in time; although I did acknowledge that the documents were interpretations of their creator and not texts that correspond to the realities they were referring to. I also relied on texts produced by other researchers conducting research on these cases, therefore, treated the documents as knowledge tools.

For the first embedded case, I predominantly used documents retrospectively; and created to discuss functionalities added, the partitioning of data controller and data processor responsibilities. For the second embedded case, documents were predominantly internal, shared with me in real-time, as they were created, discussed, and refined; some documents addressed previous regional initiatives, law change discussions, or evaluations of projects’ progress.

4.7 A Process Meta-Analysis of The Empirical Material

The initial empirical insights showed how despite data’s vast value potential, the processes of innovating with and governing personal health data unfolded across certain structures of possible forms. To study both, the processes and structures across which data are innovated with, and governed, I decided to conduct process research and uncover patterns in the empirical material (Langley & Tsoukas, 2022). Although I was unable to see clear-cut phases in the empirical material, the sequences of events in this thesis have a beginning and an end; in the empirical world, such sequence is malleable, and cases are ongoing.

I was gathering the empirical material while simultaneously doing its analysis. Initially, I did not seek to provide any higher-level theoretical understandings; instead, I tried to keep myself as close to the fieldwork, as possible. However, a year into doing fieldwork, I realized that my cases reveal insights that challenge current theoretical understandings. As I was adjusting the data collection process, I also continued shaping my understanding of the empirical material. Therefore, I decided to use my empirical material to uncover novel understanding of data innovation and governance in multi-actor environments utilizing the meta-theoretical concept of space. Throughout the analysis, I aimed to "develop a plausible understanding of a poorly understood phenomenon", as well as move from description to abstraction by generating "new concepts and novel insights" (Sarker et al. 2018b, p. 760).

The analysis of empirical material was initially conducted by categorizing, but after the first year shifted to a process-oriented analysis. Therefore, I sought connections, rather than similarities between the cases, where events were organized temporally and in their sequence of occurrence. This helped me provide a more holistic network of events at the concrete and abstract levels (Maxwell & Miller, 2008). As I was analyzing the material, I was constantly iterating between what the overall project is about, and what the individual papers are about.

My style of theorizing can be characterized as "pattern-finding" (Cornelissen, 2017), as I abstracted from particular to more general patterns. I minimized the longitudinal data into a sequence of events (ibid.) which were temporarily ordered, but there were also structures across which such events connect (ibid.). I conducted a separate sequence analysis for each case (see Figures 5 and 6), and then aggregated the findings on a more general level as a separate finding – simultaneous data innovation and governance processes. I split the empirical material between the two embedded cases. I then used some information from the interviews and official documents to construct timelines of events that I considered to be significant for the storyline I was focusing on. Later, I used the interviews and the rest of the empirical material to fill in the details about these events. After I ended up with a detailed and extensive description, I started cutting out the unnecessary details with the use of theory.

The process orientation of this case study brought in the need to make decisions on how far back, and how far in the future will I follow these cases. If I had used a narrower timeframe, other aspects could have been in focus, but using the time frame that I did, helped me see how the first period affected the second, as well as the factors that hindered the development of events in another direction. This “sampling” challenge (Dubois & Gadde, 2002) was simpler in the first embedded case, as I could follow the digital health service from its launching date. In the second case, such sampling was much more complex as I had to select events amongst multiple intertwining information systems phenomena, such as IT governance, enterprise architecture, but also potential implications to data governance. Sampling was not simply a stage in my study, but a continuous process (ibid.).

The point of departure for analyzing the cases was not based on outcomes but on the current processes observed in the case (Langley & Tsoukas, 2022). Both cases were longitudinal and retrospectively approached to map out the relevant events in history shaping their current state. This helped me see how certain conditions and their underlying processes changed over time. The first case was retrospective; I re-constructed a timeline of events by juxtaposing events, activities, and actions from the interviews with informants, their limited timeframe of being involved with the digital service of interest, as well as the additional empirical material: strategy documents, consent forms, presentations, online videos. The interviews were conducted with people who have experienced these events or have second-hand knowledge of them through their work with the digital service.

The second case was followed in real-time. The interviews were also conducted in real-time, with no insights on the outcome of the process. The temporal juxtaposing was done by setting events that have occurred one after the other in a sequence, to make sense of their connectedness. This does not indicate that the second case was solely illustrative (Thomas, 2015), as the purpose was to follow a large-scale digital initiative for sharing and innovating with personal health data, particularly person-generated healthcare data coming from remote care monitoring. As I tried to make sense of the events in the present, I also acknowledged the influence of past events on current actions and decisions. For that purpose, I then started mapping out the relevant historical timeline of events, which were significant in understanding the present state.

The patterns were emergent, as the informants were aiming towards an overarching goal without knowing the details of how those goals would be achieved; the resulting outcomes were not known to them, or to me (Langley & Tsoukas, 2022). The “a-ha” moment in uncovering the patterns was different in both cases. In the HealthNorway case, the surprise came as I realized that there is no platform ecosystem or a large-scale innovation with patient data, particularly with patient-generated healthcare data, as the risks outweigh the opportunities. In the Health South-East case, the surprise came as I realized that although the main aim was to innovate with patient (generated) healthcare data, data were just treated as a by-product of IT, and the details on how such data will be shared, governed, utilized were quite obscure in the conceptualization and procurement phase of the process platform.

Therefore, my a-priori focus on data innovation, had to be supplemented with a focus on data governance. These insights came from my first embedded case. Data collection continued at this point for both cases but was directed into encompassing data governance, and not just value creation.

Parallel to the collection of empirical material, I was also exploring concepts from assemblage theory which can help me understand data innovation and governance, guided by the insights from the empirical world. My initial intuition was to utilize the concept of “spaces of possibilities”; however, instead of solely focusing on possibilities, I decided to emphasize how data innovation and governance can unfold across certain structures of possible forms. This made me inquire more into the ontological assumptions of assemblage theory, and consequently, the ontological concept of space.

4.8 Concluding Methodological Reflections

Overall, the two embedded cases chosen for this study brought in useful insights for studying data in multi-actor environments. Choosing healthcare as a context of study was both, a strength, and a limitation of this thesis.

First, healthcare is an extreme context in highly regulated, bureaucratic, and institutional settings, and dealing with sensitive personal data about persons’ health. This resulted in cases which were progressing slowly, particularly considering the focus of this thesis – production, sharing and usage of sensitive patient data. This was particularly emphasized in the second case which I followed in real-time. At the time of submitting this thesis, the process platform, which was supposed to be the object of study, was still not implemented. Therefore, following this case for nearly three years resulted in empirical material about the preparatory work in purchasing the platform, rather than its successful implementation. This limited the insights I could provide from the case, and therefore, the added value of the case to my research.

Second, the empirical field can be characterized as a traditional sector that is not particularly data-centric. This commonly came at the expense of the cases providing limited insights relating specifically to health data, particularly to the area of patient-generated health data (the latter was initially supposed to be the central focus of this thesis). Oftentimes, my interviews would revolve around discussing the challenges in sharing patient data, rather than successful stories of innovating with and governing patient data. Particularly in the second case, a more pressing challenge was IT governance, rather than data governance. Therefore, I was oftentimes encountered with the challenge of choosing a framing of the case which fits the overall aim of the thesis, while staying true to the empirical insights.

Lastly, the Norwegian healthcare sector is a complex empirical field, which required a pre-understanding of the societal rules and norms, the overall organization of the sector, and the ongoing projects and initiatives in the area of digitalization. While I made significant progress in understanding the empirical field I am studying, learning about its specifics remains an ongoing process.

5 Meta-Analysis and The Individual Papers

In this chapter, I first describe the individual papers. Then, I provide a meta-analysis by outlining the papers' contributions in answering the two research questions. The papers included are as follows:

1. "Data governance spaces: The case of a national digital service for personal health data". Authors: Dragana Paparova, Margunn Aanestad, Polyxeni Vassilakopoulou, Marianne Klungland Bahus. Publication outlet: *Information & Organization*. Volume: 33. Issue: 1. Paper no: 100451. Year: 2023.
2. "Exploring the ontological status of data: A process-oriented approach". Author: Dragana Paparova. Publication outlet: *Thirty-first European Conference on Information Systems*. Year: 2023.
3. "Data hierarchies: The emergence of an industrial data ecosystem". Authors: Daniel Stedjan Svendsrud, Dragana Paparova. Published in edited version: *Fourty-fourth International Conference on Information Systems (ICIS)*. Year: 2023.
4. "Beyond organizational boundaries: The role of techno-legal configurations". Authors: Dragana Paparova, Margunn Aanestad, Ela Klecun. Published in edited version: *Fourty-fourth International Conference on Information Systems (ICIS)*. Year: 2023.
5. "Opening-up digital platforms to accommodate patient-generate health data". Authors: Dragana Paparova. Publication outlet: *8th International Conference on Infrastructures in Healthcare, InfraHEALTH*. Year: 2021.
6. "Governing innovation in e-health platforms: Key concepts and future directions". Authors: Dragana Paparova, Margunn Aanestad. Publication outlet: Selected Papers from the Information Systems Research Seminar (IRIS). Issue: 11. Paper no. 4. Year: 2020.

The thesis' research questions are as follows:

> *RQ1: how can processes of data innovation and governance in multi-actor environments be theoretically accounted for, utilizing the concept of space?*

> *RQ2: how are processes of data innovation and governance conditioned by data's unique nature?*

The individual papers and their contribution to the research questions are outlined in Table 4.

	Paper title	Summary
#1	"Data governance spaces: The case of a national digital service for personal health data"	In this paper, we conduct an empirical study of the changing data governance dynamics as data-sharing functionalities were added to the national digital health service HealthNorway. We conceptualize data governance around the role of data, the horizontal and vertical actor dynamics, and the forming (or not) of data governance spaces. The paper contributes to RQ1 by showing how processes of governance and innovation acquired forms and transformed across various data spaces involving public and private actors.
#2	"Exploring the ontological status of data: A process-oriented approach"	In this paper, I argue for an ontological understanding of data as dualities of structure and processes building on assemblage theory. I utilize the concepts of assemblages, multiplicities and virtuality, to show how data can be understood as irreversible historical productions which simultaneously endure and change. The paper contributes to RQ2 by arguing for a more structural understanding (spatial dimension) of data entities.
#3	"Data hierarchies: The emergence of an industrial data ecosystems"	In this paper, we conduct an empirical study in the highly specialized heavy-asset oil and gas industry. We show how an ecosystem emerged around the DigitizePlatform through industry-specific data complementarities – conceptualized as data hierarchies. We also show how the ordering of data in hierarchies was conditioned by and changed the industrial actor structures in return. The paper contributes to RQ1 by showing how different forms of data innovation and governance unfolded across the DigitizePlatform. It also contributes to RQ2 by showing how data about oil and gas assets did not simply decouple from the industrial realities. Therefore, the industrial reality conditioned the

		forms of data innovation and governance and was shaped by these processes in return.
#4	"Beyond organizational boundaries: The role of techno-legal configurations"	In this paper, we conduct an empirical study following the techno-legal configurations for sharing electronic patient record data and patient-generated health data across the regional infrastructure in South-East of Norway. We define these configurations as intra- or inter-territorial, not determined by organizational boundaries, but by technology and law. The paper contributes to RQ1 by showing how the forms in which data were shared, only corresponded to, but did not copy the structures across which data could be shared – which carry organizational, technical, legal dimensions.
#5	"Opening up digital platforms to accommodate patient-generated health data"	Supporting paper: Early-stage empirical analysis of HealthNorway. The paper contributes to RQ1 by uncovering the techno-legal complexities of sharing personal health data across multiple actors, despite the aims for recombining and utilizing these data for innovation.
#6	"Governing innovation in e-health platform ecosystems: Key concepts and future directions"	Supporting paper: Early-stage literature review, uncovering tensions in the governance of innovation in e-health platform ecosystems. The tensions presented complement the understanding on structure and change as dualities, instead of dualisms in multi-actor sociotechnical phenomena. The paper contributes to the RQs by raising the importance of data as a resource for innovation in healthcare, while also showcasing the specific complexities arising from healthcare as an empirical context.

Table 4: The individual papers and their contribution to the thesis' research questions

5.1 Paper 1: "Data Governance Spaces: The Case of a National Digital Service for Personal Health Data"

Authors: Dragana Paparova, Margunn Aanestad, Polyxeni Vassilakopoulou, Marianne K. Bahus

Research aim: In this paper, we problematize the data governance literature which commonly utilizes IT governance as a conceptual framework. We argue for reconceptualizing data governance by accounting for the distinctive role of data and the involvement of multiple actors beyond organizational boundaries in defining the governance approaches.

Empirical/conceptual: This paper consists of a longitudinal empirical study (first embedded case), following the governance of personal health data as the Norwegian national digital service HealthNorway was extended with citizen-centric functionalities. Although the collection of empirical material is focused on HealthNorway, decisions around which functionalities to add, how to partition responsibilities around data governance, and who takes responsibility for what required the involvement of various public and private healthcare actors, including the Directorate of Health, the Directorate of eHealth, National registries, Regional healthcare authorities, General Practitioners, municipalities, private vendors.

Analysis: The empirical material is analyzed using a process perspective on data governance as changing over time, as new actors, functionalities, and data were added. The analysis provided three conceptualizations. First, we emphasize the purposes for which data are being processed across multiple actors, by distinguishing between *data handling for uniform purposes*, and *handing data over for different purposes*. Second, we distinguish between two types of dynamics in governing data across multiple actors: 1) *vertical*, where one actor processes data on behalf of another who determines the purposes, rules, roles and responsibilities; and *horizontal* where each actor can process data independently, defining separate purposes, rules, roles and responsibilities for data governance. Third, we define data governance spaces as the authorized relationships among multiple actors that specify the boundaries of decision-making authority, rights, roles, and responsibilities around data processing.

Contribution: The paper conceptualizes data governance in multi-actor environments by showing how: 1) data are different than other types of assets governed, as the purposes for data processing and not the IT systems they are processed; 2) the law was not just an antecedent or an environmental context, but an actor that could delegate roles and responsibilities; and 3) decisions around data were not managerial, but extended beyond organizational boundaries.

Contribution towards thesis' research question: This paper contributes to RQ1 in conceptualizing data governance spaces through an empirical study, and the forms of horizontal and vertical data governance which unfolds as multiple actors engage in authorized relationships when sharing data. The paper also shows how the purposes for which data are processed – which can be uniform and defined by one actor, or different and defined by separate actors – condition the processes of governance (and innovation).

5.2 Paper 2: "Exploring the Ontological Status of Data: A Process-Oriented Approach"

Authors: Dragana Paparova

Research aim: In this paper, I pose the following research question: "how can data, understood as both process and structure, be ontologically accounted for?". The paper aims to move beyond understanding data as simply fluid, to also emphasizing their structural nature.

Empirical/conceptual: This paper is conceptual and aimed at understanding the ontological status of data by building on the realist, process-oriented ontology of assemblage theory.

Analysis: The paper utilizes the concepts of assemblage theory: assemblages, multiplicity, and virtuality to argue how data simultaneously acquire structures and keep on changing. The paper also ground the theoretical assumptions on examples from various IS phenomena in the data literature, such as artificial intelligence and data infrastructures.

Conclusion: The paper offers two contributions. First, it unpacks the process ontology of assemblage theory to account for data as dualities of structure and change. Second, it provides an understanding of data as irreversible historical productions that simultaneously engage in enduring and changing processes.

Contribution towards thesis' research questions: This paper contributes to RQ1 by showing how data innovation and governance can be studied as simultaneous processes as data unfold across processes and structures as dualities. The paper also contributes to RQ2 by discussing data's ontology, i.e., data's unique nature as entities.

5.3 Paper 3: "Data Hierarchies: The emergence of an Industrial Data Ecosystem"

Authors: Daniel Stedjan Svendsrud, Dragana Paparova

Research aim: This paper follows the emergence of an industrial data ecosystem in the Norwegian oil and gas industry as a data platform is introduced. Using the concept of ecosystem complementarities, the paper follows how a data ecosystem unfolds, by defining a data ecosystem as "alignment structures of interconnected, but autonomous actors, interacting around complementary data objects to materialize individual and focal value propositions".

Empirical/conceptual: The empirical study shows how by using the DigitizePlatform, actors in the industrial ecosystem could: 1) build data complementarities by recombining industrial data that were previously siloed across IT systems and organizations; 2) build actor complementarities by building apps as data products which can fulfill general and specific user needs; and 3) restructure industrial relations by creating digital twins of industrial assets, where data about these assets were ordered hierarchically.

Analysis: The analysis shows how the industrial data ecosystems emerged across two simultaneous processes: 1) the changing of actor relations; e.g., industrial actors could perform predictive maintenance based on sharing data about their digital twins, resulting in new ways of collaboration; and 2) the making of data hierarchies as industry-specific data complementarities, as data were ordered to reflect the industrial reality where assets are ordered hierarchically.

Contribution: The paper contributes to the literature on digital (data) ecosystems by showing how an industrial data ecosystem emerges around data hierarchies as industry-specific data complementarities. It also shows how in industrial settings the distinction between different types of ecosystems – physical, digital, data – should be more nuanced, due to the physical reality data refer to.

Contribution towards thesis' research questions: This paper contributes to RQ1 by showing how various data spaces were formed around the DigitizePlatform. Each space had its own set of actors, actor relationships, and data-sharing contracts, and innovated with data in distinct ways, e.g., for improving internal processes or enabling predictive maintenance. The paper also contributes to RQ2 by showing how data did not simply decouple from the real-world events they refer to – sensor data from industrial assets on oil and gas platforms. Rather, data were recombined by corresponding to the industrial reality they refer to, also encompassing the larger structures across which industrial data were shared, e.g., actor, organizational, or technical structures.

5.4 Paper 4: "Beyond Organizational Boundaries: The Role of Techno-Legal Configurations"

Authors: Dragana Paparova, Margunn Aanestad, Ela Klecun

Research aim: This paper aims at moving beyond the common architecture-governance focus in studying the evolution of information infrastructures, towards understanding the techno-legal configurations across which data can be shared across information infrastructures over time.

Empirical/Conceptual: The paper presents a longitudinal study of the evolution of a regional information infrastructure in the southeast of Norway, as the technical infrastructure and the laws were changed to facilitate the sharing of both, electronic patient records data, and patient-generated healthcare data.

Analysis: Using the concept of territorialization from assemblage theory, the case is analyzed around two phases: 1) intra-territorial configurations: sharing electronic patient records data; where legal rules for sharing these data across hospitals were harmonized, and the infrastructure homogenized the technical interconnectedness of systems; and 2) inter-territorial configurations: sharing patient-generated healthcare data; where both the technical and legal components of how patient-generated healthcare data were processed in remote care monitoring remained heterogeneous.

Contribution: The main contribution of the paper lies in understanding how territories for data sharing are not organized around organizational boundaries but around techno-legal configurations. The analysis shows how the law defined a territory that transcended organizational boundaries, and the technical components either included or excluded organizations from sharing data across specific territories.

Contribution towards thesis' research questions: This paper contributes to RQ1 by showing how the structure across which data can be shared (governed and innovated with) is defined by techno-legal configurations, instead of by organizational boundaries. The paper also shows how the actual structures across which data are shared (e.g., hospital to hospital, hospital to mobile health app vendor) only correspond to, but do not exhaust all the possible ways in which data can be shared which are enabled and constrained by legal, technical and organizational structures.

5.5 Paper 5: "Opening up Digital Platforms to Accommodate Patient-Generated Healthcare data"

Authors: Dragana Paparova

Research aim: The paper aims to uncover the barriers to accommodating patient-generated healthcare data as part of digital platforms.

Empirical/conceptual: The paper presents an early-stage empirical analysis of the first embedded case, HealthNorway, focusing on the challenges in sharing patient-generated healthcare data and accommodating them as a routine part of HealthNorway's functionalities aimed at citizens.

Analysis: The paper uncovers three main barriers. First, open up the data core using boundary resources; where boundary resources are not simply technical elements, but also organizational contracts determining access to citizens' personal health data. Second, control patient data across long chains of actors; focused on the legal implications of sharing personal health data across multiple actors and providing clarity of responsibilities once patient-generated data are stored in a public storage. Third, establish uniform rules to co-create data value; such as standards, security measures, and frameworks for verifying which digital health apps are safe for usage by citizens.

Contribution: The contribution of this paper is centered around the data-driven value creation literature, discussing how the involvement of personal health data brings specific challenges for building the necessary technical capabilities to support data innovation. However, the identified barriers tackle issues regarding data governance, thus uncovering the dual role of data as resources for innovation and governance assets.

Contributions toward thesis' research question: This paper sets the stage for understanding how data innovation is not open-ended in multi-actor environments incorporating a variety of public and private actors; instead, requires adequate governance arrangements for data sharing. The paper contributes to RQ1 by showing how innovating with data is challenged by questions related to governance, including the involvement of long chains of actors and the legal landscape across which data sharing needs to be navigated.

5.6 Paper 6: "Governing Innovation in E-Health Platform Ecosystems: Key Concepts and Future Directions"

Authors: Dragana Paparova, Margunn Aanestad

Research aim: This paper aggregates relevant conceptualizations on governing third-party innovation in platform ecosystems and adapts them to the healthcare context.

Empirical/conceptual: In this paper, we perform a structured literature review, identify three central concepts of innovating across platform ecosystems, and link them with respective tensions: 1) boundary resources as governance mechanisms: openness and control; 2) co-creating value across heterogeneous actors:

accommodation and resistance; 3) innovating across modular architectures: stability and flexibility.

Analysis: The three concepts are further transformed into three themes that require further attention: 1) patient data as a resource for innovation, specifically around the potential of recombining data from electronic patient record systems with patient-generated healthcare data from smartphone apps and wearables; 2) the role of institutions, including laws and regulations for sharing patient data; and 3) innovating across platform-oriented information infrastructures, due to the intermingling of various platform cores and peripheries in healthcare.

Contribution: The paper contributes to the literature on platform ecosystems, and platform governance by exploring the specifics of innovating through platforms in the healthcare settings.

Contributions towards thesis' research question: This paper sets the stage for RQ1 by uncovering the complexities of innovating with and governing data in healthcare, understanding these two processes as prone to tensions.

5.7 Meta-Analysis: The (Trans)formations of Data Spaces

The meta-analysis addresses the two research questions. First, it shows how the concept of space can be utilized to study processes of innovation and governance as acquiring spatial structures that are neither solely micro – individual, meso – organizational, or macro – network, ecosystem, sector or, industry. Instead, multiple spatial structures of innovation and governance can unfold across a set of actor relationships, data elements, digital technologies, organizational contracts, and laws. Second, it shows how data do not simply decouple from the realities they refer to; rather, the processes of data innovation and governance are conditioned by these realities, and shaping them in return. In what follows, I show how each of the two research questions is addressed through the empirical studies presented in the individual papers.

RQ1: how can processes of data innovation and governance in multi-actor environments be theoretically accounted for, utilizing the concept of space?

Utilizing the concept of space, as introduced from assemblage theory, processes of innovation and governance can be theorized as unfolding across multiple data spaces, as these processes change their spatial configurations from one form into another. Therefore, data spaces do not contain innovation and governance, rather, innovation and governance unfold across certain structures of possible forms, defined by data spaces.

As shown in the HealthNorway case, HealthNorway took part in various data spaces with public actors who processed citizens' data based on the Health Register Act or the Health Record Act. When processing data on behalf of the national registries, HealthNorway could only process data for the purposes determined by the national registries. When exchanging data with hospitals, each actor could copy data and determine its own purposes for data processing. Therefore, HealthNorway took on different roles as various forms of data governance were unfolding across the data spaces.

The focus was not on mapping out the elements of each space HealthNorway takes part in, or the space's outside boundaries, as data governance arrangements were continuously renegotiated once new actors, digital technologies, and data were added. Instead, in the analysis as presented in Paper #1, the focus was on showing how data governance and innovation can change forms by reaching certain thresholds. For instance, when data were governed vertically, actors were processing data for uniform purposes, and under one actor's authority and rules. When data were governed horizontally, authorities multiplied, and each actor could determine their own purposes and rules for processing data. Therefore, in the HealthNorway case, the threshold of changing from one form of data governance to another – horizontal or vertical – was defined around changing the purposes for data processing across the actors involved. The horizontal and vertical dynamics of data governance spaces are illustrated in Figure 6, as adopted from Paper #1.

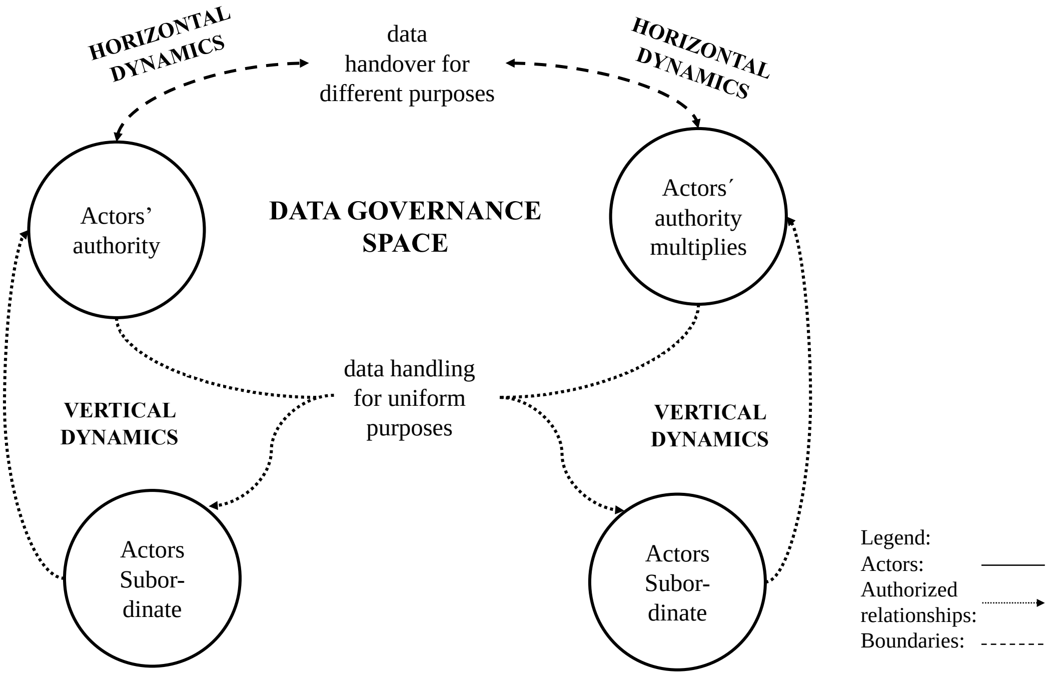


Figure 10: Data governance spaces and their horizontal and vertical dynamics

Beyond the highly regulated context of public sector actors and sensitive personal data, the individual papers also cover empirical contexts involving predominantly private companies and data about physical objects – such as assets on oil and gas platforms, as investigated in Paper #3. Multiple actors collaborated around the DigitizePlatform, giving rise to different data spaces, as data were recombined to innovate with work processes, asset optimization, or predictive maintenance. Different forms of innovation and governance with data could be formed around the DigitizePlatform. In our paper, the analysis was focused on ordering data hierarchically to correspond to the industrial reality of assets ordered on oil and gas platforms. However, the DigitizePlatform could also be used across other sectors or industries, such as power and utilities. There, the ordering of data could change, e.g., by creating relationships among data to correspond to the specific reality of the energy sector where data are organized as networks. Therefore, in this case, the threshold for innovation could be claimed to be set around how actors use data from the industry-specific data complementarities, whose usage transforms how these actors collaborate, cooperate, or compete. As shown in the paper, data hierarchies as industry-specific data complementarities resulted in new forms of data innovation, such as predictive maintenance using smart contracts.

Therefore, in both cases, data innovation and governance were unfolding across multiple data spaces which did not contain one another, but changed the spatial structures once processes reached certain thresholds.

RQ2: how are processes of data innovation and governance conditioned by data's unique nature?

The ontology of assemblage theory also brings in novel insights on how to study the relation between data as technical entities and the realities they refer to.

In the cases presented in this thesis, data do not solely decouple from the people, events, or objects they refer to. Instead, these realities structure the forms across which data innovation and governance can unfold, and are shaped by these processes in return. For instance, in Paper #3, the digital version of industrial data about physical assets was not completely detached from the industrial reality these data refer to – physical assets on oil and gas platforms and existing industrial actor relations. Instead, data were technically coupled to the assets that they referred to using unique identifiers. Moreover, data were ordered hierarchically to correspond to the hierarchical ordering of assets on oil and gas platforms. Lastly, the innovation and governance of data around the DigitizePlatform also corresponded to the existing industrial actor structures, which were cooperating, collaborating, or competing.

However, the digital reality was not a simple copy of the physical reality. Instead, as shown in the analysis, the data hierarchies also led to emergent actor relations and new industrial structures in return – smart contracts for predictive maintenance. Therefore, the realities conditioned the processes of governance and innovation, and the realities were also changed by these processes in return.

Similarly, in the case of personal health data, the actualized data governance forms were not a simple copy of the formulated laws. As shown in Paper #4, according to the Health Record Act, electronic patient data could be shared across any hospital in the region, for the purposes of providing treatment or diagnosis. However, due to the challenges of setting up technological and organizational means, authenticating and authorizing the users, or determining the transfer of responsibilities, data were shared on a lower scale. Each organization had to negotiate its data-sharing arrangements on top of these legal, organizational, technological, and sectorial structures, which enabled some forms of innovation and governance and constrained others.

As one informant stated: "*What I have experienced lately is that you should not completely rely on the definition of the terms in GDPR, with regards to who is the controller, you have to look at the whole chain to be able to see what is there, who is in actual circumstances is the closest to take the control or responsibility in a complex chain (...). The definition in the GDPR is not really fit for this complex chain of information that we operate with now." (Informant, HealthNorway)*

Therefore, personal health data did not simply decouple from the realities they refer to; instead, the person's rights, the data processing laws, the organizational contracts, and technological setups, structured the forms across which data innovation and governance can unfold.

6 Discussion

This thesis shows how the concept of space can be utilized to theorize processes of data innovation and governance as unfolding across certain structures of possible forms in multi-actor environments; and how the realities data refer to, condition the forms innovation and governance can take. In what follows, I position these insights within IS research and point out the added value of the conceptualizations proposed in this thesis.

6.1 Data Spaces and the (Trans)Formations of Data Innovation and Governance

The literature on digital innovation posits how in the open-ended value landscape "digital innovation needs to be viewed not as fixed but as fluid over time, dependent both on connections to assemblages of digital resources and on the relative engagement of individuals, firms and tools" (Henfridsson et al., 2018, p. 90). Similar views have been adopted in the literature on data innovation; authors have argued how data, as semantic entities, can be open-endedly explored as they get recombined into larger objects and commodities (Aaltonen et al., 2021) once actors assign them meaning and purposes (Alaimo, Kallinikos, & Aaltonen, 2020). In the data governance literature, which was initially organizational and framework-oriented, scholars have been arguing how data governance should not be understood as static, but as evolving over time (Benfeldt et al., 2020; Vial, 2023).

The empirical studies presented in this thesis show how, in multi-actor environments, data are seldom produced and used within single organizations. This builds on previous works arguing how data do not necessarily follow organizational boundaries, as they can be copied, reassembled, and reused for diverse purposes (Janssen et al., 2020; Vial, 2019). I introduce the concept of data spaces, building on the concept of space from assemblage theory, to argue how processes of governance and innovation can unfold across multiple data spaces, by changing their spatial configurations.

By focusing on the changing spatial configurations, instead of the forms data innovation and governance acquire, I contribute to existing debates in IS on the necessity to account for nested levels of data governance. Davidson et al. (2023) argued how "IS research to date has primarily focused at the meso (organizational) level; heightened attention to the micro-level (individual) and macro-level (industry, sector), as well as interaction across levels, is called for" (p. 03). Building on the concept of space from assemblage theory (DeLanda, 2006, 2013, 2016), I show how data innovation and governance can be "nested" across various spaces, resulting in changing actor roles, purposes for data processing, technological arrangements, standards, legal basis, which span various organizations' boundaries. With this, I argue how, due to the involvement of data, data spaces should be studied beyond the intra- and inter-organizational divide (Abraham et al., 2019; Jagals & Karger, 2021; Van den Broek & Van Veenstra, 2015). Utilizing the concept of data spaces as introduced in this thesis, IS scholars could focus less on either organizations' boundaries or macro-levels of innovation and governance. Instead, the concept of space can show how multiple forms and transformations of data innovation and governance can unfold as multiple actors produce, share, and use data on various intermediary levels that do not contain one another.

In the empirical studies in this thesis, HealthNorway engages in multiple data governance spaces as it shares data with public and private actors, across which it takes on different roles. What is significant is not mapping out all the spaces in which HealthNorway participates, but showing how data governance transforms across vertical and horizontal dynamics across these spaces, and how this entails a change in authority, rules, roles, responsibilities, and purposes for data processing. While this empirical study focuses predominantly on data governance, various data spaces can be claimed to be formed around the DigitizePlatform, as data about physical assets were shared across multiple actors; some data were shared for predictive maintenance purposes, others utilized for internal work optimization.

In both cases, there were multiple data spaces and various forms of data innovation and governance unfolding across these spaces. The concept of space, as adopted from assemblage theory, is useful to show how these data spaces did not "nest" as in containing one another as higher-level and lower-level entities. Instead, across the multiple data spaces, data governance and innovation could change their spatial structures by reaching certain thresholds. Therefore, what was significant was not mapping out each space HealthNorway or DigitizePlatform participates in, rather, to show how by taking part in different spaces, organizations engage in various data innovation or governance arrangements that allow certain data sharing but constrain another.

As raised by Davidson et al., (2023) data governance "entails new, distributed organizational forms enacted by individuals, technology vendors, data-holding (or using) organizations, and regulatory agencies" (p. 04). The concept of data spaces can move the focus from the forms data spaces acquire, or the types of spaces organizations engage in, towards how processes of data innovation and governance change their spatial configurations, i.e., spatial structures, as they reach critical points, i.e., thresholds while moving across multiple data spaces. In the HealthNorway case, for instance, such thresholds were reached when the purposes for data processing changed, once data were shared from one actor to another.

Focusing on processes of data innovation and governance as changing their spatial configurations across various data spaces is particularly useful when sharing data in multi-actor settings. For instance, data can be stored in one place, accessed, and copied in another. Thinking of data spaces as simply Euclidean is limiting as laws do not contain organizations, which contain digital technologies, which contain data as lower-level entities. As shown in the HealthNorway case, access was not simply exclusive, and authority did not simply shift from one actor to another. Instead, data could be "contained" by various digital technologies, organizations, and laws at once, as their storage, meaning, purposes, and value, were re-negotiated across multiple actors. Therefore, building on the concept of space as introduced in assemblage theory, laws could be studied as structures that are shared by many actors, constraining some data processing, and enabling another.

At the same time, thinking of data spaces as simply networked is limiting, as data are not always shared among multiple actors and open-endedly explored across organizations. Instead, some data, such as sensor data from oil platforms can be prone to organizations' intellectual property rights; other data, such as sensitive and personal health data, can be prone to strict regulations. Therefore, thinking of data spaces as simply networked does not account for the heterogeneous contexts across which data's value potential can be enabled *and* constrained. Utilizing assemblage theory, in this thesis I show how spaces can be understood as bounded in the actual, but unbounded in the virtual, and providing structures across which processes of innovation and governance can acquire certain forms.

As Haj-Bolouri et al. (2023) show, there is a whole spectrum of spatial concepts that can be utilized by IS researchers. The data space concept in this thesis is not necessarily positioned in one of the four categories of space the authors present but can be argued to have elements of e.g., both the representing and disclosing space. In the actual, the formed spaces could be representing, as in separating what is within space's boundaries, and what are the outside elements. In the virtual, space was also disclosing, as it enabled possibilities for action and interaction by connecting objects, events, places, and people. However, in this thesis, I do not place the focus on the types of data spaces the studied organizations have engaged in, rather, on the forms and transformations of data innovation and governance as they move across various spaces.

The concept of space introduced here allows for embracing how data can simultaneously exist "here" and "there", stored at multiple locations, copied across digital technologies, used for different purposes, prone to distinct governance rules – as data innovation and governance change their spatial configurations. Therefore, space provides possible forms across which processes of innovation and governance can change, instead of containing them. This view on space moves the discourse beyond designing data spaces (Geiss et al., 2023) as finished products, towards acknowledging how the nature of data, and the ongoing processes of innovation and governance form and transform (across) space(s) over time.

6.2 Data and Realities

This thesis also brings in certain ontological assumptions for data innovation and governance that arise from the unique nature of data. IS scholars have claimed that due to their properties of being editable, portable, and recontextualizable, data decouple from the realities – events, objects, persons – they refer to (Alaimo, Kallinikos, & Aaltonen, 2020). The claims for the decoupling of data from the realities they refer to come from empirical contexts e.g., building advertising audiences (Aaltonen et al., 2021; Alaimo, 2021). Alaimo, Kallinikos, & Aaltonen (2020) argue that this ontological instability of data (what they are, how they are produced) is "particularly true under conditions that entail massive data aggregation and content syndication. (…) Uncertainty is thus aggravated by the attributes of digital data and the massive ways by which they are de-contextualized, piled up and layered upon one another, without immediate concerns for loss of references" (Alaimo et al. 2020, p. 166).

Across empirical settings, there is a large variety of realities data refer to. Data are produced and used to visualize the behavior of physical objects e.g., sand deposits in oil and gas production (Østerlie & Monteiro, 2020); or monitor people's health (Grisot et al., 2019). For instance, in their study on sand monitoring across the North Sea using digital sensor technologies, Østerlie & Monteiro (2020) show how although "[d]igital representations qua symbols press the decoupling from the physical domain to the limit" (p. 03), data continue to resemble the physical realities they refer to, as "digital representations, come with procedures to closely link the digital representation with its corresponding physical referent" (p. 12).

The empirical studies investigated in this thesis show how data neither completely resemble, nor decouple from the realities they refer to. For instance, in the highly-specialized oil and gas sector, which can be regarded as a big data context, data corresponded to the physical realities they represent – data were ordered hierarchically, similarly to the ordering of assets on oil and gas platforms e.g., platform – turbine – turbine parts. In the highly regulated healthcare sector, the production, sharing, and usage of patient data were conditioned, e.g., by law, persons' consent, digital technologies, or organizational means. Therefore, by building on assemblage theory, and the insights from the empirical cases introduced in this thesis, I hereby extend the understanding of data's realities, to not solely refer to the assets, people, or events they refer to, but encompass the larger set of structures across which data come to existence.

By zooming out of the realities data refer to beyond the people, objects, and events data represent, the concept of space as introduced in this thesis, can also account for the meshed organizational, technical, organizational structures these realities are composed of. With this, this thesis builds on previous insights, arguing how the structuring of data is not solely a technical matter. "The problem of structuring data cannot be treated as a technical issue alone; instead, the data need to be understood as a human creation that is entangled with social practices and the institutional setting in which the data are used. This means that speaking of raw data as a sort of de facto natural resource is misleading as it tends to obscure organizational processes, innovations, and work involved in making data effective inside and between organizations." (Aaltonen & Penttinen, 2021, p. 5924)

In this thesis, I argue how these meshed structures of the virtual, condition how data can be innovated with, and governed in the actual. As shown through empirical work, in the highly regulated healthcare sector, data gathered in the electronic patient record systems could not be open-endedly explored e.g., advertising, or commercial usage, but could be used only for specific purposes related to treatment and diagnosis by healthcare purposes.

This thesis also has relevance for Baskerville et al.'s (2020) ontological reversal which argues for digital-first realities. The primacy of the digital was also argued by Alaimo, Kallinikos, & Aaltonen (2020) in the context of data: "[A]dvertising audiences could be described as data commodities that are first and foremost projections from the data to reality, not the other way around." (p. 168). The empirical cases presented in this thesis show how the digital (data) realities were *not necessarily first*; instead, the digital and physical realities *mutually shaped* each other. For instance, the main purpose behind gathering personal health data about patients in hospitals was not to create a digital reality of patients' diagnosis or treatment, but to provide efficient and effective healthcare service delivery "on the ground". Similarly, the sharing of industrial data was used as means that served the physical "end" where industrial actors extract and produce oil and gas; therefore, aiding the specialized industrial needs. However, the latter case also shows how the physical realities did not simply condition the ordering of digital data, but the digital reality also changed the physical reality in return. As actors created new forms of collaboration, cooperation, and competition based on data hierarchies (e.g., predictive maintenance and smart contracts), their industrial relations changed.

Understanding data's realities beyond the objects, events, and people they refer to, and as realities which data shape in return, could also help IS scholars study the context data are produced, shared, and used in, not as external to data entities, but as internal to processes of governance and innovation. Similar points have been raised by IS scholars before. For instance, Sahay (1997) argued how the context of socio-technical phenomena brings in a spatial dimension to how we study them; Mousavi Baygi et al. (2021) also argued how context should not be considered an external container *in* which things are situated. The view on data's context – as internal, where data are not being contained, but deriving purpose, meaning, relations, and structure from the multiple contexts in which they are embedded as elements, and to which they can pass on purpose, meaning and structure – also sides with Winter et al.'s (2014) view on neo socio-technical systems. The concept of space as introduced in this thesis, where the real is defined by the actual and the virtual which mutually-shape each other, can thus contribute to these debates.

6.3 Practical Implications

This thesis also offers contributions to the practical debates on building European health data spaces, by showing how these spaces should not be understood as pre-defined arenas for data innovation and governance. Instead, organizations can engage in various data spaces simultaneously. What is significant is not simply mapping out the space's outside boundaries, but setting arrangements where organizations can interact across shared digital infrastructures, legislative rules, and beyond countries' physical borders, by taking on different roles.

Moreover, the thesis also provides valuable insights for understanding the role of the law in the organizational decision-making processes. First, as per the insights of the empirical cases presented in this thesis, the law was not simply an antecedent or an external factor, but at times, also another actor that could delegate roles and responsibilities regarding data processing. Second, the thesis also shows how, in multi-actor settings, the laws are not simply "copied" in the actor relationships, e.g., the data controller and processor roles are not always as evident. Instead, each actor needs to assess its relationships with other actors independently and determine who is to take what responsibility in the specific circumstances.

Lastly, the thesis also shows how organizations should not aim to govern data by following static frameworks, as data can be simultaneously produced, copied, shared, and used beyond organizational boundaries. Therefore, I argue for a dynamic understanding of both, innovating with and governing data, which do not fit pre-determined frameworks, but should be considered as continuously evolving and changing.

7 Concluding Remarks

Overall, this thesis argues for understanding processes of data innovation and governance as changing their spatial configurations as they unfold across various data spaces. The concept of data spaces introduced in this thesis, as accommodating both processes and structures, can be useful for studying data innovation and governance as simultaneous processes in multi-actor environments and account for the nature of data. The process dimension can show how heterogeneous data can be recombined to create organizational value, not as fixed and finished forms, but gradually involving more entities as they gain functionality, purposes, and value. The structural dimension is useful to show how the realities data refer to also have meshed legal, organizational, and technical structures that impose conditions on the forms of data innovation and governance that can be actualized. Therefore, the structural dimension defines degrees of freedom along which processes of data innovation and governance can change.

7.1 Limitations and Directions for Further Research

This thesis comes with certain limitations. Although the main aim is to theorize data spaces, the main empirical context, as well as the cases chosen can be argued to not be “data-centric”. Healthcare is an extreme context with strong institutional and legal influences, where multiple public and private actors are involved, providing, utilizing, modifying a variety of large-scale systems and simper IT applications. Data are not necessarily “first” but serve as means to an end – providing healthcare services. I aimed to compensate for these limitations by backing up my theoretical arguments about data through a co-author’s case in the oil and gas industry which can be regarded as a big data context. I acknowledge that the healthcare context incorporates dynamics that are not directly applicable to private sector companies and might have a weaker focus on data-driven operations than e.g., social media platforms, online communities, telecommunication companies. However, I believe that a comprehensive theoretical understanding of data should accommodate the heterogeneous realities data refer to, including personal health data and the highly regulated and institutional healthcare environment.

Although personal health data can be claimed to be an extreme case, the theoretical ideas presented in this thesis can be generalizable across empirical settings. For instance, a common argument against understanding data as being conditioned by the realities they refer to could be how e.g., personal health data can be anonymized – therefore, decoupled from the person they refer to, and used for population-level analysis. By focusing on data innovation and governance as changing spatial configurations across data spaces, these issues can be addressed. Some spatial configurations can be claimed to be more open-ended, others resulting in more bounded forms of innovation and governance. Therefore, the decoupling of data from the realities they refer to is always *a matter of degree*; however, is never completely the case.

Further research could explore the changing structures of data spaces across various empirical settings. Researchers could particularly focus on how the realities data refer to condition the processes of innovation and governance, and how these realities are shaped by processes of innovation and governance in return. To my knowledge, such empirical studies have not been done in IS to date. I hope the theoretical ideas presented in this thesis set the stage for such further explorations.

References

Aaltonen, A., Alaimo, C., & Kallinikos, J. (2021). The Making of Data Commodities: Data Analytics as an Embedded Process. *Journal of Management Information Systems*, *38*(2), 401–429.

Aaltonen, A., & Penttinen, E. (2021). What Makes Data Possible? A Sociotechnical View on Structured Data Innovations. *Proceedings of the 54th Hawaii International Conference on System Sciences*.

Aaltonen, A., & Tempini, N. (2014). Everything Counts in Large Amounts: A Critical Realist Case Study on Data-Based Production. *Journal of Information Technology*, *29*(1), 97–110.

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.

Alaimo, C. (2021). From People to Objects: The digital transformation of fields. *Organization Studies*, *43*(7), 1091–1114.

Alaimo, C., & Kallinikos, J. (2022). Organizations Decentered: Data Objects, Technology and Knowledge. *Organization Science*, *33*(1), 19–37.

Alaimo, C., Kallinikos, J., & Aaltonen, A. (2020). Data and Value. In *Handbook of digital innovation* (p. 162–178).

Alaimo, C., Kallinikos, J., & Valderrama, E. (2020). Platforms as service ecosystems: Lessons from social media. *Journal of Information Technology*, *35*(1), 25–48.

Alvesson, M. (2003). Beyond Neopositivists, Romantics, and Localists: A Reflexive Approach to Interviews in Organizational Research. *The Academy of Management Review*, *28*(1), 13.

Alvesson, M. (2010). *Interpreting Interviews*. SAGE.

Alvesson, M., & Kärreman, D. (2007). Constructing mystery: Empirical matters in theory development. *Academy of Management Review*, *32*(4), 1265–1281.

Alvesson, M., & Sköldberg, K. (2010). *Reflexive methodology: New vistas for qualitative research* (Vol. 12). Routledge.

Asdal, K., & Reinertsen, H. (2020). *Hvordan gjøre dokumentanalyse. En praksisorientert metode* (Vol. 1). Cappelen Damm AS.

Ballard, K. E. (1960). Leibniz's Theory of Space and Time. *Journal of the History of Ideas*, *21*(1), 49.

Barbour, J. B. (1982). Relational Concepts of Space and Time. *The British Journal for the Philosophy of Science*, *33*(3), 251–274.

Barrett, M., Oborn, E., & Orlikowski, W. (2016). Creating Value in Online Communities: The Sociomaterial Configuring of Strategy, Platform, and Stakeholder Engagement. *Information Systems Research*, *27*(4), 704–723.

Baskerville, R. L., Myers, M. D., & Yoo, Y. (2020). Digital First: The Ontological Reversal and New Challenges for Information Systems Research. *MIS Quarterly*, *44*(2), 509–523.

Befring, A. K. (2021). Norwegian Biobanks: Increased Complexity with GDPR and National Law. In *GDPR and Biobanking* (pp. 323–344). Springer, Cham.

Benfeldt, O. (2017). A Comprehensive Review of Data Governance Literature. *Selected Papers of the IRIS*, 120–133.

Benfeldt, O., Persson, J. S., & Madsen, S. (2020). Data Governance as a Collective Action Problem. *Information Systems Frontiers*, *22*(2), 299–313.

Black, S., Davern, M., Maynard, S. B., & Nasser, H. (2023). Data governance and the secondary use of data: The board influence. *Information and Organization*, *33*(2), 100447.

Cecez-Kecmanovic, D. (2016). From Substantialist to Process Metaphysics – Exploring Shifts in IS Research. In L. Introna, D. Kavanagh, S. Kelly, W. Orlikowski, & S. Scott (Eds.), *Beyond Interpretivism? New Encounters with Technology and Organization* (Vol. 489, pp. 35–57). Springer International Publishing.

Cloutier, C., & Langley, A. (2020). What Makes a Process Theoretical Contribution? *Organization Theory*, *1*(1), 263178772090247.

Constantiou, I. D., & Kallinikos, J. (2015). New Games, New Rules: Big Data and the Changing Context of Strategy. *Journal of Information Technology*, *30*(1), 44–57.

Cooper, R. (2005). Peripheral Vision: Relationality. *Organization Studies*, *26*(11), 1689–1710.

Corley, K. G., & Gioia, D. A. (2011). Building theory about theory building: what constitutes a theoretical contribution?. *Academy of management review*, *36*(1), 12-32.

Cornelissen, J. P. (2017). Preserving Theoretical Divergence in Management Research: Why the Explanatory Potential of Qualitative Research Should Be Harnessed Rather than Suppressed: Preserving Theoretical Divergence in Management Research. *Journal of Management Studies*, *54*(3), 368–383.

Dainton, B. (2016). *Time and space*. Routledge.

Davidson, E., Wessel, L., Winter, J. S., & Winter, S. (2023). Future directions for scholarship on data governance, digital innovation, and grand challenges. *Information and Organization*, *33*(1), 100454.

DeLanda, M. (2000). *A Thousand Years of Nonlinear History* (Revised edition). Zone Books.

DeLanda, M. (2006). *A New Philosophy of Society: Assemblage Theory And Social Complexity* (Annotated edition). Continuum.

DeLanda, M. (2013). *Intensive science and virtual philosophy*. Bloomsbury Publishing.

DeLanda, M. (2016). *Assemblage Theory* (1st edition). Edinburgh University Press.

Deleuze, G. (2014). *Difference and Repetition*. Bloomsbury Publishing.

Deleuze, G., & Guattari, F. (1987). *A Thousand Plateaus: Capitalism and Schizophrenia* (B. Massumi, Trans.; 2nd edition). University of Minnesota Press.

D'Mello, M., & Sahay, S. (2007). "I am kind of a nomad where I have to go places and places"… Understanding mobility, place and identity in global software work from India. *Information and Organization*, *17*(3), 162–192.

Dubois, A., & Gadde, L.-E. (2002). Systematic combining: An abductive approach to case research. *Journal of Business Research*, *55*(7), 553–560.

European Commission. (2020). *European data strategy* [Text]. European Commission. https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en

Fadler, M., Lefebvre, H., & Legner, C. (2021). Data governance: From master data quality to data monetization. *ECIS 2021 Research Papers*, 16.

Fadler, m., & legner, c. (2020). Who owns data in the enterprise? Rethinking data ownership in times of big data and analytics. *ECIS 2020 Proceedings*, 17.

Farjoun, M. (2010). Beyond dualism: Stability and change as a duality. *Academy of Management Journal*, *35*(2), 202–225.

Gegenhuber, T., Mair, J., Lührsen, R., & Thäter, L. (2023). Orchestrating distributed data governance in open social innovation. *Information and Organization*, *33*(1), 100453.

Geiss, A., Moller, F., Schoormann, T., & Otto, B. (2023). DESIGN OPTIONS FOR DATA SPACES. *ECIS 2023*.

Goel, Johnson, Junglas, & Ives. (2011). From Space to Place: Predicting Users' Intentions to Return to Virtual Worlds. *MIS Quarterly*, *35*(3), 749.

Grisot, M., Moltubakk Kempton, A., Hagen, L., & Aanestad, M. (2019). Data-work for personalized care: Examining nurses' practices in remote monitoring of chronic patients. *Health Informatics Journal*, *25*(3), 608–616.

Günther, W. A., Rezazade Mehrizi, M. H., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems*, *26*(3), 191–209.

Haj-Bolouri, A., Conboy, K., & Gregor, S. (2023). An Encompassing Framework for Conceptualizing Space in Information Systems: Philosophical Perspectives, Themes, and Concepts. *Journal of the Association for Information Systems (forthcoming).*

Hanseth, O., & Rodon Modol, J. (2021). The Dynamics of Architecture-Governance Configurations: An Assemblage Theory Approach. *Journal of the Association for Information Systems*, *22*(1), 130–155.

Haugjord, B. R., & Kempton, A. M. (2022). Achieving data innovation for sustainable energy solutions. *ECIS 2022 Proceedings*.

Henfridsson, O., Nandhakumar, J., Scarbrough, H., & Panourgias, N. (2018). Recombination in the open-ended value landscape of digital innovation. *Information and Organization*, *28*(2), 89–100.

Hutterer, A., Krumay, B., & Mühlburger, M. (2023). What Constitutes a Data space? Conceptual Clarity beyond Technical Aspects. *AMCIS 2023*. Twenty-ninth Americas Conference on Information Systems, Panama, 2023.

Jagals, M., & Karger, E. (2021). Inter-organizational data governance: A literature review. 20. *ECIS 2021 Proceedings*.

Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy Artificial Intelligence. *Government Information Quarterly*, *37*(3), 101493.

Jarvenpaa, S. L., & Essén, A. (2023). Data sustainability: Data governance in data infrastructures across technological and human generations. *Information and Organization*, *33*(1), 100449.

Jones, M. (2019). What we talk about when we talk about (big) data. *The Journal of Strategic Information Systems*, *28*(1), 3–16.

Kallinikos, J., & Constantiou, I. D. (2015). Big Data Revisited: A Rejoinder. *Journal of Information Technology*, *30*(1), 70–74.

Kazemargi, N., Spagnoletti, P., Constantinides, P., & Prencipe, A. (2023). Data control coordination in cloud-based ecosystems: The EU GAIA-X ecosystem. In C. Cennamo, G. Dagnino, & F. Zhu (Eds.), *Research Handbook on Digital Strategy* (pp. 289–307). Edward Elgar Publishing.

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

Klein, H. K., & Meyers, M. D. (2022). A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems. *Management Information Systems Quarterly*, 28.

Langley, A., & Tsoukas, H. (2022). *The SAGE Handbook of Process Organization Studies*.

Lee, A. S., & Baskerville, R. L. (2003). Generalizing Generalizability in Information Systems Research. *Information Systems Research*, *14*(3), 24.

Little, D. (2016). *New directions in the philosophy of social science*. Rowman & Littlefield.

Lycett, M. (2013). 'Datafication': Making sense of (big) data in a complex world. *European Journal of Information Systems*, *22*(4), 381–386.

Massey, D. (2005). *For Space*. Sage. http://digital.casalini.it/9781446259474

Maxwell, J. A., & Miller, B. A. (2008). Categorizing and connecting strategies in qualitative data analysis. In *Handbook of emergent methods* (pp. 461–477). The Guilford Press.

Mikalsen, M., & Monteiro, E. (2021). Acting with Inherently Uncertain Data: Practices of Data-Centric Knowing. *Journal of the Association for Information Systems*, 21.

Monteiro, E., & Parmiggiani, E. (2019). Synthetic Knowing: The Politics of the Internet of Things. *MIS Quarterly*, *43*(1), 167–184.

Mousavi Baygi, R., Introna, L. D., & Hultin, L. (2021). Everything Flows: Studying Continuous Socio-Technological Transformation in a Fluid and Dynamic Digital World. *MIS Quarterly*, *45*(1), 423–452.

Northrop, F. S. C. (1946). Leibniz's Theory of Space. *Journal of the History of Ideas*, 26.

O'Leary & Cummings. (2007). The Spatial, Temporal, and Configurational Characteristics of Geographic Dispersion in Teams. *MIS Quarterly*, *31*(3), 433.

Østerlie, T., & Monteiro, E. (2020). Digital sand: The becoming of digital representations. *Information and Organization*, *30*(1), 100275.

Patel, S., Baiyere, A., & Johnsen, C. G. (2022). Beyond practice: Assemblage thinking in sociomateriality research. *ECIS 2022 Proceedings.*

Pentland, B. T., Mahringer, C. A., Dittrich, K., Feldman, M. S., & Wolf, J. R. (2020). Process Multiplicity and Process Dynamics: Weaving the Space of Possible Paths. *Organization Theory*, *1*(3), 263178772096313.

Piekkari, R., & Welch, C. (2022). *The SAGE Handbook of Qualitative Business and Management Research Methods: History and Traditions* (By pages 345-358). SAGE Publications Ltd.

Ragin, C. C. (2014). Turning the Tables: How Case-Oriented Research Challenges Variable-Oriented Research. In M. Tight, *Case Studies* (pp. 303–303). SAGE Publications Ltd.

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

Rutzou, T. (2017). Finding Bhaskar in all the wrong places? Causation, process, and structure in Bhaskar and Deleuze. *Journal for the Theory of Social Behaviour*, *47*(4), 402–417.

Rutzou, T., & Elder-Vass, D. (2019). On Assemblages and Things: Fluidity, Stability, Causation Stories, and Formation Stories. *Sociological Theory*, *37*(4), 401–424.

Sahay, S. (1997). Implementation of Information Technology: A Time-Space Perspective. *Organization Studies*, *18*(2), 229–260.

Sarker, S., Xiao, X., Beaulieu, T., & Lee, A. S. (2018a). Learning from First-Generation Qualitative Approaches in the IS Discipline: An Evolutionary View and Some Implications for Authors and Evaluators (PART 1/2). *Journal of the Association for Information Systems*, *19*, 752–774.

Sarker, S., Xiao, X., Beaulieu, T., & Lee, A. S. (2018b). Learning from First-Generation Qualitative Approaches in the IS Discipline: An Evolutionary View and Some Implications for Authors and Evaluators (PART 2/2). *Journal of the Association for Information Systems*, *19*, 909–923.

Tallon, P. P., Ramirez, R. V., & Short, J. E. (2013). The Information Artifact in IT Governance: Toward a Theory of Information Governance. *Journal of Management Information Systems*, *30*(3), 141–178.

Tarafdar, M., & Kajal Ray, D. (2021). Role of Social Media in Social Protest Cycles: A Sociomaterial Examination. *Information Systems Research*, *32*(3), 1066–1090.

Tempini, N. (2017). Till data do us part: Understanding data-based value creation in data-intensive infrastructures. *Information and Organization*, *27*(4), 191–210.

Thomas, G. (2015). *How to Do Your Case Study*. SAGE Publications.

Tuomi, I. (1999). Data Is More than Knowledge: Implications of the Reversed Knowledge Hierarchy for Knowledge Management and Organizational Memory. *Journal of Management Information Systems*, *Vol. 16*(No.3), 103–117.

Van den Broek, T., & Van Veenstra, A. F. (2015). Modes of Governance in Inter-Organizational Data Collaborations. *ECIS 2015 Proceedings*, 13.

Vassilakopoulou, P., Skorve, E., & Aanestad, M. (2019). Enabling openness of valuable information resources: Curbing data subtractability and exclusion. *Information Systems Journal*, *29*(4), 768–786.
Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems*, *28*(2), 118–144.
Vial, G. (2023). Data governance and digital innovation: A translational account of practitioner issues for IS research. *Information and Organization*, *33*(1), 100450.
Walsham, G. (1995). Interpretive case studies in IS research: Nature and method. *European Journal of Information Systems*, 8.
Walsham, G. (2006). Doing interpretive research. *European Journal of Information Systems*, *15*(3), 320–330.
Winter, J. S., & Davidson, E. (2019). Big data governance of personal health information and challenges to contextual integrity. *The Information Society*, *35*(1), 36–51.
Winter, J. S., & Davidson, E. J. (2020). *Harmonizing Regulatory Spheres to Overcome Challenges for Governance of Patient-Generated Health Data in the Age of Artificial Intelligence and Big Data*. TPRC48: The 48th Research Conference on Communication, Information and Internet Policy.
Winter, S., Berente, N., Howison, J., & Butler, B. (2014). Beyond the organizational 'container': Conceptualizing 21st century sociotechnical work. *Information and Organization*, *24*(4), 250–269.
Woerner, S. L., & Wixom, B. H. (2015). Big Data: Extending the Business Strategy Toolbox. *Journal of Information Technology*, *30*(1), 60–62.
Yin, R. K. (2017). *Case Study Research and Applications: Design and Methods*. SAGE Publications.
Zhang, Q., Sun, X., & Zhang, M. (2022). Data Matters: A Strategic Action Framework for Data Governance. *Information & Management*, *59*(4),

APPENDIX: PART II, THE PAPERS

APPENDIX I:

"Data Governance Spaces: The Case of a National Digital Service for Personal Health Data"

Dragana Paparova, Margunn Aanestad,
Polyxeni Vassilakopoulou, Marianne Klungland Bahus

Abstract

This paper investigates data governance empirically by conducting a retrospective study of the ten-year evolution of a national digital service for personal health data in Norway. We show how data governance unfolds over time as data become shared and itinerant across multiple actors. Building on our findings, we introduce the concept of data governance spaces to refer to the authorized relationships among multiple actors, which specify the boundaries of decision-making authority, rights, roles, and responsibilities around data processing. We contribute to the literature on data governance by distinguishing between a) authority multiplication, where data are handed over to other actors to serve diverse purposes triggering horizontal dynamics, and b) actor subordination, where authorities delegate data handling for uniform purposes triggering vertical dynamics. Overall, the paper extends prior research by showing how data governance unfolds beyond intra-, or inter-organizational boundaries and shifts attention to data's pivotal role, and the purposes for which data are collected, shared or used across multiple actors.

Keywords: data governance spaces, horizontal dynamics, vertical dynamics, data handling, data handover

1 Introduction

Data governance has been receiving increasing attention among information systems (IS) scholars (Abraham, Schneider, & vom Brocke, 2019; Parmiggiani & Grisot, 2020; Winter & Davidson, 2020), and also across national and international political agendas (European Commission, 2020). The IS literature has commonly conceptualized data governance by building on information technology (IT) governance (Benfeldt, 2017; Fadler, Lefebvre, & Legner, 2021; Tallon, Ramirez, & Short, 2013), or modes of governance in inter-organizational settings (Jagals & Karger, 2021; Van den Broek & Van Veenstra, 2015). However, the data governance literature has not fully taken into account the role of data as they become shared and itinerant across multiple actors. We argue that data are not "just another" organizational asset. Due to their use-agnostic and semantic nature, data can decouple from the events they refer to (Alaimo, Kallinikos, & Aaltonen, 2020) and transform into larger objects (Aaltonen, Alaimo, & Kallinikos, 2021) as part of actors' meaning-making processes. Data can also increase with use rather than being consumed, get diffused as they travel, and be shared without being depleted (Vassilakopoulou, Skorve, & Aanestad, 2019). This brings implications for data governance, particularly in the case of sensitive and personal data (Winter & Davidson, 2019), where the regulatory conditions differ from governing non-personal data. Therefore, data governance is not necessarily defined by managers within an organization, but frequently involves decisions by actors beyond organizational boundaries, including institutions regulating how such data are collected, shared and used.

We explore multi-actor data governance with personal health data as our empirical focus. Personal health data have been recognized as a key resource for innovation across healthcare services (Bardhan, Chen, & Karahanna, 2020). However, personal health data currently reside in siloed public or private actors' systems, and beyond isolated initiatives, routine sharing has not yet been achieved. Some of the core challenges in sharing personal health data lie in decisions related to data governance, including protecting intellectual property rights and privacy according to the legal provisions (Parmiggiani & Grisot, 2020), and retaining control once data are shared across multiple actors (Van den Broek & Van Veenstra, 2015). In 2018, the European Union introduced the General Data Protection Regulation (GDPR) aiming to bring more clarity to the governance of personal data (European

Commission, 2016). However, the high abstraction level of GDPR, and its interpretation by regulators pose challenges to data governance, as organizations raise questions on how to enact the regulation (Greengard, 2018). Data governance across multiple actors concerning personal data remains a challenge impeding innovation, but has been only scarcely addressed in the IS literature.

This paper aims to advance research on data governance by using data, and not IT, as the focal point and account for the involvement of multiple actors. The research questions we seek to answer are *1) how to conceptualize data governance by accounting for the role of data*, and *2) how does data governance unfold when data become itinerant across multiple actors?* To answer these research questions, we follow the data governance decision-making throughout the ten-year evolution of a citizen-facing digital health service for personal health data in Norway.

This paper offers conceptual and empirical contributions to the literature on data governance. First, this paper introduces the concept of data governance spaces, referring to the authorized relationships among multiple actors which specify the boundaries of decision-making authority, rights, roles, and responsibilities around data processing. Second, this paper shows how data governance can unfold across vertical dynamics, where actors subordinate to an authority, or horizontal dynamics, where authorities multiply and govern data separately. Third, this paper differentiates between handling data for uniform purposes and handing data over for different purposes, therefore pinpointing actors' purposes for collecting, sharing and using data. Fourth, this paper shows how data governance unfolds over time through an empirical study where multiple actors engage in decision-making around personal and sensitive health data.

The remainder of this paper is organized as follows. In the conceptual background, we problematize the existing IS literature which is commonly based on a framework-oriented understanding of data governance. We argue that data's distinctive nature requires further development of the data governance conceptualizations. In section three, we provide a description of the research design, methodology and case study. In section four, we present the findings and show how data governance dynamically unfolds as data become shared and itinerant across multiple actors. In section five, we induct concepts from the analysis of the empirical case. Finally, we discuss the added value of our concepts to the data governance literature and summarize the paper's key takeaways and limitations.

2 Conceptual Background

2.1 The Conceptual Basis of IS Literature on Data Governance

The growing body of IS literature focused on data governance commonly builds on the conceptual basis of IT governance, as Benfeldt (2017) noted. IT governance seeks to ensure that organizations utilize their IT assets to achieve strategic goals. With the increasing attention on data's potential business value, data are increasingly viewed as "assets" that also require similar governance in order to fulfill strategic purposes (Khatri & Brown, 2010; Otto, 2011). Data governance works often reference Weill and Ross' (2004) framework for allocating decision-making rights and accountabilities within several IT-related decision domains. While this framework is oriented towards governing traditional IT assets (hardware and software), Khatri and Brown (2010) proposed an alternative, but similar framework, covering a set of new decision domains relevant to data – data decisions, data quality, metadata, data access and data lifecycle – where the locus of accountability could be more or less centralized. Similarly, Tallon et al. (2013) argued for incorporating information governance as a novel decision area into the standard IT governance framework. This work draws on the governance mechanisms as defined in the IT governance literature, distinguishing between structural (relating to roles and responsibilities), processual (formal processes), and relational mechanisms (communication and coordination among stakeholders). These governance mechanisms are also central both in Abraham et al.'s (2019) conceptual framework, and in Fadler et al.'s (2021) mapping of data governance archetypes in organizations.

2.2 The Work of Establishing Inter-Organisational Data Governance

This reliance on the IT governance literature is problematic in the following ways. First, the IT governance literature is predominantly organization-focused, overseeing the empirical implications of data often flowing across organizational and sectorial boundaries (Janssen, Brous, Estevez, Barbosa, & Janowski, 2020). Governing data shared between organizations, in business ecosystems, or across public-private boundaries comes with distinct challenges from governing data within organizations. For instance, in their conceptual framework, Abraham et al. (2019) distinguish between the intra-organizational and inter-organizational scope of data governance. They discuss the need for companies to install distinct governance mechanisms in inter-organizational settings, such as data integration and usage policies, data exchange standards, processes for interaction and collaboration, service level agreements, and data sharing agreements (ibid., p. 431). In practice, there is a large variety of inter-organizational relations and governance mechanisms, but little empirical research on how this variety maps to inter-organizational data governance (Jagals & Karger, 2021). Some of the few studies on inter-organizational data governance describe the archetypes of data collaborations (Van den Broek& Van Veenstra, 2015), stakeholders' coordination (Markus & Bui, 2012) and collective actions for decision-making (Benfeldt, Persson, & Madsen, 2020; Zhang, Sun, & Zhang, 2022). However, this remains an understudied area. A better understanding of governance in the context of inter-organizational data sharing is fundamental in collaborations encompassing public and private actors that seek to address grand challenges, where data not only generate organizational value, but are also a shared resource that can create societal value.

Several empirical studies describe the work involved in establishing data governance in organizations. Vilminko-Heikkinen, Brous, and Pekkola (2016) described the tensions and conflicts associated with implementing an organization-wide master data management initiative, as the top-down logic (inherent in any governance initiative) collided with various local logics. To investigate such challenges, Benfeldt et al. (2020) applied a collective action lens in their study of data governance in public sector organizations – where the diversity of responsibilities, a fragmented IT infrastructure, different professional domains, and multiple organizational objectives created challenges for successfully mobilizing the actors. The way these studies define their research problem – how various actors with heterogeneous (possibly contrary) interests and capabilities can govern common resources – has its parallel in studies on how shared information infrastructures emerge. For instance, Constantinides and Barrett (2015) investigated the dynamics among multiple actors during the development of a regional health information infrastructure. They propose a polycentric approach to govern infrastructure development, "where multiple governing units at differing scales can exercise considerable independence to make norms and rules within a specific domain" (ibid., p. 41), i.e., a nested and layered structure of governance rather than a monolithic one. Our study is informed by the literature on the evolution of information infrastructures, which investigates the involvement of multiple actors within and across organizations (Aanestad, Jolliffe, Mukherjee, & Sahay, 2014; Bowker, Baker, Millerand, & Ribes, 2009; Grisot, Hanseth, & Thorseng, 2014; Star & Ruhleder, 1994).

2.3 Data as a Starting Point For Conceptualizing Data Governance

A second limitation in prior data governance research is the assumption that data can be considered an "asset" along with other assets, and therefore, governed likewise. In the classic IT governance framework by Weill and Ross (2004) the focus is on traditional IT assets and the data concern is diffused in the areas of IT architecture and IT infrastructure. Although the data governance literature argues that data differ from IT, data are usually considered "assets" and there is less attention to more fundamental questions about the nature of data. While there is recognition that data are context-contingent (Otto, 2011) and malleable (Abbasi, Sarker, & Chiang, 2016), these insights have not yet significantly impacted the conceptual groundings of the data governance literature. However, the nature of data has been more explicitly investigated by other literature streams within IS and such works provide valuable insights for conceptualizing data governance.

IS scholars have argued that data are conceptually different from IT, as they have a semantic (Alaimo et al., 2020) and use-agnostic nature (Alaimo et al., 2020; Constantiou & Kallinikos, 2015) and can be assigned various meanings as part of actors' value-creation. In these meaning-making processes, data transform from tokens into larger objects and commodities (Aaltonen et al., 2021), can be used in unforeseen ways (Lee, Zhu, & Jeffery, 2017; Susha, Janssen, & Verhulst, 2017), and for different purposes (Fadler & Legner, 2020). Data can also expand (increase with use rather than being consumed), diffuse (tend to travel), and be shared without being depleted (Vassilakopoulou et al., 2019). Data may also belong to various categories, where the use value, business criticality, and regulations may differ. For instance, some data may be openly shared (Bonina & Eaton, 2020), some may relate to proprietary knowledge regulated by intellectual property rights, and some may be personal data and thus subject to privacy regulations

(Parmiggiani & Grisot, 2020). In the data governance literature, there has been limited attention on this variability of data aspects.

This paper argues that governing data is distinct from governing IT and this premise should serve as a basis for conceptualizing data governance. As Zhang et al. (2022) noted, beyond decision-making roles and responsibilities, data governance also needs to account for data stewardship (Rosenbaum, 2010), data ownership (Fadler & Legner, 2020; Van Alstyne, Brynjolfsson, & Madnick, 1995), and matters of data quality, privacy, and security (Abraham et al., 2019). These aspects are particularly significant concerning personal data, which have scarcely been at the focus of data governance studies in IS. For example, in their data governance framework, Abraham et al. (2019) distinguish between traditional data (master, reference, transactional) and big data (web, social media, biometric, machine-generated, streaming). However, no category is devoted to personal data.

Personal data are typically regulated by privacy-oriented legislation aiming to protect individuals. When personal data are aggregated and utilized for different purposes, such as by social media platforms, individual level regulations are insufficient (Viljoen, 2021). Winter and Davidson (2019) explored personal data governance by focusing on personal health data. The authors show how policy makers and regulators identified that the scale and scope of data exchanges and the appropriation of data-intensive technologies make existing laws and policies inadequate to fully protect patients' data. Here, regulations were not just triggers (DalleMule & Davenport, 2017; Khatri & Brown, 2010) or antecedents (Abraham et al., 2019) but actively shaped the data governance approaches. In another paper, Winter and Davidson (2020) highlight how person-generated healthcare data as highly unregulated data are governed by an interplay of organizational, technological and regulatory spheres.

Taken together, the more nuanced conceptualization of the nature of data, their mobility beyond organizational boundaries, the diversity of actors and interests, and the complexity of the regulatory landscape suggest that it is pertinent to rethink whether studying data governance around IT decision-making frameworks, or modes of governance, is sufficient.

3 Research Approach

3.1 Case Background

HealthNorway was launched in 2011 as a public healthcare digital service aiming to serve as citizens’ single point of access to trustworthy, quality-assured health-related information. Subsequently, HealthNorway was extended with various interactive citizen-centric services, sharing data with hospitals’ electronic patient records (EPR), General Practitioners’ (GP) systems and municipal systems. Integrating with HealthNorway is very relevant for various public actors and private technology vendors. As of June 2022, 93% of citizens and residents use this service. HealthNorway is integrated with the EPR systems of 55% of GP offices, all Regional Health Trusts have made available at least one portal functionality to their citizens and the same holds for one out of every fourth municipality.

The citizen-centric functionalities provided at HealthNorway rely on the processing of personal health data; thus, they are subject to regulatory and legal frameworks HealthNorway and its collaborators must comply with. This legislation includes The Personal Data Act (regulating the processing of personal data, incorporating GDPR and special national rules for GDPR), Health Record Act (regulating the processing of health data when providing healthcare), Health Register Act (regulating the processing of health information for secondary use, such as health analysis, research, quality improvement, planning, management, and preparedness). These Acts also adopt the terms "data controller" and "data processor", as defined in the GDPR, where the data controller determines the purposes and means for processing personal data, while the data processor processes personal data on behalf of the controller.

As of 2016, HealthNorway is managed by a product board, including members from the health sector, such as the Health Directorate, Directorate of e-Health, The Public Health Institute, regional health authorities, municipalities, general practitioners, and citizens. This product board is part of the national governance structure for IT in healthcare, and the e-health board. The e-health board appoints the leader of the product board for HealthNorway and its mandate. The product strategy and roadmap for HealthNorway are developed and maintained as a collaboration between multiple public healthcare actors who define the scope of the digital service. The extension of the product with new functionalities is commonly organized as projects initiated by different public healthcare actors.

Providing citizen-centric functionalities at HealthNorway requires building trusted relationships between public and private actors for processing personal health data across authoritative sources. However, the National e-Health Strategy developed by the Directorate of e-Health (2021) raised that the unclear division of roles and rules, unstable regulatory frameworks and legal uncertainties for data sharing challenge these public-private collaborations. The coordination across multiple healthcare actors on how to govern personal health data on top of the HealthNorway services is the main empirical focus of this paper.

3.2 Collection of Empirical Material

Our study is based on an interpretive paradigm (Alvesson & Skoldberg, 2010). We conducted qualitative research to map how data were governed across the actors involved as new functionalities were added. The extension of functionalities at HealthNorway is documented in publicly available information such as project and strategy documents, public presentations, online information resources, and data processing agreements. These were the primary information source for reconstructing the timeline of HealthNorway's growth and expansion. This empirical material was complemented with 13 semi-structured interviews with key persons who were (or had been) involved with HealthNorway's decision-making and one person who provided us with written responses. The informants included: product managers, product developers, portfolio managers, functional architects, enterprise architects, lawyers, technical consultants, and senior advisors. None of the informants were part of HealthNorway's team for the whole period of time covered in this paper.

The functionalities provided at HealthNorway required collaboration with various public and private actors, providing a multi-actor perspective on data governance. To acquire a more comprehensive understanding of the actors involved, we also conducted three interviews with private vendors working with person-generated healthcare data. All vendors included in the study had earlier attempts to integrate with HealthNorway, but during this study, such collaborations remained unrealized. However, these vendors are integrated with other parts of the public healthcare service delivery. Including informants directly and indirectly related to HealthNorway's evolution helped us gain a more comprehensive understanding of the challenges of sharing data across public and private healthcare actors. The informants from the private vendors' side include the founder, co-founder, and managing director. The collection of empirical material is summarized in Table 1. The interview guides used in semi-structured interviews were adapted to fit the informants' profile, background, knowledge, and position; they also reflected our knowledge of the case at that time.

Empirical material	**Sources**	**Description**
Semi-structured interviews (HealthNorway)	13	Duration: 1 hour per interview, one interview of 1,5 hours, and one interview of 2 hours. Participants: product managers, product developers, portfolio managers, functional architects, enterprise architects, lawyers, technical consultants, senior advisors.
Semi-structured interviews (private vendors)	3	Duration: 1 hour per interview. Participants: founder, co-founder, and managing director.
Written answers to interview questions	1	Confirm/correct information with HealthNorway informant as an alternative to an interview.
Public documents	48	Product strategy documents, goal architectures, recommendations, guides for functionalities, evaluation of solutions, terms of usage and overview documents for data processing roles, private digital health apps, yearly reports of Norwegian Health Network and Directorate of e-Health.

Internal documents	5	Evaluation documents, report documents, market surveys, presentations.
Public videos	15	Length: 1-2 minutes. Short youtube videos on citizen functionalities aimed at end-users.
Video presentations	5	Length: 25, 30, 40 and 45 mins. Content on structured patient-generated data during Covid19, digital forms, Application Programming Interfaces (API) management and ecosystem vision, aimed at stakeholders.
Presentation slides	8	Video consultation, patient-generated data, digital home follow up services, API management, patient health records, municipal message exchange functionality.

Table 1: Summary of empirical material.

The interviews were executed in two batches. The first set of interviews started in June 2020 and lasted until April 2021. At this stage, we asked for participants' reflections on a broader set of topics, including the development of functionalities and data exchange arrangements. The second set of interviews was conducted from October 2021 to November 2022. In these interviews, we intended to grasp participants' reflections on recent developments towards personal data exchange, such as the Covid-19 pandemic, and get a more detailed understanding of the data governance decisions for the functionalities provided. During this phase, an interviewee read the empirical story and gave us their reflections on the narrative. The interviews were conducted online, with one exception (face-to-face), lasting approximately one hour and were fully transcribed.

3.3 Analysis of Empirical Material

The analysis of empirical material was iterated with its gathering, which helped direct our focus as we collected the empirical material. The analysis was performed in two phases. In the first phase, we conducted an inductive process analysis

(Berends & Deken, 2021). We identified significant data governance events and ordered them sequentially. These events were reconstructed as a timeline of HealthNorway's evolution, focusing on extensions of eight functionalities, as presented in Fig. 1.

We realized that as the functionalities were extended, certain data governance decisions had to be repeated, but had different outcomes over time. Such decisions include the actors involved, the delegation of roles and responsibilities (data processors and data controllers), data storage, and citizens' rights over personal data stored about them. This phase uncovered the need to move beyond a framework understanding of data governance, and to account for data governance as changing over time.

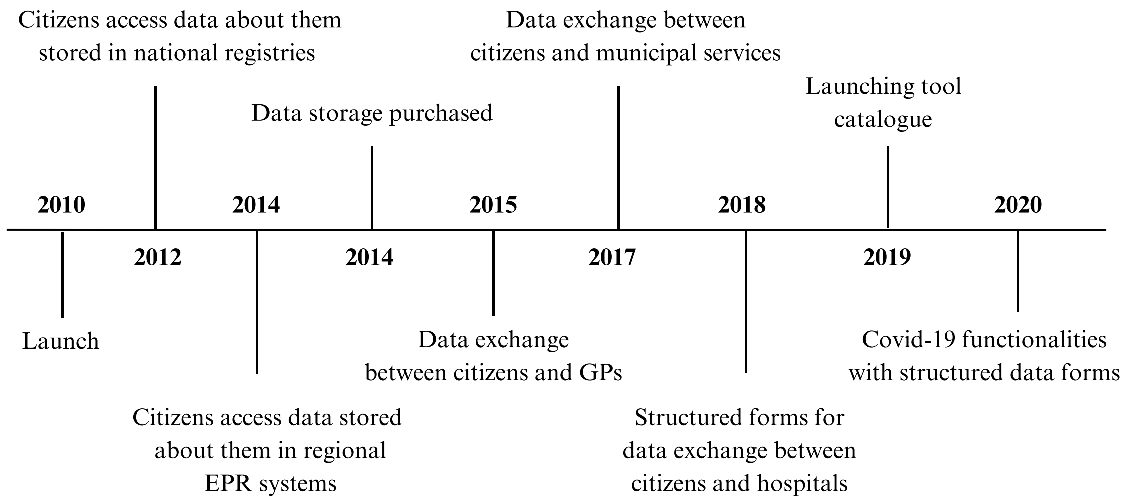


Figure 1: Timeline of citizen-centric functionalities provided at HealthNorway

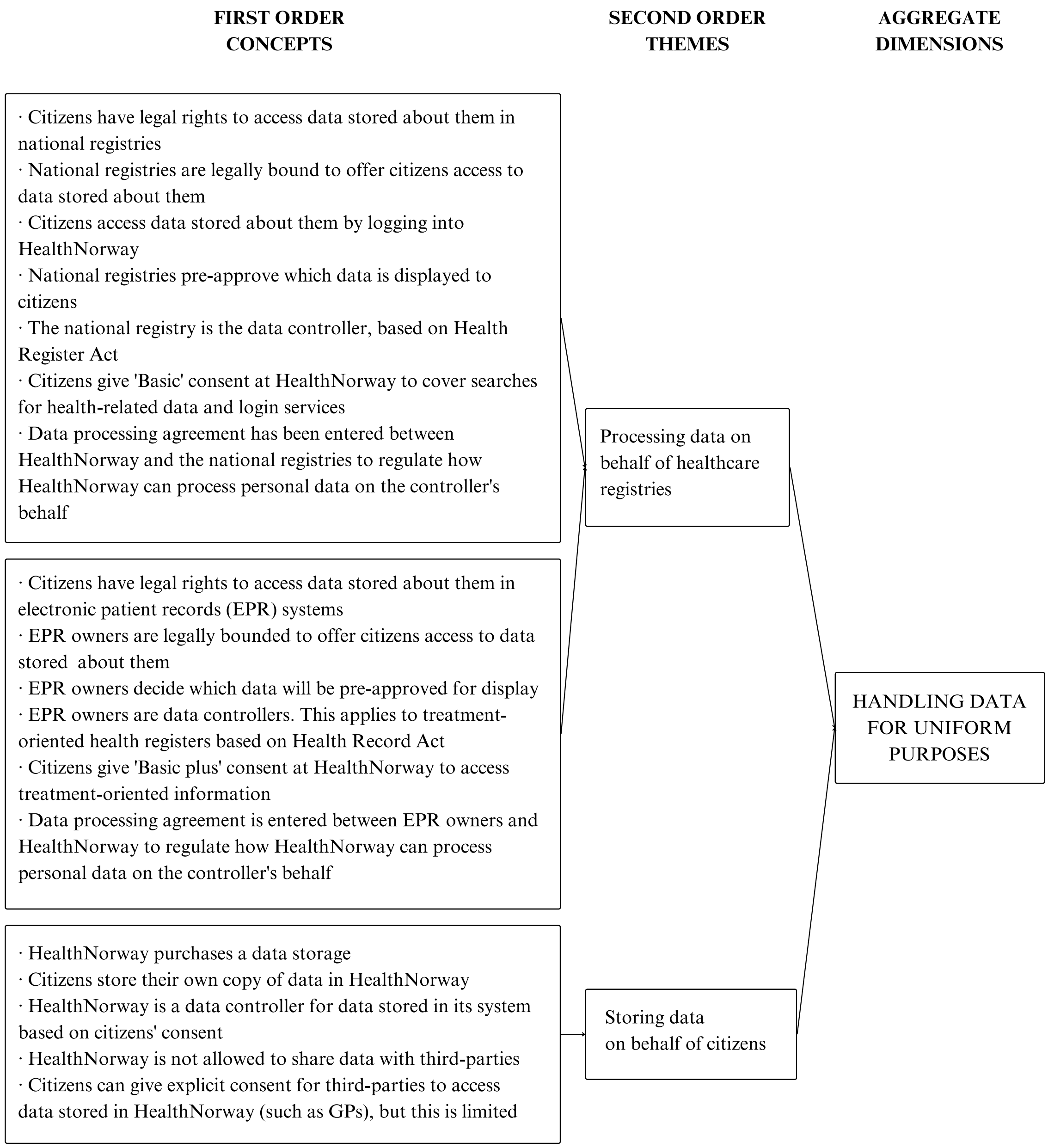
In the second phase, we used the Gioia et al. (2013) method to induct concepts on top of our process analysis and weight our empirical material against the data governance literature (see Fig. 2). We ended up with 35 first-order informant terms related to the eight functionalities defined in our initial analysis. From there, we generated five second-order themes which showed how personal data were governed across multiple actors. We then used these insights to aggregate the second-order themes into more abstract concepts. In this phase, we realized how the same data could be governed differently by multiple actors depending on whether data are processed for the same or different purposes. Thus, we ended up with two aggregate dimensions: data handling for uniform purposes and data handover for different purposes.

With the insights from these two aggregated dimensions, we returned to the existing literature on data governance to inquire whether we discovered novel concepts (Corley & Gioia, 2011). We realized that the existing literature does not account for the role of data and actors' purposes for data processing. Additionally, the assumptions in extant literature were focused on top authorities, but not on multiple authorities governing the same data. Thus, we inducted the concepts of vertical and horizontal dynamics to show how actors can subordinate to an authority, or multiply their authorities when governing data. Finally, to provide a conceptual

We realized that as the functionalities were extended, certain data governance decisions had to be repeated, but had different outcomes over time. Such decisions include: actors involved, delegation of roles and responsibilities (data processors and data controllers), data storage, and citizens' rights over personal data stored about them. This phase uncovered the need to move beyond a framework understanding of data governance, to also account for data governance as changing over time.

In the second phase, we used the Gioia et al. (2013) method to induct concepts on top of our process analysis and weight our empirical material against the data governance literature. We ended up with thirty-five 1st order informant-terms related to the eight functionalities defined in our initial analysis. From there, we generated five 2nd order themes which showed how personal data were governed across multiple actors. We then used these insights to aggregate the 2nd order themes into more abstract concepts. In this phase, we realized how the same data can be governed differently by multiple actors depending on whether data are processed for the same or different purposes. We therefore ended up with two aggregate dimensions: data handling for uniform purposes and data handover for different purposes.

With the insights from these two aggregated dimensions, we went back again to the existing literature on data governance to inquire if we have discovered novel concepts (Corley & Gioia, 2011). We realized that the existing literature does not account for the role of data and actors' purposes for data processing. Additionally, the assumptions in extant literature were focused on top authorities, but not on multiple authorities governing the same data. We therefore inducted the concepts of vertical and horizontal dynamics to show how actors can subordinate to an authority, or multiply their authorities when governing data. Finally, to provide a conceptual understanding of data governance by accounting for the role of data and the relationships between multiple actors, we generated the concept of *data governance spaces*. We elaborate on our findings and the concepts introduced in the sections that follow.



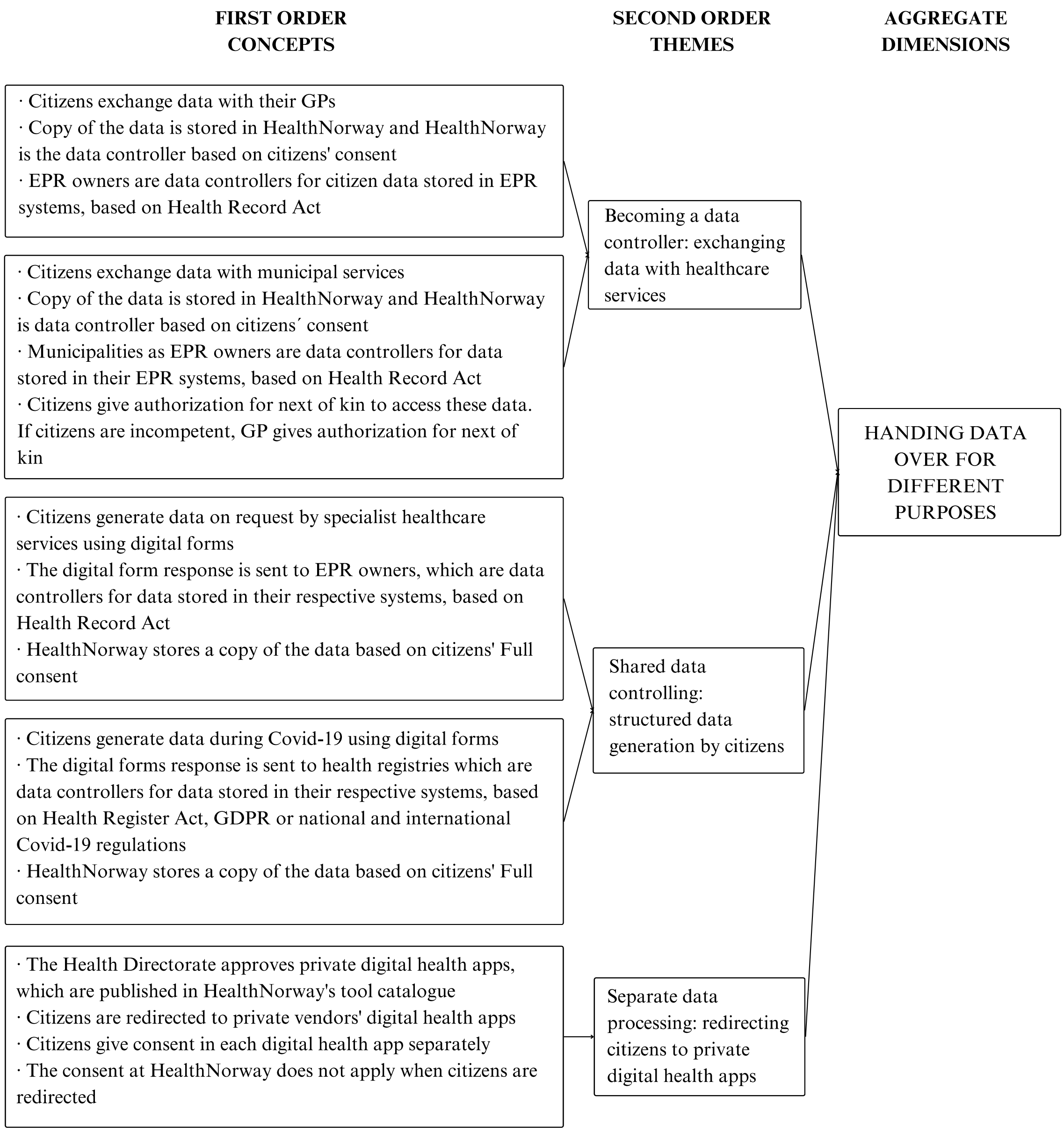


Figure 2: Inducting concepts from the empirical material (Gioia et al., 2013)

4 Findings: Governing Personal Health Data Across Multiple actors

In this section, we present how multiple actors engaged in decision-making around governing citizens' health data supported by HealthNorway's functionalities. Initially, HealthNorway provided quality-assured, general and open information about illnesses and treatments where no personal data were involved. We therefore start our account with the first login services, i.e., the services that involved personal health data.

4.1 Processing Data on Behalf of Healthcare Registries

The first personalized services HealthNorway offered were providing citizens with access to information stored about them in national health registries, such as vaccinations and drug prescriptions. For these services, HealthNorway had to collaborate with other health and care actors, such as the Public Health Institute (which owned the Vaccination Registry) and the Health Directorate (which owned the Prescription Registry). Initiating these services was triggered by citizens' legal right to access data stored about them and national registries' legal binding to provide such access. The healthcare registries decided to use HealthNorway as a channel for secure digital access. Technically, this was supported by HealthNorway introducing a login solution that identifies and authenticates the citizens before providing access to personal information stored about them. Citizens could only view, but not store a local copy of these data.

> *"Mainly we try not to store so much data because it is too complicated to process it, but HealthNorway has a lot of services where we do not store much data; we just provide access to it without seeing it. HealthNorway is saying, 'We know that you have some data at the Health Trust. We cannot open it, but we can help you see it,' so we cannot snatch the information on its way to the user. Many people think that HealthNorway knows a lot about you, but we do not. We are not allowed to read these data." (Informant, HealthNorway)*

In 2014, a second category of citizen-centric services was introduced, providing access to information stored in hospitals' electronic patient record (EPR) systems. This work was initiated by the Regional Health Trusts, and by citizens requesting access to personal data stored about them in the EPR systems. The Regional Health Trusts and citizens collaborated together with HealthNorway to map out the needs for providing these functionalities. The functionalities were available to citizens of the specific healthcare regions, but due to the systems' different maturity levels, the information citizens could access differed from region to region.

The processing of personal data in national registries was regulated by the Health Register Act; the same applies to the regional EPR systems which were regulated by the Health Record Act. In this case, the law determined and directly appointed the data controllers as it imposed a duty on the registries to collect and process personal data related to the tasks they fulfill. The law also obliged the registries to provide citizens with access to personal data stored about them. The registries could decide which information should be pre-approved for display to citizens in HealthNorway within their defined purposes and had to enter into a data processing agreement, which regulates how HealthNorway can process personal data on their behalf. However, HealthNorway also had to establish its own legal basis for processing personal data related to authentication and authorization when citizens log in. It was decided to introduce the option for citizens' consent, grouped in three categories: Basic, Basic Plus or Full consent. An informant shared some of the discussions related to defining the consent structures:

> *"When we created the structures of the consents, we were in dialogue with the data authorities, and they said that you should be able to choose not to do everything. So, the reason for us having a layered structure of consent, is because it is kind of decent to allow citizens to use parts of the functionality, but not all, and then try to find out what is possible to sort of exclude in these different layers. Then, medical records were considered something that was more sensitive to citizens than some other information. Therefore, we wanted to make a structure whereby one could say, 'Okay, I want to use these other things, but I do not want to use that.'" (Informant, HealthNorway)*

The different consent categories allowed citizens to choose whether or not they wanted their personal health data to be processed, control which information could be processed and what it could be used for. For example, providing citizens with access to data stored about them in national registries required consent to the category Basic, which covered the processing of multi-purpose health-related personal information. However, the regional EPR systems stored treatment-related personal information, and this required consent to the Basic Plus category. Therefore, for these initial functionalities the data governance authority was placed with the healthcare registries. The healthcare registries determined the purposes for processing citizen data, specified how citizens could exercise their rights to access personal data stored about them, as well as delegated roles and responsibilities to HealthNorway on how to handle data within the specified purposes.

4.2 Storing Data on Behalf of Citizens

In 2014–2015, HealthNorway started working on offering interactive services where citizens could send and receive messages with healthcare services, instead of solely accessing data stored elsewhere. This service required that HealthNorway provides a storage where citizens could keep their own copy of the messages, which would be secure, even if the healthcare services changed their EPR system providers. The storage used up to then was evaluated as being not comprehensive enough for storing this type of personal health information, so HealthNorway purchased a new data storage solution. The core principle was that the data stored in HealthNorway belongs to the citizen, not to the healthcare services. The legal basis for storing these data was citizens' consent, and HealthNorway took the role of a controller for the data stored in its system. The consent provided a legal basis for HealthNorway to store citizens' personal data, but not to share it with third parties, unless such sharing is preapproved by the citizens.

> *"The objective has been to provide the opportunity that you can, as part of what we provide, collect data that you want to share with someone, but this does not mean that HealthNorway is allowed to share it with others without your consent. So, it is always the inhabitants sharing data with someone; it is never the portal sharing data with someone without the inhabitants asking us to." (Informant, HealthNorway)*

Acquiring a storage as a technical capability, and citizens' consent as a legal basis, did not imply that HealthNorway could save a copy of just any personal data stored about citizens in external systems. Instead, whether storing a copy was allowed or not, also had to be specified within the agreements between HealthNorway and the healthcare actors it collaborated with. For example, HealthNorway could not store a copy of the data accessed from the Prescription Registry, as the data processing agreement specified that HealthNorway acts on behalf of the registry and within its defined purposes. An informant explained how storing data at HealthNorway is aimed at citizens as end users, which differs from the purposes for storing citizens' data in other public healthcare systems whose end users are healthcare personnel.

> *"HealthNorway is important because there is some information that you should be able to collect and you should be able to share, but I think it is*

> *more important to use HealthNorway for information that you are collecting for yourself, rather than for sharing information that resides in a third-party system, for instance at your healthcare provider. [...] So, it is kind of sharing of responsibilities and making sure that the portal is something that is a tool for you as a citizen, and other systems should be tools for instance for healthcare personnel who need the data." (Informant, HealthNorway)*

Therefore, the governance of data stored in HealthNorway was specified within the legal provisions of consent, and HealthNorway processed data on the citizens' behalf. The intention was that citizens would also use this storage to add or generate their own data (such as from wearables and welfare technologies) and decide whether they want to share these data with others (such as for research purposes). However, as of 2022, the primary usage of the storage was collecting copies of citizens' personal health data, due to lack of prioritization of its expanded usage. Regardless, the storage provided a significant basis for the functionalities further provided.

4.3 Becoming a Data Controller: Exchanging Data with Healthcare Services

As of 2014, many GP offices started offering interactive services for citizens. However, GP offices collaborated with different EPR vendors, but the functionalities and interfaces provided, and the security levels across the solutions differed significantly. HealthNorway started collaborating with the healthcare actors on providing common functionalities for message exchange, booking or changing appointments, and requesting or renewing prescriptions, which can bridge the differences across the back-end systems. Providing these services also required collaboration with the private EPR system vendors who had to implement the technical integrations, and the rollout was in 2015.

The processing of personal data for these functionalities had to be kept legally separate from other data processing arrangements at HealthNorway. For example, citizens could access prescriptions stored about them in the Prescription Registry, and renew prescriptions if their GP approved this through the HealthNorway functionalities. However, due to the different legal basis for processing data with national registries and GPs, these two functionalities had to be put on a separate page.

> *"In HealthNorway you can get a list of your medications, and you can see prescriptions and you can renew prescriptions. However, seeing the prescriptions is based on the regulation that handles the prescriptions, and requesting new prescriptions is based on the data processing agreement with the general practitioner and you are not allowed to mix those two. They [HealthNorway] wanted to put a button on your prescription lists where it says 'renew prescription', but they were not allowed to. They had to have those functions on separate pages because of the way the whole legal framework is built. The legal framework sort of prevents usability." (Informant, HealthNorway)*

In 2017, information exchange was also made available for citizens who received municipal health and care services (e.g. nursing services offered in elderly patients' homes). Besides the citizens, their authorized contacts (next of kin) could now also interact with the municipal health services. This required a new agreement, which can include next of kin in the digital interaction. If the citizens were competent to give an informed consent, they could enable these services themselves. If they were not competent, the next of kin could complete a form that is confirmed by the GP and be granted access at HealthNorway.

In the previous services for accessing information, HealthNorway solely acted as a data processor for data controlled by the national registries. Regarding services involving information exchange, EPR owners were data controllers only for the citizen data stored in their respective systems, and they were appointed this role based on the Health Record Act. However, a copy of the data was now also handed over to HealthNorway, and HealthNorway was the data controller for it, based on citizens´Full consent. The reason for EPR owners not being data controllers for the copy stored in HealthNorway was due to them not having formal control over the data processing taking place in external systems. An informant exemplified how data stored in HealthNorway are governed independently from data stored in other healthcare actors' systems, as in this case HealthNorway does not act on their behalf:

"Once the data is stored in HealthNorway then the data belongs not to the health provider but to the inhabitant or to us on behalf of the inhabitant, because the inhabitant should be able to delete it. The health provider cannot ask us to keep that data because they are going to use it later for reviewing their own work or something; it is the inhabitant that will be in charge of that data." (Informant, HealthNorway)

Therefore, storing the same data in different systems was subject to two authorities that could separately determine the purposes and means under which they process personal information. HealthNorway stored a copy on behalf of the citizen the data was about, where the purposes and means for data processing were regulated through citizens' consent. The purposes and means of GP and municipal systems were regulated by law, obliging them to store treatment-related information and always keep it updated.

Therefore, the same data were subject to different authorities, specifying distinct rules for data governance. The authorities could also delegate roles and responsibilities, but only for the data processed for their specified purposes. For example, some GP offices and municipalities purchased video consultation services provided by private vendors. When citizens used such video consultations services at HealthNorway, some personal information could be disclosed to the private video providers. In that case, the GPs or municipalities enter a data processing agreement with the private vendor systems they use; HealthNorway just redirects the citizen to that particular service. Therefore, the private providers of video consultations process data on behalf of the healthcare services, but not on behalf of HealthNorway.

4.4 Shared Data Controlling: Structured Data Generation by Citizens

Ever since the launch of the first interactive services in 2015, HealthNorway was collaborating with the regional health authorities to map out the needs for information exchange between specialist healthcare services and citizens. Instead of exchanging messages, specialist healthcare services raised the need for collecting structured data. The structured data forms available on the market then did not support sending or receiving data where the sender and the recipient used different systems. The healthcare actors decided that HealthNorway will provide an architecture of common components to support structured data exchange between different systems. This included standardized interfaces to integrate with private vendor forms that public healthcare actors already use, as well as a catalogue of digital forms which can be reused across healthcare services. The services were rolled out in 2018–2019, enabling a controlled way for hospitals to collect e.g., pre-consultation information, patient-reported outcome measures (PROMs) and patient-reported experience measures (PREMs).

During the Covid-19 pandemic multiple citizen-centric services were released using digital forms, such as symptoms reporting, symptoms checking, and customized reports. Some forms provided at HealthNorway were also collected by the National Institute of Public Health for better planning during the pandemic, such as deciding which citizens to prioritize for testing in the municipalities or which results to analyze first in the labs. Therefore, digital forms were used to exchange structured data with various healthcare actors who had different purposes for processing personal data.

Similarly to the previous services, HealthNorway took the role of data controller for the copy of the form stored in its components based on citizens' Full consent. The data controller responsibility for the health registries or EPR system owners as receivers of the digital forms was delegated to them by the Health Register Act or the Health Record Act, respectively. However, decisions also had to be made on how to govern data in the common components developed by HealthNorway and before these data were transferred to the health registries or EPR systems. It was assessed that there should be only one data controller for these components, which would clarify the roles and responsibilities. Since HealthNorway developed the common components, it also took the role of data controller for data processed in these components, but this role only applied until citizens' data are handed over to the health registries or EPR owners.

> *"If one solution is [only sending a personal] link for filling in a form, then they [the health registries] only use HealthNorway as a mailbox. However, if they send you something like 'This is a form,' rather than 'This is just a message, 'they could also in the metadata say something about it, 'There should be a reminder if you have not done this task in a while' [...] So, with the registries in [Region Middle] when they send you a form you are*

redirected to the form filler of the register, e.g. the health register for brain disease or something. You are in that form filler but because they cannot store your data before you say 'I want to store these data in the register', the buffer storage is done at HealthNorway. Then, when you press the final button saying, 'I want to actually store, send this to the register,' then it is stored directly in the register as well, but you then will have a copy of your data at HealthNorway." (Informant, HeatlhNorway)

Similarly to the previous services, when using digital forms provided by private vendors, citizens would be redirected instead of filling this form at HealthNorway. The private digital form vendors processed citizens' data on behalf of the national, regional or municipal public healthcare actors who determined the purposes for such processing, but not on behalf of HealthNorway.

Therefore, the digital form services involved multiple healthcare actors who exercised authority over their own data processing and could independently determine the purposes and means or delegate roles and responsibilities for the data processed on their behalf. This did not imply that two or more actors were responsible for the same data processing. Instead, there was a clear division of responsibilities on where one actor hands data over to another.

4.5 Separate Data Processing: Redirecting Citizens to Private Digital Health Apps

During 2018, multiple private digital health apps were available for citizens, but there were no national criteria to evaluate which apps were safe to use as part of the official healthcare service offering. The public healthcare authorities started discussing how to provide a place where citizens could find quality-assured digital health apps. They decided that such apps will be published on the HealthNorway website, and the service called "tool catalogue" was launched in 2019. Inclusion in the tool catalogue was determined by the Health Directorate, which assessed case by case whether the digital health apps fulfilled the criteria around information security, the health benefits of the apps' content, and the public need for including such apps as part of the official healthcare service offering.

"There are agreements with the health provider that is responsible for the app, so we have an agreement with the Health Directorate that is responsible for several of the applications. We have an agreement that on behalf of the Health Directorate this application is made available to the population in the catalogue. Most of them do not require you to log in, therefore redirection is sufficient to provide to the need." (Informant, HealthNorway)

It was decided that the Health Directorate would take the responsibility for the digital health apps, while HealthNorway would only enter into agreements with the Health Directorate, but not directly with the private vendors providing these apps. One reason for such partitioning of responsibilities was due to the large size and variance of the private vendor market, whereas HealthNorway as a product was not intended to cover the overall sector needs but provide an entry-point for citizens into their personal health information. However, another significant concern was the limited scope of control that HealthNorway has over data processed in the private vendor systems, particularly when the purpose of such processing is defined by citizens' consent on both sides. The possibility of HealthNorway taking the data controller responsibility for processing data in the digital health apps was considered too risky. An informant shared some of the possibilities discussed:

"We had a meeting where we discussed the possibility of us taking the controller responsibility for the information in the different tools that we let people out to at HealthNorway; that is one alternative. However, the consequence of that is that we have to have the control of what happens in the tools. So, the question is how realistic is that and how are we going to follow up on all these agreements that have to be put in place because we are talking about a large scale. That is one possibility; the other possibility is to have a more strict routine for maybe a self-declaration from the tool – that is something we have discussed – to ensure that they are compliant with the GDPR and that they have sufficient information security and those things. So, the self-declaration form regarding the privacy and information security together with the responsibility that lies at the Health Directorate when it comes to the clinical purpose of the tool." (Informant, HealthNorway)

The healthcare actors also discussed the potential benefits of acquiring a dedicated Act that would give HealthNorway the same legal status as the health registries whose purposes for data processing are regulated by law. The attempts to acquire a dedicated HealthNorway Act were not triggered by the need to control data processing in the digital health apps, but could strengthen the basis for HealthNorway taking on the data controller role for these services. However, as of 2022, such regulation is still not in place. An informant shared how the current consent option limits the possibilities for data processing, unlike a dedicated law:

> *"The lack of HealthNorway law has given us quite a few restrictions because we have to keep the consent from the citizen valid at all times, and with the development process we have, things are developing really quickly. It is a really hard job to make sure that we are still within the scope of the consent that citizens have given." (Informant, HealthNorway)*

Unlike the previous functionalities where the data were processed in collaboration with healthcare actors whose purposes were regulated by law, now the purposes for processing data in each digital health app were regulated by citizens' consent. Neither HealthNorway nor the digital health app vendors had the formal authority to control how data are processedonce they are handed over to an external system. Also, if HealthNorway and the digital health apps processed the same data, the compatibility of purposes would have to be evaluated case by case. Thus, it was decided that technically, the apps in the tool catalogue would be stand-alone; citizens would be redirected and give consent in the respective app. Both HealthNorway and the digital health app vendors would be the controllers for the data processed in their own systems. However, neither could delegate roles and responsibilities to the other, or impose restrictions on the purposes the data could be processed for.

5 Analysis: Reconceptualizing Data Governance

5.1 Purposes for Data Processing: Data Handling and Data Handover

Our empirical findings show how data were processed for various purposes by multiple actors, which required renegotiating data governance instead of delegating responsibilities by one actor or from the top. To answer our first research question: "How to conceptualize data governance by accounting for the role of data", we show how multiple actors can process data for uniform or different purposes by distinguishing between *data handling* and *data handover*.

Data are handled when multiple actors process data on behalf of another actor who specifies uniform purposes for data processing. In our case, data are handled when HealthNorway accesses citizens' personal data on behalf of the healthcare registries as data controllers but such processing was only within the purposes defined by the registries. *Data are handed ove*r when multiple actors copy data, and each actor can determine different purposes for its own data processing. Our findings show how in the case of message exchange, the public healthcare actors were data controllers for their data processing based on law, and HealthNorway was the data controller for the copy of the same data stored in its system based on consent.

Therefore, each actor independently determined the purposes for processing the same data. Defining whether data are being handled or handed over has implications for conceptualizing data governance, as it shows how data processing is defined around specific purposes which can be uniform or different across multiple actors.

5.2 Vertical and Horizontal Data Governance Dynamics

To answer our second research question, "How does data governance unfold when data become itinerant across multiple actors?" we distinguish between vertical and horizontal data governance dynamics.

Vertical data governance dynamics refer to the *subordination* of rights, roles, and responsibilities under the authority of a specific actor. In this case, data are handled by multiple actors on behalf of another actor who holds the authority for the data processing. The authority defines the purposes for "why" data are being processed, and other actors can act solely within these specified purposes. The "how" or the technical and organizational means on how to achieve those purposes can be delegated to other actors. However, even if responsibilities are delegated, the authority is responsible for the data processed on its behalf. Other actors can process the data only for the concrete responsibilities delegated by the authority and cannot determine their own purposes for processing the same data. For example, the Prescription or Vaccination Registries delegated the responsibility for providing access to personal data stored about citizens to HealthNorway. HealthNorway could only process data on their behalf, not for its own purposes.

Horizontal data governance dynamics refer to the multiplication of authority, rights, roles, and responsibilities across multiple actors. In this case, data are handed over from one actor to another, and each actor separately fulfills the obligations of being an authority. Therefore, multiple authorities separately determine the purposes "why" and the means "how" to achieve such purposes for processing the same data. This does not indicate that the responsibility for being an authority is delegated from one actor to another, where one actor fulfills some obligations for being an authority while the other fulfills the rest. Instead, each actor independently fulfills the responsibilities for being an authority, defines its own purposes for processing data, delegates roles and responsibilities within the specified purposes and is responsible for the data processed on its behalf. In our case, both HealthNorway and the EPR owners were data controllers for the message exchange or digital form functionalities. HealthNorway was the authority for the data copied in its own system, and the EPR owners were the authorities for the data stored in their respective systems, and responsible for the private vendor systems processing data on their behalf.

Data governance dynamics	**Actors**	**Data processing**	**Empirical example**
Vertical data governance dynamics	Authority delegates roles and responsibilities within the specified purpose	Handling data for a uniform purpose	Processing data on behalf of healthcare registries
	Actors subordinate and act on behalf of authority		Storing data on behalf of citizens
Horizontal data	Authority multiplies and each authority	Handing data over for	Becoming a data controller: exchanging

governance dynamics	specifies its own purpose for data processing Each authority delegates roles and responsibilities within its specified purpose	different purposes	data with healthcare services
			Shared data controlling: structured data generation by citizens
			Separate data processing: Redirecting citizens to private digital health apps

Table 2: Horizontal and vertical data governance dynamics.

The horizontal and vertical data governance dynamics show how actors either subordinate or multiply their authorities, rights, roles, and responsibilities when governing data (see Table 2). However, such subordination or multiplication is not static, where one actor is either an authority, or solely acting on behalf of others. Instead, as new actors, data or purposes are added, actors can take on various roles for different data processing, making data governance dynamic and changing over time.

5.3 Data Governance Spaces

To provide a comprehensive understanding of multi-actor data governance and foreground the role of data, we introduce the concept of data governance spaces. We define *data governance spaces* as the authorized relationships among multiple actors which specify the boundaries of decision-making authority, rights, roles, and responsibilities around data processing. The definition of data governance spaces consists of three pivotal parts: multiple actors, authorized relationships, and actors' boundaries. First, data governance spaces are *multi-actor,* and not single-actor; however one actor can simultaneously participate in multiple data governance spaces by taking on different roles. The same actor may act as a controller for certain data processing and as a processor for another. Second, data governance spaces specify *authorized relationships* among actors. These relationships, whether unfolding across horizontal or vertical dynamics, either subordinate or multiply actors' authorities, rights, roles, or responsibilities. Therefore, unapproved disclosure of data to a third party, such as in data breaches, does not define a data governance space. Third, in data governance spaces, actors can exercise their authority, rights, roles and responsibilities within specific *boundaries*. These boundaries are defined by the transfer of data responsibilities, where data are handled for a uniform purpose or handed over for different purposes. Therefore, the boundaries are not determined by organizations, or the IT systems processing data, but determined by actors' purposes for processing data.

Our findings show how HealthNorway was simultaneously participating in multiple data governance spaces with public healthcare actors, such as national, regional and municipal services. However, data governance spaces were not enacted between HealthNorway and the private vendors, due to the inability to establish authorized relationships, and the differences in purposes for processing personal data. The data governance spaces and their horizontal and vertical dynamics are illustrated in Fig. 3.

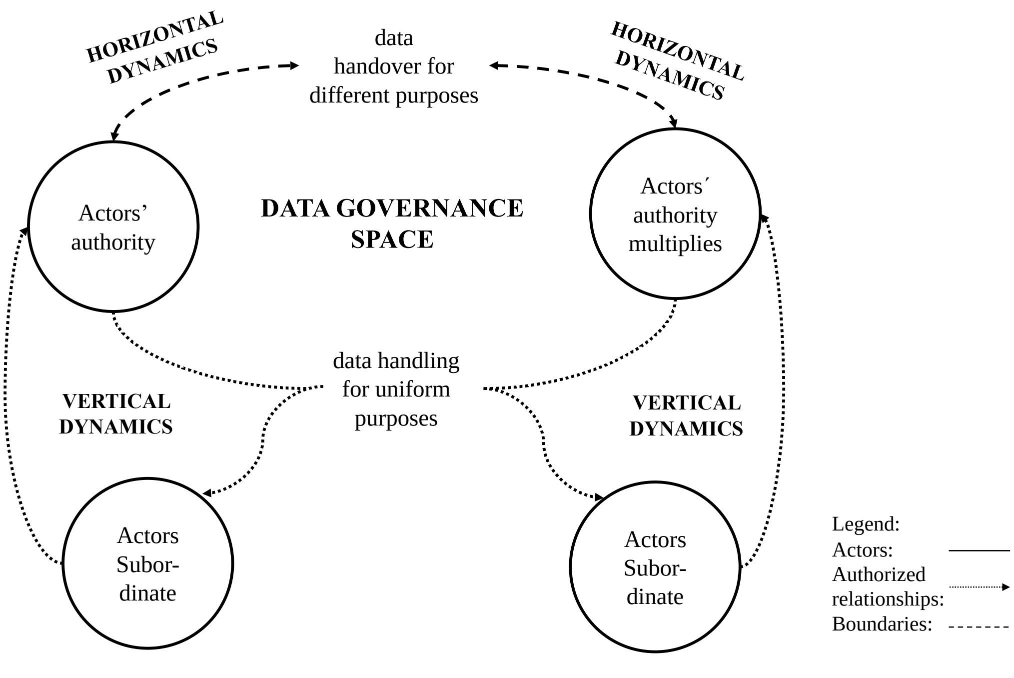


Figure 3: Illustration of data governance spaces and their horizontal and vertical dynamics.

6 Discussion

This paper contributes to the literature on data governance in the following ways. First, we contribute with an improved understanding of "what" is being governed by showing how governing data differs from governing IT. IS research on data has shown how throughout their lifecycle, data decouple from the digital technologies that produce or carry them (Alaimo et al., 2020) and get aggregated and repurposed into larger objects and commodities (Aaltonen et al., 2021). This paper argues for setting data at the core of conceptualizing data governance. By distinguishing between data handling and data handover, we show how data can be processed for uniform or different purposes by various actors. We argue that the purposes for data processing, and not the IT systems data are processed in, should be the guiding principle when conceptualizing data governance. This builds on IS works discussing how data differ from IT, as they can generate value through processes of signification, meaning-making and knowledge production, instead of by creating composite entities in the form of IT artefacts (Alaimo et al., 2020). Moreover, the data governance literature commonly refers to data as assets (Benfeldt et al., 2020; Fadler et al., 2021; Janssen et al., 2020; Nokkala, Salmela, & Toivonen, 2019; Van den Broek & Van Veenstra, 2015) implying that like other assets, data can be owned by organizations. By distinguishing between data handling and data handover, we show how data can be replicated across multiple actors, which separately determine the purposes for data use and define their own data governance rules. Furthermore, our findings show how in the case of personal data, data governance is also shaped by the rights of the data subjects they are about. This contributes to existing debates on the lack of clarity in determining data ownership (Fadler & Legner, 2020; Van Alstyne et al., 1995), as we show how data cannot be owned by organizations similarly to other types of assets governed.

Second, we contribute to the understanding of "who" governs data. Within organizations, the data governance literature commonly assigns this responsibility to the data governance leaders, councils, or offices (Abraham et al., 2019), data stewards (Rosenbaum, 2010), or data owners (Fadler & Legner, 2020). The literature on inter-organizational data governance shows how governing data requires coordination by actors around overarching goals (Jagals & Karger, 2021; Susha et al., 2017), but still assumes the presence of top authority for data governance, particularly regarding personal data. For example, Van den Broek and Van Veenstra (2015) have argued that in the context of personal health data, governance should be hierarchical due to legal implications. Our findings show how personal data can be governed by multiple authorities, which can delegate roles and responsibilities to other actors processing data on their behalf. By distinguishing between vertical and horizontal dynamics we show how actors can either subordinate to an authority or govern data independently by becoming the authority themselves. The concepts help us understand how roles and responsibilities around data are not simply delegated, access is not exclusive, and authority can shift from one actor to another.

Third, we contribute to the understanding of "how" data are governed. Our findings show that decisions about governing data are not simply managerial but can be delegated from outside organizational boundaries. Extant literature on data governance acknowledges the importance of legal frameworks and regulations, particularly in the context of personal data (Van den Broek & Van Veenstra, 2015; Winter & Davidson, 2020). However, we show how the law is not simply an antecedent (Abraham et al., 2019) or an environmental factor (Fadler et al., 2021) but another actor that actively shapes data governance, including delegates roles and responsibilities and determines the purposes for data processing. By introducing the concept of data governance spaces we show how data governance is not determined by intra- and inter-organizational boundaries but by the actors involved, their authorized relationships, and their purposes for data processing. Therefore, we move beyond conceptualizing data governance as a framework (Abraham et al., 2019) or a mode of governance (Jagals & Karger, 2021; Van den Broek & Van Veenstra, 2015), to encompass how data governance changes over time across horizontal and vertical dynamics, as new actors, authorized relationships or purposes for data processing are introduced.

This paper also has implications for practice, as it shows how data governance does not always fit pre-existing frameworks due to data's semantic and use-agnostic nature. Instead, actors must assess whom they process data with, and how and why are data processed to determine who should take the data controller or processor role in the specific circumstances. This does not indicate that data are governed without any structure but that the frameworks can direct; however, they cannot always predetermine the relationships occurring in multi-actor settings.

7 Concluding Remarks and Limitations

Personal health data – the empirical focus of this paper – are recognized as a key resource for innovation across healthcare services. Understanding data governance in such multi-actor contexts is crucial for enabling innovation in the healthcare domain and across other sectors seeking to address grand challenges requiring data-centered collaborations.

This paper contributes to the literature on data governance by introducing the concept of data governance spaces and their horizontal and vertical dynamics of change over time. The concepts are inducted from an empirical study on the evolution of data governance for digital health services in Norway, which brings certain limitations. First, it raises questions on whether the knowledge gained from a single-case study on governing sensitive and personal data in a highly-regulated environment can be applicable beyond this context. Although we conceptualize data governance through a study following the governance of personal health data in the European legal context, we believe the concepts introduced in this paper are generalizable across contexts. For instance, processing non-personal data might not be subject to data protection and privacy laws but to other types of actor agreements, such as intellectual property rights or contractual agreements. In commercial platforms for non-personal data, the platform owner may possess the intellectual property rights for data processed by third parties on its behalf, thus implying vertical data governance dynamics. However, if intellectual property rights are distributed, for instance, in an arrangement where each actor owns the data produced in its own components and separately defines the rules for governing such data when collaborating with other actors, the data governance dynamics would be horizontal.

Second, as raised in the methodology, the study's ten-year narrative was reconstructed through empirical material collected over approximately two and a half years. Therefore, the empirical story was constructed retrospectively and reflected the memory, interpretations, and opinions of informants while gathering the empirical material. We aimed to overcome such limitations by relying on official documentation encompassing the ten-year period covered in the study; however, this documentation does not fully account for the discussions and decisions occurring in real time. Moreover, collecting the empirical material was concentrated around a focal organization and a selection of private vendors, whereas the perspectives of other public or private collaborators were constructed through indirect inference. Therefore, this is another limitation of our study.

Future research can study data governance across different multi-actor settings, including perform comparative studies within and outside the highly-regulated European environment. Furthermore, in our case, data governance spaces were not enacted between the public organization studied and the private actors. Examining data governance in public-private collaborations, particularly in the context of personal data, requires exploration and could uncover novel data governance insights. The concepts this paper introduced can be developed further, and future research could show how data governance spaces unfold across different multi-actor settings.

References

Aaltonen, A., Alaimo, C., & Kallinikos, J. (2021). The Making of Data Commodities: Data Analytics as an Embedded Process. Journal of Management Information Systems, 38(2), 401–429.

Aanestad, M., Jolliffe, B., Mukherjee, A., & Sahay, S. (2014). Infrastructuring Work: Building a State-Wide Hospital Information Infrastructure in India. Information Systems Research, 25(4), 834–845.

Abbasi, A., Sarker, S., Aalto University, & Chiang, R. (2016). Big Data Research in Information Systems: Toward an Inclusive Research Agenda. Journal of the Association for Information Systems, 17(2), 1-32.

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.

Alaimo, C., Kallinikos, J., & Aaltonen, A. (2020). Data and Value. In Handbook of digital innovation (pp. 162–178).

Alvesson, M., & Sköldberg, K. (2010). Reflexive methodology: New vistas for qualitative research (Vol. 12). Routledge.

Bardhan, I., Chen, H., & Karahanna, E. (2020). Connecting Systems, Data, and People: A Multidisciplinary Research Roadmap for Chronic Disease Management. MIS Quarterly, 16.

Benfeldt, O. (2017). A Comprehensive Review of Data Governance Literature. Selected Papers of the IRIS, 120–133.

Benfeldt, O., Persson, J. S., & Madsen, S. (2020). Data Governance as a Collective Action Problem. Information Systems Frontiers, 22(2), 299–313.

Berends, H., & Deken, F. (2021). Composing qualitative process research. Strategic Organization, 19(1), 134–146.

Bonina, C., & Eaton, B. (2020). Cultivating open government data platform ecosystems through governance: Lessons from Buenos Aires, Mexico City and Montevideo. Government Information Quarterly, 37(3), 101479.

Bowker, G. C., Baker, K., Millerand, F., & Ribes, D. (2009). Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment. In J. Hunsinger, L. Klastrup, & M. Allen (Eds.), International Handbook of Internet Research (pp. 97–117). Springer Netherlands.

Constantinides, P., & Barrett, M. (2015). Information Infrastructure Development and Governance as Collective Action. Information Systems Research, 26(1), 40–56.

Constantiou, I. D., & Kallinikos, J. (2015). New Games, New Rules: Big Data and the Changing Context of Strategy. Journal of Information Technology, 30(1), 44–57.

Corley, K.G. & Gioia, D.A. (2011). Building theory about theory building: what constitutes a theoretical contribution?. Academy of management review, 36(1), pp.12-32.

DalleMule, L., & Davenport, T. H. (2017). What's your data strategy. Harvard Business Review, 95(3), 112–121.

Directorate of eHealth. (2021). Collaboration with industry in the e-health area.

European Commission. (2016). Regulation EU 2016/679 of the European Parliament and of the Council of 27 April 2016, General Data Protection Regulation. Official Journal of the European Union.

European Commission. (2020). Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions A European strategy for data. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066

Fadler, M., Lefebvre, H., & Legner, C. (2021). Data governance: From master data quality to data monetization. ECIS 2021 Research Papers, 16.

Fadler, M., & Legner, C. (2020). Who Owns Data in The Enterprise? Rethinking Data Ownership in Times Of Big Data And Analytics. ECIS 2020 Proceedings, 17.

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking Qualitative Rigor in Inductive Research: Notes on the Gioia Methodology. Organizational Research Methods, 16(1), 15–31.

Greengard, S. (2018). Weighing the impact of GDPR. Communications of the ACM, 61(11), 16–18.

Grisot, M., Hanseth, O., & Thorseng, A. (2014). Innovation Of, In, On Infrastructures: Articulating the Role of Architecture in Information Infrastructure Evolution. Journal of the Association for Information Systems, 15(4), 197–219.

Jagals, M., & Karger, E. (2021). Inter-organizational Data Governance: A Literature Review. ECIS 2021 Proceedings. 20.

Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy Artificial Intelligence. Government Information Quarterly, 37(3), 101493.

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

Lee, S. U., Zhu, L., & Jeffery, R. (2017). Data Governance for Platform Ecosystems: Critical Factors and the State of Practice. PACIS 2017 Proceedings, 12.

Markus, M. L., & Bui, Q. N. (2012). Going Concerns: The Governance of Interorganizational Coordination Hubs. Journal of Management Information Systems, 28(4), 163–198.

Nokkala, T., Salmela, H., & Toivonen, J. (2019). Data Governance in Digital Platforms. AMCIS 2019 Proceedings, 10.

Otto, B. (2011). Organizing Data Governance: Findings from the Telecommunications Industry and Consequences for Large Service Providers. Communications of the Association for Information Systems, 29.

Parmiggiani, E., & Grisot, M. (2020). Data Curation as Governance Practice. Scandinavian Journal of Information Systems, 32(1), 3-38.

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

Star, S. L., & Ruhleder, K. (1994). Steps towards an ecology of infrastructure: Complex problems in design and access for large-scale collaborative systems. Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work - CSCW '94, 253–264.

Susha, I., Janssen, M., & Verhulst, S. (2017). Data collaboratives as "bazaars"?: A review of coordination problems and mechanisms to match demand for data with supply. Transforming Government: People, Process and Policy, 11(1), 157–172.

Tallon, P. P., Ramirez, R. V., & Short, J. E. (2013). The information artifact in IT governance: Toward a theory of information governance. Journal of Management Information Systems, 30(3), 141–178.

Van Alstyne, M., Brynjolfsson, E., & Madnick, S. (1995). Why not one big database? Principles for data ownership. Decision Support Systems, 15(4), 267–284.

Van den Broek, T., & Van Veenstra, A. F. (2015). Modes of Governance in Inter-Organizational Data Collaborations. ECIS 2015 Proceedings, 13.

Vassilakopoulou, P., Skorve, E., & Aanestad, M. (2019). Enabling openness of valuable information resources: Curbing data subtractability and exclusion. Information Systems Journal, 29(4), 768–786.

Vilminko-Heikkinen, R., Brous, P., & Pekkola, S. (2016). Paradoxes, conflicts and tensions in establishing master data management function. ECIS 2016 Proceedings.

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

Winter, J. S., & Davidson, E. (2019). Big data governance of personal health information and challenges to contextual integrity. The Information Society, 35(1), 36–51.

Winter, J. S., & Davidson, E. J. (2020). Harmonizing Regulatory Spheres to Overcome Challenges for Governance of Patient-Generated Health Data in the Age of Artificial Intelligence and Big Data. TPRC48: The 48th Research Conference on Communication, Information and Internet Policy.

Zhang, Q., Sun, X., & Zhang, M. (2022). Data Matters: A Strategic Action Framework for Data Governance. Information & Management, 59(4), 103642.

Appendix I

Table A1: Summary of citizen-centric functionalities added over time.

Functio-nality	Year	Description	Citizens' rights	Actor roles	Legal basis
Accessing information in national health registries on behalf of citizens	2012	Citizens log into HealthNorway, but are redirected to Prescription or Vaccination Registry to see personal information about vaccination and prescriptions	View access	Data controller: National Health Registries Data processor: HealthNorway	Health Register Act, Personal Data Act, basic consent
Accessing information in regional health registries on behalf of citizens	2014	Citizens log into HealthNorway, but are regirected to see personal information from regional EPR systems	View access	Data controller: Regional Health Registries Data processor: HealthNorway	Health Record Act, Personal Data Act, basic consent

Storing data on behalf of citizens	2014	Citizens can save their own copy of data stored someplace else, or generate their own data	View, write, edit, save, delete	Data controller: HealthNorway	Citizens' Consent
Information exchange between citizens and GP offices	2015	Citizens can book and change an appointment, request or renew prescriptions and exchange messages	View, write, edit, save, delete (for data stored in HealthNorway)	Data controllers: GP owners for original data and HealthNorway for copy Data processors (if any): Private vendors, on behalf of GP ownerrs	Health Record Act Full consent required for message exchange, other services covered with basic or basic plus
Information exchange between citizens and municipal services	2017	Citizens or next of kin can book and change an appointment, exchange messages and get notifications	View, write, edit, save, delete (for data stored in HealthNorway)	Data controllers: Municipal services for original data and HealthNorway for copy	Health Record Act Full consent required for message exchange, other services covered with basic or basic plus

Structured data forms for data exchange between citizens and hospitals	2018	Citizenz can exchange digital forms with specialist healthcare services. The form can either be filled in at HealthNorway or citizens are redirected to a private vendor form via link	View, write, edit, save, delete (for data stored in HealthNorway)	Data controllers: Regional Health Trusts and HealthNorway for copy Data processors (if any): Private vendors, on behalf of healthcare services	Health Register Act or Health Record Act Full consent to save a copy
Structured data forms for data exchange between citizens and national services during Covid19	2020	Multiple citizen-centric services provided using structured data forms, such as: symptoms reporting, symptoms checking, book an appointment for test, book vaccination appointment.	View, write, edit, save, delete (for data stored in HealthNorway)	Data controllers: National Health Registries and HealthNorway for copy	Health Register Act, GDPR, national and international Covid regulation, full consent to save a copy

Launching tool catalogue	2019	Citizens can see a list of approved digital health apps. Redirect to use the apps.	No data exchange	No shared data processing involved	Basic consent for log in at HealthNorway Consent in the respective digital health apps

Appendix II

Table A2: Data analysis using Gioia et al. (2013) with quotes

HANDLING DATA FOR UNIFORM PURPOSES	
	”We [HealthNorway] use the National Institute of Public Health and other official registries to determine access. We have strictlyregulated access in HealthNorway, and we use a lot of audit logs for the whole chain to control which actor gets access to HealthNorway and vice versa. And we have login for the citizen, which hospital the citizen is registered at, and if there is establisheda health contact, that the citizen can contact. So, it is based on a very strict access regime.” "For the medication list, they have a right to view it, they do not have a right to get it out. So, you cannot process it for the patient, you can just read it as it is.”
Processing data on behalf of healthcare registries	”HealthNorway is legally grounded in that we get the right consent, the legal consent according to the GDPR, or if not consent, that we have the data processingagreement. [...] Based on that the consentmodels we have, we have the Basic and the Basic Plus and the full consent, which has some services involved to what kind of services you can use.” ”We were not processing, because the data was in the back-end systems. So, for e-Prescriptions, the data was not stored in HealthNorway. But we could share data with the patient. Just provide a view access.” ”HealthNorway has the ownership of the data that is processed at HealthNorway and companies or the hospitals are the controllers of the data in their respective systems.” ”I have been mostly involved in the part that you exchange data with [EPR systems] but also writing secure Text messages through HealthNorway and treating the appointments, then you both write into some part of the EPR and you get things out. So, the

	reading is one step, writing is one step, that is even a bit harder. So, for example, with Health South-East, we have only been able to get the agreements on the reading part, and that was the way to go to get all the rules and agreements on import." "The Summary Care Record is a system where the health personnel can log in to access information. What is different from HealthNorway and the Summary Care Record, is that in Summary Care Record, when you log in, you are a health personnel. So, the system will log your activity, and it will make that activity available for patients to see on HealthNorway. But, it also means that if there are cases of misuse or other things happening, then the information cannot be deleted by the inhabitants – like with HealthNorway – it is different. So, systems for use by health personnel need to respect and have different log in ways than systems used by inhabitants."
Storing information on behalf of citizens	"To be able to providea personalized experience one had to have a data repositoryand the data repository of [solution] was evaluated to not be fit for that purpose.It was basedon a completelydifferent structure and the manageability of it – as far as I remember and at least the way I reflect on it myself – was not appropriate regarding the regulations concerning this type of information." "At HealthNorway we store a lot of data on behalfof the citizen which is based on the consent given from the citizen. In some services, we store information on behalf of the sector which is based on an agreement that makes us a processor for the controller who owns the information. And when it comes to data storage outside HealthNorway, we do not have anythingto do with that. Outsideof HealthNorway is not something that is in production at this point." "Information that would be easy to share is information that you as an inhabitant gather and put there [in the HealthNorway

	storage], because that is your information, there are no other regulations with the third parties or parts of the healthcare system for saving the information there. So, that would be easy. Also, we are saving and storing your preferences regarding how you want the data to be made available. We already have a number of national registries that are using HealthNorway as a storage to save personal information. There is a register, they can have different levels of consent, it is safe [to store] at HealthNorway and the register is accessing that information. If you change it, then we notify the register that: ‘This information has now been changed, maybe you have to delete some information’.”
HANDING DATA OVER FOR DIFFERENT PURPOSES	
Becoming a data controller: exchanging information with healthcare services	”The patient side, all this information that patients e-mail and the doctors answer, will be stored in the personal health records of the patient. From the doctor’s side, this information will be stored in their patient journal systems, so there will be a copy of that dialogue on both sides. That was the motivation for having this personal health archive [at HealthNorway] in place and was the first functionality to be used.” ”GP can get access in HealthNorway, just pick out the one form that s/he needs, which is approved in advance [...]. We have the citizenwho accepts that: ‘I want to be able to share this document with my GP’, and that is registered in HealthNorway.” ”There is a video solution provided by Norwegian Health Network [owner of HealthNorway] where we transfer you from HealthNorway into the video solution as you are still authenticated and logged in. That we also for the GP solutions. For some of the other solutions it is more like a link where we are helping you to access the right video meeting but not sending, or not using your login information.” ”You can just ask for prescription, send questions, and just contact the front desk to ask to change your appointments. Some

	of the dialogue goes directly to the GP, but some also goes to the health secretary, and some is to ask for a new prescription, that is more automatic. You can also ask questions and start video consultation, and then you also have a tool that helps you to find the right consultant for you – is it best to meet physically, is it best to use the video consultation, or is it okay to just write a question." "If the patient is competent to give consent, s/he has to establish the representation himself/herself, and if s/ he is not, s/he has dementia, or is not able to understand the consequence of consent, they can have a form filled in that is confirmedby the GP, and we can grantaccess to the next of kin basedon that digital form that confirms it."
Shared data controlling: structured data generated by citizens	"A copy is stored in HealthNorway, that is the citizen's copy, and the response is sent to the controller who is responsible for maintaining the information they receive due to their legal framework. They have to be in control of how they can use this information – that is the controllers' responsibility." "We [HealthNorway] can show what prescriptionsyou have, but we cannottake that structureddata and send it to the GP– that is not part of the regulation.So, we can give you an insightinto data, and I am sure you can pull out the structured form of that and send it in an unstructured way to your GP. But, we cannot actually create the solution where it is technically easy to show you prescriptions, see which ones are soon empty, which you need a new prescription for, and send that directlyin a structured way to the GP system– that is not something that is legal to do now. It is technically easy, it will make a lot of sense for citizensand the GPs, but it just requires a change of regulation. " "If someone sends you a form, like the Multiple Sclerosis Register, then we transfer you to the form filler of that health register, and that health register will store the form at

	HealthNorway. But, it will also store it directly in the health register, so you have a copy, and they have their sort of their copy of the information." "The Public Health Institute initiated asking the population about their symptoms and they triggered the sending. So, they said 'We want that person to answer this form'. And they [the citizens] would fill in the responses in that form, and HealthNorway would send the forms back to The Public Health Institute. So, that was a service provided to The Public Health Institute by us, and by chance they were using the HealthNorway form filler. "
Separate data processing: redirecting citizens to private digital health apps	"That [the digital health apps from the tool catalogue] is available outside of HealthNorway. So, the citizen can choose which tool to use, but the termsof use for that tool are out of our control. It is the agreement from the citizen to the owner of the tool to accept the use of the tool, and how do they process [data] in the tool." "They [the digital health apps from the tool catalogue] have to verify that they follow the Norms for information security in the healthsector in Norway. We also have a third-party agreement for being integrated with the Norwegian Health Network, and the 'okay' stamp from Health Directorate that the content is clinically responsible." "The tools that are in the catalogue today are tools that different part of the healthcare sector has said that they want us to make available. For all of the tools we have made a sort of security check that they are compliant with the policies for how to treat and manage data, but it is someone else who has said 'This should be part of the public healthcare offering'. And when there is health information in the tools, then there is some part of the healthcare system that has approved that this is not harming patients 'This is something that we approve of the healthcare part of this tool'."

Appendix III

Table A3: Public document sources

Document name (translated from Norwegian)	Year	Publisher	Description
Yearly reports (2011-2022)	2011-2022	Norwegian Health Network	12 yearly reports by Norwegian Health Network (owner of HealthNorway since 2020). Describing the digital service needs of actors in the healthcare sector, including HealthNorway functionalities.
One citizen – one journal Digital services in the health and care sector	2012	Ministry of Health	Recommendation to Parliament for patient information to follow the patient lifecycle, and overview of the fragmented ICT portfolio in the health and care sector challenging such aspirations.
The primary healthcare service of the future - closeness and comprehensiveness	2015	Ministry of Health	Recommendation to Parliament from the Ministry of Health describing the necessity and importance of the digital dialogue services for GPs and municipalities at HealthNorway.
National health and hospital plan (2016−2019)	2015	Ministry of Health	Recommendation to Parliament from the Ministry of Health describing the necessity and importance of the digital dialogue

			services with specialist healthcare services at HealthNorway.
Digital citizen services in the specialist health services 2015	2015	National ICT Board	DIS Project for realizing the target image for digital citizen services with one common online health service.
HealthNorway content strategy 2017-2020	2016	Directorate of e-health	Concrete goals, quality principles and methods for evaluating the quality of the content published at HealthNorway.
Yearly reports (2017-2020)	2017-2020	Directorate of e-health	4 yearly reports of Directorate of e-health (owner of HealthNorway until 2019).
Critical health information (alert information) in the Summary Care Record	2018	Directorate of e-health	Description of principles and guidelines for registering critical information in the national Summary Care Record.
Data responsiblities	2019	Directorate of e-health	Attachment document, data responsibilities for products owned by Directorate of e-health.
Consultation response Directorate of e-health: Changes in data responsibility for the Core Summary Care Record, e-	2019	Directorate of e-health	Proposal for changes in legislation and the benefits of a dedicated HealthNorway law

prescriptions, health registries.			
Guidance in good practice for the use of Digital Dialogue for GPs	2019	Directorate of e-health	Advice and recommendations on the technical procedures, facilitation and further development of the digital dialogue services.
Description of Data Responsibility For Processing Personal Information in Residents' Use Of Services at HealthNorway	2019	Norwegian Health Network	Explaining responsibilities for data processing between HealthNorway and other healthcare actors.
Special terms of use for digital citizen services for the Norwegian Board of Health	2020	Norwegian Health Network	These terms of use complement HealthNorway - General terms of use and provide provisions for the individual services.
Special conditions of use for digital citizen services for GPs and other health personnel in the primary health service	2020	Norwegian Health Network	These terms of use supplement the General Terms of Use of HealthNorway and provide provisions for the primary health services. Primary health services that want to use services at HealthNorway must accept both HealthNorway - General Terms of Use and the special terms of use for the relevant services.

Special terms of use for digital health and care services for the municipalities	2020	Norwegian Health Network	These terms of use supplement the General Terms of Use of HealthNorway and provide provisions for the municipal services. Municipalities that want to use services at HealthNorway must accept both Helesenorge - General Terms of Use and the special terms of use for the relevant services.
Special terms of use for patient travel for the Health Trusts' services at HealthNorway	2020	Norwegian Health Network	These terms of use supplement the General Terms of Use of HealthNorway and provide provisions for the regional services. Health Trusts that want to use services at HealthNorway must accept both the General Terms of Use of HealthNorway and the special terms of use for the relevant services.
Special terms of use for the Norwegian Medical Agency's services at HealthNorway	2020	Norwegian Health Network	These terms of use supplement the General Terms of Use of HealthNorway and provide provisions for the drug prescription services. Organizations that want to use services at HealthNorway must accept both the General Terms of Use of HealthNorway and the special terms of use for the relevant services.
Special terms of use for digital citizen services for	2020	Norwegian Health Network	These terms of use supplement the General Terms of Use of HealthNorway and provide

the specialist health service			provisions for the specialist health services.
HealthNorway product strategy 2021-2026	2020	Norwegian Health Network	Five-year strategy for the development of functionalities and services at HealthNorway.
HealthNorway Roadmap	2020	Norwegian Health Network	Addition to the HealthNorway product strategy 2021-2026.
HealthNorway content strategy 2021-2026	2020	Norwegian Health Network	Description of aims for citizen-centric functionalities 2021-2026.
Target architecture for data sharing in the health and care sector	2020	Directorate of e-health	Describes the need for common components in the digital healthcare services which will facilitate data sharing between data controllers and other health personnel including the patient themselves.
Report of solution concepts: Data sharing infrastructure for digital home follow-up	2020	Directorate of e-health	Description of alternatives for national data-sharing infrastructure for digital-home follow-up to cover the need of all national, regional, municipal services.
Interim solution for appointment booking of Corona test	2021	Norwegian Health Network	Guide for solution for appointment booking during Covid19.

Time booking resource	2021	Norwegian Health Network	Guide for GP offices for appointment booking functionality at HealthNorway.
Time booking resource	2021	Norwegian Health Network	Guide for municipalities for appointment booking functionality at HealthNorway.
Temporary staff solution for GPs	2021	Norwegian Health Network	Guide for the usage of digital dialogue functionalities by temporary staff.
Digital forms at HealthNorway	2021	Norwegian Health Network	Overview of common components and functionalities related to the digital form services at HealthNorway.
General terms of use for HealthNorway	2021	Norwegian Health Network	Terms of use regulating general provisions that apply to all companies that use services at HealthNorway.
HealthNorway - Terms of use for vendors	2021	Norwegian Health Network	The terms of use for all integrations and technical interfaces between the external solution and the national e-health solutions that are in production.
Collaboration with industry in the e-health area	2021	Directorate of e-health	Recommendations and principles for collaborating with actors from the industry.
Presription of tools via the Tool Broker	2022	Norwegian Health Network	User guide for citizens and healthcare personnel in prescribing digital health tools.

Safer health apps	2022	Directorate of Health	Proposal for a national evaluation framework and model for ussage of private digital health apps.
Assessment of principles for connection between HealthNorway and other solutions in the market	2022	Directorate of e-health	Principles for providing seamless experience for citizens to use regional and municipal digital health services which interact with HealthNorway.

APPENDIX II:

“Exploring the Ontological Status of Data: A Process-Oriented Approach”

Dragana Paparova

Abstract

Information systems scholars have been inferring data as ontologically unstable and epistemologically uncertain and mobile. Data have been conceptualized as distinctive from digital technologies and possessing properties to relate with other data, digital technologies, actors, and socio-political environments. Across such relations, data stabilize into larger objects, but also change as part of actors’ value-creation processes. However, data have been predominantly understood as open-ended, and the ability of data to simultaneously acquire structures and change has not been sufficiently explored – this requires an ontological investigation. The research question this paper seeks to address is “*how can data, understood as both process and structure, be ontologically accounted for?*”. The paper offers two contributions. First, it unpacks the process ontology of assemblage theory to account for data as dualities of structure and change. Second, it provides an understanding of data as irreversible historical productions which simultaneously engage in enduring and changing processes.

Keywords: data, process ontology, assemblage theory.

1 Introduction

Data took on increasing significance in the information systems (IS) field, initially within discussions around "big data" stemming from pervasive digitalization and datafication (Lycett, 2013), and more recently concerning advanced technological developments such as machine learning and artificial intelligence (AI) (Faraj et al., 2018). The early works characterized data's volume – processing vast amounts of data; velocity – speed of processing data; variety – the heterogeneity of data sources and forms; veracity – credibility and reliability of the data sources; and value-creation, as data were used to fulfill various actor goals (Constantiou & Kallinikos, 2015; Lycett, 2013). This understanding of data came nonetheless from organizations collecting, sharing, and using social media data, (Constantiou & Kallinikos, 2015) where data were collected without a pre-determined purpose and data's value was explored a-posteriori in actors' meaning-making processes. However, as Günther et al. (2017) raised, the collection, sharing, and usage of other data types, such as personal health data, must be justified by a pre-defined purpose; therefore, data's value potential is not always open-endedly explored by actors. These works imply that there are different degrees to which data can relate to other data, digital technologies, actors and socio-political environments.

IS research has also shown how data's value potential can be constrained by technical, organizational, or legal structures. The ability of data to acquire structure, has been discussed around data's capacity to aggregate and form objects (Aaltonen et al., 2021; Alaimo & Kallinikos, 2022) but also as part of larger phenomena, such as data governance. The literature stream on data governance explored the rules, rights, roles, and responsibilities for governing data followed by conceptual (Abraham et al., 2019; Benfeldt, 2017) and empirical works (Parmiggiani & Grisot, 2020; Van den Broek & Van Veenstra, 2015). These works imply how data acquire structure as they are collected, shared, and used within and across organizations, by following specific frameworks, rules, and regulations.

Data are also a central area of interest to practitioners. For instance, as of 2019, the European Commission adopted the General Data Protection Regulation (GDPR) aiming to regulate the storage, processing, and usage of personal data (European Commission, 2016). More recently, the European Commission (2020) also encouraged the development of "common European data spaces" as shared infrastructures which can accelerate access, sharing, processing and usage of data for innovation across various industries and sectors, such as healthcare, finance, energy. Such practical developments highlight the centrality of data in today's organizational work, (cross-)sector collaborations, national and international political and regulatory contexts, where data are not solely a resource for value creation, but also resources that require dedicated governance approaches.

Data, as a distinctive IS phenomenon, have been ontologically understood as unstable (Alaimo, Kallinikos, and Aaltonen, 2020); however, the simultaneous processes across which data produce change and acquire structure have not been ontologically unpacked. The aim of this paper is to clarify the ontological status of data, by arguing for a process-oriented ontology, where fluidity and stability are dualities, instead of dualisms (Farjoun, 2010). Ontology, as "the science of what is, of the kinds and structures of objects, properties, events, processes, and relations in every area of reality" (Smith, 2003; p. 155), is concerned with the entities committed to theorizing and the relations with which such entities form larger wholes. The focus of this paper is to set the stage for re-examining what is data, by exploring the following research question: "How can data, understood as both process and structure, be ontologically accounted for?". To answer this research question, I build on the realist, process-oriented ontology of assemblage theory (AT) (DeLanda, 2006, 2016; Deleuze & Guattari, 1987), where space and time, structure and change, stability and fluidity, entities and relations are mutually enabling, instead of exclusive. By engaging with data's ontological status, this paper contributes to calls on producing novel theoretical and philosophical contributions in IS (Grover & Lyytinen, 2015), and IS debates on data (Aaltonen et al., 2021; Aaltonen & Tempini, 2014; Alaimo et al., 2020; Jarvenpaa & Essén, 2023; Tuomi, 1999) by providing an understanding of data as irreversible historical productions which simultaneously engage in enduring and changing processes.

The paper is organized as follows. Next, I provide an overview of IS research on data and the underlying ontological assumptions. In section three, I unpack data's ontology as relational and its implications for how we study data's structure and change. Section four introduces AT as a realist, process-oriented ontology, and presents the concepts of assemblages, virtuality, and multiplicities as a useful vocabulary for understanding data's ontology. Section five argues how AT's process-oriented ontology can bring an understanding of data as irreversible historical productions simultaneously engaging in enduring and changing processes. Section six discusses the main contributions of the paper. Finally, section seven highlights the implications for IS research and practice.

2 IS Research on Data

2.1 Data as Value

The early understanding of data in the IS field date back to the late 1990s, covering debates on the distinction between data, information and knowledge. As noted by Tuomi (1999), data were assumed to exist first as raw isolated facts or symbols, which are then interpreted or assigned meaning, relevance and purpose to become information. Knowledge is then extracted from data as a higher form of information. Tuomi (1999) challenged this hierarchical view of data very early on, by showing that data emerge last, only after structure and semantics are fixed to represent information. Therefore, data do not acquire structure, but are made by a structure, that is used to model, represent, and process them. Data took on larger significance in the IS discourse with the advent of big data stemming from more pervasive digitalization and datafication (Lycett, 2013), and the value creation potential of data has been central in many of those studies (Abbasi et al., 2016, 2016; Constantiou & Kallinikos, 2015; Günther et al., 2017; Kallinikos & Constantiou, 2015; Woerner & Wixom, 2015). The conceptual understandings of value creation from big data came not the least from organizations starting to use social media data. Constantinou and Kallinikos (2015, p. 54) describe the “heterogeneous, unstructured, agnostic, trans-semiotic nature of big data” – as differing from the well-structured data traditionally collected and used within a centrally controlled scheme – as social media data were captured so that they can be used a-posteriori.

As an empirical phenomenon, data and value were early on investigated in Aaltonen and Tempini's (2014) study on a mobile network operator. Their work shows how individual data tokens get re-grouped into larger audience-making events, which are incrementally formed and shaped to acquire meaning. The authors show how the actual meanings of data change over time, as “the employees perceived and acted on the assumption that there is more information in the data than that which is being actualized by the current metrics and reporting information” (ibid., p. 104). The meanings of data were also stabilized through the following mechanisms: 1) semantic closure, in which data were interpreted for a specific purpose; 2) pattern-finding or setting the parameters on how data were filtered in and out; as well as 3) framing, or reporting and presentation of the insights from such data. As the authors conclude, valuable information is only potential in the actual data, as the data pool is not deterministically useful or meaningful unless mechanisms on actualizing such value are set in place.

As empirical works followed, the focus started shifting from organizational perspectives on exploiting the potential of big data, to exploring the potential of data in environments where multiple stakeholders engage in simultaneous value creation processes, such as online platforms. For example, Barrett et al. (2016) explored how different forms of value were created in an online community for sharing healthcare data, as coevolving with stakeholders' value creation processes. Over time, value was created in different forms, such as financial, service, ethical, epistemic, reputational, or platform value; resulting in four different value propositions of the online community, such as rating, connecting, tracking, and profiling. These value propositions were not given, or pre-existing, yet emerging as new stakeholder relations were established by analyzing and repurposing data in the community.

The non-linear value creation processes of data and their simultaneous evolution with digital technologies, was furthermore explored by Tempini (2017) in the empirical setting of a social media patient community. His study shows how value gets entangled as multiple stakeholders relate across a data-intensive infrastructure; thus, bringing in non-linear and multi-dimensional value creation processes. As he notes, "the repeated updating and expansion of the web-based data-intensive infrastructure, exercised with a view to gradually improve and refine data practices across the network, repeatedly ignited cycles of value creation disruption. When an innovation disrupts shared practice, actors need to resituate data use processes in a way that is valuable according to any of the value dimensions they have stake in. Some dimensions of value creation could be enabled while others are hampered or shifted." (p. 206). These works (Barrett et al., 2016; Tempini, 2017) clearly show how creating value from data across multiple actors is non-linear and dominated by fluid and unstable processes.

The epistemic role of data was furthermore explored around data's mobility and ability to change on their "data journeys", where some data are lost, others get merged and acquire different forms and meanings (Leonelli & Tempini, 2020). However, empirical accounts also show that beyond these fluid characteristics, data can acquire stable states, as they relate to other data in different sociotechnical environments (ibid.). Data get accommodated in larger infrastructures, standardized across digital technologies and governed to manage their dissemination and interpretation across multiple actors (Aanestad et al., 2014; Parmiggiani & Grisot, 2020; Peukert et al., 2022; Winter & Davidson, 2020). I now turn to unpack works exploring data's potential to stabilize into larger objects and acquire (relative) structure.

2.2 Data as Objects

Beyond exploring data's epistemological status, IS scholars have also accounted for data's ontological status. Studies have defined data's properties, such as being editable (continuously revised, renewed and expanded), portable (shared across various digital technologies) and re-contextualizable (distanced from their origin and re-assigned meaning) (Alaimo, Kallinikos, and Aaltonen 2020). Data have also been characterized as being comprehensive (collections of tokens, behaviors), granular (can be aggregated, aligned and juxtaposed) and unbounded (have open-ended potential for acquiring meaning) (Aaltonen & Tempini, 2014). These properties provide data with a use-agnostic nature, as although they can be gathered for a purpose, their meaning is constantly explored, instead of reading pre-defined metrics. Data are, thus, not ready-made for usage, rather are often ambiguous and indeterminate, and need to be worked on, produced (Østerlie & Monteiro, 2020; Parmiggiani et al., 2021), aggregated and transformed (Aaltonen et al., 2021). As noted by Alaimo, Kallinikos, and Aaltonen (2020), these qualities bring "ontological instability" around what data are and "epistemological uncertainty" (p. 166) around how they are produced and what do they convey.

Moreover, IS scholars have also conceptualized data as being able to assemble into larger objects and commodities (Aaltonen et al., 2021; Alaimo, 2021; Alaimo & Kallinikos, 2022). For example, Aaltonen et al. (2021) show how data tokens relate according to certain criteria to be formed into objects and then assigned meaning as commodities, as they gain and lose properties. As they elaborate, data are actively produced in their journey of becoming commodities across which they constantly transform. In a study on social media data objects, Alaimo (2021) also shows how data and data objects are mutually coevolving. Data objects, as she elaborates, define what classifies as data and define interaction patterns, but such patterns are not fixed, yet constantly assembled. Another study by Alaimo and Kallinikos (2022) also shows how data objects come into being by aggregating data under a certain structure to acquire larger knowledge entities. However, these data objects are not the final output, but the intermediate step onto developing more complex organizational processes.

These studies show that data can acquire structure; however, such structure is not fixed, but relatively stable and characterized by an ongoing process of change. Additionally, these works show how data enter relations not simply with other data, but also digital technologies, and organizational environments, in which they interact across complex patterns of structures, relations, processes, entities, agencies. This requires an ontological understanding of data's spatial dimensions, i.e., the structures generated by, and generating data, and the temporal dimensions across which data transform as they form larger wholes. I now turn to ontologically unpacking data's relationality across space and time.

3 The Ontological Status of Data: Data as Relations

Scientific ontology deals with foundational beliefs of what the world we are researching is comprised of. To be more concrete, what kinds of entities, relations, processes, structures, exist in such a world. There are two opposed ontological stances on what does the structure of the world consist of – substantialism and relationism (Cooper, 2005; Dainton, 2014; Kempton, 2022). Substantialists view the world as a container in which everything else exists and occurs. The ocean contains water, fish, algae, microorganisms, and other sea life. Data pools contain data, their relations, meanings, which exist as independent entities, fixed and finished forms, and can be clearly separated from their environment. Categorizing data as independent entities, would mean that data have clear boundaries, and although they can relate with the wider environment, such relations would be conceptually treated as secondary. This paper adheres with the other view – relationism – where it is the relations between entities, rather than the entities themselves, that are central.

Relations can be defined as connections, interactions, sequences, causes and effects, or as spatio-temporal (Kempton, 2022; Pentland et al., 2020). IS research on data, particularly in qualitatively-oriented studies, has been predominantly underpinned by a relational ontological stance. However, ontological views differ on whether entities can be separated from the relations they are in, i.e., whether data change through their relations with other entities, or if such changes can (also) be caused by an internal structure (Kempton, 2022). As raised by Kempton (2022) "[a]ssuming separation and stability can be problematic when studying contemporary digital phenomena", such as machine learning and data analytics, as such technologies learn and change over time (ibid., p. 02). As he continues, "it can be difficult to establish clear boundaries between the agencies of people and the agencies of machines, as the lines between them are blurred" (ibid., p.02). For instance, IS research has shown how humans and AI do not solely substitute, or complement each other, yet can function as integrated dynamic systems which get reconfigured over time – assemblages (Grønsund & Aanestad, 2020). Therefore, data, algorithms, humans are not fixed, independent entities separated from one-another, but as a sociotechnical phenomenon, they are always in relation. Understanding data as ontologically always in relation to other entities, processes, actors, has implications for how we study data's structure and the changes they undergo, which I elaborate below.

First, data's relations over space and time. Space and time, as top-level ontological concepts (Dainton, 2014), have implications for how we study data's relationality. The dimension of space provides the conditions for the existence of some relations over others, i.e. some data relations get enabled while others get constrained. For example, GDPR enables the processing of personal data within the scope defined by law, or (informed) consent, but constrains the processing of personal data outside the purposes defined by law or consent. However, bringing in space without the dimension of time, would indicate that all possibilities in space are given at once; i.e., data are always in immediate relations with all possible data, digital technologies, actors, socio-political environments. Instead, by understanding time and space not as separate entities, but always in relation with each-other, we can show how as some relations actualize, new opportunities emerge; the relations data enter are always in the process of change. For example, different ways of processing data could be adopted, creating new opportunities for data usage, and triggering changes in the existing regulatory frameworks. Time, therefore, has an ontological significance for the movements of data in space, as it brings space in a constant process of production.

Second, data as dualities of structure and change. A relational approach to studying data also has implications for how we study data's structure and change. By putting relations, and not data as independent entities at the center, we can focus on parts and wholes, data and the forms they take, not as separate, but mutually enabling processes, i.e. dualities (Farjoun, 2010). Therefore, data's stable forms are not opposed to change, but both, a medium and an outcome of change. For instance, legal frameworks, such as GDPR, simultaneously limit and enable how actors collect, share, and use personal data by promoting coordination, setting up a common set of rules, but also constraining that data are used for purposes other than the ones they were collected for. Understanding the structures data acquire and the changes they undergo as mutually enabling, instead of exclusive, can help us show how data are constantly produced by, and simultaneously generating structures, but such structures are not static, yet change over time. As Langley and Tsoukas (2022) point out, stability and change are conceptually different, but ontologically inseparable, as they mutually interpenetrate to a point where one includes elements of the other. I now turn to elaborating on the specifics of assemblage theory as a process-oriented ontology.

4 Assemblage Theory as a Realist, Process-oriented Ontology

Realist ontologies in the social sciences have been traditionally structure-oriented and focused on causation stories, where entities with essences possess causal capacities to produce a certain outcome (Rutzou & Elder-Vass, 2019). Realism, therefore, was demanding form, order, and clarity, as opposed to chaos (Rutzou, 2017). As such, it would not deny the existence of relations between entities, but these relations would be causal and relating structure to structure. This understanding of realism has been the dominating assumption in the social sciences. However, recently, DeLanda (2006, 2016) brought in an alternative approach to the structure-process divide, as per the work of Deleuze and Guattari (1987), by introducing the ontology of assemblage theory (AT). As noted by Rutzou (2017) "this is a realism which affirms the world without necessarily affirming our representations of that world" (p. 405). This take on realism is different and important for studying data as a sociotechnical phenomenon for the following reasons.

First, the realism in assemblage theory forefronts difference, heterogeneity, and change (DeLanda, 2006; Rutzou, 2017; Rutzou & Elder-Vass, 2019). The traditional takes on realism have been advocating for clarity and simplicity. Rutzou (2017) argues that the authenticity of this traditional realism can be questioned, as it can be perceived as too abstract of a substitute for the messiness of reality. As he notes, "there are different degrees, intensities, and balances between order and chaos in phenomena just as, we might argue, there are different degrees of openness and closure in systems" (p. 404). He further on elaborates how the realism in AT is heterogeneous, but not completely chaotic, as it is also characterized by interdependence, conditions, processes, forces, and structures. The realism in AT presents an ontology in which the complex, dynamic and open world is not settled enough to be reducible to independent entities, such as things and categories. However, it still recognizes that heterogeneous parts can form relatively stable wholes. As Rutzou and Elder-Vass (2019) clarify, assemblage theory as an "ontology is a complex interplay between heterogeneity and homogeneity, dynamism and recurrence, but heterogeneity and dynamism always seem to have the upper hand" (p. 406). This was exemplified by DeLanda (2000) by referring to genes. Genes are not a blueprint for the generation of organic structure and function. Rather, genes act as constraints on a variety of processes that spontaneously generate order in organisms, in a way teasing out a form from them. Therefore, genes do not predict the structure that processes form, but provide patterns of behavior across which multiple structures can unfold. Similarly, in a data world, algorithms do not predict the data outcomes, but provide a set of variables across which various data outputs can unfold.

Second, AT focuses on becoming and formation stories, in contrast to composition and causation stories (Rutzou & Elder-Vass, 2019). In structure-oriented ontologies, such composition and causation is assigned to the role of mechanisms. Mechanisms are isolated parts which are homogeneous and generated by structures that possess causal power, but mechanisms are also producing structures. Therefore, mechanisms relate structures to structures. The central element in AT, on the other hand, are not mechanisms but multiplicities. These multiplicities are not independent entities, yet "everything is always an active production, and processes of production, that inextricably flow and bleed together in vast interconnected networks and assemblages" (Rutzou 2017; p. 407). As elaborated in the original text by Deleuze and Guattari (1987), multiplicities resemble rhizomes. They do not have points of departure like roots, nor do they have an end, yet look more like a map which is open and connectable on all of its dimensions. Multiplicities can be broken, shattered, can intersect and merge into larger wholes, but will always connect and start time and time again. The multiplicities are, therefore, not only connecting structures to structures, but continuously entangle a variety of entities, structures, processes, and forces, which are not produced by causes, but become contingent through historical evolution. DeLanda (2000) exemplifies this by referring to the unfolding of events during the industrial revolution. Technology should not be viewed as evolving in a straight line, as if the advent of steam power and factory production were the inevitable outcome of the evolution of machines. Instead, mass production techniques were only one alternative among several and the fact that they dominated the development is itself in need of explanation. Similarly, surveillance capitalism, as conceptualized by Zuboff (2019), was not the inevitable outcome of datafication and social media networks, nor was it an unfortunate accident. Instead, it unfolded through selecting choices among possible alternatives, where the deterioration of personal privacy was not a determinate outcome.

Third, AT promotes a flat ontology in which all there is, are assemblages. Structure-oriented ontologies commonly differentiate between hierarchical levels for the existence of entities. In critical realism, such hierarchy is assigned to the empirical (observable events), the actual (all events, whether experienced or not), and the real (the actual and the causal mechanisms which have not been instantiated in the actual) (Rutzou & Elder-Vass, 2019). In AT, entities are not caused by deep structures and do not simply form at two levels – the micro and the macro. Instead, heterogeneous sets of entities can interconnect, and form an emergent whole, but such wholes are not a new ontological entity, i.e. a totality, instead are a unique entity operating at a different scale. DeLanda (2000) uses the term "scales" to indicate that in AT, entities such as institutions, are not totalities which act causally on lower-level entities, e.g. organizations aiming to process data. Instead, institutions have the same ontological status as data, emerging from the relations between smaller scale entities – such as data, organizations, digital technologies – just operating at different spatio-temporal scales. Similarly, knowledge is not hierarchically produced by creating information out of data (Tuomi, 1999), yet, data, information and knowledge can relate at different spatio-temporal scales. I unpack these concepts of assemblage theory in more details in the sections that follow.

4.1 Assemblages

The main concept of assemblages translates from the French word "agencement", implying that the assemblage is not an outcome, but a process of assembling. However, recent interpretations of AT, with DeLanda (2006, 2013, 2016) as a central contributor, have adopted the term "assemblages". The process of assembling is one of a double articulation. It starts, as DeLanda (2016) elaborates, from a set of heterogeneous parts which relate to form larger wholes, which wholes are then stabilized by repeating, enduring processes. Acknowledging that wholes stabilize through enduring and recurring processes does not indicate that assemblages are unities in which the identity of parts is dependent on the whole. Instead, the parts are autonomous and can be detached from one assemblage and attached to another where different sets of relations can be established. DeLanda (2006, 2016) calls these relations of exteriority, to refer to the relations established between autonomous parts, which can change, without the identity of the whole changing.

Taking heterogeneity as a starting point is what distinguishes AT from other realist ontologies. This heterogeneity does not arise by a mere arrangement of distinctive parts and their properties. Instead, it arises due to the heterogeneity of relations such parts can establish, or their capacities. In AT, the properties of the whole are not given, but emerge as the parts interact; the whole is not formed by arranging properties, but by parts exercising their capacities. Therefore, if the parts stop interacting, the whole will also not be formed. As DeLanda (2006) explains, "relations of exteriority also imply that the properties of the components parts can never explain the relations which constitute a whole, that is relations do not have as their causes the properties of the components parts between which they are established" (ibid., p. 11). Therefore, it is the relations between the parts, and not their properties, what brings identity to the whole.

Let us translate this to data. Data are heterogeneous components. They range from social media data generated about people's behavior online, data about physical components such as pumps on oil and gas platforms, health data created to record people's healthcare, or wellness. Data also posess certain properties, such as being editable, portable, recontextualizable (Alaimo, Kallinikos, and Aaltonen 2020). These properties enable data to relate to other components, such as digital technologies, including social media platforms, sensors, electronic patient record (EPR) systems; contexts, such as finance, healthcare, energy, oil and gas. Due to their properties, data can be edited and ported across digital technologies, organizations, and contexts; however, whether such relations will take place, i.e., whether data will be shared, also depends on the interoperability of digital technologies, organizational practices, or the regulations enforced to protect personal or non-personal data, among other forces. Therefore, the relations data exercise are not caused by their properties, but also depend on the properties and capacities of other entities they interact with.

4.2 Multiplicities

As DeLanda (2000) elaborates, a realist ontology cannot only incorporate the processes which bring entities into being, but also the processes which keep their identity over time. The term multiplicity, originating from mathematics, refers to the measurements of geometrical space. In AT, a multiplicity denotes how an emergent space (consisting of part-to-whole relations) is to be measured. As such, multiplicities are a core concept in the ontology of AT, as they replace what is essentialism in other realist ontologies. Essentialist ontologies are composed of fully formed entities, called unities, which possess a core set of properties that define what they are, as well as causal capacities to constrain their parts. Multiplicities, on the other hand, structure the possibility space of the assemblage by defining "spaces of possibilities" as the possible ways in which an assemblage can change.

Therefore, multiplicities define the degree to which assemblages can be formed and change, i.e. the structure of the assemblage. The structure of assemblages is defined by the distribution of two multiplicities: 1) invariant (more stable, recurrent, shared by many parts), and 2) variant (more unstable, prone to change). DeLanda (2013) refers to these multiplicities as universal and individual singularities, respectively. The distribution of multiplicities which are both, invariant and variant, stable and changing, brings certain regularities to the structure of the assemblage, but such structure is not strata, i.e., in absolute stability, nor is it in complete fluidity, yet always oscillates between stability and fluidity. This allows the structure of assemblages to accommodate entities, relations, processes, actors, which keep on changing as they interact over certain patterns of regularities over time.

The structure gives direction, but does not predict the relations assemblages will establish, as such relations are not a copy of the structure yet correspond to it only to a certain degree. DeLanda (2006, 2016) explains how such relations unfold across simultaneous processes that stabilize or destabilize the assemblage, where the degree of stability is defined by two parameters, territorialization and coding. Territorialization refers to defining and sharpening the spatial boundaries of actual assemblages. Coding refers to increasing the degree of internal homogeneity of components and relations of an assemblage. The more territorialized and coded the assemblage is, the higher the degree of stability. At the same time, the assemblage is engaging in processes of destabilization; "[a]ny process which either destabilizes spatial boundaries or increases internal heterogeneity is considered deterritorializing" (DeLanda, 2006; p. 14). Therefore, assemblages simultaneously engage in processes of stabilization and destabilization, territorialization and deterritorialization, but never reach equilibrium, as their structure, and the relations they establish keep on changing. What is relevant is not what led to the relatively stable states, but the actual processes across which assemblages form and change.

Let us exemplify this through institutions, such as legal bodies regulating data-sharing. Institutions stabilize their identity by enacting and enforcing laws, and are more stable when they have well-defined spatial boundaries within which their jurisdiction applies. Any process which brings this jurisdiction in question blurs the spatial boundaries and destabilizes institutions, making them more prone to unlawful behavior. For instance, overlapping laws about the collection, sharing and usage of data, as in the case of cloud technologies, blur the territories within which data jurisdictions apply (Daskal, 2015). In this case, data travel across countries' borders and there is a physical disconnect between where data are stored, where they are accessed from, and who owns them. Another example for destabilization of institutions are the discrepancies between law formulation, and actual implementation. For instance, laws formulated to protect the individual-level privacy of data subjects, cannot reflect population-level data processing performed by advanced digital technologies, such as AI, where data are related to each-other (Viljoen, 2021). Therefore, institutions can be stabilized by the continuous enforcement of laws within their jurisdiction, reducing the gap between law formulation and actual implementation, and their ability to sanction unlawful behavior within defined territorial boundaries.

4.3 Virtuality

Assemblages form along a structure of possible forms, but also keep on changing; this brings the need to unpack how assemblages unfold. AT has a flat ontology, defined around the real, the actual, and the virtual. The actual includes all relations that are actualized; e.g., all the ways in which data are shared across actors. DeLanda (2013) points out that "Deleuze speaks not of realization, but of actualization and introduces a novel ontological category to refer to the status of multiplicities themselves: virtuality" (p. 24). The virtual consists of all relations that can be actualized, out of which some will and others will not; e.g., all the ways in which data can be shared across actors. As per the words of Deleuze (2014) ”the virtual must be defined as strictly a part of the real object – as though the object had one part of itself in the virtual into which it is plunged as though into an objective dimension" (p. 272). Therefore, the real consists of both, the actual and virtual; e.g., all the ways in which data are shared, can be shared, and could be shared across actors.

By introducing the concept of the virtual, AT shows how the actual can be realized in a variety of ways, which include mechanisms, but also reasons, and motives – producing non-linear causality. The non-linear causality comes from the status of multiplicities, which are distributed and meshed, and not sharply distinguishable from one-another, as essences are. In AT, the multiplicities are not given all at once – as is the case of essences – but unfold progressively, not by producing finalized forms, but by giving form to processes. Therefore, some relations actualize over others, not by being causally produced, or logically necessary, but by becoming historically contingent (DeLanda, 2006), as they could have unfolded otherwise. Due to the concept of virtuality, the focus in AT is not solely on how the actual unfolds, but also how it could unfold in conditions that may or may not occur.

Relating back to the example of the industrial revolution. The sequence of events that have taken place, such as using steam power, inventing machines, or the organization of workers into factories can be followed in the actual, but in the virtual, the possibilities for these events actualizing or not, and moving from manual to machine manufacturing, coexisted. Similarly, the sequence of events across which surveillance capitalism emerged can be followed in the actual, but in the virtual, they could have taken another route. Laws could have been adapted to reflect the technological advancements and constrain big tech companies to utilize personal data for commercial gains. Surveillance capitalism, therefore, was not caused by big tech companies exploiting data in a lawless space but emerged through complex interactions among global adoption and usage of social media, large network effects, and destabilized legal frameworks where the laws were lacking behind the rapid pace of technological advancements and innovation – among other reasons.

5 Assemblage Theory and its Implications for Data as a Sociotechnical phenomenon

The implications of the concepts from AT – assemblages, multiplicities, virtuality – discussed in the sections above, can be summarized in the following ways. First, assemblages are always heterogeneous in the actual, as they are historically produced and unique. Second, assemblages get formed around a defined set of regularities which correspond to the virtual, but are not a copy of it. Adding in the virtual is significant in showing that as historically produced, the assemblages unfold across relations which are subsequent only in the actual, but in the virtual multiple possibilities coexist on how such relations can actualize (DeLanda, 2000). As DeLanda elaborates, "the ontological status of assemblages is two-sided: as actual entities all the differently scaled social assemblages are individual singularities, but the possibilities open to them at any given time are constrained by the distribution of universal singularities, the diagram of the assemblage, which is not actual, but virtual" (DeLanda 2013; p. 41). Building on these concepts, the realist, process-oriented ontology of AT can provide an understanding of data as being: 1) irreversible historical productions; which 2) simultaneously engage in enduring and changing processes.

5.1 Data as Irreversible Historical Productions

The vocabulary of AT helps us understand data as irreversible historical productions that unfold progressively. Unfolding progressively means that not all the possible ways in which data can relate are given at all times. Instead, as some relations actualize, new possibilities open up, while others coexist. As data's forms and transformations progress in the actual, they become historically contingent, but not logically necessary, as in the virtual, they could have unfolded otherwise. Therefore, the changes data undergo cannot be logically decomposed solely by looking at the data forms in the actual; instead, multiple possible data transformations coexist, making the forms data acquire and the changes they undergo irreversible over time.

Let us consider AI-human-data assemblages. Algorithms pattern data into meaningful relationships using a set of variables. There are a variety of ways in which data can relate across these variables; for instance, the same set of data can be manipulated by the same algorithm and produce different outcomes; the same algorithmic outputs can be read by different humans and have distinct meanings. Therefore, the data outputs are neither random and chaotic, nor can they be calculated with absolute certainty. Instead, the actualized outcomes are one possibility among others, and the probability of each data output happening depends on what comes before and what can come after. This dependency can include the data used to train the algorithm, the previous set of operations performed on data, and the outputs of such operations, among other factors. Therefore, the data-human-AI assemblages resemble possible data outputs, but algorithms cannot predict the outcomes that will take place across the assemblage with complete certainty. As the outcomes are only probabilistic, data-human-AI assemblages cannot be logically followed back by solely looking at the data outputs, or the algorithmic patterns. Instead, these outputs involve historical contingencies, including the forms and meanings data have acquired, the human interpretations and alterations of algorithms, and the changing probabilities for each output happening over others. Therefore, the formations and transformations of data-human-AI assemblages become irreversible as they are not certain, but probabilistic and change among multiple outcomes over time.

5.2 Data as Simultaneous Enduring and Changing Processes

Understanding data as irreversible historical productions does not imply that data simply change over time; they also endure. By using the vocabulary of AT, we can understand data as simultaneously engaging in enduring and changing processes, where structure and change are a matter of degree. There will always be enduring processes across which data acquire forms and keep their identity over time; and changing processes across which data transform from one form to another, not by being chaotic, but by following certain patterns of regularities.

Let us, again, take data-human-AI assemblages as an example. Data, humans and AI relate across a set of parameters, programmed as algorithms. Algorithms pattern data using a set of variables, not by working as independent deep structures which cause data outputs, but by providing a set of instructions across which data outputs can unfold. However, as algorithms produce data outputs, they learn from the data and the operations they perform get altered. Some processes data engage in become contingent and enduring – certain steps in the algorithm get repeated unless instructed otherwise, or standards are developed for possible interpretations of the algorithmic outputs; other processes keep on changing – new data can come in, algorithmic rules are altered, humans re-interpret the algorithmic outputs. Therefore, over time, data acquire stable forms, as they stabilize across standards or data outputs; but also keep on transforming, as they can always be aggregated otherwise, and assigned another meaning; i.e. data engage in on-going processes of endurance and change.

Another example are data-intensive infrastructures (Jarvenpaa & Essén, 2023; Tempini, 2017). In healthcare, such data-intensive infrastructures hold legacy data, i.e. administrative or treatment-related health data about patients stored across various EPR systems. These legacy data endure over time by being standardized and interoperable across systems (Fossum et al., 2019), routinely produced and used as part of clinical work practices (Grisot et al., 2019), shared across systems, actors, organizations and regulations (Paparova et al., 2023). However, health data are not simply openly shared, as the conditions under which they are collected, processed or used is determined by healthcare personnel's official need in using such data for treatment purposes, the organizational rules and practices, legal and regulatory frameworks; i.e. processes which endure over time. At the same time, new data can become part of such infrastructures (e.g. sensor-based data from wearables), used in different ways across organizational work practices, requiring alterations of the legal frameworks, creating new possibilities for using health data for prevention and prediction of diseases; i.e., processes which change the data-intensive infrastructure over time.

Therefore, data can enter repetitive, regular processes, acquire larger forms and objects (Aaltonen et al., 2021), get standardized across digital technologies (Tempini, 2021), or patterned across algorithms (Grønsund & Aanestad, 2020); however these forms are not fixed and finished, as although data endure, the changing processes they enter always have the upper hand over time.

6 Discussion: Data, Structures and Change

This paper answers the research question: "how can data, understood as both process and structure, be ontologically accounted for?" by introducing the realist, process-oriented ontology of assemblage theory. With this, the paper contributes to calls on producing novel theoretical and philosophical contributions in IS (Grover & Lyytinen, 2015), by bringing in an ontology which can provide an understanding of data as irreversible historical productions which simultaneously engage in enduring and changing processes. The philosophical stances in IS, particularly in qualitatively-oriented studies, have been commonly centered around interpretivism and critical realism as the two poles, and their idealist and realist ontologies respectively. The former has been predominantly focused on processes and flow, and the latter on outcomes and stable entities. This paper presents assemblage theory (DeLanda, 2006, 2016; Deleuze & Guattari, 1987) as an alternative realist, process-oriented ontology, able to account for both, processes and outcomes, structures and change, as dualities (Farjoun, 2010). Assemblage theory has so far received only limited attention in IS (Hanseth & Rodon Modol, 2021; Tarafdar & Kajal Ray, 2021), and such focus was predominantly on applying its conceptual vocabulary. The ontological potential of AT has been recognized by other fields (Hodges, 2008; Rutzou, 2017; Rutzou & Elder-Vass, 2019), but has not been discussed in the IS discipline so far. Bringing in the ontological assumptions of AT in IS can drive the field further by providing an ontology which accounts for complex sociotechnical phenomena, which deal simultaneously with structures, processes, relations, entities, actors, oscillating across multiple dimensions (Rutzou & Elder-Vass, 2019), instead of solely from structures to structures.

The predominant assumptions in IS have so far been focused on the instability (Parmiggiani et al., 2021) and open-ended potential of data in actors' value-creation processes (Alaimo et al., 2020; Barrett et al., 2016; Tempini, 2017). Indications of data's ability to acquire structure have also been present, such as in studies on data objects (Alaimo, 2021; Alaimo & Kallinikos, 2022), or data governance (Abraham et al., 2019; Van den Broek & Van Veenstra, 2015; Winter & Davidson, 2020). More importantly, Leonelli and Tempini (2020) conceptualized data as historical productions, or data lineages as "not static objects whose significance and evidential value are fixed, but objects that need to be transformed in order to travel and be re-used for new goals" (p. 07). The authors foreground data's mutability and transformations as they travel, although they acknowledge how across those journeys data can have various degrees of stability.

This paper builds on this work (Leonelli & Tempini, 2020), and argues for understanding data as irreversible historical productions which simultaneously engage in enduring and changing processes. With this, the paper contributes to IS debates on data (Aaltonen et al., 2021; Alaimo et al., 2020; Parmiggiani et al., 2021), and the call by Jarvenpaa and Essén (2023) who encouraged novel theoretical approaches on "data sustainability" as "data's capacity to endure across technological and human generations" (p.10). This paper show how AT as a realist, process-oriented ontology can accommodate time and space, process and structure, fluidity and stability, endurance and change, to understand the transformations data undergo, and the forms they acquire over time.

Moreover, the process-oriented ontology of AT could help IS researchers to account for data as a sociotechnical phenomenon which has varying degrees of structure and change across different organizational, technological and legal contexts. For instance, the processes data enter can be more enduring or prone to change, depending on whether data are personal or non-personal; sensitive health data about patients, or maintenance data about physical components in oil and gas platforms; regulated by law or organizational contracts. By understanding data as having varying degrees of structure and change, we could study data as being more open-ended if shared without restrictions – such as open-government data; or following specific rules if they are sensitive and regulated by law – such as personal health data. This could bring a more comprehensive understanding that goes beyond data's properties of being editable, recontextualizable (Alaimo et al., 2020), or mobile (Leonelli & Tempini, 2020), to also encompass the structures across which data enter processes with other data, actors, digital technologies, socio-political environments.

This paper does not argue that AT as an ontology should be the preferable choice of IS scholars aiming to study data as a sociotechnical phenomenon. However, the concepts presented could be useful for scholars aiming to study data along the interplay of organizational, technological, legal contexts where data fluctuate across stability and instability over time.

7 Implications for IS Research and Practice

As noted by Little (2016), it is not possible to research a domain well if we do not know what things or processes it consists of. Metaphysical debates, such as "what is data?", and "are data separate, independent, fixed entities, or are they always in relation?", could provide rich foundations for developing the IS field's own metaphysics (Hassan et al., 2018). This paper explores the ontological status of data by arguing for a realist, process-oriented ontology which can accommodate structure and change as dualities. The ideas raised in this paper could contribute to IS researchers and practice in the following ways.

Data have become a central debate in IS, commonly conceptualized across larger IS phenomena such as data platforms (Alaimo & Kallinikos, 2017), data infrastructures (Tempini, 2017), data governance (Abraham et al., 2019; Parmiggiani & Grisot, 2020), data network effects (Gregory et al., 2021), artificial intelligence (Faraj et al., 2018; Grønsund & Aanestad, 2020). Understanding data's ontology could bring clarity around the role of data as part of these larger phenomena, but also to data as a phenomenon in itself. This could stimulate additional works focused on the distinctiveness of data as an IS phenomenon (Aaltonen et al., 2021; Alaimo et al., 2020), instead of solely treating data as a by-product of IT governance, digital platforms, or digital ecosystems. Moreover, understanding data's ontology as relational and produced through dualities of structure and change, could bring distinct research streams in IS closer, such as data-driven value creation (Alaimo et al., 2020; Tempini, 2017) and data governance (Abraham et al., 2019; Benfeldt, 2017), and encourage them to learn from each-other. Furthermore, the understanding of data's relations as historically contingent and irreversible, but not logically necessary, could ontologically ground phenomena such as algorithmic unfairness and bias (Schulze et al., 2022), as IS scholars and practitioners question the degree to which decisions can be automated and delegated to algorithms with or without human supervision.

Moreover, this paper provides contributions that bring the IS field closer to practical debates around data. One example is the work done by the European Commission on building trusted data spaces (2016, 2020) aiming to encourage actors to explore data's value potential across sectors and industries, while preserving the European laws, rules and regulations. The ideas presented in this paper could help IS researchers and practitioners understand data as simultaneously being governed by rules and regulations, and creating value, instead of treating governance as opposite to value. Furthermore, by focusing on data's relations, instead of data as independent entities, this paper could help practitioners understand data not as finished, fixed products, but as resources whose value needs to be worked on and continuously produced, as data, digital technologies, organizational practices and socio-political contexts are assembled.

References

Aaltonen, A., Alaimo, C., & Kallinikos, J. (2021). The Making of Data Commodities: Data Analytics as an Embedded Process. Journal of Management Information Systems, 38(2), 401–429.

Aaltonen, A., & Tempini, N. (2014). Everything Counts in Large Amounts: A Critical Realist Case Study on Data-Based Production. Journal of Information Technology, 29(1), 97–110.

Aanestad, M., Eike, M. C., Grünfeld, T. B., & Undlien, D. (2014). Building an Infrastructure for Personalized Medicine: Connecting the local knowing in practice to global "circuits of knowledge.". International Conference on Organizational Learning, Knowledge and Capabilities, Oslo, Norway.

Abbasi, A., Sarker, S., Aalto University, & Chiang, R. (2016). Big Data Research in Information Systems: Toward an Inclusive Research Agenda. Journal of the Association for Information Systems, 17(2), 01-32.

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.

Alaimo, C. (2022). From people to objects: the digital transformation of fields. Organization Studies, 43(7), 1091-1114.

Alaimo, C., & Kallinikos, J. (2017). Computing the everyday: Social media as data platforms. The Information Society, 33(4), 175–191.

Alaimo, C., & Kallinikos, J. (2022). Organizations Decentered: Data Objects, Technology and Knowledge. Organization Science, 33(1), 19–37.

Alaimo, C., Kallinikos, J., & Aaltonen, A. (2020). Data and Value. In Handbook of digital innovation. 162–178.

Barrett, M., Oborn, E., & Orlikowski, W. (2016). Creating Value in Online Communities: The Sociomaterial Configuring of Strategy, Platform, and Stakeholder Engagement. Information Systems Research, 27(4), 704–723.

Benfeldt, O. (2017). A Comprehensive Review of Data Governance Literature. Selected Papers of the IRIS, 120–133.

Constantiou, I. D., & Kallinikos, J. (2015). New Games, New Rules: Big Data and the Changing Context of Strategy. Journal of Information Technology, 30(1), 44–57.

Cooper, R. (2005). Peripheral Vision: Relationality. Organization Studies, 26(11), 1689–1710.

Dainton, B. (2016). Time and space. Routledge.

Daskal, J. (2015). The Un-Territoriality of Data. The Yale Law Journal, 326–559.

DeLanda, M. (2000). A Thousand Years of Nonlinear History. Zone Books.

DeLanda, M. (2006). A New Philosophy of Society: Assemblage Theory And Social Complexity Continuum.

DeLanda, M. (2013). Intensive science and virtual philosophy. Bloomsbury Publishing.

DeLanda, M. (2016). Assemblage Theory. Edinburgh University Press.

Deleuze, G. (2014). Difference and Repetition. Bloomsbury Publishing.

Deleuze, G., & Guattari, F. (1987). A Thousand Plateaus: Capitalism and Schizophrenia. University of Minnesota Press.

European Commission. (2016). Regulation EU 2016/679 of the European Parliament and of the Council of 27 April 2016, General Data Protection Regulation. Official Journal of the European Union.

European Commission. (2020). European data strategy. European Commission.

Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. Information and Organization, 28(1), 62–70.

Farjoun, M. (2010). Beyond dualism: Stability and change as a duality. Academy of Management Journal, 35(2), 202–225.

Fossum, K., Fossum, S. M., Hanseth, O., & Sanner, T. A. (2019). Adaptive Networked Governance of E-Health Standards: The Case of a Regional Health Information Infrastructure in Norway. ECIS 2019 Proceedings, 16.

Gregory, R. W., Henfridsson, O., Kaganer, E., & Kyriakou, H. (2021). The role of artificial intelligence and data network effects for creating user value. Academy of management review, 46(3), 534-551.

Grisot, M., Moltubakk Kempton, A., Hagen, L., & Aanestad, M. (2019). Data-work for personalized care: Examining nurses’ practices in remote monitoring of chronic patients. Health Informatics Journal, 25(3), 608–616.

Grønsund, T., & Aanestad, M. (2020). Augmenting the algorithm: Emerging human-in-the-loop work configurations. The Journal of Strategic Information Systems, 29(2), 101614.

Grover, V., & Lyytinen, K. (2015). New State of Play in Information Systems Research: The Push to the Edges. MIS Quarterly, 39(2), 271–296.

Günther, W. A., Rezazade Mehrizi, M. H., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. The Journal of Strategic Information Systems, 26(3), 191–209.

Hanseth, O., & Rodon Modol, J. (2021). The Dynamics of Architecture-Governance Configurations: An Assemblage Theory Approach. Journal of the Association for Information Systems, 22(1), 130–155.

Hassan, N. R., Mingers, J., & Stahl, B. (2018). Philosophy and information systems: Where are we and where should we go? European Journal of Information Systems, 27(3), 263–277.

Hodges, M. (2008). Rethinking time’s arrow: Bergson, Deleuze and the anthropology of time. Anthropological Theory, 8(4), 399–429.

Jarvenpaa, S. L., & Essén, A. (2023). Data sustainability: Data governance in data infrastructures across technological and human generations. Information and Organization, 33(1), 100449.

Kallinikos, J., & Constantiou, I. D. (2015). Big Data Revisited: A Rejoinder. Journal of Information Technology, 30(1), 70–74.

Kempton, A. M. (2022). The digital is different: Emergence and relationality in critical realist research. Information and Organization, 32(2), 100408.

Langley, A., & Tsoukas, H. (2022). The SAGE Handbook of Process Organization Studies.

Leonelli, S., & Tempini, N. (2020). Data Journeys in the Sciences. Springer Nature.

Little, D. (2016). New directions in the philosophy of social science. Rowman & Littlefield.

Lycett, M. (2013). ‘Datafication’: Making sense of (big) data in a complex world. European Journal of Information Systems, 22(4), 381–386.

Østerlie, T., & Monteiro, E. (2020). Digital sand: The becoming of digital representations. Information and Organization, 30(1), 100275.

Paparova, D., Aanestad, M., Vassilakopoulou, P., & Bahus, M. K. (2023). Data governance spaces: The case of a national digital service for personal health data. Information and Organization, 33(1), 100451.

Parmiggiani, E., & Grisot, M. (2020). Data Curation as Governance Practice. Scandinavian Journal of Information Systems, 38.

Parmiggiani, E., Østerlie, T., & Almklov, P. G. (2021). In the Backrooms of Data Science. Journal of the Association for Information Systems, 64.

Pentland, B. T., Mahringer, C. A., Dittrich, K., Feldman, M. S., & Wolf, J. R. (2020). Process Multiplicity and Process Dynamics: Weaving the Space of Possible Paths. Organization Theory, 1(3), 263178772096313.

Peukert, C., Bechtold, S., Batikas, M., & Kretschmer, T. (2022). Regulatory Spillovers and Data Governance: Evidence from the GDPR. Marketing Science, mksc.2021.1339.

Rutzou, T. (2017). Finding Bhaskar in all the wrong places? Causation, process, and structure in Bhaskar and Deleuze. Journal for the Theory of Social Behaviour, 47(4), 402–417.

Rutzou, T., & Elder-Vass, D. (2019). On Assemblages and Things: Fluidity, Stability, Causation Stories, and Formation Stories. Sociological Theory, 37(4), 401–424.

Schulze, L., Trenz, M., Cai, Z., & Tan, C.-W. (2022). Algorithmic Unfairness on Digital Labor Platforms: How Algorithmic Management Practices Disadvantage Workers.

Smith, B. (2003). Philosophical Ontology. In Blackwell Guide to the Philosophy of Computing and Information.

Tarafdar, M., & Kajal Ray, D. (2021). Role of Social Media in Social Protest Cycles: A Sociomaterial Examination. Information Systems Research, 32(3), 1066–1090.

Tempini, N. (2017). Till data do us part: Understanding data-based value creation in data-intensive infrastructures. Information and Organization, 27(4), 191–210.

Tempini, N. (2021). Data curation-research: Practices of data standardization and exploration in a precision medicine database. New Genetics and Society, 40(1), 73–94.

Tuomi, I. (1999). Data Is More than Knowledge: Implications of the Reversed Knowledge Hierarchy for Knowledge Management and Organizational Memory. Journal of Management Information Systems, Vol. 16(3), 103–117.

Van den Broek, T., & Van Veenstra, A. F. (2015). Modes of Governance in Inter-Organizational Data Collaborations. ECIS 2015 Proceedings, 13.

Viljoen, S. (2021). A Relational Theory of Data Governance. The Yale Law Journal, 82.

Winter, J. S., & Davidson, E. J. (2020). Harmonizing Regulatory Spheres to Overcome Challenges for Governance of Patient-Generated Health Data in the Age of Artificial Intelligence and Big Data. TPRC48: The 48th Research Conference on Communication, Information and Internet Policy.

Woerner, S. L., & Wixom, B. H. (2015). Big Data: Extending the Business Strategy Toolbox. Journal of Information Technology, 30(1), 60–62.

Zuboff, S. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power (1st edition). PublicAffairs.

APPENDIX III:

“Data Hierarchies:
The Emergence of an Industrial Data Ecosystem”

Daniel Stedjan Svendsrud, Dragana Paparova

Abstract

In this paper, we empirically investigate the emergence of an industrial data ecosystem in the heavy-asset oil and gas industry in Norway, as a business-to-business data platform gets introduced. For that purpose, we build on the notion of complementarities, as assets, activities, capabilities, software, data, recombined to co-create value across ecosystem actors. Our specific focus lies in exploring the role of data complementarities in ecosystem emergence – referred to as data ecosystems. We define data ecosystems as alignment structures of interconnected, but autonomous actors, interacting around complementary data objects to materialize individual and focal value propositions. Our findings offer two contributions. First, we show how industrial data ecosystems emerge as actor and data complementarities restructure existing actor relations. Second, we conceptualize data hierarchies as specific types of data complementarities which correspond to the physical reality of heavy-asset industries, and specific to the oil and gas context.

Keywords: data complementarities, data ecosystem, data hierarchies

1 Introduction

In information systems (IS) studies, digital ecosystems have been commonly construed as a platform ecosystem. Scholars have investigated the evolution of complementary third party apps surrounding core platform technologies, such as social media platforms (Claussen et al., 2013), the ecosystem-based capabilities of platform owners and third parties (Schreieck et al., 2021; Selander et al., 2013), and their governance and architectural set ups (Eaton et al., 2015; Ghazawneh & Henfridsson, 2013; Wareham et al., 2014). While a broadly shared definition does not exist, the notion of a digital ecosystem largely echoes the general ecosystem concept (Autio & Thomas, 2020; Sawy et al., 2010; Selander et al., 2013; Yoo et al., 2010) as an interdisciplinary phenomenon spanning technological and organizational boundaries (Adner, 2017; Autio & Thomas, 2020; Jacobides et al., 2018; Thomas & Autio, 2019).

The cornerstone of the ecosystem concept can be placed around the nature of complementarities; how heterogeneous components – assets, activities, capabilities, software or economic items – when combined constitute an ecosystem. Related research streams in IS have been making similar claims. For instance, the literature on digital innovation contends how digital ecosystems emerge when digital and physical components are recombined to produce novel outputs (Nambisan et al., 2017; Yoo et al., 2010; Yoo et al., 2012). The ecosystem literature argues that what makes ecosystems specific in comparison to other organizational arrangements is the capacity to provide generative innovation (Zittrain, 2008) afforded by digital technologies (Thomas & Autio, 2019). As such, digital components enable the formation of digital ecosystems by being used as shared complementary resources by interconnected actors who recombine them to cocreate value (Henfridsson et al., 2018). Although it is generally accepted how ecosystems emerge when actors form complementarities as they align around a shared structure (Adner, 2017; Ansari et al., 2016; Moore, 1993), empirical findings on ecosystem emergence are still lacking (Jacobides et al., 2018; Thomas & Autio, 2019), particularly in exploring the role of data in ecosystem emergence.

The production, distribution and consumption of data offerings have transformed various industries into data economies. In the business-to-consumer context, where e.g., Facebook sells user profiles to marketing firms, this is not a novel phenomenon. However, traditional industrial business-to-business (B2B) markets are becoming data-driven as well; traditional products get embedded with sensors producing data e.g., about products' performance. The central role of data in these changing industries makes it necessary to understand how different actors' capabilities, technologies and goals are purposefully recombined around data objects (Alaimo & Kallinikos, 2022; Alaimo, Kallinikos, & Aaltonen, 2020; Alaimo, Kallinikos, & Valderrama, 2020). While there is some empirical evidence of the formation of data service ecosystems (Alaimo, Kallinikos, & Valderrama, 2020), and the conceptual role of data resources in digital ecosystems (Alaimo & Kallinikos, 2022; Alaimo, Kallinikos, & Aaltonen, 2020), there is little empirical insight into how an industrial data ecosystem emerges around data complementarities, especially in traditional product-based industries.

In this paper we investigate how a data ecosystem emerges in oil and gas heavy-asset industry. Our research question is: how do data complementarities unfold in a data ecosystem in the context of heavy-asset industries? For that purpose, we empirically follow the changes in industrial actor relations in the oil and gas industry in Norway, as a B2B data platform gets introduced. We contribute to the literature on ecosystems by: 1) showing how industrial data ecosystems emerge as actor and data complementarities restructure existing actor relations; and 2) conceptualizing data hierarchies as a specific types of data complementarities representative of our industrial context.

The paper precedes as follows. In the ensuing section, we present our conceptual background, namely the ecosystem concept and the notion of complementarities. Then, we present the research approach and a case description. In section four, we present three coexisting phases of ecosystem emergence in our industrial context; followed by a case analysis in section five. In section six we discuss the main contributions of the paper and conclude with suggestions for future research.

2 Conceptual Background: Digital Ecosystems and Complementarities

2.1 The Ecosystem Concept

The ecosystem concept originates from the field of strategic management, initially introduced as a metaphor for how firms can be embedded in a business environment transcending several industries and the subsequent strategic implications for a focal firm (Moore, 1993). While being part of both practical and scientific discourses for decades, the ecosystem phenomenon was only recently theorized into a coherent interdisciplinary concept (Adner, 2017; Jacobides et al., 2018; Thomas & Autio, 2019).

Jacobides et al. (2018) define an ecosystem as "a set of actors with varying degrees of multilateral, non-generic complementarities that are not fully hierarchically controlled". Central to this definition is that an ecosystem is an economic system with a production and consumption side characterized by different forms of complementarities. Complementarities, originating from the economic sciences, are a systemic concept denoting how heterogeneous components yield larger economic value when combined, as opposed to being kept separate (Katz & Shapiro, 1994). What defines an ecosystem is that complementarities exist both on the production and the consumption side (Jacobides et al., 2018). On the production side complementarities take the form of being co-specialized (Teece, 1986) among hierarchically independent actors. A complement in this instance is when two or more individual business firms are dependent on coordinating their actions, assets and investments to create value out of an offering. For example, when an app developer uses the application programming interfaces provided by Apple to develop mobile applications, the value of Apple's offering increases, resulting in the complementary offering of an iPhone. According to, on the consumption side complementarities are specific and non-generic as they are being assembled by the end-user; in our case, the user of an iPhone. These complementarities form a larger system where the different actors coordinate around recombining components to create value.

Adner (2017, p. 42) goes more in-depth on the role of actors in creating and maintaining complementarities, by defining an ecosystem as "the alignment structure of the multilateral set of partners that need to interact for a focal value proposition to materialize". Here, complementarities constitute a particular and holistic actor configuration with defined value creating roles and activities. According to Adner (2017), an ecosystem becomes evident once there is a significant change in the activities of the actor configuration stemming from a novel value proposition. For instance Ansari et al. (2016) uncovered how the established actor complementarities in the U.S. television ecosystem was disrupted by the technological innovation of the digital video recorder. The introduction of digital video recorder, championed by the technology firm TiVo, brought a completely novel value proposition in terms of TV-watching that changed the existing actor relations significantly; firms now complemented each other on software as opposed to hardware. Moreover, several other studies in organization and management science have studied how complementarities are made among heterogenous actors in the intent of creating an offering (Adner & Kapoor, 2010; Dattée et al., 2018; Gawer & Henderson, 2007; Hannah & Eisenhardt, 2018).

2.2 Digital (Data) Ecosystems

The notion of complementarities is also present in IS research, albeit labeled under a different concept and mostly developed within the digital innovation literature. In digital innovation, complementarities are often referred to as "digital resources"; the digital technologies that take part in the creation and capture of informational value (Henfridsson et al., 2018). Placed under an ecosystem lens, digital resources generate two sociotechnical complementarities in need of explanation: 1) complementarity at a technological level; and 2) complementarity at a data level.

2.2.1 Complementarity at a Technological Level

Digital technologies have a set of unique properties; they are reprogrammable, have the ability to homogenize data and have a self-referential nature (Yoo et al., 2010). Due to these properties, when digital components are recombined in a layered modular architecture, they afford the generative innovation potential of a digital ecosystem. When physical products are embodied with modular architectures, such as cars, the complementarities consist of physical modules where each module performs a specific function. However, when physical products get digitized, the form decouples from the function; digital components organized around modular architectures can be recombined in a variety of ways to perform different sets of functions; not strictly bound by their physical materiality. Therefore, digital ecosystems are not product-specific, but product-agnostic as they can perform various functions and be recombined in a variety of ways.

The bulk of IS research has relied on the fundamental assumption about the generativity of digital ecosystems when studying platform ecosystems (Autio &

Thomas, 2020). Several works have shown how platform owners enable generative innovation potential by providing boundary resources and incentive mechanisms to third-party developers to contribute with complementary applications (Ghazawneh & Henfridsson, 2013; Wareham et al., 2014).

2.2.2 Complementarity at a Data Level

A line of research has put forward the idea that data have a unique set of properties making data a distinctive form of digital resource separate from digital technologies; thus, in need of being analyzed on its own terms (Aaltonen et al., 2021; Alaimo & Kallinikos, 2022; Alaimo, Kallinikos, & Aaltonen, 2020). According to these works, data are editable (can be re-aggregated, filtered, re-ordered), portable (across settings, platforms, organizations) and recontextualizable (can acquire meaning beyond their origin) (Alaimo, Kallinikos, & Aaltonen, 2020). While data are a resource that can be combined with other data and made into a data-based product or service, once organized into an architecture, they do not embody functions as software does. Data are likewise a medium and carrier of facts, sign tokens and meaning-making, as they are made within a particular sociotechnical context. Therefore, value creation from data should not be understood solely as recombinant innovation of digital components, but as a process of discovery, learning and knowledge-making (Alaimo, Kallinikos, & Aaltonen, 2020).

This requires a rethinking of a digital ecosystem underpinned by data complementarities. The innovation potential of data comes from data as sign tokens being de-coupled from the reality they are supposed to represent and refer to. The user in a data ecosystem does not just combine and recombine offerings; the user recontextualizes and reinterprets the complementarities that make up data offerings. As one data item (datum) is assembled with another data item, the result is an emergent data object (Alaimo & Kallinikos, 2022)as opposed to a mere aggregate. Every encoding and de-coupling that an agent does along a data value chain is novel as the new data object depends on the situational knowledge and meaning-making predispositions of the agent in question (Alaimo & Kallinikos, 2022; Alaimo, Kallinikos, & Aaltonen, 2020) In their study of the evolutionary process of TripAdvisor from a travel search engine to a central hub in a digital travel service ecosystem, Alaimo, Kallinikos and Valderrama (2020) show how the ecosystem formation was based on the production and use of various types of data that were deemed as complementary by the actors involved.

2.3 Synopsis

Research on data ecosystems – and the central role of data – is still scarce. Data are commonly reified as technological components, as part of digital ecosystems. Digital ecosystems are characterized by their technological nature and composition of complementarities; a divide can also be made between digital ecosystems and non-digital ecosystems. Therefore, defining digital ecosystems strictly according to how complementarities take the form in an architecture does not sufficiently account for the inherent dynamic in an ecosystem connected to actor's motives and decisions (Nambisan, 2018). In fact, Nambisan (2018) recommends a coupling ecosystem studies (Adner, 2017; Jacobides et al., 2018), and digital innovation research (Henfridsson et al., 2018; Yoo et al., 2010) by emphasizing actors' agency in recombining digital resources.

In this paper, we also argue for actors' agency as being crucial to the notion of data complementarities. By using the term data ecosystems, we seek to conceptualize the distinctiveness of creating data complementarities in digital ecosystems. In data ecosystems, data are produced, used and shared across heterogeneous actors with interconnected, yet autonomous goals and interests. Thus, in line with the existing literature (Alaimo, Kallinikos, & Valderrama, 2020; S. Oliveira et al., 2019) we define data ecosystems as alignment structures of interconnected, but autonomous actors, interacting around complementary data objects to materialize individual and focal value propositions.

3 Research Approach

3.1 Case Description

Digitize was established in January 2017 by Indus – an oil and gas (O&G) company, as part of the Indus ASA consortium. Indus ASA is an industrial investment company with ownership interests concentrated in heavy-asset industries, such as power and utilities, renewable energy, seafood and marine biotechnology. Indus established Digitize as an industrial software company aiming to facilitate sharing of data generated by sensors installed on physical equipment on O&G platforms. Although Digitize is an independent company, as of March 2023 Indus ASA holds 50,5 percent of its shares and is Digitize's main shareholder.

Traditionally, the O&G industry was characterized by various older physical installations, where information about the equipment and its maintenance were commonly not recorded, or transformed into digital data. Data were at times "memorized" by fieldworkers, or stored on a piece of paper that would be saved in a folder or thrown away if the equipment seems to be working properly. Some physical equipment on the O&G platforms, such as pumps, or oil extractors, had dedicated IT systems connected to their sensors, but were predominantly used for storing, and not sharing data about the equipment. Therefore, data in O&G were siloed across various IT applications, suppliers and operators, which created a complex landscape of standalone applications used in the daily operations of industrial organizations.

The main product of Digitize is a software platform, hereafter referred to as DigitizePlatform, working as a core functionality that can enable data sharing by: 1) extracting data from the siloed legacy systems and copying it in the cloud; 2) creating meaningful connections between the data using data models – data contextualization; and 3) providing IT applications upon which the value from data can be realized in a business context. The O&G industry is characterized by sensor data connected to physical assets, creating large and continuous streams. DigitizePlatform can perform various analysis upon these industrial data, including: 1) timeseries, by providing real-time and historical data about sensors; 2) maintenance data, such as capturing events of maintenance incidents and predict future behavior; 3) digital twins, or digital representations of physical equipment; 4) process diagrams, including pipeline and instrumentation; 5) granular data sources related to equipment, such as documents, 3D models and images; 6) as well as apps which improve the workflow of fieldworkers and operators of e.g. O&G platforms.

Initially, DigitizePlatform was used only internally by Indus, but over time it got adopted by Indus's existing alliance partners, and other strategic partners within and across the O&G industry. This resulted in the emergence of an industrial data ecosystem with restructured actor relations. The unfolding of this industrial data ecosystem is the main empirical focus of this paper.

3.2 Research Design

We conducted a qualitative (Alvesson & Sköldberg, 2017) interpretive case study (Walsham, 1995; Walsham, 2006) gain an in-depth understanding of the phenomenon of data ecosystems. The research design, including the research questions were not fixed up-front, but defined progressively, as the insights gained from the empirical material were confronted with theory (Alvesson & Sköldberg, 2017). Due to this abductive nature of this study, we used a set of pre-defined concepts, such as: digital platforms, digital ecosystems, data-driven value creation to guide the collection and analysis of empirical material.

3.3 Collection of Empirical Material

The initial gathering of empirical material started by conducting 29 semi-structured interviews. The interviews were used as a tool to capture the participants' perceptions, understandings, opinions about the data ecosystem around Digitize. The participants held managerial or technical roles in the organizations, such as architects, managers, product owners, project managers. The interview guides were adjusted according to the background, job position, and the informants' role. The interviews elicited the interviewees' accounts of the need for sharing industrial data, DigitizePlatform's functionalities, the development of ecosystems, strategic partnerships and collaboration with various industry partners. The sources are summarized in Table 1.

Empirical material	Amount	Duration	Description
Interviews	29	Approx. 1h	Participants: architects, managers at various organizational levels, product owners, project managers…
Documents	5	/	Smart contract between Indus and Alpha, company presentations.

Podcasts	2	Approx. 1h	Trends in the oil and gas industry, such as Internet-of-Things and the value from sharing industrial data
Press releases	31	/	Events related to the company and technology development

Table 1. Summary of the gathering of empirical material.

3.4 Analysis of Empirical Material

We conducted a process analysis (Langley & Tsoukas, 2016) where we minimized the longitudinal data into a sequence of events (Cornelissen, 2017) in three steps (Langley & Tsoukas, 2016): 1) identify the events of the sequence; 2) identify the relations that cluster those events; 3) characterize the pattern across which these events unfold.

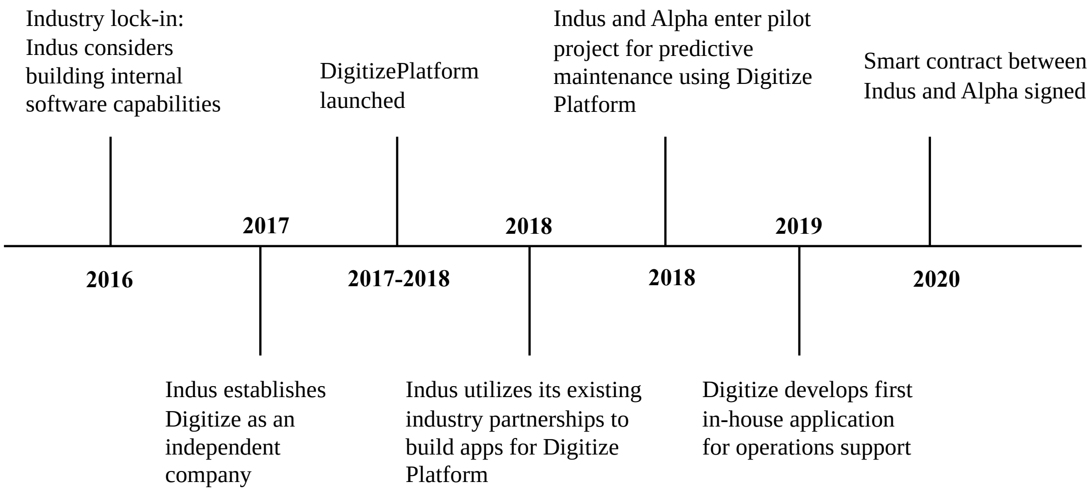


Figure 1: Timeline with significant events in the emergence of the industrial data ecosystem.

First, we inductively defined 7 events from the empirical material that we considered significant for the emergence of the data ecosystem and ordered them in a timeline, as represented in Figure 1. We bounded our analysis to the establishment of, and evolution of Digitize, as our primary object of study. We then consulted the digital ecosystems and data-driven value creation literatures to understand the relations across these events. We identified three coexisting phases across which the data ecosystem unfolded, as represented in the findings: 1) building data complementarities: the need for recombining industrial data; 2) building actor complementarities: aligning specialized and general needs; and 3) restructuring industrial relations through data hierarchies. We realized that our empirical case uncovers insights about industrial data ecosystems which did not correspond fully to the existing digital ecosystems and data ecosystems literature. For instance, beyond raising the need for (data) complementarities, existing research did not account for the changes in actor relations while engaging in complementary data relations. Moreover, there was limited understanding of the specifics of industrial (data) ecosystems and the specialized context they operate in. To account for these novel findings, we analyzed two processes across which industrial data ecosystems unfold: 1) the making of data hierarchies; 2) the shaping of industrial actor relations.

4 Findings: The Emergence of an Industrial Data Ecosystem

4.1 Phase 1: Building Data Complementarities: The Need for Recombining Industrial Data

As of 2016, many operators in the O&G industry were locked in by the suppliers of physical equipment. For instance, the compressors which were installed on oil plants or rigs came with certain IT applications. Operators had to buy both, the physical equipment, and the IT applications, and pay a license to be able to extract data about the equipment's performance through various proprietary application programming interfaces (APIs). Moreover, the suppliers and operators would enter annual contracts pre-determining the frequency at which consultants arranged by the suppliers would check the state of the equipment. Therefore, the equipment, IT applications and maintenance services were provided the suppliers, and sharing data about the equipment was not possible unless all services were purchased by the same supplier.

Indus realized the need to share data across the various equipment and IT applications used on their O&G platforms. Sharing data was perceived to automate the maintenance of physical equipment, but also help companies operating in the industry make better data-driven decisions. For instance, Beta's (pseudonym) equipment used to break by vibrations of equipment from other suppliers installed next to it. However, Beta only had access to, and shared data about their own equipment, but could not detect vibrations from other supplier equipment before they occur. Sharing data could help e.g. with preventing such damages of physical components, but also inform maintenance services by the actual state of equipment, instead of by consultants' calendars. However, as of then, there was no existing solution on the software market which could facilitate data-sharing across IT applications provided by different suppliers and operators.

"We get compressor from Beta, we install it on the plant. And they say: 'OK, if you want to use my application – my code, to really understand how we should operate the compressor, you also need to buy the whole ecosystem that we sell you. All the tools, the user interface the software itself. And we say 'no, we only want the code, we only want the machine learning, we only want to know how to detect the problems on the compressor' – then the answer was 'we are never going to work that way, we are Beta and then you need to buy everything from us if you want'." (Indus Digitalization Manager)

As of 2016, Indus started exploring the option of building internal software development capabilities. However, developing an internal software platform was assessed to be expensive, and challenging when attracting IT talent. For those, among other reasons, in 2017, Indus established Digitize as an independent software company. The main product of Digitize was an industrial data platform – DigitizePlatform – which enabled Indus to share operational data about the physical components installed on their oil and gas (O&G) platforms. Indus shared various data types related to their operations, such as timeseries, documents, events, which were then contextualized in DigitizePlatform.

> "We had ambition to, 'OK we need to share data', because we realized that there was no common platform across the entire Exploration and Production organization; we decided to found and finance Digitize to create DigitizePlatform" (Indus Digitalization Manager)

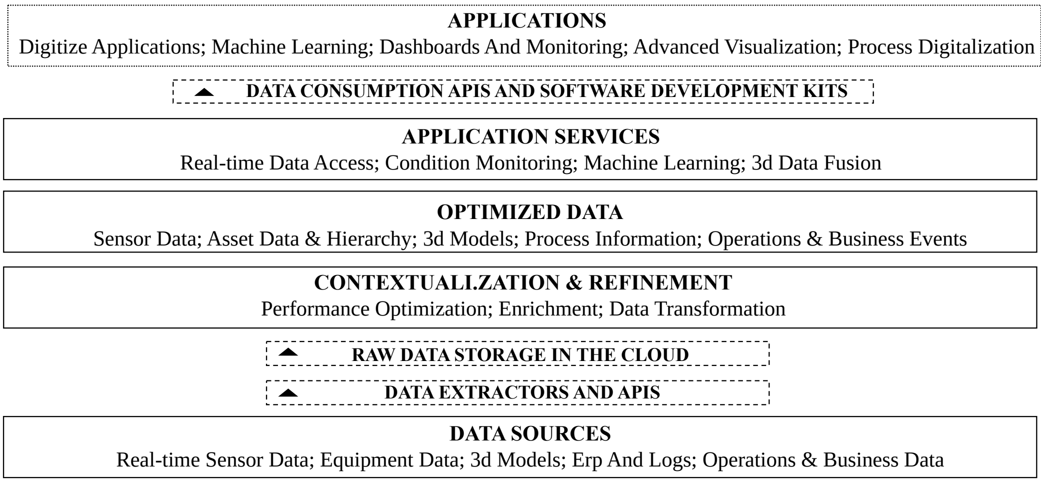


Figure 2: DigitizePlatform architecture

The architecture of DigitizePlatform extracted data from the existing data sources, such as pumps on O&G platforms, and copied these data in the cloud. Once copied, data got categorized using data models and connected in meaningful ways using algorithms and machine learning (ML) technologies. The data models could show real-time and historical view of the behavior of individual data assets, regardless of the applications used in the backend. DigitizePlatform also functioned as core on top of which various applications can be built, such as visualizations, dashboards and reporting used by organizations to make sense of the data. For instance, maintenance workers could access data about the performance of equipment and record structured data. These data could be re-used and visualized as timeseries in various front-end applications. The platform architecture of DigitizePlatform is represented in Figure 1.

Sharing data through DigitizePlatform provided new opportunities for Indus as an industrial operator. First, while there was a lot of data in the O&G industry, some were not repeatable enough to train a model that can predict future behavior. For instance, a pump that has been running for 30 years may have been broken two times, but due to different causes. In this case, there was not enough data to make predictive maintenance analysis, because the pump had been destroyed a few times. By facilitating the sharing of data, and getting access to data from various source IT systems, suppliers and operators, Digitize could train their algorithms and predict the behavior of physical components using historical or real-time data from different data sources and create value that isolated actors could not achieve on their own.

> "Often when something is destroyed it is because something next to it vibrates. Beta's equipment vibrates and vibrates. Then it becomes good again. Machine learning models will think it is good that it vibrates because it becomes good again afterwards. We, on the other hand, know that the reason why the equipment stopped vibrating was because maintenance was done and a part of Beta's equipment was switched. By combining different data sources [maintenance data with sensor-data from Beta's equipment], the more the probability is that one can do predictive maintenance of equipment." (Digitize Partner & Alliances Director)

Second, before Digitize, data from physical components in the O&G industry was lacking structure. For instance, in a traditional maintenance process, the equipment performance would be checked every 12 months, and if something does not function as intended, it would be recorded as an error. However, the recorded data were unstructured, as each installation was based on different reading values of the equipment indicating whether the performance is within the minimum and maximum value. Data were stored in different formats, categories, and using multiple identifiers to refer to the same equipment. The architecture of DigitizePlatform analytically decomposed data from the source systems by providing them structure which keeps data about each asset as raw and separate, while making meaningful connections.

At last, data from the physical components were sometimes owned by the operators, other times owned by the suppliers of IT systems. To resemble such ownership, Digitize kept data about their customers architecturally segregated when copying data into the cloud. Therefore, Digitize allowed customers to share data with each-other, and only used customers' data for testing and improving the DigitizePlatform technology. For instance, Indus could give Beta access to all data about equipment surrounding theirs – Beta could analyze these data and find the cause of equipment breakdown. Moreover, Beta could give Indus maintenance information, showing who has done maintenance on the equipment, and what errors have occurred. Therefore, Digitize provided the technical components for data sharing, but the data continued to be owned by its customers. Customers could control what data they share, who has access to it, and what they can do with it, using the APIs.

4.2 Phase 2: Building Actor Complementarities: Aligning Specialized and General Needs

Indus's plan was to use DigitizePlatform as an industrial software platform that diversifies their overall business, but also spans across various industries. However, at first, Digitize was predominantly used to digitalize Indus's internal business operations. Initially, DigitizePlatform was a product composed of a set of APIs with no user interface; this made it difficult to comprehend what DigitizePlatform is, and how it can be used. Indus realized that if Digitize was to become an industrial platform used outside of Indus, it had to cover a more generic need. However, the O&G industry was highly specialized; operators, suppliers, fieldworkers, IT vendors had diversified needs. Indus decided that beyond the ability to share data, DigitizePlatform should also showcase how these stakeholders can create direct value from the sharing of data. This required developing data products as front-end applications that can be consumed directly by customers.

Digitize was unable to cover the large variety of requirements in the industry by building apps internally; they decided to provide APIs and software development kits (SDKs) so that third parties can build specialized apps externally. However, the process of attracting third parties moved slower than expected; this was due to various reasons. First, O&G industry actors were not used to developing apps themselves, but purchased software-as-a-service where all apps are available upfront. Second, the industry was immature in terms using AI and ML to make value from data. Third, IT vendors perceived Digitize as their competitor in developing IT applications; Digitize was also perceived as having a competitive advantage to develop algorithms, as they had access to a large data set to train their algorithms on. This brought in the need for Digitize to consider developing some applications in-house.

In 2019, Digitize launched the first in-house application built on top of DigitizePlatform offering operation support to fieldworkers. This app gave fieldworkers access to data on a handheld device, making it possible to search for equipment using 3D models, scan or tag equipment on the field using pictures and videos, upload and view documents related to that equipment, as well as connect onshore using video calls. Fieldworkers could get historical and real-time data where they could see if oil needs to be changed, or what type and amount of oil goes with which equipment. These data were existing in the O&G industry for many years, but stored and viewed across many systems – with the use of Digitize, they were structured into one application. This enabled fieldworkers to maintain the equipment more effectively; data generated by one asset, e.g. a pump on an O&G platform could be viewed in context with other assets and equipment.

However, Digitize continued their attempts in attracting application developers externally. For that purpose, they decided to utilize Indus's existing alliances and strategic partnerships. At that time, Indus was engaged in six different alliance models, and eight different strategic partnerships, together with a selected set of supply-chain partners. Within a year and a half, Indus managed to attract various large suppliers; Beta developed applications for monitoring the conditions of the assets using dashboards and reporting. Gamma provided applications for ML and advanced analytics; Delta was providing 3D and advanced visualization techniques; the key ingredient was access to DigitizePlatform's data.

In general, IT vendors that had a strong connection to Indus shared data with Digitize; companies that had a stronger independent position on the market refused to give access to their data. In contrast to application developers who perceived Digitize as a competitor, O&G operators had larger incentives to share data with DigitizePlatform, as they wanted to get better applications and improve their business operations, e.g. learning how to maintain their turbines in a better way. Therefore, although Digitize attracted strategic partners, they still had to develop some applications in-house. For instance, customers needed an app to do a root-cause analysis by comparing different timeseries – why something happened and what can be learned about it in the future. For that purpose, Digitize developed a charting application, which allows fieldworkers to search for documents, get a graph to discover when the pump stopped functioning and analyze this in context with other parameters.

Therefore, the primary focus of Digitize was facilitating data-sharing in the industry – not compete with IT vendors in developing applications. If there was a general need on the market, Digitize developed the app themselves. If there was an existing solution on the market, Digitize aimed to integrate with it, Moreover, Digitize utilized the expertise of existing suppliers by treating them as implementation partners when integrating an app with DigitizePlatform – this allowed them to execute the integrations faster by utilizing the domain expertise of suppliers. Therefore, Digitize aimed to create a collaboration arena among operators, suppliers, IT vendors, consultancy services, instead of competing with them on the service or application levels.

> "Digtize plays a facilitator role [in an ecosystem]. Not only are customers able to share the data, but also share the insight so they get a total overview of the equipment. And since we are a neutral part with other incentives than equipment-suppliers it is easier for us to take a facilitator role." (Digitize manager)

However, due to the highly specialized industry it operated in, Digitize had to continuously balance between providing generic products that can be utilized by various customers, and providing products that fit specific customer needs. For instance, the app for operation support was developed for Indus, but could be bought by other customers that needs the same functionality. However, Digitize also developed custom-made applications intended to cover Indus' unique needs, which applications could not be re-used by another O&G operator. As of 2020, Digitize had around 5 in-house applications developed for all customers, and around 40 custom-made solutions. Moreover, there were around ten applications developed by independent software providers that were integrated with DigitizePlatform, e.g. a dashboard developed by Beta and used by Indus.

4.3 Phase 3: Re-structuring Industrial Relations Through Data Hierarchies

Over time, Digitize realized that more comprehensive digital representations of physical assets, such as digital twins, could provide the possibilities to test physical assets digitally, avoid unnecessary equipment check-up by engineers, and lower the costs and margins of human errors. For that purpose, they needed more comprehensive data about the physical assets and started scanning installations using drones and robots for data extraction. Gradually, they could map the connection between separate equipment and sensors into "asset hierarchies", showing how different assets are connected to each-other, e.g. platform–turbine–turbine parts. Therefore, Digitize coupled timeseries data with 3D models showing the physical location of the equipment, creating relationships among the assets, as well as recording documents coupled to that equipment – building digital twins of industrial assets. The digital twin is illustrated in Figure 3.

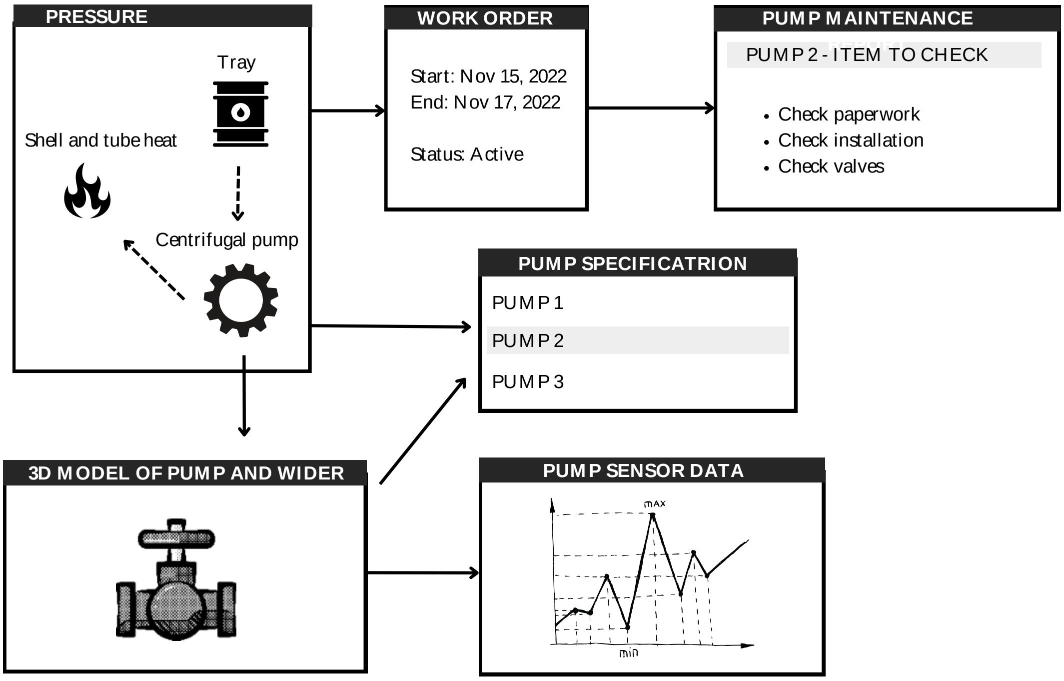


Figure 3: Illustration of an industrial digital twin of a pump, showing sensor timeseries data, 3D models, specifications, maintenance reports.

The most significant resource for developing digital twins was data. The digital twin was essentially a data model, or a knowledge graph, of all data about a given asset. However, these data could be utilized by operators or suppliers in various way. Some customers used the digital twin internally for maintenance and optimization of the products' performance, such as represent pumps installed on the O&G platforms in geometry, shape, color, see operational data and maintenance status – how much the pumps vibrate, how hot they are, the last time maintenance was done. That way, they would get historical and real-time data about how their equipment performs, and how it is expected to perform over time.

However, some customers used the digital twin to share data about their physical assets with other operators or suppliers. For instance, Indus's digital twin contained operational data about all physical equipment installed in their O&G platform. Using DigitizePlatform, Indus could authorize suppliers to get access to operational data about the specific equipment they are delivering to monitor their conditions, or understand their performance in context to other equipment. This allowed Indus to enter a pilot collaboration on predictive maintenance where they shared data with a pump supplier using DigitizePlatform. Traditionally, the pump supplier Alpha owned the test data and the domain expertise, provided support during the installation of the pumps and a calendar-scheduled or on-call services; however, they did not have data about how the pump operates on the field. Indus as an operator of the O&G platform owned the operational data about the pump, such as temperature, pressure, flow, but they could not predict the future behavior of the pump, calculate equipment breakdown ahead of time and schedule the appropriate intervention. As of 2018, Indus could give Alpha access to a digital twin of their pump, based on which Alpha could support Indus in its pump operation and maintenance; moreover, using live operational data, Alpha could improve their domain expertise.

As of 2020, the pilot was lifted into a new maintenance contract based on algorithms and sensor data – smart contract. The contract aimed to maximize uptime, optimize pump performance and avoid breakdown by reducing calendar-driven maintenance. As part of this contract, Indus started operating each pump according to commonly agreed range of indicators with Alpha, and could perform basic maintenance, such as replacement of filters and monitoring of oil levels. Alpha was expected to use the conditional data on their equipment plus the operational data supplied by Indus to reinforce their ML models and strengthen their algorithms for predictive maintenance. Based on domain knowledge and expertise, Alpha was expected to advise Indus how to improve equipment performance by changing the way they operate the pumps, such as when to change the filters.

> "Our [Digitize's] contribution is to be the data bridge between Alpha and Indus, so that Alpha can get real-time data from their pumps and sit in their offices and understand how the pump operates and consider whether or not to go out [on the offshore platform] and perform an activity or not. And the result of this is that Alpha is much less frequently out on the offshore platforms and the number of operations per pump has decreased significantly. That has enabled Alpha to get more insight into how the pumps function in the real life. Instead of just working on the basis of test

data, one has data from actual operations so the suppliers can build more knowledge on the state of the equipment." (Digitize manager)

The contract was signed for a duration of six years, with an option for additional six years. The parameters were set around: diagnosis, troubleshooting, service, parts and total maintenance for Alpha's seawater lift pumps; non-foreseen breakdowns were not addressed under this contract. The price was fixed, and based on performance attached to a list of defined key performance indicators. For instance, there was a bonus related to increasing accuracy of Alpha's algorithm. Therefore, by sharing data using DigitizePlatform, the industrial relation between Indus and Alpha changed.

5 Analysis: Towards The Emergence of an Industrial Data Ecosystem

Our findings show how building data and actor complementarities required constant renegotiations. To explore how these complementarities influenced one another, we analyzed our case through the unfolding of two processes: 1) the making of data hierarchies; and 2) the shaping of industrial actor relations.

5.1The Making of Data Hierarchies

Our case shows how the making of data complementarities was shaped by the physical industrial reality and actor relations. In the physical reality, assets were organized hierarchically on an O&G platform, e.g. turbine parts – turbine – platform. Data about these assets were generated by sensors using electronic signals to turn events from the physical reality – pressure, temperature, vibrations of physical assets – into digital data points.

The DigitizePlatform extracted these data, and performed various analysis on them, including: 1) timeseries data, such as real-time and historical data about the asset's performance; and 2) work-process data such as user manual documents and process diagrams used by field workers. By combining these data sources, Digitize generated a first-order data hierarchy of a specific asset. This data object could represent the individual equipment on the platform, such as a pump.

Further on, these data could be combined with: 1) 3D models about the asset, 2) data generated from maintenance processes, such as events showing equipment shutdown; and 3) electronic resource planning systems storing operational data about the organizational context surrounding the asset. By combining these data objects, DigitizePlatform created a second-order data hierarchy of 3D assets and their organizational context. This data object could represent e.g. the pump as component part of the O&G platform.

Finally, the 3D assets could be combined with data originating from outside the organizational context, such as weather data, satellite images and navigation data. This created a third-order data hierarchy as a digital twin about the asset and its wider context. This data object could e.g. represent the pump, the O&G platform, and the wider organizational and environmental context they operate in.

The making of data hierarchies shows how the industrial data ecosystem did not simply emerge by sharing data from heterogeneous sources or creating end-user applications to showcase value from such data. Instead, data were ordered hierarchically in a way that resembles the industrial reality they operate in.

By ordering data hierarchically according to physical assets, Digitize could create data complementarities not only by recombining heterogeneous data sources from various actors, but also by representing existing industrial relations. In the industrial reality, a set of actors supplied equipment, others operated the O&G platforms, provided IT services, consultancy services for implementation etc. Therefore, actors' physical assets were organized hierarchically – affecting actors' collaborations – bringing the need for ordering data hierarchically.

5.2The Shaping of Industrial Actor Relations

Our case shows how the industrial data platform was provided by Digitize; however, various actors played different roles while collaborating, cooperating and competing for the industrial data ecosystem to emerge. For instance, the DigitizePlatform orchestrated the sharing of data across industrial actors; Indus orchestrated the industrial actors, such as operators, suppliers, IT vendors around the DigitizePlatform; the IT vendors provided end-user applications.

As various industrial actors were orchestrated around the DigitizePlatform, new industrial relations emerged. For instance, Indus as an O&G platform operator and Alpha as a supplier started sharing data about the asset they collaborate around – the pump. The complementarity of operational data and test data about this asset led to the emergence of a new smart contract between the two actors. The basis for this contract was data about physical assets, extracted, contextualized and shared by Digitize. The smart contract transformed how the industrial actors collaborate when it comes to predictive maintenance and optimized asset performance, which no longer was calendar-driven, but data-driven. Therefore, as industrial actors organized themselves around the DigitizePlatform, the industrial reality in which they operate changed.

6 Discussion

This paper seeks to answer the following research question: "how do data complementarities unfold in a data ecosystem in the context of heavy-asset industries?". We answer this question in two ways: 1) by showing how data ecosystems emerge through actor and data complementarities; and 2) by conceptualizing data hierarchies as industry-specific data complementarities.

6.1 The Emergence of an Industrial Data Ecosystems Through Actor and Data Complementarities

The ecosystem literature posits that an ecosystem emerges when complementarities are formed (Adner, 2017; Adner & Kapoor, 2010; Jacobides et al., 2018; Thomas & Autio, 2019). Moreover, Adner (2017, p. 44) argues that "in mature industries, much of the ecosystem is latent most of the time", stating that in an established industrial setting there is a structural configuration of actors oriented towards the materialization of a specific value proposition. As heterogeneous actors attempt to turn technological innovations into customizable market offerings, they can also disrupt the existing alignment structure that makes up industrial relations (Ansari et al., 2016; Moore, 1993). For Adner (2017), the change of alignment structure is an indication of an ecosystem; a radical change in the overall existing industrial relations that goes beyond incremental change occurring between dyadic relations. However, when such change takes place – one alignment structure transforms into another – the role of technology is left implicit.

Our case – although not representative of the overall international context of oil and gas industry – reveals how a specific value proposition changes due to actor and data complementarities. Traditionally, the value proposition in the industry was characterized by suppliers providing physical equipment, software systems and associated data analytics services to operators. The Digitize platform was a deliberate attempt by Indus ASA to engage in an ecosystem model where third party actors contribute with complementary resources and modularize their offerings (Jacobides et al., 2018). The data output of this ecosystem model – which we refer to as data hierarchies – disrupted the existing industrial actor configurations, as interdependent actors performed novel activities to realize a new value proposition (Adner, 2017) related to data contextualization.

In our case, the interaction between Indus and Alpha changed fundamentally, as now Alpha did not only supply pumps, but also domain knowledge about how Indus should do predictive maintenance while onshore. This change was not simply dyadic, as Indus started doing predictive maintenance based on the data analysis performed by Alpha using DigitizePlatform. This change in actor relations indicates a larger industrial change, where the knowledge practices underlying the physical equipment maintenance process must be explained by the ecosystem logic themselves (Alaimo & Kallinikos, 2022). Using DigitizePlatform, Indus and other operators in the oil and gas industry were able to take on different roles in the ecosystem by breaking the traditional value proposition based on supplier lock-in, to developing complementary knowledge about the physical state of equipment. As such, the data complementarities, composed of heterogeneous data sources (e.g. operational and supplier test data), contributed by various actors and formalized into a new contract arrangement, brought in the emergence of an industrial data ecosystem.

6.2Data Hierarchies as Industry-Specific Data Complementarities

The introduced concept of data hierarchies extends existing research on data ecosystems by exploring the notion of data complementarities (Alaimo, Kallinikos, & Valderrama, 2020) in an industrial heavy-asset context. Research on data ecosystems is still nascent; existing conceptualizations of data ecosystems tend to overlook the particularities of data objects, subsuming data under the more general concept of digital resources. One notable exception is the study by Alaimo, Kallinikos and Valderrama (2020), that specifically theorizes the formative and constitutive role of data complementarities in the context of social media ecosystem formation. The authors show how a data ecosystem is based on complementarities between data types, technological functionalities, and the economic motives of heterogenous actors. These insights advance our understanding of the sociotechnical nature of ecosystems beyond a pure economic and management-based perspective which tends to dominate ecosystem concepts (Alaimo, Kallinikos, & Valderrama, 2020). In this paper, we build on and further extend the authors' work (Alaimo, Kallinikos, & Valderrama, 2020) by exploring data complementarities in heavy-asset industries.

Data complementarities ontologically account for the existence of data ecosystems – based on the nature of complementarities different ecosystems can be said to exist. In our case data hierarchies as multi-level data objects represent the complementarities that exists between physical components in the industrial reality; the physical components are ordered hierarchically. Alaimo, Kallinikos and Valderrama (2020, p. 43) claim how the emergence of a data-based ecosystem is not dependent on the pre-existing physical complementarities underlying traditional products and services. Our findings expose how the emergence of an industrial data ecosystem is as equally dependent on the pre-existing complementarities underlying physical products and industrial actor relations, as it is on data complementarities. Based on this logic, by identifying the output (physical item, digital resource/data object) and the associated value creating complementarities, one would be able to identify different forms of ecosystems, in our case a physical (non-digital) and a data ecosystem. A physical ecosystem consists of complementarities that are asset-specific (Autio & Thomas, 2020), and have a product-specific architecture (Yoo et al., 2010) A data ecosystem, on the other hand, is product-agnostic due to the properties of data, such as being portable, editable, recontextualizable and generative across industrial and organizational contexts.

Our findings indicate that in a heavy-asset industrial context, contrasts between different types of ecosystems should be more nuanced. When analyzed at the level of data complementarities, the emergent data hierarchies cannot be explained by clearly being placed within the productive apparatus of a digital ecosystem or non-digital ecosystem. Data hierarchies are not a simple digital representation of physical assets, or homogenizations of data shared across organizational boundaries, where the ecosystem emerges by combining different data types into a novel data object. Instead, the digital data complementarities aim to "mirror" the actor complementarities in creating value through physical assets. In industrial data ecosystems, the existing, specialized industrial structures embed the emerging data ecosystem as actor relations change. Therefore, data hierarchies do not simply transform open-endedly – as the properties of data might suggest – but are produced, shared and used according to the industrial actors' specialized goals and needs.

The transformed data objects still have to give a clear representation of the milieu characterizing the physical reality of oil and gas field workers; the main ecosystem output is physical objects (oil, gas, production equipment), not data. Hence, an industrial heavy-asset data object is product-specific to a degree, making the ecosystem asset-specific to a degree. With this, we contribute the layered modular architecture framework of (Yoo et al., 2010)The authors state that at one extreme end of a continuum is a traditional modular architecture based on a fixed product boundary – non-digital ecosystem – with nested and fixed components. At the other extreme end is an ideal form of a layered modular architecture with a product boundary that is not fixed where the components being digitized are product-agnostic – digital ecosystem (Yoo et al., 2010). The heavy-asset data ecosystem reported on in this case can be placed somewhere along this continuum, having elements of both.

7 Conclusion

Our paper shows how industrial data ecosystems emerge through the actor and data complementarities which resemble, but do not correspond to the existing industrial relations. We conceptualize data hierarchies, as data objects specific to our industrial context. In other industries, these data complementarities will be probably ordered differently, in a way which reflect the specific industrial actor relations. The reliance on one case is, thus, a limitation of our study. Further research could try and explore data complementarities in other industrial contexts and show how they shaped industrial actor relations.

References

Aaltonen, A., Alaimo, C., & Kallinikos, J. (2021). The Making of Data Commodities: Data Analytics as an Embedded Process. Journal of Management Information Systems, 38(2), 401-429.

Adner, R. (2017). Ecosystem as structure: An actionable construct for strategy. 43(1), 39-58.

Adner, R., & Kapoor, R. (2010). Value creation in innovation ecosystems: How the structure of technological interdependence affects firm performance in new technology generations. Strategic Management Journal, 31(3), 306-333.

Alaimo, C., & Kallinikos, J. (2022). Organizations Decentered: Data Objects, Technology and Knowledge. Organization science, 33(1), 19-37.

Alaimo, C., Kallinikos, J., & Aaltonen, A. (2020). Data and value. In Handbook of Digital Innovation. Edward Elgar Publishing.

Alaimo, C., Kallinikos, J., & Valderrama, E. (2020). Platforms as service ecosystems: Lessons from social media. Journal of Information Technology, 35(1), 25-48.

Alvesson, M., & Sköldberg, K. (2017). Reflexive Methodology: New Vistas For Qualitative Research. Sage.

Ansari, S., Garud, R., & Kumaraswamy, A. (2016). The disruptor's dilemma: TiVo and the US television ecosystem. Strategic Management Journal, 37(9), 1829-1853.

Autio, E., & Thomas, L. D. (2020). Value co-creation in ecosystems: Insights and research promise from three disciplinary perspectives. In Handbook of Digital Innovation. Edward Elgar Publishing.

Claussen, J., Kretschmer, T., & Mayrhofer, P. (2013). The Effects of Rewarding User Engagement: The Case of Facebook Apps. Information systems research, 24(1), 186-200.

Cornelissen, J. P. (2017). Preserving Theoretical Divergence in Management Research: Why the Explanatory Potential of Qualitative Research Should Be Harnessed Rather than Suppressed. Journal of Management Studies, 54(3), 368-383.

Dattée, B., Alexy, O., & Autio, E. (2018). Maneuvering in Poor Visibility: How Firms Play the Ecosystem Game when Uncertainty is High. Academy of Management Journal, 61(2), 466-498.

Eaton, B., Elaluf-Calderwood, S., Sørensen, C., & Yoo, Y. (2015). Distributed Tuning of Boundary Resources The Case of Apple’s iOS Service System. MIS quarterly, 39(1), 217-244.

Gawer, A., & Henderson, R. (2007). Platform Owner Entry and Innovation in Complementary Markets: Evidence from Intel. Journal of Economics & Management Strategy, 16(1), 1-34.

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.

Hannah, D. P., & Eisenhardt, K. M. (2018). How firms navigate cooperation and competition in nascent ecosystems. Strategic Management Journal, 39(12), 3163-3192.

Henfridsson, O., Nandhakumar, J., Scarbrough, H., & Panourgias, N. (2018). Recombination in the open-ended value landscape of digital innovation. Information and Organization, 28(2), 89-100.

Jacobides, M. G., Cennamo, C., & Gawer, A. (2018). Towards a Theory of Ecosystems. Strategic Management Journal, 39(8), 2255-2276.

Katz, M. L., & Shapiro, C. (1994). Systems competition and network effects. Journal of economic perspectives, 8(2), 93-115.

Langley, A., & Tsoukas, H. (2016). The SAGE handbook of process organization studies. Sage.

Moore, J. (1993). Predators and Prey: A New Ecology of Competition. Harvard business review, 71, 75-86.

Nambisan, S. (2018). Architecture vs. ecosystem perspectives: Reflections on digital innovation. Information and Organization, 28(2), 104-106.

Nambisan, S., lyytinen, k., Majchrzak, A., & Song, M. (2017). Digital Innovation Management: Reinventing Innovation Management Research in a Digital World. MIS quarterly, 41.

S. Oliveira, M. I., Barros Lima, G. d. F., & Farias Lóscio, B. (2019). Investigations into Data Ecosystems: a systematic mapping study. Knowledge and Information Systems, 61(2), 589-630.

Sawy, O., Malhotra, A., Park, Y., & Pavlou, P. (2010). Seeking the Configurations of Digital Ecodynamics: It Takes Three to Tango. Information systems research, 21, 835-848.

Schreieck, M., Wiesche, M., & Krcmar, H. (2021). Capabilities for value co-creation and value capture in emergent platform ecosystems: A longitudinal case study of SAP’s cloud platform. Journal of Information Technology, 36(4), 365-390.

Selander, L., Henfridsson, O., & Svahn, F. (2013). Capability Search and Redeem across Digital Ecosystems. Journal of Information Technology, 28(3), 183-197.

Teece, D. J. (1986). Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. Research policy, 15(6), 285-305.

Thomas, L. D., & Autio, E. (2019). Innovation ecosystems. Available at SSRN 3476925.

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

Walsham, G. (2006). Doing interpretive research. European Journal of Information Systems, 15(3), 320-330.

Wareham, J., Fox, P. B., & Giner, J. L. C. (2014). Technology Ecosystem Governance. Organization science, 25(4), 1195-1215.

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.

Yoo, Y., Richard J. Boland, J., Lyytinen, K., & Majchrzak, A. (2012). Organizing for Innovation in the Digitized World. Organization science, 23(5), 1398-1408.

Zittrain, J. (2008). The Generative Internet'(2006). Harvard Law Review, 119, 1974.

Appendix IV:

“Beyond Organizational Boundaries: The Role of Techno-legal Configurations”

Dragana Paparova, Margunn Aanestad, Ela Klecun

Abstract

In this paper, we explore how techno-legal configurations shape the evolution of an information infrastructure (II) by focusing on data as its critical components. We define techno-legal configurations as assemblages, which are technologically determined by the functionalities for data storage, processing, sharing and usage, and legally determined by the basis for data processing, such as consent, data-processing agreements or laws. To study IIs evolution we conduct an 11-year study of a regional II in Norway, by following its techno-legal configurations as electronic patient record data and patient-generated healthcare data were shared among various organizations. Our findings show how data were shared along intra-territorial and inter-territorial configurations, where the territories were not defined around organizational boundaries, but by the configurations of technology and law. We contribute to the II literature by raising the importance of law as an actor in its own right shaping II’s evolution.

Keywords: information infrastructures, techno-legal configurations, patient (generated) health data

1 Introduction

Information infrastructures (II) in healthcare have been extensively studied among information systems (IS) scholars (Aanestad et al., 2017; Bygstad et al., 2017; Bygstad & Øvrelid, 2020; Grisot et al., 2014) focusing on their architecture-governance configurations (Bygstad & Hanseth, 2016; Hanseth & Rodon Modol, 2021), aims to homogenize a complex portfolio of IT systems (Bygstad & Hanseth, 2016), or accommodate the needs of various public and private stakeholders (Kempton et al., 2020). As of recently, II studies have also started foregrounding the role of data, exploring phenomena such as data-intensive infrastructures (Tempini, 2017), or data infrastructures (Jarvenpaa & Essén, 2023), implying how data, not IT, have a central role in IIs – however, this remains an underexplored topic.

In this paper, we zoom-in on data as critical components of IIs, whose interplay with digital technologies, users, organizational practices, institutions shape II's evolution. To study the role of data in shaping II's evolution, we focus on the configurations of technology and law, hereafter referred to as techno-legal configurations. We define techno-legal configurations as assemblages, which are technologically determined by the functionalities for data storage, processing, sharing and usage, and legally determined by the basis for data processing, such as consent, data-processing agreements or laws. Scholars in IS have showed how sharing data requires legal compliance (Khatri & Brown, 2010), but also how techno-legal entanglements are actors in their own right (Gualdi & Cordella, 2022). In this paper, we explore how the heterogeneous, open, and evolving installed base (Hanseth & Lyytinen, 2010) unfolds across techno-legal configurations by focusing on data-sharing. We argue how, due to the critical role of data, the evolution of IIs is not simply open-ended, but can unfold within a limited set of possible forms defined by techno-legal configurations.

Our empirical motivation comes from the recognition that patient-generated health data (PGHD) is a valuable, but underutilized resource in healthcare. This category of data includes health data related to symptoms, treatment, lifestyle choices, generated by patients using mobile health apps, but also reporting of measurement data from medical equipment including sensors and wearables in remote care monitoring services (RCM). PGHD differ from the data commonly stored in electronic patient record (EPR) systems, as they are generated by patients, outside of hospitals' physical boundaries; they also provide real-time and continuous data about patients' health. These continuous streams of data are particularly valuable for treating and monitoring chronically ill patients (Bardhan et al., 2020), where healthcare professionals employed in various organizations, such as general practitioners, municipal services, specialist services need to collaborate around patients' treatment. However, as of now, PGHD are seldomly utilized beyond the treatment they are generated for, and rarely integrated with the installed base of patient data from EPR systems (Tiase et al., 2021). Moreover, PGHD are unregulated (Winter & Davidson, 2020) and often reside in vendors' storages in the cloud. This architectural and legal uncertainty around PGHD brings challenges on how these heterogeneous, novel, continuous data streams are to be produced, used and shared across the pre-existing conditions of the installed base in IIs in healthcare.

Our research aim with this paper is to understand the technical and legal challenges associated with the production, sharing and utilization of PGHD in healthcare by exploring them in line with other phenomena of significance to IS scholars – namely, information infrastructures. The research question we seek to answer is: "how can techno-legal configurations shape data-sharing in IIs in healthcare". For that purpose, we empirically follow the 10-year evolution of a regional information infrastructure in the specialist healthcare sector in the South-East region of Norway. Our empirical story encompasses both, the installed base of EPR data and the increasing adoption of remote care monitoring supported by PGHD. To study how techno-legal configurations shaped the evolution of the II over time, we use assemblage theory (AT) (DeLanda, 2006, 2013, 2016) as a theoretical lens, and the concepts of assemblages, territorialization, and thresholds. Overall, this paper contributes to the literature on II by showing how data-sharing in II unfolds across intra-territorial or inter-territorial configurations, whose boundaries are not determined by organizations, but the configurations of technology and law. It also offers a practical contribution in highlighting the technical and regulatory complexities in producing, sharing and utilizing PGHD as part of the routine healthcare service delivery.

This paper is structured as follows. Next, we review the literature on IIs and show how the interplay of technology and law rarely has been at the focus of previous II studies. In section three, we introduce AT and the vocabulary of assemblages, territorialization and thresholds as a theoretical lens in analyzing the case. In section four, we describe the research approach. Section five presents the findings and shows two phases of change in the II based on techno-legal configurations: 1) intra-territorial configurations; and 2) inter-territorial configurations. In section six, we analyze the findings and show how, in our case, the law was determining what was internal and what was external to territories, while the technology was defining the heterogeneity or homogeneity of components across internal and external territories. In section seven, we discuss the main contributions of the paper.

2 Research Background: Information Infrastructures

An information infrastructure is defined as a shared, evolving, open, standardized, and heterogeneous installed base (Hanseth & Lyytinen, 2010; Tilson et al., 2010). This definition implies how IIs are shared by the members of a community, including vendors, users and staff; evolving, as they are not designed upfront but continuously changing; open, as the functions or uses they fulfill have no clear boundaries; standardized, allowing for interoperability and interconnection of components; heterogeneous, as they encompass different elements; technology, users, organizations. The literature on IIs in IS is extensive; for the purpose of this paper, we divide it into two research streams. The first stream explicitly drawing on the notion of the installed base as a set of heterogeneous IT systems, users, practices, organizational structures and its role in the evolution of IIs. The second stream discussing the institutional pressures, resistance and accommodation of various actors' needs in large-scale infrastructure projects and programs (such as an implementation of shared electronic patient records). This research stream emphasizes the cross-organizational nature of IIs, as well as the continuous gaps between central formulation of policies, strategies, rules and the discrepancies in their local implementation.

First, the central notion in IIs research is that of the installed base. The installed base refers to the pre-existing set of technological capabilities, organizational practices, user communities, institutional resources that enable and constrain the evolution of the II. Therefore, the evolution of IIs cannot be attributed to a single source but is a result of inter-related socio-technical elements (Henfridsson &

Bygstad, 2013; Sahay et al., 2013); it is not created de novo, but built or rather grown on the installed base of technical systems, organizational structures, and practices that influence their formation and adoption (Aanestad & Jensen, 2011; Hanseth et al., 1996; Sahay et al., 2009). Researchers have argued how the installed base needs to be kept stable enough to allow for new connection to happen, but also flexible enough to accommodate change (Tilson et al., 2010).

Growing infrastructures is seen is a long process involving different potential paths, alternative solutions and step-by-step advances and reversions Hanseth (2022). For example, Sahay et al. (2013, p. 294) suggest that "with each step new socio-technical configurations are created which not only shape subsequent steps, but also redefine the content of the artifact". Those configurations are accomplished through translations involving the interplay of infrastructure, software, and institutions. Various strategies have been identified for growing the installed base, including expanding it with novel components, complementing the old with the new, as well as substituting the old components (Aanestad et al., 2017).

A central focus in studies on the evolution of IIs has been on their governance and architectural arrangements. When it comes to governance, scholars championed bottom-up developments and distributed control, e.g. in the form of polycentric governance to satisfy needs of multiple parties (Constantinides & Barrett, 2015; Vassilakopoulou et al., 2018). When it comes to architecture, scholars have argued how modular architectures can balance between generic and specialized user needs (Grisot et al., 2014). However, several studies have considered architecture and IT governance in tandem (Bygstad & Hanseth, 2016; Hanseth & Rodon Modol, 2021; Rodon & Silva, 2015). For instance, it was argued how decentralized control and loosely coupled architecture are more likely to lead to successful scaling of e-health IIs (Aanestad & Jensen, 2011). Some propose that the choice of governance (centralized vs. decentralized) should depend on the stability of the elements of the architecture (Bygstad & Hanseth, 2016), where hierarchical structure develops with centralized locus of control, and modular architecture is linked to decentralized structure (Rodon & Silva, 2015). Hanseth and Rodon Modol (2021) have also argued how architecture and governance are intrinsically related; thus should be conceptualized as unified entity or an assemblage (A-G) that simultaneously shapes the evolution of II and changes due to the II evolution.

A second research stream of IIs examines the effects of institutional factors on the evolution of II infrastructures, such as shared electronic health records and health management information systems (Currie, 2012; Currie & Finnegan, 2011; Currie & Guah, 2007; Klecun et al., 2019; Mekonnen & Sahay, 2008, p. 200). Researchers illustrate how formal institutions (e.g. healthcare structure and policies), informal practices and institutional logics interplay to produce specific outcomes. For example, in the context of Ethiopia, Mekonnen and Sahay (2008) illustrate how different agencies influenced standards development and HIS scaling through their efforts to define health indicators and uniform reporting formats. Currie (2012) and Klecun et al. (2019) analyze the effects of institutional pressures on the development of nation-wide EHR, highlighting the political context and problems of translating policies into practice. Currie and Seddon (2022) extend such analysis to the initiative for developing supra-national health information technology across European Member States.

More recently, in a study of a national welfare technology program in Norwegian municipalities, Kempton et al. (2020) also show how the divergent perspectives, strategies, regulatory frameworks and policy agendas of various autonomous public and private actors (governmental agencies, municipalities, IT vendors) resisted and accommodated a proposed architecture for welfare technologies. The authors raise how challenges emerged as IT vendors for remote care monitoring were storing data in cloud services which collided with laws and regulations. However, the role of laws and regulations was touched upon only marginally and not been further explored in this study.

In summary, while the term ‘information infrastructure’ indicates the central role of information, the majority of literature focuses on the technology (IT architecture) - governance nexus, rather than on techno-legal configurations shaping IIs in healthcare. We raise the importance of studying techno-legal interlay due to the personal and sensitive nature of patient data and the highly-regulated nature of the healthcare context (Paparova et al., 2023). Previous studies on architecture-governance configurations marginally touch upon the role of data in II evolution, subsuming data within concerns about IT architecture, or day-to-day decisions of users in data work practices (e.g. Parmiggiani and Grisot 2020). Data capture, storage and analysis have also been addressed in relation to data requirements of different stakeholders (e.g. Sahay et al., 2009), privacy and security (e.g. see Pouloudi et al., 2016) or standards and data quality (e.g. Mekonnen and Sahay 2008). However, laws and regulations have rarely been acknowledged as actors in their own right shaping the II evolution, with limited exceptions (Gualdi & Cordella, 2022). Our work builds on studies of IIs in healthcare by zooming in on (patient) health data and the techno-legal configurations shaping II evolution. For that purpose, we turn to using assemblage theory (AT) as a theoretical lens.

3 Assemblage Theory as a Theoretical Lens

Assemblage theory (DeLanda, 2006, 2013, 2016) is a realist ontology focusing on difference, heterogeneity, and change (Rutzou & Elder-Vass, 2019), as opposed to traditional views on realism focusing on forms, order and structures. AT, instead, focuses on processes and structures, simultaneously, arguing how there are different degrees of order and chaos, heterogeneity, and homogeneity, in the dynamic, social world, which do not connect structures to structures, but encompass a variety of entities: structures, processes, forces, relations, agents. Assemblages, as its central concept, refers to the process of fitting together a set of heterogeneous components that form larger wholes, but keep on changing. Assemblages are defined along two dimensions: horizontal – processes across which the heterogeneous parts relate; vertical – processes that stabilize the heterogeneous parts to form larger wholes and keep its identity over time. These simultaneous vertical and horizontal processes differentiate AT from other realist ontologies; the former keep the cohesion of the whole, the latter simultaneously keep the relations between its components changing.

The degrees of dynamism and structure of assemblages is defined by their degree of territorialization. The territorialization of assemblages can be determined by two parameters: 1) the sharpness of boundaries defining what is internal and what is external to the assemblage' territory; and 2) the internal homogeneity of components and their relations, such as sorting processes including or excluding certain components from the assemblage's territory. While territorialization helps the assemblage keep its identity over time, the heterogeneous parts continue relating – assemblages are simultaneously deterritorialized. The more blurred the boundaries between the internal and external territory are, and the more heterogeneous the components of the assemblage are, the more deterritorialized an assemblage is. Therefore, assemblages have a certain degree of dynamism and recurrence, heterogeneity and homogeneity, territorialization and deterritorialization, but dynamism, heterogeneity and deterritorialization always prevail.

At a given point, the processes of territorialization or deterritorialization can reach a critical limit, and the assemblage can undergo a transition from one state to another – reaching a threshold. Once the assemblage reaches a threshold, it changes its form, and the degrees of territorialization and deterritorialization need to be negotiated de novo.

The vocabulary of AT has been used by IS researchers studying IIs before (Hanseth, 2022; Hanseth & Rodon Modol, 2021). In this paper, we use the vocabulary of AT to understand IIs as composed of various heterogenous parts, such as data, digital technologies, organizational processes, users, institutions. As part of IIs, data are fitted together with other components, however the territories within which data can interact with these other components are defined by techno-legal configurations. Therefore, the territories are not organizational, but determined by technology and law. These techno-legal configurations can perform various functions, such as sharing data within healthcare organizations, across healthcare organizations, and externally with private vendors. However, the coming of PGHD brings in the need for reassembling the techno-legal configurations across which data interact with other components. Therefore, IIs reach a certain threshold such as shifting from sharing EPR data, to also sharing PGHD, and new territories across which data can be shared are defined.

4 Research Approach

4.1 Case Background

Our empirical study was conducted in the South-East Health Region in Norway which offers specialist health services to 57% of the total population in Norway. The Regional Health Trust (RHT) is the administrative body overseeing 11 public hospital trusts, 5 private, non-commercial hospital trusts and its own IT company (HospitalPartner) that works together with the vendors and hospitals in implementing the necessary digital technologies. Moving services outside of the hospital was emphasized in the RHTs Strategic Development Plan towards 2035.

This comprises both temporary home-based cases using connected medical equipment (so-called home hospital services), long-term monitoring with sensor technologies (called digital home follow-up), and more episodic communication services such as video and chat. In this paper, we regard all these services as remote care monitoring (RCM). The strategic emphasis on moving services to the home aligns with national policy as well as general trends. Going back to 2011, national strategy documents called for provision of digitally mediated care in the patients' homes. A national implementation program for so-called welfare technologies (also known as ambient assisted living or telecare) saw many municipalities implement technologies in patients' homes. Some of the municipalities also implemented digital home follow-up services for patients with chronic diseases such as diabetes, heart failures or chronic-obstructive lung disease (COPD).

The technologies in use included sensor devices, patient-reporting of data, and digital consultations. Several of the hospital trusts in the South-East health region had already initiated various home hospital projects (e.g., to allow patients with cancer or on long-term antibiotics treatment to stay at home) and digital home follow up services (e.g., to support early discharge of newborns). Some of these were in pilot phase and others had been implemented in routine service. There was, however, no dedicated II in place that could support the deployment of RCM at scale, and each initiative had conducted their own procurement and service design process. In the autumn of 2020, the HSE started work to consolidate the fragmented portfolio of RCM services. This was connected to a larger initiative which aimed to provide a shared infrastructure that would enable the HSE to scale up RCM beyond the stand-alone projects, through implementing a new process platform. We aligned our study with this digital infrastructure initiative, starting the study in October 2020 when the infrastructure project was in its initial concept phase, up until its purchasing and implementation start as of 2023.

The coming of RCM triggered significant changes in the installed base of EPR systems. This made us realize that we also needed to include a retrospective component to our study. Thus, our paper follows these significant changes in the South-East health region's information infrastructure over a decade.

4.2 Gathering of Empirical Material

We conducted a qualitative study and primarily relied on interviews as a data gathering method (Alvesson & Sköldberg, 2010). We conducted 12 semi-structured interviews with key participants in the process platform initiative (some participants were interviewed twice). We included core project members from the regional authority (5 persons), innovation and technology experts from the regional hospitals with pre-existing services in the area of DHF (7 persons), as well as vendor companies (2 persons). The interviews with the project team were group interviews, containing two, or three participants at the same time. The interviews with hospitals and private vendors were either group (containing two participants) or individual. Including more than one participant was either suggested by us (such as with the project team), but most often the participants we contacted would suggest that another person from their organization takes part. It was common that the interviews would include both technical and management people from the same organization. We could thus get an overview of both, the technical and organizational implications of the regional initiative for the stakeholders involved.

Through these interviews, we elicited the interviewees' accounts of the rationale for the process platform project, its progress, challenges and achievements. We started the data gathering by interviewing informants from the regional authority, which shared their project vision, planned scope, and expectations. As the project moved into the procurement phase in December 2021, we shifted to interviewing informants from the regional hospitals, as well as vendor companies that offered DHF services and had active installations in the region. Our study included three hospitals that were involved in the concept phase and two other hospitals that were assigned pilot user status.

However, the interviews were focused on the process platform as an upcoming large-scale infrastructural project; empirically we were puzzled by remote care monitoring (RCM), and particularly, patient-generated health data (PGHD). We realized that RCM as a use case was posing new challenges for the installed base of data, IT applications, users, institutions. For that purpose, we decided to collect documents which do not only justify the rationale around the process platform project, but also encompass the historical evolution of the II over time. To do this, we relied on official documents encompassing the regional architecture, its data-sharing capabilities, the status quo, the legal changes in data sharing, remote care monitoring as a service. We also analyzed presentations commonly shared with us by participants, and observed events of importance such as steering group meeting and regional innovation events. The sources are summarized in Table 1.

Data gathering	**Amount**	**Duration**	**Description**
Group interviews	5	60 min, 90 min, 100 min, 120 min	Participants: regional project leaders, or hospital managers
Individual interviews	7	45 min, 60 min and 90 mins	Participants: project leaders, hospital managers/ innovation directors, or private vendors
Document analysis	21	/	4 concept phase documents for process platform, API platform and digital home follow-up 4 internal regional documents on process platform, API platform and digital home follow-up 2 tender documents for process platform 5 documents on national data sharing architectures, including message exchange, document sharing and structured data exchange 4 documents on changing the Health Register Act and Health Record Act PGHD report from Health Directorate 1 document on structuring the electronic patient health record across regions

Presentations	5	/	2 presentations for steering group meetings and 3 individual presentations
Meeting observations	2	/	Steering group meeting and innovation network event
Video presentations	1	/	Hospital presentations on digital home follow-up
Webpages/ press releases	7	/	RCM – ongoing initiatives nationally and in regional hospitals

Table 1: Summary of the data gathering process

4.3 Analysis of Empirical Material

We conducted an abductive (Dubois & Gadde, 2002), process analysis (Berends & Deken, 2021). Our starting point was empirical, as we were puzzled by the nature of PGHD, as heterogeneous, unregulated and novel types of data sources and their interaction with the pre-existing conditions in across the installed base. To understand this, we created a timeline of 10 key events which we found relevant to the current state of the II – starting from the process platform initiative and PGHD (as represented in Figure 1). We ordered events chronologically and identified three key processes which were significant throughout time: 1) consolidating the IT portfolio; 2) sharing data within and across organizational boundaries; and 3) changes in laws which regulate patient health data.

With these insights, we realized that a focus on IIs and the installed base could be useful in grasping the complexity of the empirical material. Therefore, we consulted the literature on IIs to help us understand how these events led to infrastructural changes over time. However, we realized that the literature on IIs is commonly focused on architecture-governance configurations of IT systems, but does not account fully for the techno-legal configurations which are instrumental when sharing data. For that reason, we decided to consult assemblage theory (AT) and the concepts of assemblages, territorialization and thresholds to make sense of the complex empirical reality. By using the concept of thresholds, we defined two phases across which II expanded over time: sharing data across EPR systems and sharing data from RCM. By using the concepts of assemblages and territorialization, we defined two types of techno-legal configurations: 1) intra-territorial– covering changes in the II to facilitate the sharing of EPR data; and 2) inter-territorial– following changes in the II to facilitate sharing of PGHD. Defining these configurations as phases does not imply that they are consequential; instead, they overlap over time.

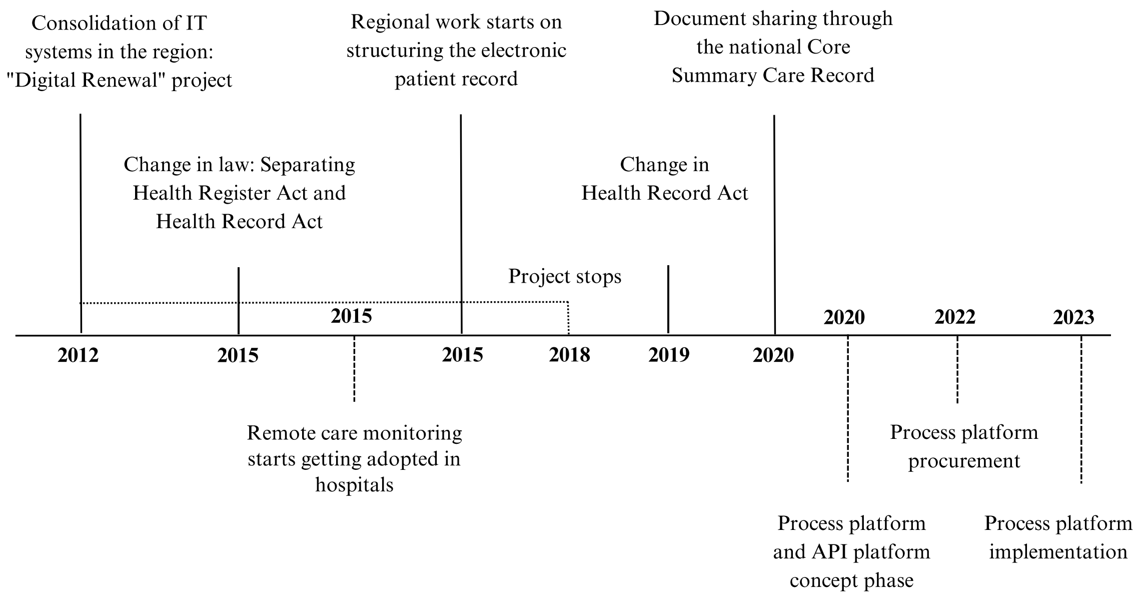


Figure 1: Evolution of II in Health South-East. Events above the line showing changes around EPR data; events below the line show changes regarding RCM

5 Findings

5.1Phase 1: Intra-Territorial Configurations: Sharing EPR Data

As of 2012, patient data recorded in the EPR systems could only be accessed internally, by healthcare personnel employed within one organization. Externally, data about patients could be shared using message exchange, such as through referrals. However, such sharing only covered data directly relevant to the referral, but did not cover sharing of any data from the electronic patient record systems. Moreover, by sharing data as messages, data were transferred from one organization to another and stored in both the originating and recipient organizations' EPR systems.

5.1.1 Harmonizing Laws For EPR Data-Sharing

At that time, storing, processing, and using patient data was regulated by the Health Register Act, which regulated both patient records, but also secondary uses of data. The Act restricted access to data in patient record systems to healthcare personnel that were members of the organization. The rationale was that data processor authority and managerial authority of employees had to overlap in order to ensure the necessary information security levels. Health personnel could therefore not access data from patient records in external organizations, even if they were treating the same patients as health personnel in the other organizations.

In 2014, various healthcare actors, the Ministry of Health and Care and the government started discussions on changing the Health Register Act to allow for patient data-sharing across organizational boundaries. The rationale for changing the Act was to support the flow of patient data across healthcare personnel and provide better diagnostics, treatment and follow-up of patients. The changes were in effect starting 1 July 2015, and the Health Register Act was split, regulating two separate areas; 1) Health Register Act, regulating health registers, and the processing of health information related to health analysis, population health management, and other types of secondary use of data; and 2) Health Record Act, regulating the processing of health information in treatment-oriented health registries, i.e., electronic patient records.

The Health Record Act in principle opened possibilities for healthcare personnel to access and search for patient data stored in external EPR systems. This change allowed the establishment of a national Summary Care Record, where critical information about citizens could be recorded. Such access could be granted based on healthcare personnel's official need. The possibility for patients consenting to the data sharing was assessed as time consuming, and instead, it was decided that patients should have the right to object to the data sharing. Moreover, the Health Record Act opened up a possibility for healthcare personnel employed in different organizations to establish formal organizational collaboration that involved data sharing. If the organizations established such a formal joint entity, health personnel could add patient data in the same treatment-oriented health register. The change of the law therefore enabled novel data sharing to be done in two ways: 1) by establishing a joint patient record; or 2) by using the national solutions to document information, such as the Core Summary Care Record. The joint or national records were not intended to work as a substitution, yet more as complementary records to the existing patient records used within organizations.

As of 2019, the Health Records Act was updated again in order to simplify access to patient information to healthcare personnel. With this change, the rules for accessing data internally and externally across healthcare organizations was harmonized. According to these changes, health data could be made available to healthcare personnel for the purposes of providing healthcare, but also for quality assurance, self-recording and training. According to previous practice, access to healthcare personnel was controller-based on, among other things, their connection to organizations and departments, professional IT systems, connection to patients, professional roles, task and responsibilities. As per the new law, data controllers could provide automatic access on an organizational level which did not need to cover one patient at a time. This did not mean that access could be given to whole organizations or departments, as the law only allowed for accessing as much information as necessary to provide health and care. Instead, legally the rules for distinguishing between internal access and external were removed, but it was still up to the organizations to determine what requirements they place when access is requested and determine the specific data responsibilities according to the legal provisions. However, it was also pointed out that due to challenges with integration between systems between GPs, psychologists, municipalities and health organizations alike, access would still need to take place manually.

Therefore, the changes in the law allowed for various forms of data sharing, but it was to a large degree up to the individual organizations to determine how to preserve security and privacy in their IT solutions, determine the technical and organizational means under which access can be granted and establish the necessary data-sharing agreements.

5.1.2 Building Regional Integration Services

In 2012, the Parliament released a white paper "One citizen – one record" suggesting that patient information should follow the patients along the course of treatment, with the rationale that this would support patient-oriented healthcare services. At that time, HSE was facing a heterogenous portfolio of siloed IT systems purchased to cover specific needs, which were not able to share data with each-other. Therefore, HSE initiated regional projects aiming to: 1) consolidate the IT portfolio of systems; and 2) build the necessary APIs for sharing data across systems and organizations.

First, as of then, hospitals in HSE had did not only have different systems, but also different installments and configurations of the same systems; this resulted in various one-on-one integrations. HSE started initiating projects aimed at standardizing the regional IT portfolio, so that the same EPR, laboratory, radiology, or medications systems would be used across hospitals. At that time, only one hospital in the region did not use the dominant EPR vendor; therefore, work was initiated on replacing its existing EPR system. HSE also initiated additional projects on purchasing a regional medications solution that provides a clinical overview of data about individual patients (medical chart and vital signs information such as pulse, temperature, blood pressure, medication, lab results); as well as a joint laboratory management system, a joint radiology system, and a multimedia solution for storing and using multimedia data, such as images, video, ultrasound.

Second, consolidating the IT portfolio was a step towards homogenizing the II; however, it was not enough for sharing data across these systems. As of 2010, HSE also started working on updating the regional integration services, including developing the necessary APIs. The integration services are intended to connect the IT systems but keep the hospitals legally separate. Therefore, the regional authority built separate integration platforms for each Hospital Trust; including 1) local APIs for sharing data within hospitals, for example from the EPR system to lab and radiology systems, allowing healthcare personnel to order tests and see results from the EPR system; 2) regional integration services allowing for data-sharing across hospitals in the region; 3) regional integration services towards the national solutions, other regions, or primary healthcare services.

Regardless of the technical consolidation of systems and the integration services developed, the connections between EPR systems in the region remained heterogeneous. Three different technical platforms were used, and the EPR vendor had 9 installations and databases for the different Hospital Trusts. The various systems used in the region were integrated through a mix of proprietary and open

APIs. The systems continued to have strong one-on-one interdependencies and there was no overview over which APIs were used against which systems. Moreover, except for some patient administrative areas, the EPR record was based on free text and there was no structure in the data that could be represented in the APIs.

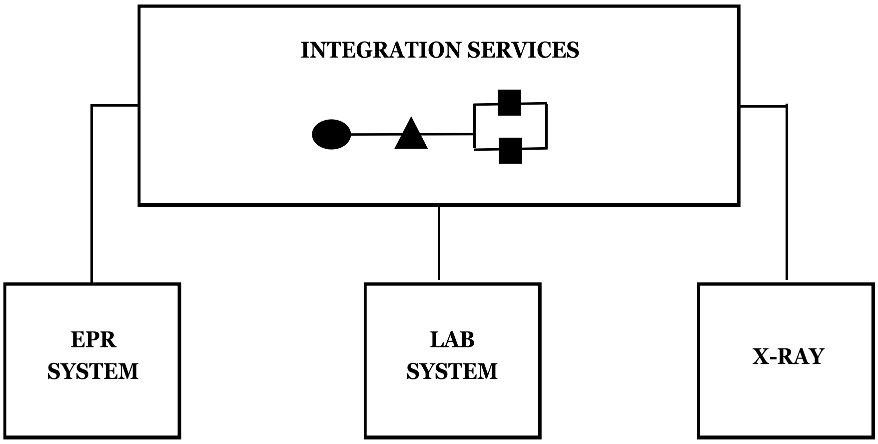


Figure 2: Consolidating the regional portfolio to the same ERP, radiology, and laboratory systems and connecting them to the integration services of individual hospitals.

However, the changes in law and the homogenization of IT systems allowed for various forms of sharing data that were not possible before, such as sharing documents across primary and secondary healthcare services through the national components. For instance, as of 2020, HSE collaborated with the national Core Summary Care Record to facilitate access to selected documents from the hospital EPR systems. Such access was partitioned across four categories: 1) all users of the Core Summary Care Record; 2) doctors and psychologists; 3) patients' GPs; and 4) selective access to e.g. doctors from emergency rooms in municipalities or private hospitals. Sharing documents through the national components allowed healthcare personnel to re-use patient data for various purposes beyond the organizations they were employed in, but within the provisions defined by law.

5.2Phase 2: Inter-Territorial Configurations: Sharing PGHD

As far back as 2015, hospitals in the region started adopting RCM services, which were project based and each hospital signed individual contracts with the RCM vendors. These services were commonly adopted for treating chronically ill patients, such as chronic obstructive pulmonary disease, diabetes, cancer patients. For instance, in 2017, HospitalEast, together with a private vendor, HospitalPartner, the Cancer Association, and other hospitals, started a collaborative project on developing a patient-facing app supporting patients with cancer. The following vignette describes how the app worked to facilitate following up a patient at home.

Healthcare personnel log into the app with their EPR login details. In the app, nurses can see an overview over all her/his patients and click on each patient to see the specific measurements. If there is a discrepancy in the measurements, it is colored red, and a smaller discrepancy is colored yellow. The nurse can use the app to get statistics on the reported data, define the thresholds on when to be notified about a patient, schedule the repetition of these digital forms, and set reminders for the patients. The nurse can also chat with the patients. The app is integrated with the EPR system, and the information is stored in the EPR system as a dynamic document which is updated in real-time. In the app, the nurse can set up chat groups with other healthcare personnel, e.g. create groups around the same patient. Patient log in using BankID and generate data from their homes. (source: video presentation by HospitalEast)

Remote care monitoring was based on PGHD, including fever measurements, blood pressure, oxygen. Working with PGHD required amending the techno-legal configurations across which these data could be shared and integrated with the installed base of patient data stored, shared and used across the EPR systems.

5.2.1 Internalizing PGHD Through Data-Processing Agreements

PGHD raised a legal uncertainty as to whether RCM were to be regarded as treatment-oriented health registers, whether they should be treated as extended components of the EPR system and what requirements the law places on the management of data in these solutions. This was due to PGHD not being regulated by the Health Record Act, but regulated by the consent patients give in the respective RCM solutions. To establish a legal basis for accessing the data stored in RCM, hospitals entered data-processing agreements with the individual vendors they were using. The data-processing agreements regulated how the RCM vendor processed data on behalf of the healthcare organizations, and the vendor cannot use these data for their own purposes.

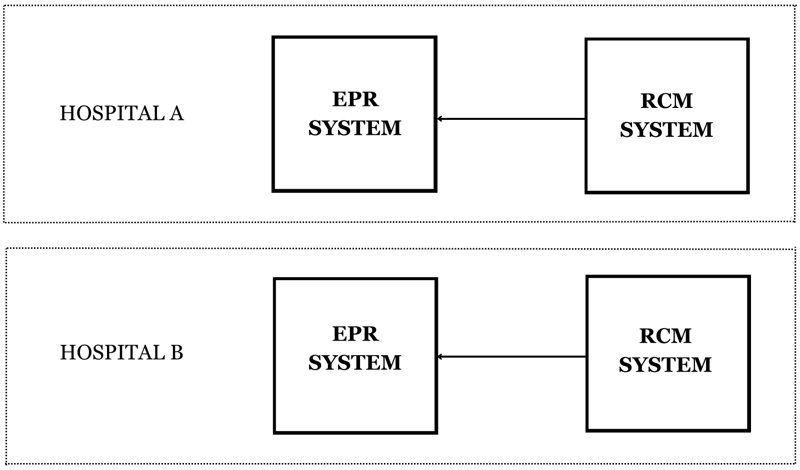


Figure 3: One-on-one integrations of remote care monitoring and electronic patient record systems.

By regulating RCM through data-processing agreements, healthcare personnel could access, edit, write in the RCM records, and user rights were determined based on their roles in the organizations. Therefore, PGHD were legally internal to the hospital organizations, which acted as data controllers for the data processing in the RCM solutions. However, PGHD were predominantly stored in the cloud. The hospitals thus needed to assess vendors' logical and physical security in the cloud solutions; each data-processing agreement had to independently determine the routines for deleting data, data transfer from RCM to ERP systems, or transfer of data if the vendor system is discontinued. To conduct the necessary risk assessments, vendor negotiations, and service design was time-consuming and challenging to do for each individual hospital, as information security competence was scarce.

5.2.2 From Isolated Integrations To a Process Platform

PGHD were commonly structured data reported using digital forms, while EPR data were often free text and unstructured. This made it architecturally challenging to transfer PGHD to the EPR systems and store them there. Instead, vendors were storing data in the cloud, and could transfer limited data as a summary document that can be stored in the ERP systems; in other circumstances, data would be transferred manually through copying some of the measurements or event descriptions into the EPR system. Moreover, the architecture of the EPR systems did not consider remote care monitoring as a service which is internal to hospitals. The patients were not formally admitted to the hospital, and creating a novel category in the EPR system for home-based patients was not trivial. An informant from the HospitalPartner gave an example: "You have a contact between the hospital and the patient, but it is not inpatient, it is not outpatient – it is homecare. If they {the EPR vendor} have picked a database where you can reconfigure a lot of different types of contacts, then it would be relatively easy to do, but this contact information and the values may exist in 20 integrations in {EPR vendor} already. So, putting all in contact which is called 'homecare', then you have to do quality check on all these 20 integrations with the other systems – do they still function.".

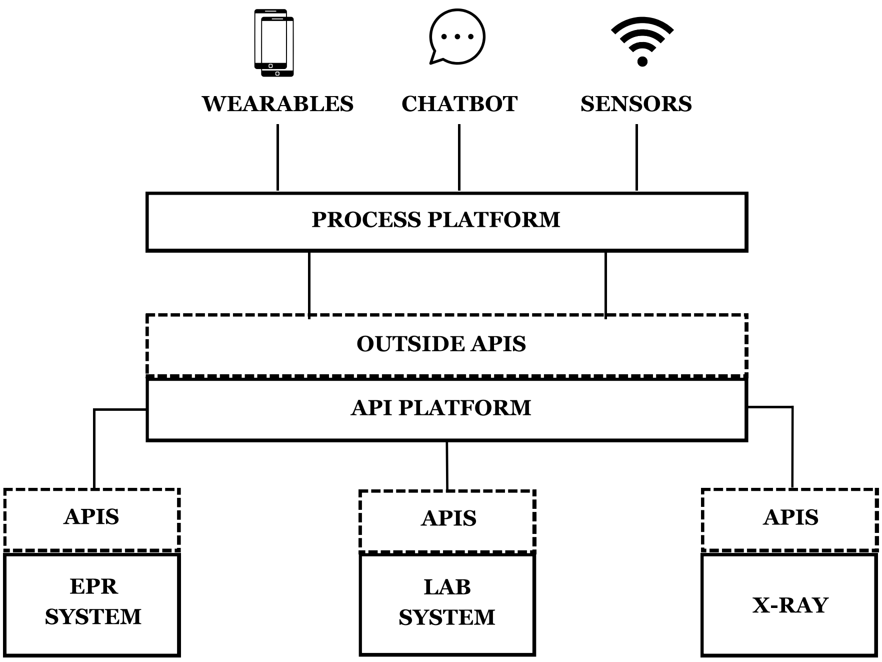


Figure 4: Process platform architecture.

Over time, the adoption of RCM apps, sensor devices and wearable technologies was increasing, but the connections with hospital EPR systems continued to be handled individually by hospitals. Healthcare personnel needed to login separately in the RCM and the ordinary hospital IT systems, using multiple user identities and passwords. They had to register the patient in different systems, to create a personalized plan in the RCM system which was not available in the EPR system. To document care in the EPR system (which is a legal obligation), data from the

RCM system had to be transferred manually. Moreover, healthcare personnel were not able to see RCM data in context with other data about patients in the other hospital IT systems.

To manage this complexity, in the autumn of 2020, the RHT started working on a new regional initiative – the process platform. The process platform was a large architectural project in which RCM was chosen as a central use case. An informant explained: “Today we do not have a lot of APIs and good integration services on the system layers. We are talking with the vendors of different kind of systems to make them to create standardized APIs from {EPR vendor}, for instance. (…) Those APIs from {EPR vendor} will be the bottom APIs, the inner APIs, and we will create regional APIs to get the loose coupling on the middle API management.” (Enterprise architect, Hospital Partner).

The process platform aimed to work as a mediator between the EPR systems and RMC, where the integrations would be based on APIs, instead of direct integrations. Therefore, EPR systems would need to make their data available through inner APIs; RCM data would be consumed through outer APIs. In between these APIs there is an API mediation service, which orchestrates the outside and inner APIs. The sharing of data was to happen from EPR to RCM on access, and RCM to EPR on copy. Moreover, the process platform promised a “low code, no code” functionality, which would enable hospitals to design new services in an easy way. They could for instance build structured forms for patient data capturing themselves (not relying on vendors), such as forms to collect patient-reported outcome measures (PROMs) and patient-reported experience measures (PREMs). Another effect of moving the RCM services onto a joint infrastructure, was this it would allow for standardizing the terminologies used across hospitals for specific diseases, as well as have central coordination of best practices when defining treatment plans using structured data forms. Not the least, a centralized architecture would offload hospitals of the work of vendor negotiations and risk assessment to ensure that information security concerns were met. The process platform architecture is illustrated in Figure 4.

In the spring of 2023, the process platform was procured; meanwhile, two hospitals were chosen for the pilot implementation. The focus areas that were prioritized were video consultations and patient-reported outcome measures; the prioritized diseases were chronic obstructive pulmonary disease, children with diabetes and cancer patients. The work processes and treatment plans established in the pilot hospitals were later to be expanded across other hospitals in the region. While the process platform provided an opportunity to homogenize the IT portfolio of RCM, it did not resolve the legal issues related to the status of PGHD. Instead, PGHD were expected to continue remaining in the cloud, bringing in discussions around Schrems II, and the General Data Protection Regulation.

6 Analysis

We define techno-legal configurations as assemblages, which are technologically determined by the functionalities for data storage, processing, sharing and usage, and legally determined by the basis for data processing, such as consent, data-processing agreements or laws. Our findings show how the sharing of PGHD and the sharing of EPR data required different techno-legal configurations, which we defined as: 1) intra-territorial; and 2) inter-territorial. By using the term territory, we do not refer to organizational boundaries, but boundaries determined by technology and law. The findings and analysis are summarized in Table 2.

We regard techno-legal configurations as intra-territorial when the territory across organizations is legally harmonized and homogeneous technical interconnections. In our case, the techno-legal configurations for sharing EPR data were intra-territorial. The Health Record Act harmonized the rules for sharing EPR data across healthcare organizations; technically, shared integration services across the installed base of IT systems were created. The changes in law opened possibilities for various forms of sharing patient data; however, the degree to which data were actually shared was still determined by organizational and technical means. For instance, access continued to be handled by the individual organizations, or EPR systems, based on healthcare personnel’s job positions (e.g. nurse, doctor, administrative personnel), their tasks and responsibilities, their connections to patients, or the departments they were employed in. Therefore, even though EPR data had the same legal status across organizations, some territories remained external due to the lack of technical integrations.

Findings	Empirical illustration	Techno-legal configuration
Intra-territorial configurations	Health Record Act harmonizing rules for EPR data sharing across healthcare organizations	Boundaries between internal and external organizations are legally harmonized
	Regional projects aiming to consolidate the IT portfolio and build shared integration services	Technical components are homogenized (to a degree) with shared interconnections
Inter-territorial configurations	Data-processing agreements regulating data processes in RCM	Boundaries between internal and external organizations are distinguishable
	RCM and EPR as siloed, or based on one-on-one integrations; process platform brings potential for interconnectedness	Technical components are heterogeneous and have separate integrations

Table 2: Summary of findings and illustrations from empirical material.

We regard techno-legal configurations as inter-territorial when the territory across organizations is distinguished with sharp legal boundaries and heterogeneous technical interconnections. In our case, the techno-legal configurations around PGHD were inter-territorial. Legally, healthcare organizations entered data-processing agreements with RCM vendors, allowing them to access and copy data from RCM vendors. Technically, EPR systems and RCM vendors were creating one-on-one integrations, and healthcare personnel were given access to PGHD stored in the cloud based on their organizational role. Therefore, the data-processing agreements and siloed technical integrations made PGHD internal to hospital organizations. However, the technology, as well as the legal practices around PGHD differed across the region. The process platform was expected to standardize the technical interconnections between RCM and EPR systems. However, the legal status of PGHD remained unresolved.

7 Discussion and Conclusion

This paper seeks to answer the following research question: ”how can techno-legal configurations shape data-sharing in IIs in healthcare”. We explore this research question by conducting an empirical study of an II in the highly-regulated healthcare context dealing with personal and sensitive healthcare data. Our research is in line with other works exploring the need for harmonizing organizational, technological and regulatory spheres around PGHD (Winter & Davidson, 2020). In this paper, we contrast the techno-legal configurations of the installed base of EPR data and increasing adoption of PGHD, and show how data-sharing in II unfolds across 1) intra-territorial configurations and 2) inter-territorial configurations. Our contribution to the literature on IIs is two-fold.

First, we show how the territories for data-sharing in IIs are not defined by organizational boundaries, but by techno-legal configurations. This builds on previous works showing how techno-legal interplay transcends multiple actors when sharing sensitive personal healthcare data (Paparova et al., 2023). This paper extends previous research by unpacking the individual roles of technology and law in sharing data across organizations. Namely, the law defining what is internal and what is external to a specific territory; the technology providing capabilities for including or excluding components in internal and external territories.

With this, we contribute to works discussing the institutional forces influencing large-scale IT projects in IIs in healthcare. Scholars have shown how policy formulation and their practical implementation can be shaped by institutional pressures (Currie & Seddon, 2022; Klecun et al., 2019). Studies on governing large-scale welfare technology programs have also raised how private RCM vendors' interests influence policy goals bringing emerging opportunities and challenges to the program (Kempton et al., 2020). By arguing how internal and external territories were defined by the interplay of technology and law, we do not imply that organizational agency does not matter. Instead, our case clearly highlights how organizations were not simply seeking compliance (Khatri & Brown, 2010), but could determine local implementation of law through their organizational and technical means. However, our study raises the importance of law, as an actor in its own right, defining the forms across which data sharing across organizations can take place.

For instance, the techno-legal configurations in our case enabled and constrained different forms of data-sharing. As the Health Record Act changed and allowed for EPR data to be accessed across healthcare organizations, the Core Summary Care Record was established, enabling document-sharing across national healthcare services. Before the law was amended, healthcare personnel were not allowed to write in EPR records outside of the organizations they were employed in. Therefore, the law defined a territory which transcended organizational boundaries, and the technical components either included or excluded organizations from this territory. However, our case also shows how due to the unregulated status of PGHD and their heterogeneous technical integrations, the forms of sharing PGHD are more reliant on organizations' agency. RCM are currently regulated through data-processing agreements limiting the possibilities to share PGHD across organizational boundaries. The data-processing agreements allow healthcare personnel in specific organizations to write in and access data generated by patients through RCM. However, allowing access to, or writing in RCM by healthcare personnel employed by different organizations is currently not possible. As of now, RCM vendors need to keep the data-processing on behalf of different hospitals separate; this limits the possibilities for establishing cross-organizational forms for data sharing and facilitating collaboration of healthcare personnel around PGHD.

Second, we contribute to II literature exploring the degree to which the components in the installed base should be kept stable or change. For instance, Grisot et al. (2014) differentiate among three architectural innovations in IIs – innovation of infrastructures, where the existing infrastructure is re-engineered; innovation in infrastructures where existing components are replaced; and innovation on infrastructure where new modules are added on top of what is already there. In our case, PGHD started out as innovations on top of the existing installed base of EPR data, as they were considered as peripheral components to healthcare organizations. However, the wide adoption of PGHD across individual hospitals made RCM central components in providing care to chronically ill patients. Therefore, PGHD became innovations of infrastructures, as they brought in a sub– cloud-based– infrastructure that was now to be connected to the installed base of EPR systems.

Previous works argue how IIs should be underpinned by modular architectures which allow for loose coupling between stable and unstable components (Grisot et al., 2014), where the core components (such as EPR systems) should be kept stable, while the peripheral components (such as RCM) should be kept changing (Bygstad & Hanseth, 2016). Our case shows how the process platform aims to technically loosely couple RCM to the installed base of ERP systems. However, we argue that in order to allow RCM technologies to change, PGHD need to be decoupled from the technologies that generate them and acquire a stable legal status that can support further technological change. As of now, PGHD are not just heterogeneous technical components, but also legally heterogeneous components. If stored in EPR systems they are regulated by the Health Record Act, if stored in the cloud, they are regulated by patients' consent and the data-processing agreements between organizations. This complexity transcends territories and requires legal adjustments that cannot be overcome through modular architectures. The study by Kempton et al. (2020) shows how a national welfare program proposed a hub architecture where PGHD were transformed into a format suitable for ERP systems, instead of creating connections to the cloud. We argue how such a hub-architecture could benefit the process platform initiative, as it would provide legal stability to PGHD, by making them parts of the patient record systems; thus, regulated by the Health Record Act.

Our paper also contributes to the IS field by showing how the vocabulary of assemblage theory can be empirically utilized to study socio-technical phenomena as having a degree of change, characterized by both, dynamism and endurance, heterogeneity and homogeneity. Scholars have already explored the potential of AT in studying the dynamics of architecture-governance configurations in a healthcare II (Hanseth & Rodon Modol, 2021). We show how AT's vocabulary can be useful in studying the configurations of technology and law. At last, we contribute to practice by arguing how the degree to which organizations can determine different forms of sharing sensitive personal data in highly regulated environments such as healthcare, is dependent on the interplay of technology and law. Moreover, we argue how legally homogenizing PGHD could allow for more cross-organizational forms of sharing PGHD and improve healthcare personnel's collaboration when providing follow-up for chronically ill patients.

References

Aanestad, M., Grisot, M., Hanseth, O., & Vassilakopoulou, P. (2017). Information Infrastructures within European Health Care. Springer International Publishing.

Aanestad, M., & Jensen, T. B. (2011). Building nation-wide information infrastructures in healthcare through modular implementation strategies. The Journal of Strategic Information Systems, 20(2), 161–176.

Alvesson, M., & Sköldberg, K. (2010). Reflexive methodology: New vistas for qualitative research (Vol. 12). Routledge.

Bardhan, I., Chen, H., & Karahanna, E. (2020). CONNECTING SYSTEMS, DATA, AND PEOPLE: A MULTIDISCIPLINARY RESEARCH ROADMAP FOR CHRONIC DISEASE MANAGEMENT. MIS Quarterly, 16.

Berends, H., & Deken, F. (2021). Composing qualitative process research. Strategic Organization, 19(1), 134–146.

Bygstad, B., & Hanseth, O. (2016). Governing e-Health Infrastructures: Dealing with Tensions. ICIS 2016 Proceeding, 19.

Bygstad, B., Hanseth, O., Siebenherz, A., & Øvrelid, E. (2017). PROCESS INNOVATION MEETS DIGITAL INFRASTRUCTURE IN A HIGH-TECH HOSPITAL. 14.

Bygstad, B., & Øvrelid, E. (2020). Architectural alignment of process innovation and digital infrastructure in a high-tech hospital. European Journal of Information Systems, 29(3), 220–237.

Constantinides, P., & Barrett, M. (2015). Information Infrastructure Development and Governance as Collective Action. Information Systems Research, 26(1), 40–56.

Currie, W. L. (2012). Institutional Isomorphism and Change: The National Programme for IT – 10 Years On. Journal of Information Technology, 27(3), 236–248.

Currie, W. L., & Finnegan, D. J. (2011). The policy-practice nexus of electronic health records adoption in the UK NHS: An institutional analysis. Journal of Enterprise Information Management, 24(2), 146–170.

Currie, W. L., & Guah, M. W. (2007). Conflicting Institutional Logics: A National Programme for IT in the Organisational Field of Healthcare. Journal of Information Technology, 22(3), 235–247.

Currie, W. L., & Seddon, J. J. (2022). Stakes, positions and logics: An institutional field analysis of cross-border health IT policy. Journal of Information Technology, 37(2), 165–187.

DeLanda, M. (2006). A New Philosophy of Society: Assemblage Theory And Social Complexity (Annotated edition). Continuum.

DeLanda, M. (2013). Intensive science and virtual philosophy. Bloomsbury Publishing.

DeLanda, M. (2016). Assemblage Theory (1st edition). Edinburgh University Press.

Dubois, A., & Gadde, L.-E. (2002). Systematic combining: An abductive approach to case research. Journal of Business Research, 55(7), 553–560.

Grisot, M., Hanseth, O., & Thorseng, A. (2014). Innovation Of, In, On Infrastructures: Articulating the Role of Architecture in Information Infrastructure Evolution. Journal of the Association for Information Systems, 15(4), 197–219.

Gualdi, F., & Cordella, A. (2022). Techno-legal entanglements as new actors in the policy-making process. Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2022.298

Hanseth, O. (2022). When Stars Align. The Interactions and Transformations of e-Health Infrastructure Regimes. Historical Social Research/Historische Sozialforschung, 47(3), 40–80.

Hanseth, O., & Lyytinen, K. (2010). Design Theory for Dynamic Complexity in Information Infrastructures: The Case of Building Internet. Journal of Information Technology, 25(1), 1–19.

Hanseth, O., Monteiro, E., & Hatling, M. (1996). Developing Information Infrastructure: The Tension Between Standardization and Flexibility. Science, Technology, & Human Values, 21(4), 407–426.

Hanseth, O., & Rodon Modol, J. (2021). The Dynamics of Architecture-Governance Configurations: An Assemblage Theory Approach. Journal of the Association for Information Systems, 22(1), 130–155.

Henfridsson, O., & Bygstad, B. (2013). The Generative Mechanisms of Digital Infrastructure Evolution. MIS Quarterly, 37(3), 907–931.

Jarvenpaa, S. L., & Essén, A. (2023). Data sustainability: Data governance in data infrastructures across technological and human generations. Information and Organization, 33(1), 100449.

Kempton, A. M., Grisot, M., Braa, K., & Aanestad, M. (2020). Infrastructural Tuning in Public-Private Partnerships. ECIS 2020, 16.

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

Klecun, E., Zhou, Y., Kankanhalli, A., Wee, Y. H., & Hibberd, R. (2019). The dynamics of institutional pressures and stakeholder behavior in national electronic health record implementations: A tale of two countries. Journal of Information Technology, 34(4), 292–332.

Mekonnen, S. M., & Sahay, S. (2008). An institutional analysis on the dynamics of the interaction between standardizing and scaling processes: A case study from Ethiopia. European Journal of Information Systems, 17(3), 279–289.

Paparova, D., Aanestad, M., Vassilakopoulou, P., & Bahus, M. K. (2023). Data governance spaces: The case of a national digital service for personal health data. Information and Organization, 33(1), 100451.

Parmiggiani, E., & Grisot, M. (2020). Data Curation as Governance Practice. Scandinavian Journal of Information Systems, Vol. 32(1), 38.

Rodon, J., & Silva, L. (2015). Exploring the Formation of a Healthcare Information Infrastructure: Hierarchy or Meshwork? Journal of the Association for Information Systems, 16(5), 394–417.

Rutzou, T., & Elder-Vass, D. (2019). On Assemblages and Things: Fluidity, Stability, Causation Stories, and Formation Stories. Sociological Theory, 37(4), 401–424.

Sahay, S., Monteiro, E., & Aanestad, M. (2009). Configurable Politics and Asymmetric Integration: Health e-Infrastructures in India. Journal of the Association for Information Systems, 10(5), 399–414.

Sahay, S., Sæbø, J., & Braa, J. (2013). Scaling of HIS in a global context: Same, same, but different. Information and Organization, 23(4), 294–323. https://doi.org/10.1016/j.infoandorg.2013.08.002

Tempini, N. (2017). Till data do us part: Understanding data-based value creation in data-intensive infrastructures. Information and Organization, 27(4), 191–210.

Tiase, V. L., Hull, W., McFarland, M. M., Sward, K. A., Del Fiol, G., Staes, C., Weir, C., & Cummins, M. R. (2021). Patient-generated health data and electronic health record integration: A scoping review. JAMIA Open, 3(4), 619–627. https://doi.org/10.1093/jamiaopen/ooaa052

Tilson, D., Lyytinen, K., & Sørensen, C. (2010). Research Commentary —Digital Infrastructures: The Missing IS Research Agenda. Information Systems Research, 21(4), 748–759.

Vassilakopoulou, P., Pesaljevic, A., & Aanestad, M. (2018). Polycentric Governance of Interorganizational Systems: Managerial and Architectural Arrangements. 16.

Winter, J. S., & Davidson, E. J. (2020). Harmonizing Regulatory Spheres to Overcome Challenges for Governance of Patient-Generated Health Data in the Age of Artificial Intelligence and Big Data. TPRC48: The 48th Research Conference on Communication, Information and Internet Policy.

APPENDIX V:

"Opening-up Digital Platforms to Accommodate Patient-generated Healthcare Data"

Dragana Paparova

Abstract

This paper investigates the process of opening-up digital platforms to accommodate patient-generated healthcare data (PGHD) and argues that in data platforms, barriers arise due to the entanglement of technology and policy. The empirical setting for the study is the opening up of a Norwegian eHealth platform for PGHD captured by external vendor technologies. The possibility to accommodate PGHD opens up new innovation arenas by recombining data from multiple sources and actors. However, such process is prone to a unique set of challenges when innovation is centered around data, instead of functionalities, such as: 1) open-up the data core using boundary resources; 2) control patient data across long chains of actors; 3) establish uniform rules to co-create data value. The findings show that the proves of opening up data platforms faces legislative barriers which should be overcome in a way that technology and policy enable each-other.

1 Introduction

Novel sources of patient-generated healthcare data (PGHD) captured through medical devices, sensors and smartphone apps are entering the healthcare landscape, holding the potential to transform the way patient information is "generated, collected and analyzed in healthcare practices and used in clinical decision making" (Grisot et al., 2020). PGHD can support the needs for circular interaction between patients and healthcare professionals, changing the role of patients from passive recipients to active "prosumers" (Barrett et al., 2016), as they consume and produce data using dispersed devices and under diverse circumstances. This opens up new arenas for data innovation by interconnecting patient data across a variety of actors and from dispersed sources, which can be recombined along multiple value-creation pathways.

One way of supporting innovation across multiple actors is by organizing their interactions around digital platforms. Digital platforms are underpinned by modular architectures, interconnecting core and peripheral modules using boundary resources as standardized interfaces. The boundary resources support the use and re-use of common components and can facilitate innovation on a larger scale by transferring design capabilities to external actors (Ghazawneh and Henfridsson, 2013). However, opening up the boundary resources can also lead to greater fragmentation and loss of control over the emerging innovation network in the platform periphery (Gawer and Cusumano, 2014). Therefore, the boundary resources need to be governed in a way that balances the trade-offs between expanding the platform with complementary components, but at the same time keeping control by setting up uniform rules, standards and shared institutional logics which govern the behavior of distributed actors (Autio and Thomas, 2020; Eaton et al., 2015)

In data platforms instead, the boundary resources connect the core and periphery, but those relations are established based on digital data, rather than functionalities, tools or applications (Tempini, 2017). The common core modules are full of data, and the external actors innovate with these data in the platform periphery (Bonina and Eaton, 2020). However, data are not components and do not embody functions in the same way as components do (Alaimo et al., 2020). Rather, data are captured as events, computed into tokens and then assigned meaning when they are used by actors, due to their semantic nature. Therefore, innovation in data platforms does not necessarily follow the same recombinant logic of modular architectures, as it happens at a more granular levels than assembling together a set of components. Although data innovation does not happen independently of the components, actors recontextualize these data across their value-creation processes and assign them meaning, rather than constructing a functionality (Aaltonen et al., 2021). Thus, data innovation distinguishes from recombinant innovation with modules, as it takes place in the way data is ported and used across actors' value-creation trajectories (Alaimo et al., 2020).

Research by Grisot et al. (2020) already shows that the re-combination of components can lead to multiple alternative value pathways around PGHD by aligning the underlying digital infrastructure and work practices to tailor the diverse needs for data exchange between patients and healthcare professionals. Besides this architectural perspective, research also reveals the complexity of data-based value-creation due to their use-agnostic character. For example, Tempini (2017) shows the multidimensionality of PGHD in the value-creation trajectories of actors organized around a social media platform. Similarly, Barrett et al. (2016) follows the evolution of an online community, and discloses the tensions stakeholders face as their goals and the meaning they assign to data change over time. However, while previous research has been focusing on the generation of patient data using approved devices in an outpatient clinic (Grisot et al., 2020), and in peer-support communities (Barrett et al., 2016; Tempini, 2017), I hereby put my focus on the process of opening up the boundary resources to support larger scale innovation centered around PGHD from devices and technologies developed by external actors.

The research question I seek to address is: what are the barriers in opening-up digital platforms to accommodate patient-generated healthcare data? The empirical case follows the Norwegian national eHealth platform and reveals barriers in the process of opening up the data core for PGHD generated by external actors' components. The paper is organized as follows. In the methods section, I elaborate on the methodology used, introduce the case and explain the data analysis process. Next, I introduce the main findings of the study and elaborated on them in more details in the discussion section. At last, I present the main conclusions of the conducted study.

2 Research Methodology

2.1 Case description

The empirical setting for this study is the Norwegian healthcare context. The case is conducted by including: 1) a national eHealth platform in Norway "HelseNorge", which undergoes a process of opening up its boundary resources for external development and accommodate PGHD; as well as 2) three private vendors solutions for PGHD, which are part of the public digital ecosystem, but are not integrated with the national platform. HelseNorge provides citizens access to information stored about them in several health registries in the public sector. The platform was launched June 15th 2011, driven by the need to create a single point of entry portal for citizens, instead of letting them "search Google for health" across many websites. The platform has grown considerably throughout the years, integrating with several systems across the public infrastructure, including 11 approved digital health apps which are part of its tool catalogue. Currently, HelseNorge is undergoing a process of opening up its boundary resources for external development and incorporate PGHD as part of its ecosystem (Directorate for eHealth, 2019).

The case also incorporates three digital health tools for PGHD. 1) Mobile medical record system which is built jointly for patients and healthcare professionals and shares medical patient data using secure messaging, video consultation, photos, forms and other health information, stored in the cloud. 2) Shared patient diary, a solution for information exchange between healthcare professionals, service recipients and their relatives. The app shares information in the form of text, images, and video includes a common calendar for real-time updates and combines both medical and lifestyle data about a patient. 3) Outpatient clinic tool, used for specialist care which collects structured medical data about patients, stored systematically in hospital systems, from where healthcare professionals can extract the data and monitor patient's health status.

2.2 Data collection and analysis

The study was explorative and conducted using a qualitative method (Sarker et al., 2018a, 2018b). The data was gathered via 10 semi-structured interviews, using a snowball approach where participants recommended potential suitable candidates further. 7 interviews were conducted with representatives from the national HelseNorge platform, and 3 interviews with private vendors. To gain more insights into the context, online information, including websites, presentations, and strategy documents were also collected. The data gathering process lasted for 5 months, June-Nov 2020, although the case was followed in retrospect dating back to 2010, when the development of the national eHealth platform started to take place. Participants were technical and managerial staff, working with the national eHealth platform and private software vendors. The background of participants varies between: software architects, software developers/medical doctors, lawyers, consultants, data scientist, providing a wider perspective over the cross-disciplinary nature of the research problem. The interviews lasted for approximately one hour and were afterwards transcribed to analyze the data. The data gathering process was guided by the research interest on capturing and interpreting informants' meanings, and their understanding of the decisions that need to take place towards opening up the national platform to incorporate PGHD (Dubois and Gadde, 2002). The interview guides included questions on the challenges for extending HelseNorge's core functionalities with external solutions, the role of policy in making decisions about incorporating PGHD within the ecosystem and opinions on what is the way forward towards integrating PGHD as part of the ecosystem.

The data was analyzed in an abductive and iterative way (Alvesson and Sköldberg, 2009; Dubois and Gadde, 2002). The data collection and analysis were informed by existing theoretical concepts in the digital platforms literature (Aanestad et al., 2017; Ghazawneh and Henfridsson, 2013), which were used to initially define and re-articulate the research problem (Dubois and Gadde, 2002). The theoretical concepts informed the data gathering process, and were used to categorize the empirical data, coordinate the findings, as well as to direct and redirect the study as new insights emerged. The analysis moved iteratively between asking questions, generating the findings, making comparisons with existing knowledge and refining this again. The data was coded to establish categories for grouping the information gathered from the empirical work (Maxwell and Miller, 2008) and analyzed from the perspective of the platform owner. The findings were first organized around 6 key decisions of opening up the platform for PGHD and grouped around 3 barriers which need to be overcome throughout this process. Although the data collection was triangulated with information from official documentation and strategy documents by the Directorate of eHealth, these documents were only used as a first step to provide a contextual understanding over the case, and to inform the data collection process. However, they were not systematically analyzed to generate the findings of the study, as the findings were bases solely on the empirical data.

3 Findings

The findings are organized around three main barriers for opening up the digital platform to accommodate PGHD captured through external components, defined as: 1) open up the data core using boundary resources; 2) control patient data across long chains of actors; 3) establish uniform rules to co-create data value. The findings are elaborated in more details as follows.

3.1 Open-up The Data Core Using Boundary Resources

HelseNorge.no was launched on 15th of June 2011, as a “Citizen Portal” working as a single point of entry for Norwegian citizens, which were previously “searching Google for health” across many websites. At the beginning some team members thought that platform-thinking could be beneficial in the long-run. However, this vision was abandoned due to the pressing deadline for delivering the first version and all functionalities were developed in-house. After the initial launch, discussions have been on-going on what functionalities should be supported next, what parts should the core connect with and which components should be developed externally. So far, decisions about expanding HelseNorge were supported when the team would identify a functionality they need and then 1) build it inside; 2) get a third-party vendor to build it; 3) or integrate with an existing vendor solution on the market providing these functionalities. Such an approach has resulted with many vendor-specific APIs adapted to the requirements and functionalities of the external solutions. What is currently set in place for vendors who want to connect with HelseNorge is a requirements list for certain areas, such as: video conferencing, appointments, journal patient health record, message exchanges between healthcare professionals and patients. Therefore, HelseNorge as of now has a very limited set of APIs exposed to third-parties, very few consumers that send data packages to it and even fewer which retrieve data and are creating solutions based on access to data from the core.

The lack of published APIs has resulted with some of the vendors having to "self-resource" boundary resources themselves in order to enter the ecosystem. Others decided to step back from the integration process, due to delays in API provision, followed by bureaucracy and large documentation. "It would be much easier if the system was ready to share APIs with the private companies like us, if there was a system and rules that are there for that and APIs. We have a feeling that we have to fight for every access we get. (…) If there were more information and advice from the government on how to do things and more APIs ready from the beginning, then we would do things differently and we would save a lot of resources, money and time." (Informant, Vendor 1). Decisions about opening up HelseNorge are currently made around two options. 1) Contained environment, where smaller vendors who need a stable platform to support them can re-use a lot of the core capabilities, with lower thresholds for innovation and fast innovation cycles. 2) Uncontained environment, in which external actors have a stable standalone platform, build all the functionalities themselves and connect to HelseNorge. However, the technical side of publishing APIs is considered "the easiest", as challenges arise since APIs need to validate the party which gets access to patient data, verify the user and make sure that patient data is handled in accordance to policy and law. Therefore, these APIs need to work as digitalized contracts between multiple parties which do not only guide what functionalities to be developed, but also what data elements are processed, and how to regulate the behavior of multiple parties across the ecosystem.

3.2 Control Patient Data Across Long Chains of Actors

Initially, data in HelseNorge was stored in a single storage solution called the Personal Health Archive (PHA). Although there were discussions about data being stored with responsible entities interacting with each-other, versus laying in a single storage, the sense for platform architecture which should do connections instead of storage was not there from the beginning. As HelseNorge started integrating with GP ERP and hospital EHR systems, patient data was getting exchanged outside of HelseNorge's control. Therefore, decisions had to be made on how to govern patient data exchanged across multiple systems. It was decided that once citizens are handed over to another party, the respective systems are the owners of this data and take the responsibility as a data controller. Therefore, HelseNorge only provides access to view this data, and in some cases stores a copy in the PHA. “Many people think ‘Helsenorge knows a lot about me’, but we are not allowed to read this data. (…) HelseNorge is saying ‘we know that you have some data at the Health Trust, we cannot open it, but we can help you see it’, so we cannot snatch the information on the way to the user” (Informant, HelseNorge). Currently, a lot of the data provided by HelseNorge is not owned by the platform. The platform is organized as consent-based and takes responsibility as a data processor only for the data processed inside its core components. Patient consent has to be registered in HelseNorge and the patient can stop sharing at any time. Although citizens can choose to use different digital health tools via HelseNorge, they have to accept the terms of use in the particular tool. This is due to the lack of control HelseNorge has over how patient data is handled throughout the entire chain of actors and whether this process is legal all the way.

One possible approach towards controlling the chain is the aspiration of the HelseNorge team for having a “dedicated HelseNorge law”, which states that the platform is the official national provider of healthcare services. In the absence of law which might provide an independent basis for data processing, they have to find complex solutions within the restrictions given by legal regulations. The lack of control has made it challenging to keep consent valid at all time and make sure that it is still within the scope of the time patients provided it. “What I have experienced lately is that you should not completely rely on the definition of the terms in the GDPR, with regards to who is the controller, but you have to look at the whole chain to be able to see what is there, what is in factual circumstances that the working part has expressed. You have to see who is in the factual situation the closest to take the control or responsibility in a complex chain.” (Informant, HelseNorge). Therefore, it is challenging to provide clarity of responsibilities when data is exchanged across many actors who often fail to understand what their legal responsibility is and what the terminology means.

As data is re-written and re-copied across multiple systems, the data also gets stored in decomposed solutions across these long chains. Although the team is assertive that “the time has passed for single storage solutions”, all parties wish access to data, but they all avoid the responsibility to store it due to the strict legislation. Smaller vendors either keep data in the cloud to enhance scalability or store it on a third-party server and public actors do not want to use data kept in a storage they cannot trust. “If the patients think that this is critical data and upload it in a storage, they think that someone will look into this. But that is not the case if you do not have an agreement in it, because no doctor will look into it before there is an appointment, or a reason for looking at it. So, it is a dangerous misconception if the patient thinks that 'I gave this to the healthcare service in Norway, someone should react if something is wrong' (Informant, HelseNorge). Therefore, discussions are on-going on whether HelseNorge should provide a storage solution for vendors who do not want to store data themselves, and at the same time allow flexibility for larger vendors who want to store and keep data in their own storage solution.

3.3 Establish Uniform Rules to Co-create Data Value

Decisions about opening up HelseNorge also need to encompass establishing criteria on which PGHD captured by external vendors are allowed into the platform ecosystem. Such decisions need to be informed by setting universal rules and standards which reflect patient data privacy and security in the digital health tools developed by external vendors. At the time HelseNorge was launched, most of the standards which are in use today were either non-existing or premature. From 2016, the team started working with HL7 FHIR with the aim to standardize data exchange by not bringing in too much data, but keep it as small as possible, yet still within the clinical context. However, such standards are not mandatory for all vendors in the public infrastructure, leading to lack of understanding on what standards do different solutions for PGHD have to comply with. "They {the municipality} said you need to have high security level, because it is sensitive data, and when we did and lost all of our users, I remember I called {the person} and I said: 'you asked us to do this, and now you have to tell your workers that it is safe to log in with bankID, because we are losing everyone'." (Informant, Vendor 2).

One way of dealing with this fragmented portfolio is for the platform to establish universal criteria for evaluating digital health tools for PGHD. In 2015, the Directorate of eHealth started working on a framework for assessing digital health tools to make sure they bring benefit to the healthcare service and are safe to use in clinical practice, but such attempts have been dropped. What is currently set in place as a "screening procedure" to assess the external tools are testing and approvement queries to verify that the tool is compliant with GDPR as well as follows the Code of conduct for information security in the health sector in

Norway. The tools also need to get a confirmation from the Directorate of eHealth that the content is clinically responsible. After that, they sign an agreement for third-party data processing. HelseNorge also has a publicly published tool catalogue as a library of digital health tools which can be prescribed to the citizens. However, the aim is for privacy and security to be embedded across all digital health tools in the ecosystem, instead of assessing individual cases. One-way forward is to provide self-declarations issued by an authenticated governmental body which approves the use of the PGHD tool and finds it to be trustworthy for patient treatment. However, as of yet, such process is still not set in motion.

4 Discussion

Research already shows that innovation facilitated by digital platforms can result with "unpredictable innovative contributions by large, uncoordinated audiences" (Autio and Thomas, 2020), but this paper goes further by showing the added complexity of opening up when the re-combinative innovation is centered around data, instead of functionalities. Innovation in data platforms starts with boundary resources, but the way core patient data are used is more difficult to control due to their semantic nature. This shifts the focus of data innovation from the modular architectures in which they are generated, towards the ways actors use them across their value-creation trajectories. The findings show that the combined effects of innovating with functionalities and data (Tempini, 2017) challenge the process of opening up the boundary resources, as data also bring in the emergent role of policy and legislation. In data platforms, the opening up of boundary resources does not solely mean transferring design capabilities to external actors (Ghazawneh and Henfridsson, 2013), but they also need to manifest themselves as "invisible data rights" in the background. This creates difficulties in regulating access over data and controlling how data is used and re-used by external actors in the platform periphery.

Once data leave the core, they can be re-copied across multiple actors which assign them meaning. Due to data's use agnostic nature, it is hard to track the data interactions across all actors, as the owners of the components are not necessarily the owners of data. In actors' value trajectories, data decouple from the components that carry them and can be re-copied and assigned meaning on top of those components (Aaltonen et al., 2021; Alaimo et al., 2020). Thus, data can be re-used across long chains of actors where the partitioning of data rights is not always as straightforward once data leave their natural source. Instead, there is a need to look at the actual chains to determine who holds responsibility for what and how data rights are dispersed. Current laws such as GDPR do provide certain clarity on the roles of data controllers and processors, but when opening up for innovation on a large scale, data do not always flow from one actor to another, but across long chains of actors. This brings a new set of challenges on how to regulate the patient data use and keep the chains legal all the way.

The inability to control these long chains works as a barrier to support external data innovation. The lack of control rises as it is hard to determine a-priori how data will be used and track the interactions of data across actors' value creation trajectories. Data are not components and do not embody functions, but are recombinant resources which acquire meaning as they are collected, stored and used by actors (Alaimo et al., 2020). The way data are used, recombined and aggregated can trigger a new set of interactions which are not reflected in the existing law. Therefore, instead of thinking about complying with legislation when opening up the boundary resources in a linear way, the actors also need to make sure that the process of data use and re-use is legal across the chains in an on-going manner. Such a turbulent environment requires that new laws are created to provide more clarity, as well as old laws are constantly revised to reflect the changed circumstances. Therefore, the lack of clarity on how to reflect legislation in complex environments where data is exchanged across multiple actors, works as a constraint to data innovation.

Ensuring compliance with legislation and controlling the long chains can also be facilitated by setting up upfront criteria on which data bring value to the platform and its ecosystem. Such criteria can orchestrate actors by making sure that verified technologies, provided by legitimate actors catch relevant data which is of value to the healthcare service. However, the lack of universal criteria and legislative instruments to regulate the ecosystem actors in such an automated way, also suppresses the opportunities for patient data innovation on a larger scale. This paper contributes to the literatures on PGHD and digital platforms by showing that although platforms' modular architectures can facilitate the process of accommodating PGHD in digital platforms, innovations centered around data entangle technology and policy. Although the paper shows that accommodating

PGHD in eHealth data platforms may be premature at the stage, the potential is promising towards that direction.

5 Conclusion

This paper investigates the process of opening-up an eHealth platform to accommodate PGHD captured through external vendor technologies and shows that barriers arise due to the entanglement of technology and policy. The data-intensive environment brings in an increasing complexity in regulating data use and re-use across the long-chains of actors in the periphery, which suppress the innovation potential with patient data in the platform periphery. This research also has certain limitations, as the empirical study is based on a limited number of interviews and the case chosen is one where PGHD from external technologies is still not accommodated in the digital platform. Further research can go beyond by showing how to tackle the interplay of technology and policy and do that in a way that enables innovation with PGHD in data-intensive environments.

References

Aaltonen, A., Alaimo, C., Kallinikos, J., 2021. The Making of Data Commodities: Data Analytics as an Embedded Process. Journal of Management Information Systems 47.

Aanestad, M., Grisot, M., Hanseth, O., Vassilakopoulou, P., 2017. Information Infrastructures and the Challenge of the Installed Base, in: Aanestad, M., Grisot, M., Hanseth, O., Vassilakopoulou, P. (Eds.), Information Infrastructures within European Health Care, Health Informatics. Springer International Publishing, Cham, pp. 25–33.

Alaimo, C., Kallinikos, J., Aaltonen, A., 2020. Data and value. Handbook of Digital Innovation.

Alvesson, M., Sköldberg, K., 2009. Reflexive Methodology: New Vistas for Qualitative Research, Second edition. ed. SAGE Publications Ltd, Los Angeles ; London.

Autio, E., Thomas, L.D.W., 2020. Value co-creation in ecosystems: insights and research promise from three disciplinary perspectives. Handbook of Digital Innovation.

Baldwin, C.Y., Woodard, C.J., 2009. The Architecture of Platforms: A Unified View. Platforms, markets and innovation.

Barrett, M., Oborn, E., Orlikowski, W., 2016. Creating Value in Online Communities: The Sociomaterial Configuring of Strategy, Platform, and Stakeholder Engagement. Information Systems Research 27, 704–723.

Bonina, C., Eaton, B., 2020. Cultivating open government data platform ecosystems through governance: Lessons from Buenos Aires, Mexico City and Montevideo. Government Information Quarterly 37, 101479.

Directorate for eHealth, 2019. Sharing health data with the patient.

Dubois, A., Gadde, L.-E., 2002. Systematic combining: an abductive approach to case research. Journal of Business Research 55, 553–560.

Eaton, B., Elaluf-Calderwood, S., Sørensen, C., Yoo, Y., 2015. Distributed Tuning of Boundary Resources: The Case of Apple's iOS Service System. MIS Quarterly 39, 217–244.

Gawer, A., Cusumano, M.A., 2014. Industry Platforms and Ecosystem Innovation: Platforms and Innovation. J Prod Innov Manag 31, 417–433.

Ghazawneh, A., Henfridsson, O., 2013. Balancing platform control and external contribution in third-party development: the boundary resources model: Control and contribution in third-party development. Information Systems Journal 23, 173–192.

Grisot, M., Lindroth, T., Islind, A., 2020. Digital Infrastructures for Patient Centered Care: Examining Two Strategies for Recombinability. pp. 289–300.

Maxwell, J.A., Miller, B.A., 2008. Categorizing and connecting strategies in qualitative data analysis, in: Handbook of Emergent Methods.

Sarker, S., Xiao, X., Beaulieu, T., Lee, A.S., 2018a. Learning from First-Generation Qualitative Approaches in the IS Discipline: An Evolutionary View and Some Implications for Authors and Evaluators (PART 2/2). JAIS 19, 909–923.

Sarker, S., Xiao, X., Beaulieu, T., Lee, A.S., 2018b. Learning from First-Generation Qualitative Approaches in the IS Discipline: An Evolutionary View and Some Implications for Authors and Evaluators (PART 1/2). JAIS 19, 752–774.

Tempini, N., 2017. Till data do us part: Understanding data-based value creation in data-intensive infrastructures. Information and Organization 27, 191–210.

APPENDIX VI:

"Governing Innovation in E-health Platform Ecosystems – Key Concepts and Future Directions"

Dragana Paparova, Margunn Aanestad

Abstract

This paper conceptualizes knowledge in the IS literature on governing innovation in platform ecosystems using boundary resources. Platform innovation arises when platform owners realize the need to expand the functionalities and invite external actors with specialized knowledge to do so. We conduct a literature review to identify the relevant concepts on governing innovation in platform ecosystems in IS and adapt them to the specific settings of the eHealth context. As most relevant concepts, we identify: 1) boundary resources as governance mechanisms: openness vs. control; 2) co-creating platform innovation across heterogeneous actors: accommodation vs. resistance; and 3) platform innovation within the underlying architecture: stability vs. flexibility. We then derive areas that should be prone to further research in eHealth, defined as: 1) patient data as a resource for eHealth platform innovation; 2) the role of institutions in eHealth platform innovation; and 3) innovating within platform-oriented eHealth information infrastructures. This paper contributes with expanding the understanding in the current state of knowledge in IS and provides basis for further research adapted to the eHealth context settings.

Keywords: platform innovation, boundary resources, third-party development, eHealth

1 Introduction

The healthcare sector is of central societal importance and has been a central empirical domain also for IS scholars (Fichman et al., 2011). Currently, the digital transformation in healthcare is driven by an increasingly central role for patients, parallel to the "consumerization" that accompany digital transformation in other sectors (Agarwal et al., 2010). Patient-centric healthcare systems seek to empower and engage patients to care for their own health, and this is enabled by the growth of easily available and affordable medical devices and software to work with smartphone applications, welfare technologies and wearable devices. This transformation involves not only patients themselves as more central actors, but also novel technology actors, both in the device, software and analytics industries. These actors are not necessarily part of the established health IT landscapes, and the need for harnessing the innovation potential from this segment has increased the interest of understanding the role of platforms as stimulators for third-party innovation.

The importance of platforms lies in their capabilities to enable modularization, where functionality is distributed between the core and complementing modules provided by third-party actors (Karhu et al., 2018). In this paper, we rely on Tiwana's (2013) definition of platforms as the "extensible codebase of a software-based system that provides core functionalities shared by apps that interoperate with it, and the interfaces through which they interoperate" (Tiwana, 2013). Platform architectures have been found to facilitate generative third-party development (Ghazawneh and Henfridsson, 2010; Eaton et al., 2015, de Reuver and Sørensen, 2018), but the generative potential still needs to accompany a certain level of control over the platform externalities.

Platform innovation emerges when platform owners realize the need to extend the platform functionalities in an area they are not specialized in (Bygstad, 2015) and invite third-party actors to contribute with components which have not been foreseen in the initial platform design phase (de Reuver et al., 2020, 2018). Although this extension is enabled by boundary resources which orchestrate the development of modules in the platform periphery (Tiwana, 2013), it can still result with ill-performing apps (de Reuver et al., 2020), large fragmentation of functionalities that are not compliant with the overall vision of the platform

(Bygstad, 2015), or lead towards platform lock-in, where the core becomes increasingly dependent on third-party functionalities. To avoid such scenarios, platform owners need to find the proper balance between using boundary resources to "open up" the platform to external actors, while also exercising control over third-party innovations in the periphery.

Third-party innovation becomes increasingly intermingled in eHealth platform ecosystems, where the partitioning of decision rights between platform owners and third-parties is not always clear, as diverse stakeholders appear on both sides of the platform – the core and the periphery. Due to the divergence of institutions involved at a national, regional and local level, multiple governmental bodies with overlapping jurisdictions, as well as private third-party vendors as part of the ecosystem, governance decisions spread across the entire healthcare information infrastructure, and not just within standalone platforms. Such differences impose the need to adjust decision-making about innovation in eHealth platform ecosystems towards addressing the socio-technical complexity of the healthcare context settings.

We stick to defining boundary resources as "the software tools and regulations that serve as interfaces for the arm's-length relationship between the platform owner and the application developer" (Ghazawneh and Henfridsson, 2013). Such resources can consist of, but are not limited to: application programming interfaces (APIs), software development kits (SDKs), contract agreements, app distribution channels, and similar tools that increase the value for third-party developers.

This paper aggregates relevant conceptualizations on governing third-party innovation in eHealth platform ecosystems and adapts them to the healthcare context. While we find previous literature from all disciplines as fundamental and relevant, the purpose of the paper is to identify the current state of knowledge in IS and understand the distinct settings of eHealth platform ecosystems. Our research question is: what are the conceptual approaches of governing innovation in eHealth platform ecosystems? To answer this question, we conducted a systematic literature review to summarize the relevant concepts that are present in the IS literature and adjust them to the healthcare settings.

For the purpose of this paper, we put our focus on the technology-oriented perspective of platforms, where platforms are defined as “a set of stable components that support variety and evolvability in a system by constraining the linkages among the other components” (Schreieck et al., 2016). Such a perspective provides an understanding of the distinct settings of eHealth platforms, by not treating them as multi-sided markets connecting buyers and sellers (de Reuver et al., 2018), but as ”coordinating devices among innovators” (Gawer, 2014). Anyhow, we do not exclude publications discussing the relevant concepts from a market-oriented perspective, as we acknowledge that these perspectives interact and should not be considered in isolation.

This review will be useful to researchers in IS and eHealth as a basis for conceptualizing the governance of platform ecosystems using boundary resources as facilitators for innovation. Our findings provide useful insights for practitioners in both, the public and private sector for making more informed decisions about governing platform ecosystems. The paper is organized as follows. In the next section, we describe our literature review design and paper selection process. Moving further, we highlight the relevant concepts in the IS field up to date and apply them to eHealth. In the discussion, we adjust the concepts to the eHealth context and point towards areas that should be prone to further research. At last, we highlight the contributions and limitations of our study.

2 Literature Review Design

To understand the relevant concepts on governing innovation in eHealth platform ecosystems, we conducted a systematic literature review (Webster and Watson, 2002) using a hermeneutic approach (Boell and Cecez-Kecmanovic, 2014), to aggregate present knowledge in the IS field. We searched the following databases: Web of Science (Web of Knowledge) and AIS Electronic Library (AISeL) to retrieve relevant publications.

Initially, we searched for variations by combining keywords within the following keyword groups: “boundary resources”, “third-party development”, “platform innovation” and combined them with “eHealth”. Since the results retrieved a low volume of articles on eHealth platform innovation and also publications which we considered irrelevant to our research focus, we decided to extend the search by not limiting it to the eHealth context. That way, we got a broader perspective on the relevant concepts investigating platform innovation in the IS literature and contemplated it with publications focusing especially on the eHealth field.

We included publications that were peer reviewed, written in English and published in journals, conferences and books in the following date range 2000-2020. We excluded publications that were duplicates and irrelevant to the keywords and research area. On Web of Science, we additionally refined our search to include only the following document types: article, abstract of published item, book and book chapter. The search was limited to the following categories: computer science information systems, computer science interdisciplinary applications, multidisciplinary sciences, medical informatics, health care sciences services, management and business.

Our keyword search retrieved 649 publications. The main selection process involved two rounds. In the first round, we primarily selected publications based on their title, abstract, and keywords. At this stage, we only included publications that addressed all three keyword groups simultaneously, as defined above: “boundary resources”, “third-party development”, and “platform ecosystem”. Such an approach helped us use boundary resources as the main unit of analysis for conceptualizing platform innovation (de Reuver et al., 2018). Therefore, we eliminated publications that focused on discussing these keyword groups in isolation. At this point, we also eliminated articles discussing the concepts in complementary fields, such as: computer science, hardware computing, or software engineering.

We suspected that some publications may not have used “boundary resources”, “third-party development”, or “platform ecosystem” as exact keywords in the title, abstract, or keywords, but are still discussing these concepts in the main text. Thus, in the second round, we inspected the full texts and searched for all three keywords and their variations across the publication. This way we made sure to not eliminate relevant publications that use synonyms, and yet investigate the concepts of our interest. At this stage, we eliminated articles where the primary focus is on complementary issues, such as patient-generated healthcare data, integration of silo heavyweight systems, information infrastructures, artificial intelligence, machine learning.

Database	Search Keywords	Hits	Selected	Final
Web of Science	((“boundary resource” OR “boundary resources” OR “application programming interfaces” OR “API” OR “APIs” OR “SDK” OR “SDKs” OR “software development kits” OR ”third-party development” OR ”third party development” OR ”third-party applications” OR ”third party applications” OR “third-party developers” OR “third party developers” OR "lightweight technology" OR "lightweight technologies") AND (“platform” OR “digital platform” OR “digital platforms” OR “ehealth platform” OR “eHealth platform” OR “ehealth platforms” OR “eHealth platforms” OR “platform innovation” OR "platform eco-systems" OR "platform ecosystems" OR "platform eco-system" OR "platform ecosystem"))	379	16	10
AISEL		270	5	3
Backwards citations	/	/	/	9
Papers known to us	/	/	/	14
Summary				36

Table 1. Summary of the literature review process.

Therefore, we ended up with 21 publications that were to be read in full and ranked based on their relevance. To enrich this review with a solid volume of relevant publications, we further augmented 9 articles found via backwards research and additional 14 publications known to us from previous work. After reading them in full, we eliminated 8 articles that we considered irrelevant to our research focus and ended up with 36 final publications that are of interest in this literature review. The literature review process is summarized in Table 1.

We used an inductive coding process to code the retrieved publications based on the data provided from the search. Even though we chose this exploratory approach, we had previous knowledge on the topic to guide us from the start. Based on the retrieved publications, three main concepts emerged. Across the three concepts, we also identified tensions which encompass decisions about governing platform innovation.

The first concept focuses on boundary resources as governance mechanisms for platform innovation, investigating the tensions between openness and control. The second concept explores the role of heterogeneous actors in co-creating innovation in platform ecosystems, through the tension of accommodation and resistance to change. The third concept provides an understanding on supporting platform innovation within the underlying architecture, shaped by the tension of stability and flexibility.

We also added two additional parameters to understand the context of the three concepts. The first parameter regarded whether the publications discussed the relevant concepts in the general IS field, or in the eHealth context. The second parameter distinguishes between the main unit of analysis in the publication, distinguishing between: 1) boundary resources; 2) platform ecosystems; or 3) platform architecture. Although all publications encompass all three units of analysis, this coding shows which is the predominant one. The summary of publications included and their coding are represented in Table 2 below.

Publication Title	Authors	Outlet	eHe-alth	Unit of analysis		
				BR	PE	PA
JOURNAL PUBLICATIONS						
The Digital Platform: A Research Agenda	de Reuver et al. (2018)	JIT			X	
Distributed Tuning of Boundary Resources: The Case of Apple's iOS Service System	Eaton et al. (2015)	MIS Quarterly		X		
Complementors as connectors: managing open innovation around digital product platforms	Hilbolling et al. (2020)	R&D Management		X		
Digital platform ecosystems	Hein et al. (2019)	Electronic Markets			X	
Coherence or flexibility? The paradox of change for developers' digital innovation trajectory on open platforms	Brunswicker and Schecter (2019)	RP			X	
Configurations of platform organizations: Implications for complementor engagement	Saadatmand et al. (2019)	RP			X	
Cultivating Third Party Development in Platform-centric Software Ecosystems: Extended Boundary Resources Model	Msiska (2018)	AJIS	X	X		
Innovation, Openness, and Platform Control	Parker and Van	Management Science			X	

	Alstyne (2018)					
Differential effects of formal and self-control in mobile platform ecosystems: Multi-method findings on third-party developers' continuance intentions and application quality	Goldbach et al. (2018)	Information & Management			X	
P for Platform. Architectures of large-scale participatory design	Roland et al. (2017)	SJIS	X			X
Design and governance of eHealth data sharing	Vesselkov et al. (2019)	CAIS	X		X	
Doing Infrastructural Work: The Role of Boundary Objects in Health Information Infrastructure Projects	McLoughlin et al. (2016)	SJIS	X		X	
Innovation Of, In, On Infrastructures: Articulating the Role of Architecture in Information Infrastructure Evolution	Grisot et al. (2014)	JAIS	X			X
Architectural alignment of process innovation and digital infrastructure in a high-tech hospital	Bygstad and Øvrelid (2020)	EJIS	X			X
Balancing platform control and external contribution in third-party development: the boundary resources model: Control and contribution in third-party development	Ghazawneh and Henfridsson (2013)	ISJ		X		

Appraising the impact and role of platform models and Government as a Platform (GaaP) in UK Government public service reform: towards a Platform Assessment Framework (PAF)	Brown et al. (2017)	GIQ			X	
Open Platform Strategies and Innovation: Granting Access vs. Devolving Control	Boudreau (2010)	MSJ		X		
Platform Desertion by App Developers	Tiwana (2015)	JMIS			X	
Research Commentary—Platform Evolution: Coevolution of Platform Architecture, Governance, and Environmental Dynamics	Tiwana et al. (2010)	ISR			X	
On The Roles of APIs in the Coordination of Collaborative Software Development	de Souza and Redmiles (2009)	CSCW		X		
Co-Creating Platform Governance Models Using Boundary Resources: a Case Study from Dementia Care Services	Farshchian and Thomassen (2019)	CSCW	X			X
Between Personal and Common: the Design of Hybrid Information Spaces	Vassilakopoulou et al. (2018)	CSCW	X			X

Technology Ecosystem Governance	Wareham et al. (2013)	OSJ			X	
CONFERENCE PUBLICATIONS						
Collaborative Innovation in Healthcare: Boundary Resources for Peripheral Actors	Aanestad et al. (2019)	ICIS	X	X		
Building National eHealth Platforms: the Challenge of Inclusiveness	Vassilakopoulou et al. (2017)	ICIS	X		X	
Governing third-party development through platform boundary resources	Ghazawneh and Henfridsson (2010)	ICIS		X		
Governing eHealth Infrastructures: Dealing with Tensions	Bygstad and Hanseth (2016)	ICIS	X			X
Infrastructural tuning in public-private partnerships	Kempton et al. (2020)	ECIS	X		X	
The Coming of Lightweight IT	Bygstad (2015)	ECIS	X		X	
Design and governance of platform ecosystems – key concepts and issues for future research	Schreieck et al. (2016)	ECIS			X	
Extending eHealth Infrastructures with Lightweight IT	Øvrelid and Bygstad (2016)	SCIS	X		X	

Innovation Readiness in Healthcare Information Infrastructures: Key Resources to Enable Collaborative Digital Innovation	Aanestad and Vassilakopoulou (2019)	SHI	X			X
BOOK PUBLICATIONS						
Platform Governance	Tiwana (2014)	Book chapter, Elsevier			X	
The Architecture of Platforms: A Unified View	Baldwin and Woodard (2008)	Book chapter, MIT				X
Maintaining the Pharmacy Model: The Catalan Electronic Prescription Infrastructure	Modol (2017)	Book chapter, Springer	X		X	
The Swedish Patient Portal and Its Relation to the National Reference Architecture and the Overall eHealth Infrastructure	Sellberg and Eltes (2017)	Book chapter, Springer	X		X	

Table 2. Summary of publications, authors, outlets, context and unit of analysis (BR – boundary resources; PE – platform ecosystem; PA – platform architecture).

3 Findings

Our review identifies three conceptualizations on governing innovation in platform ecosystems and couples them with designated tensions, as follows: 1) boundary resources as governance mechanisms: openness vs. control; 2) co-creating innovation across heterogeneous actors: accommodation vs. resistance; and 3) platform innovation within the underlying architecture: stability vs. flexibility. We have assigned these concepts to a particular scope across the platform ecosystem, respectively: 1) platform governance; 2) platform ecosystem; and 3) platform architecture. The findings are summarized in Table 3 and the concepts are described in more details as follows.

Scope	**Lens**	**Concept**	**Tension**	**eHealth extension**
Platform governance	Socio-technical	Boundary resources as governance mechanisms	Openness vs. Control	Patient data as a resource for eHealth platform innovation
Platform ecosystem	Socio-technical	Co-creating innovation across heterogeneous actors	Accommoda-tion vs. Resistance	The role of institutions in eHealth platform ecosystem innovation
Platform architecture	Technical	Platform innovation within the underlying architecture	Stability vs. Flexibility	Innovating across platform-oriented eHealth information infrastructures

Table 3. Summary of the findings, identified concepts, tensions and eHealth extension.

1) Boundary resources as governance mechanisms: Openness vs. Control. Within the platform governance scope, there is an ongoing tension over the optimal level of openness and control using boundary resources as governance mechanisms for platform innovation. We find this tension to be of central importance in the IS literature. Some researchers investigated it by looking at private platforms of

dominant players in the IT industry (Eaton et al., 2015; Ghazawneh and Henfridsson, 2013, 2010), but this tension was also investigated in eHealth where boundary resources allocate openness and control across eHealth platform ecosystems (Modol, 2017; Sellberg and Eltes, 2017).

2) Co-creating innovation across heterogeneous actors: Accommodation vs. Resistance. Within the platform ecosystem scope, we have identified the tension of heterogeneous actors resisting to changes and accommodating them. In IS this tension was introduced by following the evolution of boundary resources in private platforms where the distinction between platform owners and third-party developers is clearly defined (Eaton et al., 2015). In eHealth, the tension was adapted to the larger information infrastructure (Aanestad et al., 2019; Kempton et al., 2020; Vassilakopoulou et al., 2017), where configurations of boundary resources were shifting across a more complex set of actors, such as: governmental bodies, institutions and private third-party vendors, which are not necessarily individual developers.

3) Platform innovation within the underlying architecture: Stability vs. Flexibility. Within the platform architecture scope, we identified the tension of keeping stability in the interfaces, while enabling flexibility in the periphery. We found the concept of loosely-coupled architectures to be one solution towards balancing this tension in both the IS and eHealth field (Bygstad and Hanseth, 2016; Grisot et al., 2014). We explain our findings in more details further in the text.

3.1 Boundary Resources as Governance Mechanisms: Openness vs. Control

Third-party innovation arises when platform owners realize the need to extend platform functionalities in an area they have to expertise in, and invite external parties with specialized knowledge to do so (Bygstad, 2015). Platform owners can support third-party innovation by providing standardized interfaces which facilitate the development of applications in the platform periphery. The key potential of boundary resources lies in transferring design capabilities to external actors, thus getting exposed to their specialized knowledge to build modules and functionalities which complement the platform core (Ghazawneh and Henfridsson, 2010).

Although standardized interfaces "open up" the platform for external parties, thus stimulating them to contribute with novel functionalities, "too much" openness can result with platform owners losing control over the ecosystem and its evolution. On the other hand, opening-up the platform "too little" can suppress innovation by making it difficult for external actors to contribute with novel functionalities, if they do not have access to the core platform modules.

A platform is more "open" if it places fewer restrictions on third-parties for producing novel add-ons, plug-ins and platform functionalities (Parker and Van Alstyne, 2018). However, too much openness in the periphery can make the platform too fragmented to serve as a platform (Bygstad, 2015). Therefore, the need to incorporate new expertise provided by external actors, at the same time requires a delicate balance in control over third-party modules. While low levels of control can stimulate third-party innovation, this can result with diverse applications in the periphery that are not interoperable with the core, or are not compliant with the overall vision of the platform (Boudreau, 2010). On the contrary, high levels of centralized control can result with lack of flexibility in the periphery, thus making the ecosystem lose its ability to generate external innovation (Bygstad, 2015).

Our review shows that there are conflicting views among IS researchers over the optimal level of openness and control in platform ecosystems. Putting our focus on boundary resources as enablers for third-party innovation, we look at interfaces as the most stable parts of the platform. Thus, control over the interfaces amounts to control over the platform and its evolution (Baldwin and Woodard, 2008). While platform owners are usually seen as dominant actors in platform ecosystems, there is a continuous debate over how much control and autonomy third-party actors should have (Hein et al., 2019; Saadatmand et al., 2019).

It is generally accepted that control points should be dispersed across all actors, but it is still unclear how and at what degree control should be allocated. Some researchers argue that boundary resources as control points should be designated evenly across third-parties, to encourage greater third-party engagement with platforms (Saadatmand et al., 2019). Therefore, they point towards the need for autonomy of external actors to choose their desired level of control (Goldbach et al., 2018; Wareham et al., 2013), or negotiate control based on the perceived value (Hein et al., 2019; Tiwana et al., 2010).

Although studies shows that self-control of third-parties can result with higher application quality in the periphery (Goldbach et al., 2018), giving up too much control can make it harder to achieve cohesion between the complements and the focal platform, thus jeopardizing innovation (Boudreau, 2010). If external actors have too much control, they can use unofficial APIs to build novel functionalities and get the platform enveloped by another platform (Hilbolling et al., 2020). On the contrary, if faced with strict boundary resources, developers might look for alternative ways to open up the platform and "self-resource" new boundary resources themselves (Eaton et al., 2015).

Decisions about openness and control are even more complex when applied to the eHealth field, as they are allocated amongst multiple national, regional and local governmental bodies and different bureaucratic levels. For example, following the evolution of the Swedish Patient Portal, Sellberg and Eltes (2017) show that core components were owned by multiple national and local authorities, with overlapping jurisdictions. In order to coordinate such complex partitioning of decision rights, the project team used a National Architecture Framework for eHealth services as a coordination mechanism, providing SDKs, APIs and guidelines to support the development of third-party modules across the platform ecosystem (Sellberg and Eltes, 2017).

The governance of the Catalan e-prescription solution, on the other hand, is an example where the Department of Health as a governing body had full ownership over the initiative at first, but gradually shifted towards an interoperability framework to open up and include third-party services. Even though third-parties shared ownership over the complementary services, the governing body still kept control over the services using accreditation mechanisms and quality certificates to orchestrate the development of applications that are trustworthy and interoperable with the platform core (Modol, 2017).

This concept shows that there is an on-going discussion in the IS literature over how should control points be dispersed across platform ecosystems, using boundary resources as tools to facilitate openness and control. Next, we investigate the role of multiple heterogeneous actors in tuning boundary resources and shaping the evolution of platform ecosystems.

3.2 Co-creating Innovation Across Heterogeneous Actors: Accommodation vs. Resistance

Platform ecosystems encompass dynamic relationships emerging among multiple heterogeneous actors (Brown et al., 2017). To support such a diverse environment, platform owners need to respond to the different goals and objectives of third-party actors in the ecosystem. Although standardized interfaces play a central role in orchestrating third-party actors towards a common platform goal (Tiwana, 2014), it is not always clear how such changes will shape the evolution of the platform (de Reuver et al., 2020) and affect other actors in the ecosystem. Therefore, many IS researchers have put their focus on investigating how boundary resources evolve in diverse socio-technical environments, where platform owners and their ecosystems mutually shape each-other's goals.

Looking at boundary resources from the platform owners' perspective, IS researchers conceptualize them as tools for "resourcing" external platform functionalities and "securing" the platform core to control third-party innovation (Ghazawneh and Henfridsson, 2013). This typology was further adapted to eHealth, where third-party actors from within and outside hospitals, the public and private sector, as well as citizens are integrated into the innovation cycle (Aanestad et al., 2019). In such settings, the focus extends towards the third-party developers' perspective, who use boundary resources to "discover" the limitations and possibilities of the core, and "vest" the benefits through copyrights, ownership and data exploitation (Aanestad et al., 2019).

Treating the evolution of boundary resources as a cyclical process, some IS researchers explained it through the prism of "tuning", where boundary resources evolve as a constant tension between third-parties resisting to change and accommodating it. Some used this tension to follow the evolution of boundary resources across private platform ecosystems (Eaton et al., 2015), while others adapted it to eHealth and extended it across the entire eHealth information infrastructure to investigate how public-private platforms emerge (Kempton et al., 2020). Looking at how boundary resources were "tuned" in Apple's iPhone platform, Eaton et al. (2015) showed that once the introduced set of boundary resources met resistance by third-party actors, the resources were either shaped by Apple additionally opening up the platform, or by third-party developers changing their goals and strategies to enter the platform ecosystem.

This tension was also used to follow shifts in decisions across large public-private platforms in eHealth infrastructures. Investigating the tuning of a welfare technology initiative in Norway, Kempton et al. (2020) found that decisions were constantly transitioning between the Directorate of Health's wish to overcome infrastructural silos, municipalities trying to make independent decisions about local investments and the need for autonomy of third-party actors. They propose a hub solution, whereby defining a minimal core and incremental changes to widen it, the circle between resistance and accommodation becomes tighter.

The process of tuning is shown to be highly dependent on the differences in power of third-party actors, to influence the trajectory of changes in the boundary resources' design (Eaton et al., 2015). This issue is also of significant importance in healthcare, where a multitude of actors, including governing bodies, third-party vendors, as well as institutions come into play and influence the evolution of the platform ecosystem (McLoughlin et al., 2016). While in the case of Apple's iPhone platform (Eaton et al., 2015) Apple as a central actor controls the surrounding environment, in eHealth platform ecosystems the distinction between the role of platform owners and third-party developers is not always as clear, as a multitude of intertwined powerful actors appear on both sides and the platform owner is not necessarily the most dominant one.

Following the e-prescription initiative in Catalonia as an example, the pharmacy association managed to obtain a key role in governing the pharmacy IT system, which was initially part of the larger e-prescription solution, governed by the Directorate of Health (Modol, 2017). APIs were "tuned" only when the association determined that the new feature adds enough value to pharmacies, or when it was mandatory or required by law.

Implications from practice suggest that if the owner is less powerful than the peripheral actors, third-party actors with higher influence are more likely to dominate the platform, and make the platform less attractive to other actors in the periphery (Saadatmand et al., 2019). Therefore, some IS researchers point towards a more balanced, cooperative and less hierarchical power relationship between platform owners and third-party developers (Bygstad, 2015; Goldbach et al., 2018; Saadatmand et al., 2019). To create and maintain a coherent identity for the platform, complementors need to balance the pursuit of their own interests with the interests of other players in the ecosystem (Saadatmand et al., 2019).

This section shows that in eHealth, there is a multitude of actors appearing on both sides of the platform core and periphery. This social diversity increases the complexity of managing actor relationships and govern how eHealth platform ecosystems emerge. Next, we use a technical lens to understand how the arrangement of components within the architecture supports platform ecosystem innovation.

3.3 Platform Innovation Within the Underlying Architecture: Stability vs. Flexibility

Governing innovation in platform ecosystems is closely reliant on the underlying platform architecture (Kempton et al., 2020; Tiwana et al., 2010). The fundamental architecture behind platforms consists of stable “core components” with low variety, flexible “peripheral components” with high variety, and design rules or “standardized interfaces” that connect the complements with the core (Baldwin and Woodard, 2008). Interfaces act as the most stable parts of the platform, since they determine how the different components coordinate and work together (Baldwin and Woodard, 2008). Therefore, preventing changes in the interfaces is essential to keep the interoperability and compliance between the core and the periphery (Wareham et al., 2013), as architectures need to incorporate interfaces that are stable, versatile and evolve over time (Baldwin and Woodard, 2008).

Our review shows that balancing the tension between keeping stability in the core and enabling flexibility in the periphery is of central importance to support platform innovation. Contrary to understanding stability and flexibility as conflicting forces or “dialectics”, some researchers look at them as “dualities” which are mutually reinforcing and interdependent (Bygstad and Hanseth, 2016). Peripheral components require a standardized platform core that can enable scaling, while the core needs flexible applications that can adapt to emerging local needs (Bygstad and Hanseth, 2016; de Reuver et al., 2018).

IS researchers commonly rely on the concept of loosely-coupled architectures, to balance the tension between stability and flexibility (Bygstad and Øvrelid, 2020; de Reuver et al., 2018; Saadatmand et al., 2019). When architectural components are loosely-coupled, changes in one application in the periphery do not necessarily result with changes within the core, or the other applications in the periphery, which can still remain stable or non-affected (MacCormack et al., 2010). In such a detached portfolio, boundary resources decouple the core from the distributed ecosystem of third-party apps (Brunswicker and Schecter, 2019), acting as coordinative tools between the dispersed components.

The concept of loosely-coupled architectures to govern platform innovation is also adopted in eHealth (Bygstad and Hanseth, 2016; Grisot et al., 2014). One view from Grisot et al. (2014) is to build external components as a new architecture, that is loosely-coupled to the installed base and let it evolve in a bottom-up approach. That way, the innovation results in a flexible solution that is easily modified without disturbing the core and incorporates new functionalities "on top of what exists" (Grisot et al., 2014). Looking at the relationship between innovative lightweight technologies and stable heavyweight systems, Bygstad (2017) also supports that they should interact with each other, instead of being integrated as a whole (Bygstad and Øvrelid, 2020). Thus, loosely-coupled components in platform ecosystems result with lower needs for coordination from the core and greater autonomy in the periphery (Bygstad, 2017).

As loose-coupling enables greater autonomy, it also has some positive effects on developers' motivation to innovate within the periphery (Brunswicker and Schecter, 2019; Goldbach et al., 2018). De Souza and Redmiles (2009) point out that while developers expect boundary resources to evolve, they still do not expect interfaces to be prone to frequent changes. And if frequent changes happen, they should not severely affect the interoperability and compliance of the app with the core (de Souza and Redmiles, 2009). Frequent changes in the interfaces can also enforce developers' decisions about "app desertion", or leaving the platform ecosystem, since they require constant effort from third-parties to stay interoperable with the core (Tiwana, 2015).

Contrary to decisions about app desertion, findings show that developers are more likely to make frequent, iterative changes to a certain application which is coherent with their past knowledge and expertise (Brunswicker and Schecter, 2019; Vesselkov et al., 2019). However, developers' motivation for innovating is stronger if they can learn novel skills and make flexible changes to applications in the platform periphery (Brunswicker and Schecter, 2019).

Balancing this tension between platform's macro and micro architecture is necessary to keep the interfaces stable, while allowing third-parties to create agile changes in applications that are interoperable with the platform core. Next, we adapt the three concepts to the eHealth context and extend them towards areas that should be prone to further investigation, as follows in the discussion section below.

4 Discussion: Governing Innovation in eHealth Platform Ecosystems

The proliferation of lightweight technologies, smartphone apps and wearable devices have placed the patient at the heart of healthcare service delivery, where patient data acts as a key driver for digital transformation in eHealth information infrastructures. Patient data associated with lightweight technologies has shifted the way healthcare is being delivered, how patients interact with caregivers and how information is exchanged and coordinated across the healthcare system (Bardhan et al., 2020). As a large set of haphazard and diverse patient data is gathered across dispersed apps and devices, developed and used outside hospital environments (Vassilakopoulou et al., 2017), there is a critical need to understand the role of patient data in leveraging the generative potential of eHealth platform ecosystems (Kempton et al., 2020; Vassilakopoulou et al., 2018; Vesselkov et al., 2019).

While this literature review uses IS knowledge as the basis to develop the findings, the purpose of this section is to adapt the concepts identified in the findings to the eHealth context and uncover areas that are understudied and require further attention. Relying on our previous knowledge of platform ecosystems in IS and eHealth, and not solely on the publications included in the literature review, we extended the three identified concepts in the findings and the tensions arising within them, with key areas that should be addressed in eHealth. 1) First, we highlight the importance of patient data as a key resource for eHealth platform innovation. 2) Second, we address the role of institutions in eHealth platform innovation. 3) At last, we extend innovation across platform-oriented eHealth information infrastructures. The transformation towards these three concepts is explained in more details in the sections that follow.

4.1Patient Data as a Resource for eHealth Platform Innovation

Lightweight technologies distribute a large set of standardized and unstandardized data across eHealth platform ecosystems (Constantiou and Kallinikos, 2015), generated and used by patients themselves. This heterogeneous set of data challenges the traditional ways of storing patient data in clinical and hospital systems, with access provided only to healthcare professions. The potential to cull patient generated data collected outside hospital environments with traditional medical sources, uncovers missing opportunities to learn more about diseases, adhere to personalized treatment, predict treatment outcomes (Bardhan et al., 2020), as well as use patient data as a key resource to innovate within eHealth platform ecosystems (Aanestad and Vassilakopoulou, 2019; Bygstad and Øvrelid, 2020; de Reuver et al., 2018).

While some applications can be developed using only test data, others may require access to real patient data (Aanestad and Vassilakopoulou, 2019), which is prone to stricter regulatory frameworks. It is hard for third-party vendors to innovate, if they do not have access to core data modules (Kempton et al., 2018). Therefore, researchers need to understand the novel approaches, mechanisms and tools to regulate access and control over such a heterogeneous patient data set (Kallinikos and Constantiou, 2015) and the importance of boundary resources to govern the decision-making process. Further research needs to explore the types of boundary resources to support data access and control (Aanestad et al., 2019) and the optimal approaches on balancing the tension between openness and control over sensitive patient data in eHealth platform ecosystems.

4.2The Role of Institutions in eHealth Platform Innovation

The complexity of heterogeneous actors in eHealth platforms does not encompass only national and regional government bodies, local municipalities and software vendors, but is also influenced by institutional logics, laws and regulations, politics and concentration of power that shape the evolution of the platform ecosystem

(Eaton et al., 2015; Hein et al., 2019). Regulators and interest organizations can seek to influence actors toward a certain behavior, exercise power to protect their own interests, as well as shape public opinion on trust in sharing patient data, which are directly related to decisions about patient data privacy and security in eHealth platform ecosystems (Eaton et al., 2015).

The novel landscape of lightweight technologies developed and used outside hospitals incorporates heterogeneous actors participating in a shared platform ecosystem, which cannot be separated from the political context in which it is embedded (Eaton et al., 2015). Therefore, the dialectics between accommodation and resistance in eHealth platform ecosystems, spread across conglomerates of the public and private sector, prone to institutional pressures. This imposes the need to further investigate and understand the cyclical process of how institutional power shapes boundary resources as influenced by decision-making on patient data sharing, privacy and security, as well as how institutions are shaped to accommodate such changes in response. Therefore, further research needs to understand how different actors come and shape decisions about data sharing, privacy and security in eHealth platforms ecosystems (Bygstad and Øvrelid, 2020; Øvrelid and Bygstad, 2016), as well as their motivation to contribute, stay within, or leave an eHealth platform ecosystem.

4.3 Innovating Across Platform-Oriented eHealth Information Infrastructures

The large arena of lightweight technologies has resulted with multiple platforms and isolated data repositories within eHealth information infrastructures, which do not exchange data with each-other. Patient data is captured and stored in many independent databases, using different patient identifiers, which imposes the challenge of deciphering the interrelationships of the data cube and provide critical insights (Baesens et al., 2016). It is not yet clear how to integrate data gathered from such a diverse set of sources, and in different types, such as: text, images, video, audio, which raise questions about semiotic compatibility and interoperability across disparate systems (Constantiou and Kallinikos, 2015), and do that on an infrastructural level. These platform silos have separate architectures arranging their components, as well as multiple boundary resources guiding their interaction with other platforms across the eHealth information infrastructure (Hanseth and Bygstad, 2018).

Our review shows that there is an increasing need to understand the integration of platforms into extensive information infrastructures, where core and peripheral components, as well as standardized interfaces are combined and mashed together, to improve healthcare service deliery (de Reuver et al., 2018). In eHealth, the tension between keeping the boundary resources stable, versus enabling flexibility in the periphery needs to be addressed across the entire information infrastructure, and not just within standalone platforms (Bygstad and Øvrelid, 2020). Our review shows that decisions about data sharing in fitness mHealth platforms have greater implications on platform generativity, than the actual design and architecture of the platform (Vesselkov et al., 2019). In order to break down these silos and exchange patient data, researchers need to understand how to balance stability and flexibility of multiple boundary resources dispersed across different institutions, with overlapping jurisdictions, while at the same time supporting the need for common standardized interfaces that can standardize data exchange across eHealth information infrastructures.

5 Conclusion

In this paper, we aggregate IS knowledge on governing innovation in platform ecosystems, by answering our research question: what are the conceptual approaches in governing third-party innovation in eHealth platform ecosystems? We address this question by developing three concepts on governing innovation in platform ecosystems within the IS literature which are shaped by tensions and adapting them to the complexity arising of the specifics in the eHealth settings.

By conceptualizing the existing literature and adjusting it to the eHealth context, we contribute in several ways. First, this review integrates IS knowledge on governing innovation in platform ecosystems which was not previously adapted to the complexity of the cross-disciplinary eHealth settings. Second, we define concepts which are not addressed in the present literature and should be prone to further research. We conclude that while IS literature is relevant for conceptualizing innovation in platform ecosystems, we still need to adapt these concepts to the emerging healthcare landscape where external innovations are developed outside hospitals and diverse actors across multiple levels, such as national, regional and local governmental bodies, with overlapping jurisdictions and different goals are included in the decision-making process around eHealth platform ecosystem.

We also acknowledge the limitations of our paper in two ways. 1) Inconsistent terminology: boundary resources and third-party development are referred to using different terms across IS. Therefore, this review may not retract all relevant published literature in the IS field. 2) Cross-disciplinarity: colliding IS and eHealth might neglect some of the generalizations in the current state of knowledge, as what works across the private sector, may not be applicable to the public-private partnerships arising in eHealth information infrastructures.

References

Aanestad, M., Vassilakopoulou, P., 2019. Innovation Readiness in Healthcare Information Infrastructures: Key Resources to Enable Collaborative Digital Innovation. Proc. 17th Scand. Conf. Health Inform., Linköping Electronic Conference Proceedings 6.

Aanestad, M., Vassilakopoulou, P., Øvrelid, E., 2019. Collaborative Innovation in Healthcare: Boundary Resources for Peripheral Actors. Proc. Int. Conf. Inf. Syst. 14.

Agarwal, R., Gao, G. (Gordon), DesRoches, C., Jha, A.K., 2010. Research Commentary: The Digital Transformation of Healthcare: Current Status and the Road Ahead. Inf. Syst. Res. 21, 796–809.

Baesens, B., Bapna, R., Marsden, J.R., Vanthienen, J., KU Leuven, BELGIUM, Zhao, J.L., 2016. Transformational Issues of Big Data and Analytics in Networked Business. MIS Q. 40, 807–818.

Baldwin, C.Y., Woodard, C.J., 2008. The Architecture of Platforms: A Unified View 32.

Bardhan, I., Chen, H., Karahanna, E., 2020. Connecting systems, data and people: A multidisciplinary research roadmap for chronic disease management 16.

Boell, S.K., Cecez-Kecmanovic, D., 2014. A Hermeneutic Approach for Conducting Literature Reviews and Literature Searches. Commun. Assoc. Inf. Syst. 34.

Boudreau, K., 2010. Open Platform Strategies and Innovation: Granting Access vs. Devolving Control. Manag. Sci. 56, 1849–1872.

Brown, A., Fishenden, J., Thompson, M., Venters, W., 2017. Appraising the impact and role of platform models and Government as a Platform (GaaP) in UK

Government public service reform: Towards a Platform Assessment Framework (PAF). Gov. Inf. Q. 34, 167–182.

Brunswicker, S., Schecter, A., 2019. Coherence or flexibility? The paradox of change for developers’ digital innovation trajectory on open platforms. Res. Policy 48, 103771.

Bygstad, B., 2017. Generative Innovation: A Comparison of Lightweight and Heavyweight IT. J. Inf. Technol. 32, 180–193.

Bygstad, B., 2015. The Coming of Lightweight IT, in: ECIS 2015 Completed Research Papers Proceedings. Presented at the ECIS 2015, p. 17.

Bygstad, B., Hanseth, O., 2016. Governing e-Health Infrastructures: Dealing with Tensions, in: Proceedings of International Conference of Information Systems. Presented at the Thirty Seventh International Conference on Information Systems, Proceedings of International Conference of Information Systems, Dublin, p. 19.

Bygstad, B., Øvrelid, E., 2020. Architectural alignment of process innovation and digital infrastructure in a high-tech hospital. Eur. J. Inf. Syst. 1–18.

Constantiou, I.D., Kallinikos, J., 2015. New Games, New Rules: Big Data and the Changing Context of Strategy. J. Inf. Technol. 30, 44–57.

de Reuver, M., Sørensen, C., Basole, R.C., 2018. The Digital Platform: A Research Agenda. J. Inf. Technol. 33, 124–135.

de Reuver, M., van Wynsberghe, A., Janssen, M., van de Poel, I., 2020. Digital platforms and responsible innovation: expanding value sensitive design to overcome ontological uncertainty. Ethics Inf. Technol.

de Souza, C.R.B., Redmiles, D.F., 2009. On The Roles of APIs in the Coordination of Collaborative Software Development. Comput. Support. Coop. Work CSCW 18, 445–475.

Eaton, B., Elaluf-Calderwood, S., Sørensen, C., Yoo, Y., 2015. Distributed Tuning of Boundary Resources: The Case of Apple’s iOS Service System. MIS Q. 39, 217–243.

Farshchian, B.A., Thomassen, H.E., 2019. Co-Creating Platform Governance Models Using Boundary Resources: a Case Study from Dementia Care Services. Comput. Support. Coop. Work CSCW 28, 549–589.

Fichman, R.G., Kohli, R., Krishnan, R., 2011. Editorial Overview—The Role of Information Systems in Healthcare: Current Research and Future Trends. Inf. Syst. Res. 22, 419–428.

Gawer, A., 2014. Bridging differing perspectives on technological platforms: Toward an integrative framework. Res. Policy 43, 1239–1249.

Ghazawneh, A., Henfridsson, O., 2013. Balancing platform control and external contribution in third-party development: the boundary resources model: Control and contribution in third-party development. Inf. Syst. J. 23, 173–192.

Ghazawneh, A., Henfridsson, O., 2010. Governing third-party development through platform boundary resources. Presented at the the International

Conference on Information Systems (ICIS), AIS Electronic Library (AISeL), pp. 1–18.

Goldbach, T., Benlian, A., Buxmann, P., 2018. Differential effects of formal and self-control in mobile platform ecosystems: Multi-method findings on third-party developers' continuance intentions and application quality. Inf. Manage. 55, 271–284.

Grisot, M., Hanseth, O., Thorseng, A., 2014. Innovation Of, In, On Infrastructures: Articulating the Role of Architecture in Information Infrastructure Evolution. J. Assoc. Inf. Syst. 15, 197–219.

Hanseth, O., Bygstad, B., 2018. Platformization, Infrastructuring and Platform-oriented Infrastructures. A Norwegian e-Health Case 39.

Hein, A., Schreieck, M., Riasanow, T., Setzke, D.S., Wiesche, M., Böhm, M., Krcmar, H., 2019. Digital platform ecosystems. Electron. Mark.

Hilbolling, S., Berends, H., Deken, F., Tuertscher, P., 2020. Complementors as connectors: managing open innovation around digital product platforms. RD Manag. 50, 18–30.

Kallinikos, J., Constantiou, I.D., 2015. Big Data Revisited: A Rejoinder. J. Inf. Technol. 30, 70–74.

Karhu, K., Gustafsson, R., Lyytinen, K., 2018. Exploiting and Defending Open Digital Platforms with Boundary Resources: Android's Five Platform Forks. Inf. Syst. Res. 29, 479–497.

Kempton, A., Grisot, M., Brænden, K., Aanestad, M., 2020. Infrastructural tuning in public-private partnerships. Presented at the ECIS.

Kempton, A.M., Aanestad, M., Grisot, M., Br, K., 2018. Facillitating value creation and digital innovation in public-private platforms 9.

MacCormack, A., Baldwin, C., Rusnak, J., 2010. The Architecture of Complex Systems: Do Core-periphery Structures Dominate? 37.

McLoughlin, I.P., Garetty, K., Wilson, R., Dalley, A., Ping, Y., 2016. Doing Infrastructural Work: The Role of Boundary Objects in Health Information Infrastructure Projects. Scand. J. Inf. Syst. 31.

Modol, J.R., 2017. Maintaining the Pharmacy Model: The Catalan Electronic Prescription Infrastructure, in: Aanestad, M., Grisot, M., Hanseth, O., Vassilakopoulou, P. (Eds.), Information Infrastructures within European Health Care: Working with the Installed Base, Health Informatics. Springer International Publishing, Cham, pp. 55–72.

Msiska, B., 2018. Cultivating Third Party Development in Platform- centric Software Ecosystems: Extended Boundary Resources Model 10, 18.

Øvrelid, E., Bygstad, B., 2016. Extending e-Health Infrastructures with Lightweight IT, in: Lundh Snis, U. (Ed.), Nordic Contributions in IS Research, Lecture Notes in Business Information Processing. Springer International Publishing, Cham, pp. 43–56.

Parker, G., Van Alstyne, M., 2018. Innovation, Openness, and Platform Control. Manag. Sci. 64, 3015–3032.

Roland, L.K., Sanner, T.A., Sæbø, J.I., Monteiro, E., 2017. P for Platform. Architectures of large=scale participatory design. Scand. J. Inf. Syst. 29.

Saadatmand, F., Lindgren, R., Schultze, U., 2019. Configurations of platform organizations: Implications for complementor engagement. Res. Policy 48, 103770.

Schreieck, M., Wiesche, M., Krcmar, H., 2016. Design and governance of platform ecosystems – key concepts and issues for future research, in: 24th European Conference on Information Systems. p. 21.

Sellberg, N., Eltes, J., 2017. The Swedish Patient Portal and Its Relation to the National Reference Architecture and the Overall eHealth Infrastructure, in: Aanestad, M., Grisot, M., Hanseth, O., Vassilakopoulou, P. (Eds.), Information Infrastructures within European Health Care: Working with the Installed Base. Springer, Cham (CH).

Tiwana, A., 2015. Platform Desertion by App Developers. J. Manag. Inf. Syst. 32, 40–77.

Tiwana, A., 2014. Platform Governance, in: Platform Ecosystems. Elsevier, pp. 117–151.

Tiwana, A., 2013. Platform Ecosystems: Aligning Architecture, Governance, and Strategy. Newnes.

Tiwana, A., Konsynski, B., Bush, A.A., 2010. Research Commentary—Platform Evolution: Coevolution of Platform Architecture, Governance, and Environmental Dynamics. Inf. Syst. Res. 21, 675–687.

Vassilakopoulou, P., Grisot, M., Aanestad, M., 2018. Between Personal and Common: the Design of Hybrid Information Spaces. Computer Supported Cooperative Work CSCW 27, 1085–1112.

Vassilakopoulou, P., Grisot, M., Jensen, T.B., Sellberg, N., Eltes, J., Thorseng, A.A., Aanestad, M., 2017. Building National eHealth Platforms: the Challenge of Inclusiveness 14.

Vesselkov, A., Hämmäinen, H., Töyli, J., 2019. Design and Governance of mHealth Data Sharing. Commun. Assoc. Inf. Syst. 299–321.

Wareham, J.D., Fox, P.B., Cano Giner, J.L., 2013. Technology Ecosystem Governance. SSRN Electron. J.

Webster, J., Watson, R.T., 2002. Guest Editorial: Analyzing the Past to Prepare for the Future: Writing a literature Review 11.