1 Introduction

1.1 Data as Driving Force for Industrial Businesses

Practitioners and researchers have long understood that data are essential for companies. Inventions for database management systems date back to the 1960s (Codd, 1981, p. 110), and with increasing digitalization, it became apparent during the 1980s that "data communications is increasingly becoming an important area of management information systems" (Stamper, 1988). Today, many manufacturers see data as an essential element of competitiveness, for example through improved decision-making, early identification and resolution of problems, gaining insights into processes that enable innovation not only in products and processes, but also how people work (Manufacturing Leadership Council, 2024, p. 24). Next to efficiency gains, cost savings, productivity increases, data from supply chains allow to draw relevant information for resiliency (Zamani et al., 2022) and material data enables circular processes that result in resource savings (World Economic Forum, 2023). However, the challenges have become more complex and no longer concern only single data sets and their analysis; they include interorganizational exchange and the flexible operation of large-scale networks. The value of data depends on how it can be created, shared, and used (OECD, 2022, p. 6). As stated by the European Commission's vice president for the digital single market, the "main challenge now for all of us is to guarantee the secure and free flow of data, develop technical standards, and ensure interoperability" (Ansip, 2016). Data and data flows are economically essential for the production of many goods and services (OECD, 2022, p. 8).

In the industrial domain, numerous opportunities arise from the use of data, and the focus is mostly on analytical data. Analytical data form the key ingredient for machine learning and artificial intelligence (AI), with the goal of enabling intelligent services (cf. Dehghani, 2022, p. 3 & 12). The relevant data sets of industrial interest are primarily generated by connected equipment or sensors and are termed "Industrial Internet of Things (IIoT) data." IIoT data are characterized by large volume, real-time transmission and velocity, heterogeneous types ranging from structured to unstructured data, and being connected through various platforms and communication links (Dai et al., 2020, p. 1279). These characteristics make their handling particularly complex and challenging. The original goals of analyzing these data sets are productivity increases and efficiency gains, so analyses are used to improve quality, reduce warranty costs, enable benchmark analysis, and improve yield, or conduct predictive maintenance (Lade et al., p. 74). However, data-driven services are increasingly being integrated into the industrial value chain to determine the value of physical products. As a result, data-driven services challenge traditional business models while enabling new ones. This means that data have a significantly greater impact than just efficiency gains or optimization purposes. Thus, industrial companies are forced "to react and to adapt to the newly emerging data-savvy environment" (Voigt et al., 2021, p. 138).

Data are a key enabler for circular models that enable significant sustainability goals and material savings (World Economic Forum, 2023). Material flows in the circular economy, which encompass the processes of reduction, reuse, and repair, require relevant data about the materials in order to take informed decisions and enable safe reuse. Material savings are not only associated with less waste for environmental reasons, but also have a political dimension. In 2023, the European Union (EU) rated 34 materials as critical and exposing Europe to structural geopolitical dependencies linked to all industries and ranging across supply chain stages (European Commission, 2023b). Further, data-driven technologies like AI or big data analytics can be used to increase resilience in supply chains by serving supply chain transparency, visibility, and responsiveness, and by identifying sources of disruption (Dubey et al., 2021; Zamani et al., 2022, p. 618–619; Zouari et al., 2021). The need for resilience has become apparent in recent years, not only for industry but also for society, as demonstrated by the effects of the Covid-19 pandemic, the energy crisis resulting from the Russian invasion of Ukraine, and global political and economic changes. In the coming years, the geopolitical and business policy landscape, including political developments in the United States, will challenge companies and how they utilize, manage, and share data in global interorganizational data infrastructures.

In Europe, manufacturing is the "backbone of the economy"; it provides 32 million jobs directly and more than 20 million indirectly (Lichtblau et al., 2015, p. 9 & 11). The future direction of the manufacturing industry and its opportunities and risks will be shaped inherently by data-driven technologies (Deloitte, 2024). From the perspectives of industry, science, business policy, and geopolitics, the situation gives rise to opportunities, challenges, and unresolved issues that accompany the large-scale sharing and utilization of data and necessary interorganizational technological data infrastructures.

1.2 Motivation from Industrial Practice

1.2.1 Challenges in Practice

The challenges faced by chief information officers (CIOs) today are no longer limited to the scope of individual systems but also encompass their interplay (Fürstenau et al., 2019, p. 1319; Henfridsson & Bygstad, 2013). The data landscape is increasingly distributed, disparate, and diverse, and there is a shift away from "manually integrated point solutions" to cohesive ecosystem approaches, with increased use and combinations of different cloud and multi-cloud approaches (Gartner, 2023). Additionally, a single-company focus is no longer sufficient to understand and capture the value added by data, and an interorganizational perspective on industrial data ecosystems is required (Castro et al., 2021). In global information networks, the sharing of data and interorganizational data flows has become increasingly important. This raises new questions about data exchange, integration, and trust. Traditionally, static communication channels and tight integration with each partner in an interorganizational process were non-negotiable requirements for collaborations, with the goal of establishing a long-term business relationship. Today, processes and systems must rapidly respond to changes in their environment and be inherently flexible to react, for example, to supply chain changes (Aulkemeier et al., 2019, pp. 597–598). There is a need to quickly integrate new players and systems into processes and change them repeatedly. Further, instead of considering only one actor in a "winner-takes-all" manner, there is increased demand for fair and balanced approaches that create benefits for a complete network (P. Wang, 2021, p. 421). In this context, special attention falls on cloud service providers, who play a central role in the storage and transmission of data, as well as the provision of essential services. Especially large, non-European cloud service and platform providers are reputed to create lock-in situations for data due to the lack of standardization of fundamental interfaces, technologies, and semantics (Arce, 2022, p. 501).

The topic of connected industry is the subject of intense discussion and investigation by industrial companies and public institutions, which are looking at the opportunities for value creation and digital transformation, but also at the risks of disruption (World Economic Forum, 2015). This has resulted in a multitude of approaches and technical architecture models, one of the most prominent in the German and European economic landscape being the Reference Architectural Model Industrie 4.0 (RAMI4.0) (Plattform Industrie 4.0, 2018, pp. 29-30). Other approaches include the Industrial Internet Reference Architecture (IIRA)

(Industry IoT Consortium, 2022, p. 7), the Asset Administration Shell (AAS) by the Industrial Digital Twin Association (IDTA) (IDTA, 2024a), and various platform approaches.

he RAMI4.0 model addresses digitized and connected industrial value creation as the main targets of Industry 4.0, which is a concept coined and rooted in a program driven by the German federal government and accompanied by research and industry associations, as well as large companies (Plattform Industrie 4.0, 2018, pp. 29-30). The concept of Industry 4.0 was first presented and promoted in 2011, and since then, many investments have been made and initiatives launched, but the concept has also faced challenges and skepticism (Drath & Horch, 2014, p. 58; Veile et al., 2020, pp. 992–994). Another outcome of the German initiative *Plattform Industrie 4.0* is the AAS, which is now being further developed by the IDTA. The AAS provides a standardized digital representation of physical assets and, which is summarized as digital twin (IDTA, 2024a, 2024c). The corresponding information model enables the creation of manufacturer-independent digital twins (IDTA, 2024b), and it is accompanied by an interface and set of services for managing the information model. Among these services is the AAS registry, which enables companies to obtain information about available digital twins from other companies. Another approach is provided by the IIRA, created by the Industry IoT Consortium, which is an alliance consisting of industrial companies without the major participation of public bodies (Industry IoT Consortium, 2025). The IIRA model "highlights the important architectural concerns commonly found in IIoT systems across industrial sectors" (Industry IoT Consortium, 2022, p. 7). The various solutions from different consortia show the high demand for information and communication technology-based networking, as well as for services to implement interorganizational governance and management mechanisms. They also show the heterogeneity of the various technologies. A comprehensive solution, in the form of an underlying data infrastructure, must have the capacity to connect to a wide range of solutions and technological heterogeneous systems.

In addition to the developments that have emerged from various initiatives, a number of platforms have addressed the goals of ecosystem building and interorganizational data sharing, such as Snowflake Inc. (2024a), Databricks (2024a), Azure Data Share (Microsoft, 2024), Amazon RedShift (Amazon Web Services, 2024), Google Analytics Hub (Google Cloud, 2024), and Alibaba Resource Sharing (Alibaba Cloud, 2024). However, these platforms and service providers are privately owned and focus on the interests of single customers rather than alliances, large consortia, or open networks. Typically, platforms are associated with vendor lock-in situations and dependencies. Further, their core focus is on various services that add value to data directly by analyzing or processing mechanisms, and not the wider ecosystem itself.

There is a growing need for a data ecosystem that provides a level playing field for data value chains and supports both business and social objectives. In data ecosystems, conditions are created for collaboration between diverse participants or marketplace competition between connected and interdependent participants (Curry, 2020, p. 7). Data ecosystems work on the assumption that no member is able to create the ecosystem or data value creation on its own so they need to team up, with each member contributing to the benefits, and data are addressed as a strategic resource that is exchanged and monetized (Capiello et al., 2020, p. 90).

Data space technologies explicitly address the needs of data ecosystem settings, managing uncertain and flexible relationships and loosely connected data sources (Franklin et al., 2005, p. 27; Hedeler et al., 2010, p. 115). Data spaces relieve developers of the burden of connecting these loosely connected data sources, as no upfront integration is required; rather, a data coexistence approach based on a set of services is used. They enable the addition of services in a "pay-as-you-go" fashion, including sophisticated operations, integrations, and queries (Franklin et al., 2005, p. 28). In this way, data spaces encourage "flexible organizational forms" for the trusted sharing of data (Möller et al., 2024, p. 41). In data spaces, special attention is given to the concept of data sovereignty and its technical implementation. Data sovereignty is closely associated with self-determination and the ability of data providers to retain control over shared data (Hellmeier & Scherenberg, 2023, p. 8). Data spaces may even enable public spaces for data sharing that have the potential to shape the global digital transformation. In light of their importance, there is a need to better understand data spaces using empirical and theory-based insights (Beverungen et al., 2022, pp. 498–499).

1.2.2 Industry Policy Challenges

In recent years, political goals and conditions have become key drivers for the development of data ecosystems and their related technologies, particularly in Europe. Data spaces are an important element of the European data strategy (European Commission, 2020a). The European data strategy fosters a single market for data to realize Europe's potential in the data economy and strives to make more data available “while keeping the companies and individuals who generate the data in control” (European Commission, 2020a). The EU works to boost cross-sectoral data flows for the benefit of businesses, public administrations, and research organizations under European rules and values, especially privacy and data protection (European Commission, 2024c). It also addresses a recovery plan for the economic crises due to the Covid-19 pandemic in 2020. While there are legislations that demand mandatory data sharing, the data spaces envisioned by the EU are based on voluntarily provided data. The original 10 data spaces have grown to 14 planned data spaces in strategic economic sectors and domains of public interest (European Commission, 2024d, p. 3), with 75 related projects (bitkom e.V., 2024). Different funding programs have boosted the development of various technologies related to data spaces (IDSA, 2024a).

In addition to intensive research, public funding, and the promotion of data spaces, two pieces of legislation support the goals of the EU in relation to data space technologies. First, the EU Data Governance Act (DGA) provides a legal basis for interorganizational data-sharing mechanisms, which form the basis for operative data spaces. Importantly for industrial organizations, the DGA emphasizes data intermediation services and outlines rules for neutrality and transparency so that trust among data space participants can be created. Further, data intermediation services must adhere to strict measures of data protection and user control, and enable the management of data exchanges without gaining access to sensitive data or claiming rights or ownership. Furthermore, the DGA supports data altruism by enabling individuals and organizations to donate their data for a greater common good (Data Governance Act, 2022; European Commission, 2024b).

The second legislation is the EU Data Act. The act aims to ensure fairness, promote data access and use, and drive data-driven innovations. Therefore, connected products and services must be designed and manufactured in a way that allows users, which may be human consumers or companies, to easily secure, access, use, and share the generated data (Data Act, 2023; European Commission, 2024a). This means that manufacturers are obliged to design their products and services in such a way that the created data can be passed on to other service providers. Data spaces enable this data sharing in a neutral way.

Apart from the common data spaces driven by the EU, further data spaces and developments in data space technologies continue to evolve. Prominent examples that stand out for their maturity and operability, as well as their large scale, are the German initiative Catena-X, which deals with the automotive industry and supply chain challenges Catena-X (2025); the German Mobility Data Space (MDS) and French data space Eona-X Eona-X (2025), which address the tourism domain; Energy Data-X, which deals with energy data Energy Data-X (2025); and the healthcare data space HEALTH-X dataLOFT (2025).

While the development of data spaces is rooted mostly in Europe, other countries have also adapted the data space concept to foster data sharing, particularly in Asia. In China, the Chinese National Data Administration aims to create a wide range of data spaces (Zheng, 2024). Interest in data spaces is growing in Japan, particularly with regard to the automotive sector and coping with laws such as legal requirements for battery manufacturers (Nomura Research Institute, 2024). The IDSA Data Space Radar provides insight into a multitude of data spaces from a global perspective. As of December 2024, it included 184 data spaces, most of which had several specific application scenarios for the data shared. An overview of the sectoral distribution of data spaces is provided in Figure 1. As shown, many data spaces are in the category of “others” as they cannot be clearly assigned to specific categories but are cross-sectoral. Manufacturing has the largest unique share, closely followed by related domains such as energy and transportation, which can also be linked to the data spaces in the circular economy (IDSA, 2024a).

In recent decades, various approaches have emerged to meet the challenges of data and information management in industry. However, an overarching perspective that goes beyond the boundaries of single information systems and considers the challenges of multiple data spaces and data ecosystems with respect to shared governance or interoperability issues has not been sufficiently addressed. The required attention, concepts, functionalities, and integrations remain lacking in the literature.

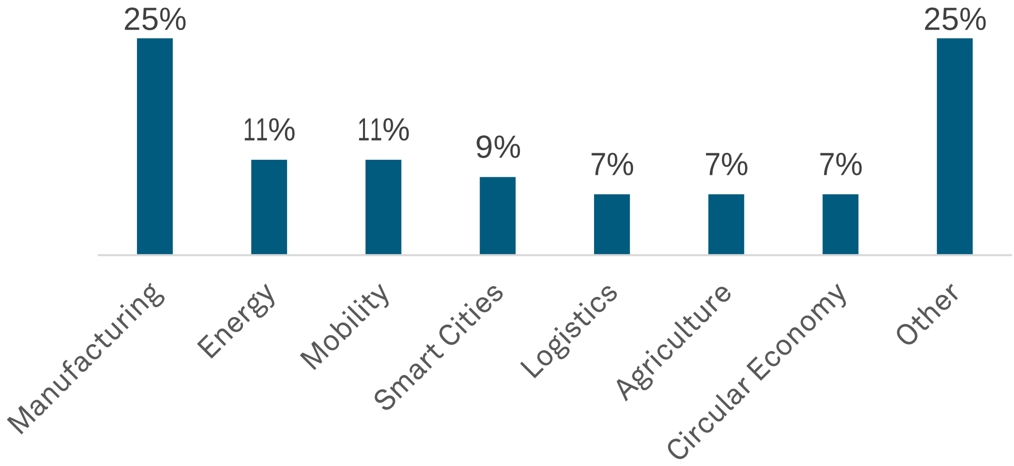

Figure 1 Sectoral Distribution of Data Spaces (IDSA, 2024a)

1.3 Motivation from Academic Research and Research Question

From an academic research perspective, the motivation stems from the data ecosystem and data space research field, as well as from the digital infrastructure research stream and complexity science. From a data ecosystem perspective, many studies on data ecosystems consider only some parts of the ecosystem and not the complete network. Additionally, there are limited explanations for the behavior, design, and management issues of data ecosystems (Oliveira et al., 2019, p. 614). An investigation of data ecosystems reveals that their structural differences from other value systems are striking. Ecosystems are fundamentally different from established value systems focused on single companies, such as those studied by Porter (1980), but also from buyer–supplier relationships, integrated hierarchies of firms, and strategic networks (Jacobides et al., 2018, p. 2258). Even in research on the technologies of data spaces, many observations tackle only individual data spaces and neglect the perspective of numerous data spaces. The interoperability and exchange of data sets between "overlapping" data spaces holds the potential to create a large, encompassing data ecosystem (cf. Möller et al., 2024, p. 41). With regard to the unit of analysis in research, this means that a focus on a single closed information system is insufficient to adequately address the loose coupling in data ecosystems, as are individual considerations of selected interconnected systems. The role of infrastructure has been mentioned but not explicitly highlighted and examined in terms of consequences or implications for its design (Capiello et al., 2020, p. 66). Thus far, there has been no research on digital infrastructure that explicitly considers the requirements of data ecosystem settings or specifically targets industrial ecosystems (Schleimer & Duparc, 2023).

When digitizing technologies are applied to a broader social and institutional context, they are rendered infrastructural and imply a new relationship between the service model and underlying capabilities (Tilson et al., 2010). In 2010, Tilson et al. (2010) pointed out significant gaps in research with his seminal work on digital infrastructures in IS research. Tilson et al. (2010, p. 749) argued that (digital) infrastructures were still "the orphan of IS research," and that there was a need to "theorize about the evolution of digital infrastructures in ways that recognize the salience of the paradoxes of change and control" (p. 749) and understand how actors "establish generative platforms by instituting a set of control points acceptable to others in a nascent ecosystem" (p. 754). Even if there "has since been a flowering of research on digital infrastructure" (Swanson, 2021, p. 13)

Taking the internet as an example of a successful digital infrastructure, the benefits for businesses and society are evident when infrastructural development works. However, failures in infrastructure design are more common than successes, and they can result in huge investment losses, opportunity costs, and even political and social problems (Hanseth & Lyytinen, 2010, pp. 1–2). One specific challenge of infrastructural research is "the difficulty of translating vivid empirical descriptions of [information infrastructures'] evolution into effective socio-technical design principles that promote their evolution, growth and complexity coordination" (Hanseth & Lyytinen, 2010, p. 2). Infrastructures come with competing requirements and particular characteristics, which "create a fascinating design challenge—even a new science" (Star & Ruhleder, 1996, p. 32).

Moreover, the internet-enabled transformation shifts the managerial discourse from single firms and industry-specific value systems to global networks and increases interest in complexity issues (Merali, 2006, p. 217). Complex systems adapt, evolve, and self-organize over time, and are the object of complexity science (cf. Braa et al., 2007; Ciborra et al., 2000). Complexity is traditionally viewed as a concern that needs to be reduced. However, systems with increased complexity and unstable states can provide advantages, such as agility and the emergence of novelty (Kauffman, 1995; Stacey, 2007; Wang & Vidgen, 2007). Order within complex systems is created without a distinct design but through emergence and organic unfolding analogous to cells in an organism (Braa et al., 2007, p. 384; Edwards et al., 2009). Therefore, complex systems are commonly considered to be not actively shapeable or designable. However, a wide range of complex systems in today's digitized economy require dedicated design efforts and interventions to achieve economic and societal goals. Large-scale data infrastructures and data ecosystems aim to create long-term benefits for individuals, the private sector, the digital economy, and society (Capiello et al., 2020; Swanson, 2021). For example, health information infrastructures can leverage health data and improve the well-being of citizens (Braa et al., 2007).

Despite the challenging properties of complex systems, there is a need for their design, and exploring effective methods for achieving design goals is important. Additionally, large-scale data ecosystems and data infrastructures will never emerge and operate without a significant degree of complexity due to their connected and organic nature, which makes attempts to reduce complexity obsolete and shifts focus to questions of how to handle this complexity. The insights gained from research into these dynamics can provide valuable information for the design of complex systems in general. This major need for action in the design of industrial data ecosystems and data spaces, as well as the aforementioned design challenges, infrastructure dynamics, and properties of data, leads to the following research question:

> How to design data infrastructures for industrial data ecosystems?

1.4 Objectives

1.4.1 Goals

The goal of this study is to create a nascent design theory for data infrastructures as enablers of data spaces and thus data ecosystems. Its structured design knowledge can provide guidance for future designers and implementors. Its key aspects are openness, autonomy of the competing and uncoupled actors, technological heterogeneity and complexity, and the specifics of handling data in industrial scenarios. The theory is based on the existing scientific body of knowledge and is guided by contemporary industry developments and collaboration challenges in the sector of secure data sharing and cloud service technologies. Thus, industrial practice and real-world phenomena drive this investigation.

Recently emerging industry consortia and standardization efforts, such as the Eclipse Dataspace Working Group with member organizations like the IDSA, Catena-X, iShare, and Gaia-X and their work trust protocols and communication standards (Eclipse Dataspace, 2025), serve as a trigger phenomenon to investigate their drivers and goals in more detail, and to integrate them on a scientific basis and advance them further. Further knowledge is derived from the automotive and mobility application industries. This imprints a focus on the industrial context and the characteristics of their data assets and data-sharing challenges. The drivers of these developments show that there are already existing efforts to deploy targeted measures and resources in data infrastructures so that the desired goals and target states are achieved. For example, public funding program or standard developments. This means that no completely free evolution has taken place, but concrete properties have already been implemented. The design principles of this study acknowledge the complex nature of data infrastructures and assess how they can be influenced, despite not being completely manageable by top-down approaches and building on existing paths.

1.4.2 Target Audience

The recipients of this work can be categorized into different groups. First, IS researchers in the field of data ecosystems and data sharing are provided with insights and design solutions for large-scale, sustainable sociotechnological infrastructure systems that enable fair data ecosystems. Additionally, this study can benefit researchers in the field of digital infrastructure by enhancing knowledge about infrastructural design and shedding light on the particularities of data and software-based infrastructures. The results and intermediary supplements can also inform other design science research projects that take place in an ecosystem context. Practitioners in the field of data management in the industrial sector, especially strategists with long-term plans to realize large-scale solutions, are equipped with tactics for effectively engaging in data infrastructure activities. They can benefit from information about the nature and challenges of such endeavors. Participants in standardization activities and project steering committee members who make decisions about design steps and tactics can benefit from insights about dynamics and prescriptive knowledge, as can community managers and system designers in projects that create collaborative solutions for interorganizational data exchange and drive standardization activities. The findings also aid project managers and consortia leaders in prioritizing and structuring their tasks. Similarly, the findings can guide the design process of different sub-parts and components of larger solutions. The last group comprises public agencies, such as legislators or compliance bodies, that create incentives and barriers for data ecosystems. By learning about the distinct dynamics and having practice-based knowledge at hand, they can govern and facilitate the process of creating a fair and sustainable data ecosystem.

1.5 Structure of the Work

Chapter 1 covers an introduction to the importance of data for industrial businesses (1.1), the motivation from industrial practices for dealing with data infrastructures (1.2), the motivation of the research to follow the research question (1.3), and the objectives of this work (1.4).

Chapter 0 deals with the research design. First, a short introduction to the design science paradigm is given (2.1), which guides this study. Next, the case study methodology is introduced (2.2), followed by the data analysis and design process (2.3). The analysis is guided by a conceptual research framework, which is introduced in detail (2.4), before the quality criteria and evaluation mechanisms of the research approach are examined (2.5).

Chapter 3 introduces and defines relevant concepts and presents the current state of the scientific body of knowledge. Since this work follows the design science paradigm, particular attention is paid to the design challenges. Data ecosystems are explained (3.1), along with the concepts of data spaces (3.2), digital infrastructures, and data infrastructures (3.3). To summarize the state of knowledge on design challenges, the chapter finishes with an overview of all design challenges described (3.4).

Chapter 4 presents the case study analysis, which is divided into steps. First, the problem and existing solution approaches are refined (4.1), followed by some remarks on coding, instrument testing, and the pilot elements (4.2). Key decision points are identified and synthesized by emphasizing a harmonized inventory and descriptive coding (4.3), followed by a conceptual analysis (4.4) of the main topics. The findings of this analysis are then summarized into a set of knowledge nuggets (4.5).

Chapter 0 deals with the design principles. It begins by transparently explaining the creation path in terms of the underlying kernel theory and the process for creating the design principles (5.1). Kernel theory is then utilized to create four design principles, demonstrated using the characteristics of two related cases (5.2). The evaluation and incremental improvement are described (0), and the evaluation framework of the overall study is assessed and discussed (5.5).

In Chapter 6, the focus is on the interpretation of the results and design objectives (6.1). A reflection on the design objectives follows (6.2), before the limitations of the study are addressed (6.3), and the findings and contributions are described (6.4). The study is concluded in Chapter 7.

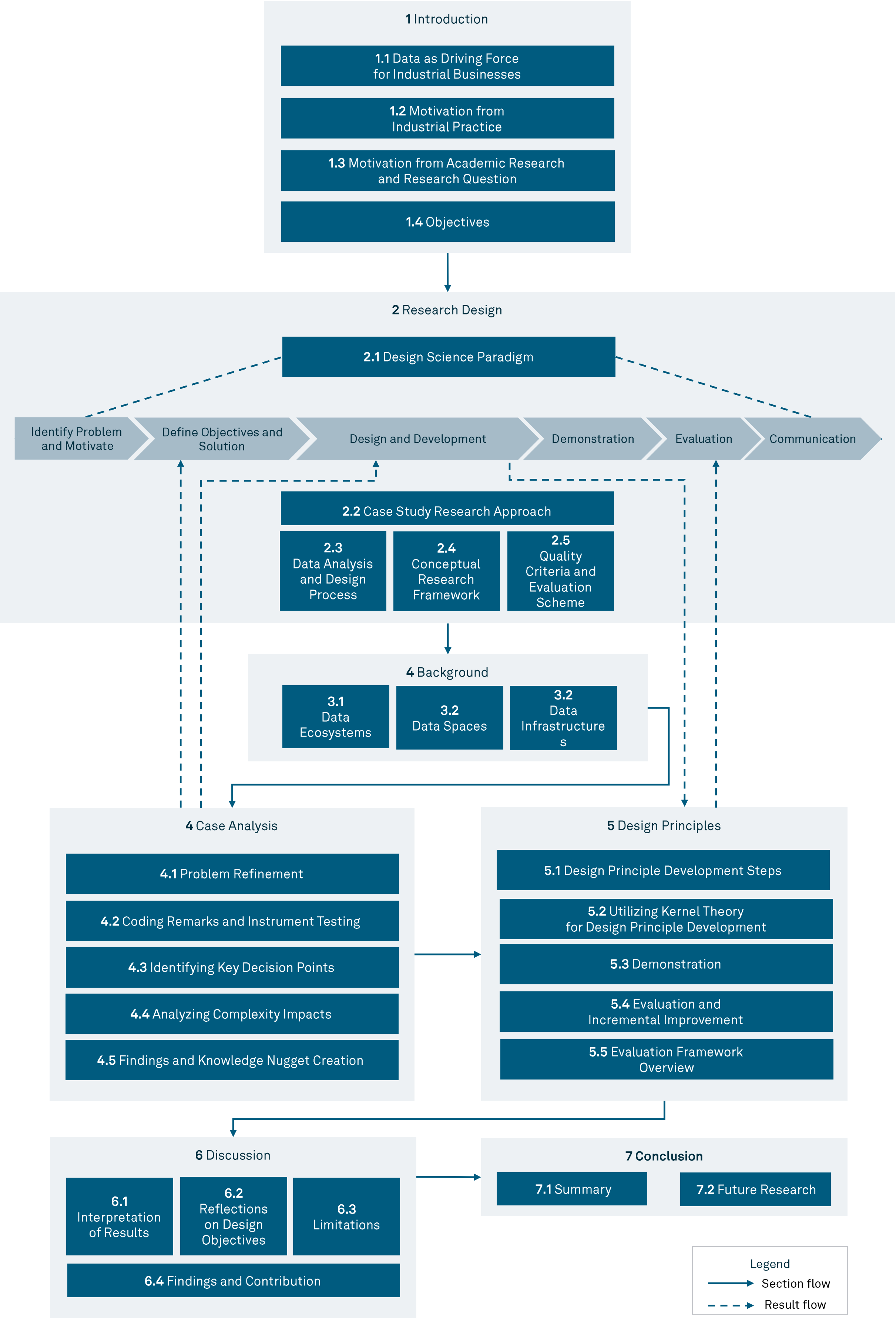

Figure 2 Structure of the Work