

## Next Generation Data Infrastructures

### Towards an Extendable Model of the Asset Management Data Infrastructure as Complex Adaptive System

Brous, Paul; Janssen, Marijn; Herder, Paulien

**DOI**

[10.1155/2019/5415828](https://doi.org/10.1155/2019/5415828)

**Publication date**

2019

**Document Version**

Final published version

**Published in**

Complexity

**Citation (APA)**

Brous, P., Janssen, M., & Herder, P. (2019). Next Generation Data Infrastructures: Towards an Extendable Model of the Asset Management Data Infrastructure as Complex Adaptive System. *Complexity*, 2019, Article 5415828. <https://doi.org/10.1155/2019/5415828>

**Important note**

To cite this publication, please use the final published version (if applicable). Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

## Review Article

# Next Generation Data Infrastructures: Towards an Extendable Model of the Asset Management Data Infrastructure as Complex Adaptive System

Paul Brous , Marijn Janssen , and Paulien Herder

*Delft University of Technology, Jaffalaan 5, 2628 BX Delft, Netherlands*

Correspondence should be addressed to Paul Brous; [p.a.brous@tudelft.nl](mailto:p.a.brous@tudelft.nl)

Received 28 September 2018; Accepted 17 January 2019; Published 24 February 2019

Academic Editor: Dimitri Volchenkov

Copyright © 2019 Paul Brous et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Organizations are increasingly looking to adopt the Internet of Things (IoT) to collect the data required for data-driven decision-making. IoT might yield many benefits for asset management organizations engaged in infrastructure asset management, yet not all organizations are equipped to handle this data. IoT data is collected, stored, and analyzed within data infrastructures and there are many changes over time, resulting in the evolution of the data infrastructure and the need to view data infrastructures as complex adaptive systems (CAS). Such data infrastructures represent information about physical reality, in this case about the underlying physical infrastructure. Physical infrastructures are often described and analyzed in literature as CASs, but their underlying data infrastructures are not yet systematically analyzed, whereas they can also be viewed as CAS. Current asset management data models tend to view the system from a static perspective, posing constraints on the extensibility of the system, and making it difficult to adopt new data sources such as IoT. The objective of the research is therefore to develop an extensible model of asset management data infrastructures which helps organizations implement data infrastructures which are capable of evolution and aids the successful adoption of IoT. Systematic literature review and an IoT case study in the infrastructure management domain are used as research methods. By adopting a CAS lens in the design, the resulting data infrastructure is extendable to deal with evolution of asset management data infrastructures in the face of new technologies and new requirements and to steadily exhibit new forms of emergent behavior. This paper concludes that asset management data infrastructures are inherently multilevel, consisting of subsystems, links, and nodes, all of which are interdependent in several ways.

## 1. Introduction

Public utilities infrastructure is developed over many years, and decisions regarding this infrastructure have to be made in the midst of a good deal of uncertainty regarding the future [1]. Technology, policy, and stakeholders are some examples of influences that can change over time, greatly increasing the potential risks of infrastructure failure [1], and, due to these fluctuating influences, infrastructures have been seen to have extraordinary complexity [2]. As seen in Industry 4.0 developments [3, 4], more and more, modern organizations tasked with the management of infrastructure are relying on data to improve their decision-making capability [5]. The benefits of Internet of Things (IoT) are often found in the provision of data which form the foundation of the information required by organizations for the improvement

of their decision-making capabilities. However, many organizations are not yet equipped to manage this data and many benefits of IoT are often not achieved. The objective of the research is therefore to develop an extensible model of asset management data infrastructures which helps organizations implement data infrastructures which are capable of evolution and aids the successful adoption of IoT. We seek to achieve this objective by answering the following research questions: (1) “what does a model of asset management data infrastructures look like”, (2) “what are the benefits of the model for the implementation of data infrastructures”, and (3) “what are the benefits of taking a complex adaptive system view”?

The scope of this research is asset management of public utilities in the infrastructure domain. According to Mohseni [6, p. 962], asset management (AM) is, “a discipline

for optimizing and applying strategies related to asset life cycle investment and work planning decisions". As such, this research helps practitioners identify how IoT may be used to improve their asset management decision-making in complex environments. According to Herder, Bouwmans, Dijkema, and Stikkelman [7], public utility infrastructures are complex systems, however, although public utility infrastructures are often approached as Complex Adaptive Systems (CAS), their underlying data infrastructures hardly are. In this research we follow Barnes, Matka, and Sullivan [8, p. 276] who define CASs as "open systems in which different elements interact dynamically to exchange information, self-organize, and create many different feedback loops, relationships between causes and effects are nonlinear, and the systems as a whole have emergent properties that cannot be understood by reference to the component parts", whereas the conceptualization of physical infrastructures is focused on the evolution and adaptive nature of infrastructures, data infrastructures are often conceptualized as being static having limited change. Since asset management data infrastructures are deeply embedded in the asset management organization, they not only are subject to rapid technological changes, but also have to keep up with institutional and economic developments, such as reduced budgets, increasing digitalization, and structural changes to business processes [9]. The challenge is to provide technical flexibility and budget flexibility to ensure that the system can adapt the initial design to these changing requirements.

In this paper we develop a model of asset management data infrastructures using a CAS lens. The contribution originates from (1) being the first one to use a CAS lens to design asset management data infrastructures, (2) deriving the essential elements of data infrastructures as coded in the propositions, (3) demonstrating the value of taking a CAS lens by showing that this results in more robust designs, and (4) showing that IoT demands new types of data infrastructures that need to evolve with changes in the environment.

Two research methods were used to answer the research questions and derive the requirements of the asset management data infrastructure model. First, a systematic literature review identifies theoretical elements and behaviors of asset management data infrastructures as CAS. Analysis of the literature led to the formation of four propositions for modelling asset management data infrastructures. In the second method, these propositions were confirmed using a real world IoT case study. This paper reads as follows: in Section 2 the research and design methods are described; the results of the literature review are discussed in Section 3; the results of the case study are described in Section 4; in Section 5 the asset management data infrastructure model is described; the propositions are discussed in Section 6; and conclusions are drawn in Section 7.

## 2. Research Methods

*2.1. Literature Review Method.* The aim of the literature review is to derive initial requirements of the asset management data infrastructure and identify the theoretical elements

and behaviors of asset management data infrastructures. As we adopt the CAS lens in this research in the development of the asset management data infrastructure model we focus on papers related to descriptions of data infrastructures from a CAS perspective. Webster and Watson [10] criticize the Information Systems (IS) field for having very few theories and quality literature review, believing that a dearth of proper literature reviews has, in the past, hindered theoretical and conceptual progress in information systems research [10, 11]. In this research we follow the method proposed by Webster and Watson [10]. There is only limited research on asset management data infrastructures [12] and data infrastructure models for the adoption of IoT in asset management are missing. The literature review was completed in June 2018. During the selection of literature, we found that many articles mentioned CAS as a theory without elaborating on the body of it. They were therefore not included in this analysis. Eighteen articles were selected based on the criteria that they included a theoretical discussion of the characteristics of CAS theory in data infrastructure context.

Following Webster and Watson [10], the literature review was developed concept centrally. During the reading phase, we compiled a matrix of concepts into which the literature was grouped. According to Denyer and Tranfield [13], the aim of analysis of literature is to break down individual studies into constituent parts. An important purpose behind this activity is to analyze consistency of interpretation and definitions [10]. We therefore followed the recommendations of Wallace and Wray [14] and Denyer and Tranfield [13] and collated the literature according to a series of questions as listed below.

- (i) What are the general details of the study?
- (ii) What type of study is this?
- (iii) What are the broad aims of the study?
- (iv) In which context was the study conducted?
- (v) What are the key findings?

Context is important in a systematic review [13] so we grouped the key concepts of data infrastructures as CAS according to focus areas identified within the broad aims of the studies. Based on these groupings we derived four propositions. In the literature, theoretical precepts are often discussed, but there are few systematic accounts of the application of data infrastructures in infrastructure management in practice and how data infrastructures evolve. We therefore tested the validity of the propositions by means of an exploratory case study as described below in Section 2.2.

*2.2. Case Study Method.* As the literature review did not provide us with definitive results as to the elements of asset management data infrastructures and their relationships, the second step was to conduct a case study to gain a deeper understanding of the manifestation of data infrastructures in a real world setting. The paper uses case study research as a methodological approach to examine CAS characteristics of asset management data infrastructures within the contemporary phenomenon of IoT adoption in asset management

TABLE 1: Case study overview.

Attribute	Value
Organization	RWS
Name	Weigh-In-Motion
Domain	Road Management
Number of Measuring Stations	18

organizations. The research design follows the case study methodology proposed by Yin [15]. The design of case study research includes the research questions, the propositions for research, the unit of analysis, the logic which links the data to the propositions, and the criteria for interpreting the findings [15]. Propositions for the case studies were finalized based on the findings from the literature review. The unit of analysis, the asset management data infrastructure, sets the boundaries for the case regarding the generalizability of its results. From an IoT adoption standpoint, the asset management organization is characterized by the intensive use, management and maintenance of large scale, and public utility infrastructure. The paper studies how an asset management organization uses IoT. Along with a clear understanding of the unit of analysis, case selection is crucial for building theory from case studies because it is case selection that determines the external validity of the case study and the limits for generalizing the findings [15]. The case of IoT adoption investigated was selected based on the criticality of its use and importance to the organization. The chosen case of IoT adoption was the automatic measurement of the weight of vehicles over the Dutch National Highways, “Weigh-In-Motion” (WIM). Table 1 presents an overview of the case.

To prepare the organization for the case study research project, RWS was provided with information material outlining the objectives of the project. Following Yin [15], multiple data sources were used. Although our unit of analysis is the organization, by interviewing persons within the cases it helped to better understand and capture the model requirements. RWS allowed the researchers unrestricted access to subject matter experts and internal documentation for all the cases. This helped ensure the construct validity of the case study [15]. Interviewees were selected on the basis that they were intimately involved in the project as early adopters. Interviewees were selected from three levels in the organization, namely, the strategic, tactical, and operational. Triangulation of characteristics of data infrastructures as CAS found within the cases was made by listing data infrastructure characteristics found in internal documentation and comparing these to the data infrastructure characteristics exposed in the interviews. There were several iterations throughout the research as the literature and case introduced new data infrastructure characteristics. During the research the characteristics of data infrastructures as CAS found in literature were listed and compared with the evidence of data infrastructure characteristics pertaining to IoT adoption found in the case study analysis.

*2.3. Model Design Method.* According to Janssen and Verbrack [16], the best models seek to reduce the semantic

gap between the units of analysis and the model constructs, whilst, as suggested by Curtis, Kellner, and Over [17], removing unnecessary detail and highlighting the essence of the problem. Following Weijnen et al. [18], this paper argues that the sociotechnical complexity of infrastructure systems calls for the combination of object-oriented and agent-oriented perspectives. Model designs of CASs often use the concept of agents for interacting with elements in the system [7, 18]. According to Weijnen et al. [18], the “cross-over” modelling technique forces the modeler to consider problems from the agent perspective, whilst providing insight into known and unknown variables such as the relationship between agents. Object oriented environments require communication between objects [16]. Implementing an agent within an object orientation by developing the objects as agents allows an agent to comply with the common characteristics of agents, such as autonomy, communication, and behavior. This approach breaks up the asset management data infrastructure into reusable parts without forcing limitations on extensibility.

The model is built using the Resource Description Framework (RDF) as specified by the World Wide Web Consortium (W3C). RDF was originally designed as a metadata model and has come to be used as a general method for conceptual description or modeling of data [19]. The RDF data model is based upon the idea of making statements about resources in the form of subject–predicate–object expressions, known as triples [19, 20]. According to Harth and Decker [21], this theoretically makes an RDF data model better suited to knowledge representation than other relational or ontological models. RDF data is still often persisted in relational databases or native representations known as “triple stores”.

### 3. Results of Literature Review

*3.1. Asset Management Data Infrastructures as Complex Adaptive Systems.* By conceptualizing data infrastructures as CAS, policy-makers and decision-makers can gain a better understanding of the dependencies involved [22], ensuring that typical characteristics of CAS are taken into account as suggested by Herder, Bouwmans, Dijkema, Stikkelman et al. [7]. In this research we follow the definition of Wollmann and Steiner [23] (p. 2) and define CAS as “systems made up of components (agents) that interact with one another according to a set of rules. The evolution of the system is the result of interactions between agents, where each of them acts in response to the behavior of the other agents in the system, which ensure it has its own dynamic”. CASs are often described as systems of interactive, mutually interdependent, individual elements which merge over time into coherent forms, adapting and organizing themselves without any singular entity deliberately managing or controlling them [24]. CASs are “dynamic systems” which are able to adapt and evolve to changing circumstance [25].

According to Auyang [26], within CASs, individual agents adapt as they interact with each other and their environment. CASs are often described by their characteristics, made up by elements and behaviors. CAS elements are sets of physicalities which, working together, make CASs different

from other systems. Similarly, CAS functions and operations make the overall behavior of CASs unique. There have been a number of calls for attention to this topic [22, 27] as few researchers have made this distinction when defining CAS in the information systems domain. The contribution of this research is to clarify the characteristics of CASs with regard to data infrastructures by cataloguing them according to their elements or behaviors. These elements and behaviors are identified below in *italic*.

**3.2. CAS Elements of Data Infrastructures.** CASs consist of relatively stable and simple *components* [23, 28–30], building blocks which are the constituent parts of the system. The overall behavior of a CAS emerges from the activities of lower-level components. This emergence is the result of an organizing force that can overcome a variety of changes to these components although, typically, a complex system will die when an essential component is removed [31]. Brous et al. [12] have identified three essential components of data infrastructures, namely, *data*, *people*, and *technology*. Technology can also be further separated into *hardware*, the collection of physical components that constitute an information system, and, *software*, that part of an information system that consists of computable instructions. People and, increasingly, technology are impacting the data infrastructure through *agency*. An agent is something or somebody that “can be viewed as perceiving its environment through sensors and acting upon that environment through actuators” [32, pp. 75]. When people or things act (or react) on an environment, that environment can be changed in unexpected ways [33]. Generally, actors follow a *schema* or “shared rules” such as norms, values, beliefs, and assumptions [34]. But when internal or external actors act, the *environment* in which data infrastructures exist may change often and quickly, forcing the data infrastructure to evolve and adapt to these changes. This section continues by further discussing each of these *elements* of data infrastructures in brief.

*Data* has long been recognized as a core component of information systems and has been generally defined as the measure or description of objects or events [35, 36]. The term “data” is often used to refer to either raw data or to information [37–39]. However, the two are not the same [36]. As such, the scope of data infrastructures is difficult to define. The term “data” is often distinguished from “information” by referring to data as raw data and referring to information as data put in a context or data that has been processed [40, 41].

The inherent challenge with these definitions occurs when data and information is registered and digitalized. From an information systems (IS) perspective, data and information can both take digital forms and, in these forms, are often, in practice, collectively referred to as data. For example, in an IoT environment, sensors such as temperature gauges make observations or measurements about an object or its environment, which may be registered in a system and is often referred to as raw data. This data can also often be enriched with other descriptors which help identify an object or thing, or the environment, infrastructure, system, or network in which the sensors, object, or thing can be found. An example of this would be a name given to a

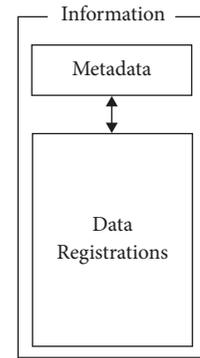


FIGURE 1: The relationship between data elements and information.

person or object. In practice, these identifiers are often referred to as “master data” [42, 43]. Data can also enter a data infrastructure as the description of an event, such as commercial credit card purchases, stock market trades, or HTTP requests to a web server. This type of data is often known as “transactional data” [44].

But for information to be gained from all this data, context is required. This contextual data is gained from data which describes the data that is being created, often referred to as “metadata”. Often, metadata also provides data about the sensor itself or about the object or thing that is being sensed. Metadata is often defined as data about data [37, 45]. As such we must also recognize that metadata is also data. According to Khatri and Brown [37], metadata describes what the data is about and provides a mechanism for a concise and consistent description of the representation of data, thereby helping interpret the meaning or “semantics” of data. Metadata is often stored in a registry or repository [45]. Khatri and Brown [37] describe different types of metadata as being physical, domain independent, domain-specific, and user metadata which play roles in the discovery, retrieval, collation, and analysis of data. According to Khatri and Brown [37], physical metadata includes information about the physical storage of data; domain-independent metadata includes descriptions such as the creation or modification of data and the authorization, audit, and lineage information related to the data; and user metadata includes annotations that users may associate with data items or collections.

Figure 1 shows that information can be gained by combining data (from the registration of observations, measurements, decisions, or transactions) with metadata (data which provides context). In practice, this information is often stored in within data stores such as data warehouses [46] and visualized in the form of reports. The buildup of this information over time becomes knowledge which is also often stored digitally within knowledge management systems [47]. The lines of responsibility may often become blurred as multiple users combine multiple data sources and data types to create multiple information products.

*Technology* is required to manage data. This technology must support the data management process [48]. An often-faced problem faced by the data analysts is that a vast quantity of data is available, but the format, quality, and

precise location are often not known, making retrieval and use difficult [48–50]. According to Yue et al. [51] the core of the IoT lies with the sharing of information between things and things or between people and things. Yue et al. [51] summarize the basic characteristics of things as comprehensive perception, reliable transmission, and intelligent processing. Comprehensive perception includes the acquisition of observations or measurements by using perception, acquisition, and measurement technology such as RFID [52] and two-dimensional code and sensors. Reliable transmission includes ensuring that the objects have access to information networks and can realize reliable information interaction and sharing through communications networks. Intelligent processing is the analysis of sensor data by using a variety of intelligent computing technology, to achieve intelligent decision-making and control [51]. As such, data infrastructures are increasingly being migrated to cloud solutions [53] whereby service providers provide the hardware and software necessary to manage the data resources [54]. According to Vaquero et al. [54], infrastructure providers manage a large set of computing resources, such as storing and processing capacity and are able to split, assign, and dynamically resize these resources to build ad hoc systems as demanded by customers. This is commonly known as the Infrastructure as a Service (IaaS) scenario [55]. Cloud systems can also provide the software platform where systems run on [52]. This is known as Platform as a Service (PaaS) [54, 55]. Finally, there are services which run applications such as online alternatives of word processors or spreadsheets. This scenario is often called Software as a Service (SaaS) [54, 55].

In a CAS, multiple *agents* often interact with one another in large variety of ways. Agents are entities that have the ability to intervene meaningfully in the course of events [34]. Data infrastructures include people as agents. People are seen as a key element in data infrastructures as people are responsible for the decision-making, design, implementation, and use of the data infrastructure [27, 56, 57]. Knowledge management is often of importance [58]. Local knowledge is often central to the ongoing maintenance of data, particularly in the face of unanticipated and unpredictable changes in local context and practice [58] as people have a direct influence on the role of organizational culture within data infrastructures, and effective data infrastructures are developed and applied around commonly felt needs [59]. Significantly, artificial intelligence is becoming more and more prevalent in service oriented environments, especially in the form of software commonly known as “bots” [60]. As such, artificial intelligence and robotics as agents are beginning to play an important role in the development of data infrastructures as more and more infrastructure management processes become automated. Agents have degrees of connectivity with other agents, and this connectivity follows a schema that determines the states and rules of the agents’ behavior [34].

*Schema* refers to shared rules [34], and agents use rules to make decisions and have frames of reference (schemata) which help them interpret and evaluate information [61]. Roles and rules are negotiations and gambits in the competition to define and construct meaning between agents. In fact, CASs can have competing stakeholders [62] and

competing schemata. The schemata that prove to be the most resilient and significant are the ones that are enforced. In this research we identify the schema of data infrastructure as being data governance, following Brous, Herder, and Janssen [63] and Malik [64]. According to Khatri and Brown [37], data governance refers to what decisions must be made so that proper management and use of data is organized (decision domains) and determines who makes the decisions (locus of accountability for decision-making). Wende and Otto [39] suggest that, in the past, organizations have often given responsibility for data management mostly to IT departments, often ignoring critical organizational issues. Data governance is a complex undertaking and many data governance initiatives in public organizations have failed in the past. Principles of data governance include organization of data management, ensuring alignment with business needs, ensuring compliance, and ensuring a common understanding of data [63].

However, the organization of data governance should not be a “one size fits all” approach and data governance must be institutionalized through a formal organizational structure that fits with a specific organization. Data governance should also ensure that data is aligned with the needs of the business. This includes ensuring that data meets the necessary quality requirements, ensuring alignment can take the form of defining and monitoring and enforcing data policies (internal and external) throughout the organization.

Establishing and enforcing policies regarding the management of data is important for an effective data governance practice. According to Dawes [65], data policies often reflect organizational choices about how data should be managed. Dawes [65] suggests that applying data principles to data management provides broad general guidance and helps to organize data management processes. However, organizations also need to properly understand what the data to be managed means, and why it is important to the organization [63, 66].

There is a need for interoperability through standardization due to a variety of data formats, protocols, and data types. Standards are “an agreed upon set of rules that are established by an authority” [67, p. 1090]. Many researchers (e.g., [27, 56, 67]) believe that standards play an important role in data infrastructures. Their importance lies within their endorsement by authorities and the compromise that these authorities have reached [67]. This endorsement encourages the wide implementation of the standards, improving interoperability, and supporting the data management process, including key processes such as collection of raw data, storage and maintenance of data, and user retrieval and manipulation of data [68–70].

A data infrastructure, as CAS, both reacts to and creates the *environment* it is operating in [12, 34]. In this way, a data infrastructure is inseparable from its environment. A CAS and its environment interact and create dynamic, emergent realities. The environment forces changes in the CAS, which in turn induces changes in the environment. Choi et al. [34] explain this phenomenon with the example of a team. As team members grow more cohesive, they collectively become more distant from the outside environment, and vice versa. Such

TABLE 2: CAS elements of asset management data infrastructures.

Elements	Sub-Elements	Literature Sources
Components	Data, technology, agents	[27–30, 34, 61, 62, 71, 74–77]
Schema (data governance)	Organization, alignment, compliance, common understanding	[22, 27, 29, 30, 61, 71, 76, 78, 79]
Environments	Internal (e.g., systems of systems), external (e.g., physical, political, economic)	[71, 77, 78, 80]

interdependencies ensure that a great deal of dynamism will exist in the environment. A dynamic environment surrounds complex adaptive systems that adapt and evolve to maximize some measure of fitness to their environment [34, 71]. When individual components contribute in different ways in a tightly coupled manner, an optimal state may become difficult to find, as many local optima may exist [34]. Most systems are nested within other systems and many systems are systems of smaller systems [22]. A system is a set of interrelated elements [72]. Each of a system's elements is connected to every other element, directly or indirectly. A system of systems is a collection of systems which creates a new, more complex system offering more functionality and performance than simply the sum of the constituent systems. While the individual systems constituting a system of systems can be very different and operate independently, their interactions typically expose and deliver important emergent properties [73]. Due to the many peaks, or possible states, the environment of a CAS is often considered to be “rugged”.

Based on the discussion above, Table 2 summarizes the CAS elements related to asset management data infrastructures.

### 3.3. CAS Behaviors of Asset Management Data Infrastructures.

CAS behavior emerges because many of the simple components interact simultaneously. The whole system is different from the sum of its parts [81] which means that CASs cannot be adequately analyzed by examining their parts separately. The greater the variety within the system, the stronger it is [22]. CASs rely on ambiguity, paradox, and contradictions to create new possibilities. According to Rupert, Rattrout, and Hassas [29], the variety of skills and strategies of agents within a CAS ensures dynamic and adaptive behavior. For example, it is difficult for a single agent to evolve and become more useful in an isolated context [30], so there is a constant exchange of information and needs between the components and the actors in the system [27]. The relationships are complicated and massively entangled because the components are numerous and highly interrelated [81]. Also, many CASs are driven by many *interdependent* variables and behavior is often influenced by a wide variety of factors. In addition to being numerous, these variables are often nonlinear and discontinuous and some variables vary in their influence over time. They may, for example, lie dormant for long periods until a certain control parameter reaches a critical value and activates them.

Nonlinearity is the property in which the emergent behavior of a CAS is the result of a nonproportionate response to its stimulus. That means the behavior resulting from

the interactions between agents is more complicated than a simple summation of the simple agents [29, 76]. Thus, the system cannot be predicted by simply understanding how each component works and behaves. As such, data infrastructures, as CASs, are *dynamic*, and because of the number of agents, their interdependence, and their openness to external influences, changes constantly and discontinuously. Constant change in a data infrastructure is driven by the number of agents, their association with their own rules of behavior, and the interdependence between the agents and their environments [29, 81]. Under normal circumstances, a complex system maintains a quasi-equilibrium state, balancing between complete order and incomplete disorder [34]. This balance point, the “edge of chaos” [22], allows the system to maintain order while also enabling it to react to qualitative changes in the environment. The most productive state for a data infrastructure to be in is at the edge of chaos where there is maximum variety and creativity, leading to new possibilities. The system is normally attracted to its original pattern of behavior if it is affected by an environmental variable as once a data infrastructure reaches the state of being good enough, it will trade off efficiency with greater effectiveness [22]. However, sensitivity to environmental changes increases as the environment pushes the system farther and farther away from the point of quasi-equilibrium.

Janssen et al. [79] show that *coordination and connectivity* are important characteristics of data infrastructures. The term “coordination mechanism” denotes the way interdependencies between activities are managed [82]. A coordination mechanism determines how information is obtained and used in decision-making and manages the demand and need for data, for example, through self-organization, feedback [27], and planning, such as direct supervision [83] or mutual adjustment and standardization [84] and contracting [85]. Data infrastructures as CASs often take the form of a patchwork of functional components working together [12]. Allocation of resources is a very prominent kind of coordination mechanism within data infrastructures, since various resources are needed to perform a task [85, 86]. Contracting emphasizes the allocation of human capital and expertise to different tasks and the ways in which the agents in a CAS connect and relate to one another is critical to the survival of the system, because these connections form patterns and disseminate feedback [76]. The relationships among the agents are often considered more important than the agents themselves [22, 79]. These patterns of interactions can be explained in terms of a network of interconnections.

*Evolution* is a process of change and agility for the whole system [28]. In a CAS, agents are interconnected so that

the behavior of an agent is influenced by the behavior of other agents in the system. As one agent evolves, so does the other [28]. This process is often referred to as “coevolution” [34, 61, 79]. At a macrolevel, data infrastructures exist within their own environments, and they are also part of that environment. As their environments change, data infrastructures need to change to ensure a good fit with their environments. However, as they change they also enforce changes in their own environments in a continuous, reciprocal process of evolution [79].

*Adaptation* can be described as a change in the system’s structure (strategy) resulting from the system’s experience [24]. CASs are able to adjust and adapt themselves to external influences [27, 87, 88] and a data infrastructure will change constantly because of the continuous interactions and interdependence between its agents and its environment [29]. But the adaptive behavior of the system cannot be the result of completely random dynamics and [24] describes the evolution of these systems as the result of a strategy which combines exploration (to maintain a certain diversity) and exploitation (to reinforce promising tracks), which encourages adaptation. Behavior in a data infrastructure is induced not by a single agent, but rather by the simultaneous and parallel actions of agents within the system itself. In this way, behaviors emerge. In other words, new structures, patterns, and properties arise without being externally imposed on the system [34, 79, 89]. In this regard, macroscopic properties of a data infrastructure arise from the heterogeneity of its elements and its relevant properties [80]. The system displays a set of properties that is distinct from those displayed by any subset of its elements.

*Aggregation* is the behavior by which agents form groups that in turn can recombine to a higher level leading to the complex system [29]; it is the basis for identity [59, 74, 75]. According to [30], there are two important modes of aggregation in data infrastructures: (1) objects and (2) components. Forming components from objects and forming systems from components are higher-level aggregation. Meta-agents, such as an enterprise, are formed of aggregates of lower agents such as systems which are formed of aggregates of components, which are formed of aggregates of objects [30].

Cilliers [87] defines *self-organization* as a process in which a system can develop a complex structure from unstructured beginnings. Agents learn and adapt to actions of other agents [90] which results in the structure and dynamics of a data infrastructure [61]. In a data infrastructure, there is often no centralized control that dictates the system’s overall behavior [27]. Rather, order may emerge as agents learn to govern their own rules of behavior and adapt to their environment as suggested by Rupert, Rattrout, and Hassas [29]. Formal order is not externally imposed from outside of the data infrastructure, but rather *emerges* from interactions between agents [91]. In this regard, data infrastructures are inherently multilevel as the order is seen as an emergent property which results from lower levels of aggregate behavior as suggested by Anderson [92]. For example, as seen in the case below, the case is reliant not only on the IoT system which is developed specifically for this case, but also on multiple data sources originating from other systems. Furthermore, the

data created in the case below is also utilized for purposes other than those for which the system was designed.

*3.4. Literature Review Conclusions and Propositions.* Current data models of asset management data infrastructures are often static, a characteristic which poses constraints for extensibility and adaptability in the face of adoption of new data sources such as IoT and changing requirements. However, data infrastructures, like their real world infrastructures are complex and should be treated as CAS. As CAS, data infrastructures are made up of relatively stable *components*. Typical components found in data infrastructures are *data*, *technology*, and *agents*. This leads us to our first proposition which reads as follows: (1) IoT data infrastructures are composed of data, agents, and technology.

*Data governance* refers to the decisions that need to be made to ensure proper management and use of data, including establishing who determines the requirements for data quality. Data governance should be institutionalized through a formal organizational structure; however, the organization of data governance should not be a “one size fits all” approach and should fit with a specific organization. This leads us to our second proposition which reads as follows: (2) agents operating in IoT data infrastructures are guided by schema which is defined by data governance.

A data infrastructure is inseparable from its environment, both reacting to and creating the *environment* it is operating in. In this way, a data infrastructure and its environment interact and create dynamic, emergent realities. Due to the many possible states, the environment of a data infrastructure is often considered to be “rugged”. This leads us to our third proposition which reads as follows: (3) IoT data infrastructures develop within environments that are often in a constant state of flux.

Data infrastructure *behavior* emerges because many of the simple components interact simultaneously. The diversity of skills, experiments, strategies, and rules of different agents within a data infrastructure ensures its dynamic adaptive behavior and there is a constant exchange of information and needs between the components and the actors in the system. The relationships are complicated and massively entangled because the components and agents are numerous and highly interrelated. This leads us to our fourth proposition which reads as follows: (4) IoT data infrastructures emerge, evolve, and adapt over time.

As read above, conclusions taken from the literature have resulted in four propositions which we partly test by means of an exploratory case study described in the following sections.

## 4. Case Study Results

The main goal of the case study is to test the propositions developed on the evidence provided by the literature review. RWS focusses on having the capability to make the right choices with regards to management and maintenance of their assets [93]. However, these choices are not always straightforward, as, for example, during maintenance procedures, roads often still need to be accessible. RWS therefore requires data that can be trusted to conform to a certain

quality. Characteristics of data infrastructures as CAS discovered during the literature review provided the basis for the propositions. We investigated whether these characteristics occurred and were adhered to in practice. The case study research involved the use of multiple methods for collecting data. The propositions guiding the case studies are derived from conclusions taken from the literature review and are described in Section 3.4 above. In the following sections the case is described and discussed in relation to the case study propositions.

*4.1. Case Study: Overloading of Vehicles (Weigh-In-Motion).* At present, RWS estimates that at least 15 percent of freight traffic on the Dutch national road network is overloaded. Overloading of heavy vehicles causes road pavement structural distress and a reduced service lifetime [94, 95]. Effectively reducing overloading reduces the damage to the road infrastructure, lengthening the road's lifetime, and reduces the frequency of maintenance. The damage to pavements and installations by overloaded trucks in 2008 was estimated to be at least 34 million euros per year. In addition, the extra maintenance required creates a significant amount of traffic disruptions. These disruptions are estimated to cost several million euros per year. The ambition of RWS is to increase the operational efficiency and effectiveness of the approach of overloading and thus reduce maintenance costs. Traditional enforcement of laws and regulations regarding overloading involved the use of physical measuring stations. This included manual checks by the police in which many vehicles were selected where overloading was suspected but uncertain. This often led to unnecessary inconvenience to citizens as vehicles were often stopped unnecessarily. Until 2010, The Netherlands had 5 measuring stations nationwide. It was suspected that many carriers could avoid these stations by choosing alternative routes whilst retaining their economic gain. In response, RWS created a national network of monitoring points, the "Weigh in Motion" (WIM) network. The WIM system is one of the most advanced overloading measurement systems in the world. In the period 2010-2013, RWS built a nationwide network of WIM stations, a total of 22 measuring stations. In addition to sensitive sensors, cameras are also part of the WIM systems. The WIM network, consisting of measuring stations in the road on which the axle loads of heavy traffic is weighed, is used to support the enforcement of overloading by helping the enforcement agency to select overloaded trucks for weighing in a static location.

**Proposition 1.** *IoT data infrastructures are composed of data, agents, and technology.*

Data on overloaded vehicles on the road are automatically sent from WIM to the Real-Time Monitor (RTM) web application which processes, stores, and publishes the data of all weigh points. The Inspectorate for the Living Environment and Transport (ILT) is then able to perform supervision and enforcement actions on overloaded vehicles in near-real time (within 10 seconds), improving the overall flexibility of the services as ILT and RWS can decide where and when offenders are controlled. The network provides

access to information about the actual load of the main road and about peak times when it comes to overloading. This provides RWS and ILT with the ability to collect information concerning the compliance behavior of individual carriers as, in addition to sensors, cameras are also part of the WIM systems. Via camera footage, the ILT can identify the license plates of vehicles that are overloaded and therefore the detect owner and / or licensee and address. The strategy is to tackle overloading by integrating roadside enforcement along with targeting carriers according to behavior based on the information from the system.

An enforcement chain is a mission critical system, where accuracy and reliability are essential. RWS faces and has faced a variety of impediments and challenges during the implementation and maintenance of the WIM network with regard to accuracy and reliability of the data. Configuration of the system is a delicate process. The WIM system can differentiate between the vehicle and the load, but not all vehicles weigh the same. Not all number plates are placed in the same place on the vehicle, and not all drivers have the same driving style. It is necessary to be able to account for drivers who drive very close to other vehicles, or those who change lanes during inspection (and thus have wheels in two different lanes). The configuration is closely monitored, but, according to RWS officials, a structured learning cycle with regard to data quality is still required. Some sources have questioned whether the reliability of the data is sufficiently well equipped and some interviewees raised questions about the quality of the data. According to an RWS official, "the quality of the data needs to be quantified, and solving data quality issues is incident driven". RWS project managers also cited several technological challenges due to IT infrastructure limitations which needed to be overcome and which no single market partner could supply at the time. IoT generates large amounts of data and this data needs to be processed near to real time so that inspectors can quickly identify trucks for roadside inspection. Based on this discussion we conclude that Proposition 1 can be confirmed.

**Proposition 2.** *Agents operating in IoT data infrastructures are guided by schema which is defined by data governance.*

Rijkswaterstaat is a process-based organization in which an executive board member is responsible for managing a particular primary process. The information management process at Rijkswaterstaat is managed by the Chief Information Officer (CIO) who, besides his executive role, is also the managing director of the Information Services division. The CIO is supported by a Chief Data Officer (CDO) and a Chief Technical Officer (CTO), both having advisory roles.

RWS has developed a data management organization to maintain their asset management data. This organization implements and enforces uniform data entry. All of the agents operating in the data management organization can be described by extending the agent class in the model. For example, divisions of RWS are organized according to geographic location, and each division is an independent agent which manages standardized processes in their own

way. Within the divisions, each individual person, in his turn, can act as an independent agent.

RWS has adopted an integral approach to managing its network of assets and has adopted a variety of coordination mechanism. According to a RWS official, “an integral approach to managing the network of assets helps us know better the quality that we desire from the performance of the assets”. This includes the Information Delivery Specification (IDS) which guarantees a uniform exchange of information on structures between the different partners. Planning by means of annual portfolio plans is also an important coordination mechanism. Based on this discussion we conclude that Proposition 2 can be confirmed.

**Proposition 3.** *IoT data infrastructures develop within environments that are often in a constant state of flux.*

RWS and ILT have been able to improve the efficiency of regulations as they are able to perform administrative enforcement through administrative fines for repeat offenders which are far in excess of the penalties for individual offenders. WIM can differentiate between the load and the vehicle. It is possible to identify not only the transporter, but also the owner of the load. Enforcement of regulations is therefore greatly improved. One of the initial challenges of the WIM project was the definition of the service and the identification of possible solutions. Initial proof of concepts used a combination of intermediate products to approximate the final solution. This led to several interoperability and integration issues which needed to be overcome. RWS noticed that the adoption of WIM has led to new products being offered by companies who may not necessarily be established partners of RWS and to the loss of old products being offered by more established partners. This has led to new streams of revenue for private parties. New revenue streams for the government also became clear as fines for overloading are automatically sent directly to offenders. This has led to new streams of revenue as, according to RWS officials, implementing WIM has led to a higher chance of catching actual offenders and better effectiveness of controls. The duality of achieving new revenue streams is that, not only implementation costs, but also maintenance costs of the WIM network are high as the sensors often come loose in the asphalt and the repair of the damage is very expensive.

The conflicting market forces created by the new demand have meant that RWS needed to rethink their approach to framework agreements with established parties. There are different perceptions of the level of ambition pursued by the WIM project. The WIM function has gradually changed from being a tool used to apprehend offenders to being a tool used for digital inspection. Analysis of the stored measurement data shows patterns, improving forecasting, and trend analysis. There is obviously something wrong with vehicles that are frequently flagged in the system. That may be reason to perform roadside inspections in a subsequent inspection or to visit the parent company for an inspection. The duality attached to the gain provided by being able to identify offenders, is the necessity for ensuring data privacy and data security. Any images or other data created by the

system which are made publicly available need to ensure anonymity. Furthermore, security of the data is of vital importance due to the importance of being able to prove offence. It must not be possible in any way to tamper with the “evidence” provided by the data. It is not yet possible to entirely automate the enforcement process, as physical testing is still required to legally prove overloading. The Dutch legal system does not yet fully trust WIM to provide legally conclusive evidence with regard to overloading. The interviewees believe that as an instrument to help roadside enforcement WIM works well, but there are difficulties in using WIM to legally prove offence. A new legal framework is required before this system is legally acceptable in The Netherlands. Based on this discussion we conclude that Proposition 3 can be confirmed.

**Proposition 4.** *IoT data infrastructures emerge, evolve, and adapt over time.*

The ability to detect overloaded trucks is based on data and it is possible to ensure owners of the carriers and load are also identified and thus enforce regulations at source. With regard to improving planning and maintenance, RWS's strategy was to outsource the operational side of WIM to external contractors which meant that divisions which previously did the work of weighing and monitoring vehicles needed to be reorganized to do other work. RWS initially outsourced the management of the system. However, RWS has since rescinded that tactical decision due to clashes in planning with other processes such as traffic management. According to a RWS Director, “in order to effectively manage the technology, it is important to have sufficient mandate to manage the entire chain”. Managing only the technology or parts of the system produces inefficiencies and can disrupt other processes, such as traffic management, if the overview of the system is not considered when planning maintenance.

Innovation was required in order to be able to ensure the required precision of the data required. Tensions arose as to where responsibility for innovation lay. As public sector organization, RWS did not wish to give market advantage to a single private sector party, but was also unwilling to develop the innovation internally. Introducing new technology to the market empowered citizens to develop new products and created new business opportunities. But the duality was that a RWS Director expressed concern with regard to the impact of the adoption of WIM by RWS on the private sector and conflicting market forces which WIM has introduced. As there were few private organizations capable of implementing WIM, if RWS would provide innovation opportunities to a single party, this would have provided that party with an unfair market advantage. The RWS Director explained that it is important to develop a procurement strategy with regard to IoT adoption. In this case, cooperation with the universities was sought to develop the required innovation. With the help of universities in the Netherlands, a proof of concept was developed, the results of which were made publicly available. Based on this discussion we conclude that Proposition 4 can be confirmed.

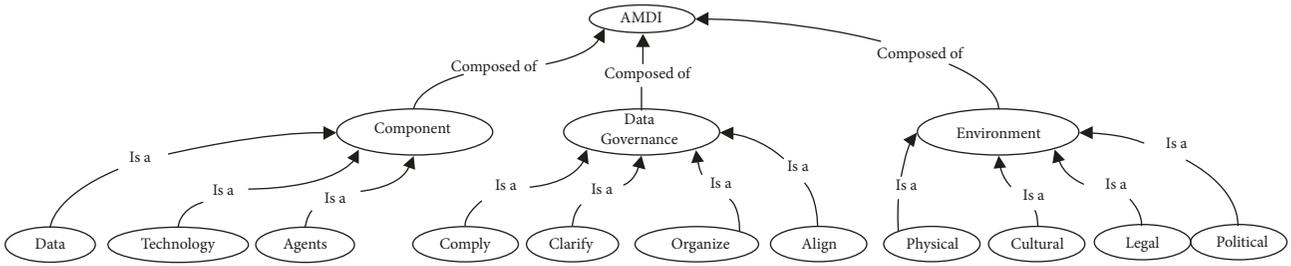


FIGURE 2: The main elements of the asset management data infrastructure data model.

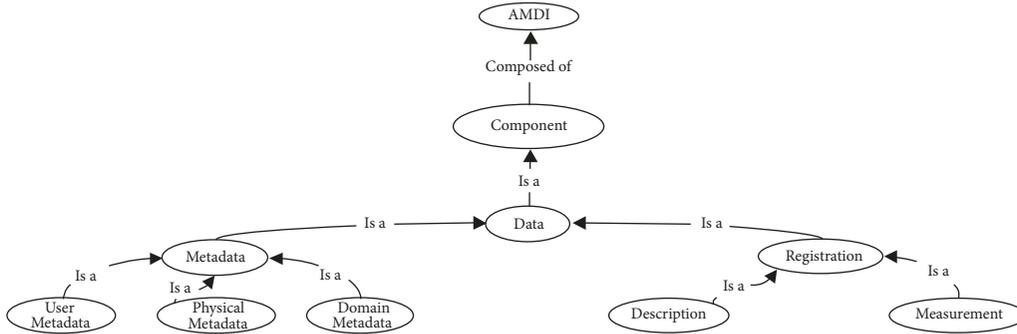


FIGURE 3: The data class model.

The results of the case study confirm the propositions which were developed on the basis of theoretical characteristics which were synthesized from the literature review. The following sections utilize the results of the literature review and case study to develop a data model of asset management data infrastructures as CAS.

### 5. Asset Management Data Infrastructure CAS Model

As demonstrated above in the case study, confirming the suggestions of de Man [59], the goals of data infrastructures are to facilitate and coordinate exchange, sharing, accessibility, and use of data and encompass complexes of interacting institutional, organizational, technological, human, and economic resources. For example, the goal of WIM is to improve the efficiency, speed, and quality of data collection by automating the weighing and inspection of freight trucks and improve the access and sharing of data between road inspectors, law enforcement, and road managers. The potential of data infrastructures to facilitate access to and sharing and communication of data may be subject to existing cultural, political, and societal factors [59]. For example, actors may want to maintain their powerful positions and prevent others from direct access to the data infrastructure, thus making it a means of domination and exclusion. As seen in the case study, IoT systems which produce trusted data are difficult to develop, and asset data is regularly considered to be lacking in quality [96]. Addressing this issue requires an approach which describes the sociological as well as the technological components [12]. The following sections describe the asset management data infrastructure data

model as CAS. Section 5.1 describes the classes of *components* of the data infrastructure as confirmed in Proposition 1 of the case study. Section 5.2 describes the classes of “schema” of data infrastructures, identified in this research as *data governance* (Proposition 2), and Section 5.3 describes the classes of *environments* (Proposition 3). Proposition 4, which proposes that data infrastructures evolve, adapt, and emerge over time is dealt with in the model by breaking up the asset management data infrastructure into reusable, logical parts without imposing a limitation to the extensibility of an element. Figure 2 depicts the main elements of data infrastructures as identified in the literature review and confirmed in the case studies.

5.1. *Components of Asset Management Data Infrastructures (Proposition 1).* Data has often been recognized as an important element in data infrastructures and has been generally defined as the measure or description of objects or events [12, 27, 35, 36]. In our model we follow Kettinger and Li [36] and consider data to be a set of interrelated data items that describe the attributes of subjects, objects, or events. These data elements as components of data infrastructures are encapsulated in the data class as seen in Figure 3.

The technology class includes the collection of Information Technology (IT) components, used in the production of data or in the development of information, such as data analysis or data management. According to Broadbent and Weill [97], a business driven IT means prioritization based on a strong understanding of an organization’s business strategy, often a challenge for many asset management organizations [98] due to conflicting interests between divisions. The technology class is depicted in Figure 4.

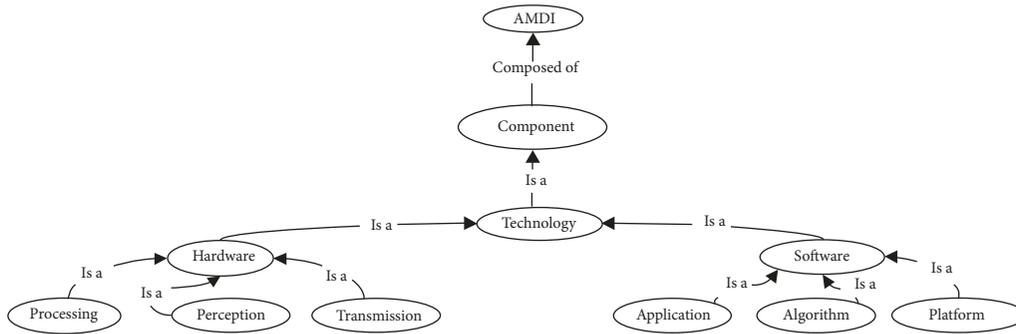


FIGURE 4: The technology class model.

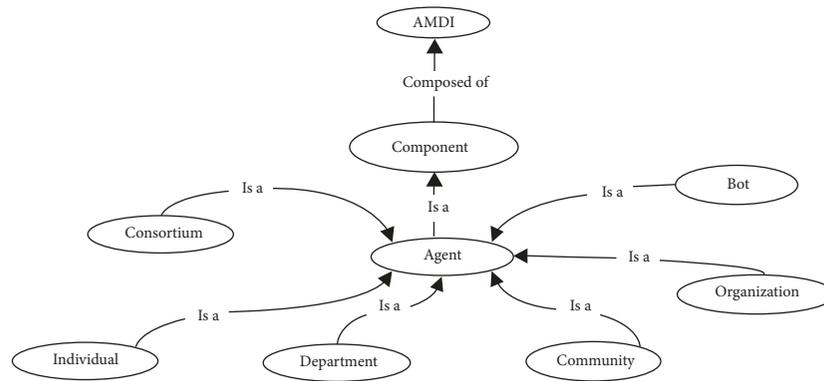


FIGURE 5: The agent class model.

As seen in Figure 5, in our model all independent actors are viewed as agents. We adopt Janssen and Verbraeck's [16] definition of an agents as being “autonomous, goal driven entities that are able to communicate with other agents and whose behavior is the consequence of their (1) observations, their (2) knowledge, and their (3) interactions with other agents” [16, p. 375]. CAS theory holds that multiple interactions between agents result in structural changes of the system at an aggregate level [75]. Changes to data infrastructures are structural changes that require the interaction of agents to coalesce around both technical changes to the data infrastructure, as well as social change to reflect social values as drivers of change. For example, whilst the rules of the system may be set at the strategic level or tactical levels, by overarching governance bodies, it often comes down to individuals to interpret and implement these policies at the operational level.

According to Janssen and Verbraeck [16], in many multiple agent architectures problems are often decomposed, with subproblems being assigned to specific agents. This resolves the greater problem through the inclusion of multiple agents. Within the model, each agent has a role to play in the implementation of the data infrastructure, based on their position within the organization and the underlying processes [99]. We thus further develop the data model by examining the characteristics of agents with regard to the underlying schema as described by the data governance class as seen in Section 5.2.

5.2. *Data Governance within Asset Management Data Infrastructures (Proposition 2)*. The data governance class defines the nature of agents and shapes the agents' behavior. Agents receive inputs and act on the environment, behavior often being viewed as a manifestation of intelligence [100]. Data governance provides the guidelines which guide the actions of the agents. In the model, the data governance class determines the behavior of the agent and how the agent chooses to organize their activities. The behavior is modeled in terms of the tasks that need to be accomplished given its position [16, 101]. The behavior of the agents dictates which technology is implemented and which data is developed and also dictates how the data and the technology are maintained. The data governance class is depicted in Figure 6.

Data governance is a complex undertaking. Principles of data governance should include the data management function and assigning roles and responsibilities, ensuring alignment with business needs, ensuring compliance, and ensuring clarification of how the data infrastructure has been set up, including definition of terms [63]. Data governance should also ensure that data is aligned with the needs of the business. This includes ensuring that data meets the necessary quality requirements which align with the rules and requirements of the business. Ensuring alignment can take the form of defining, monitoring, and enforcing data policies (internal and external) throughout the organization. Establishing and enforcing policies regarding the management of data is important for an effective data governance

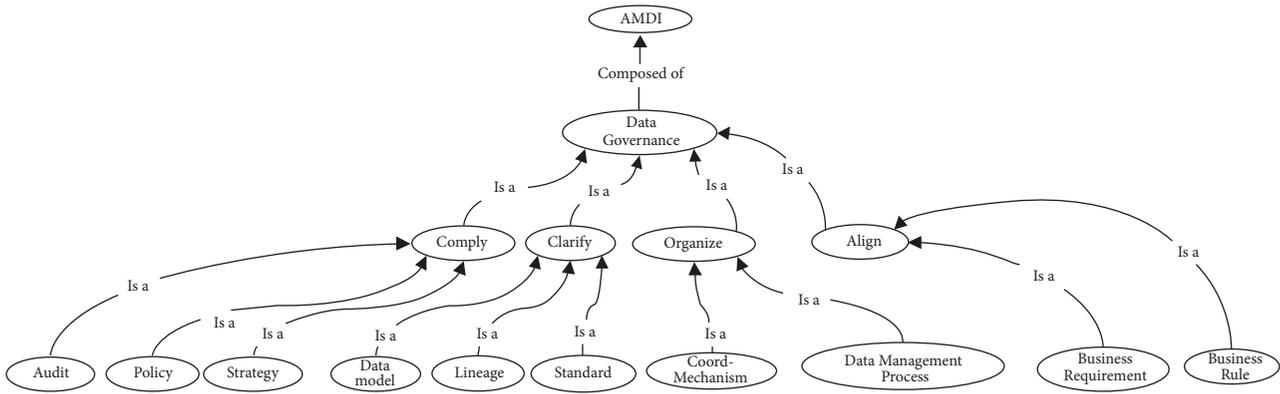


FIGURE 6: The data governance class model.

practice. But governing data appropriately is only possible if it properly understood what the data to be managed means, and why it is important to the organization [63]. Although data governance should be recognized as the schema which guides actors operating within and acting on the data infrastructure, it should also be recognized that data governance should be practiced in accordance with the environments within which the data infrastructure finds itself. As such, the organization of data governance should not be a “one size fits all” approach and the data governance organizational structure should fit with a specific organization. We therefore further develop the model by examining the environments of asset management data infrastructures as discussed in Section 5.3.

**5.3. Environments of Asset Management Data Infrastructures (Proposition 3).** The potential of data infrastructures to facilitate access to and sharing and communication of data may be subject to existing cultural, political, and societal factors. For example, actors may want to maintain their powerful positions and prevent others from direct access to the data infrastructure, thus making it a means of domination and exclusion. Taken a step further, Kim and Kaplan [77] believe that the weak cause-and-effect linkages and nonlinearity evidenced within data infrastructures is due in no small part to the reflexive interpretation of context and interests and actors should not be regarded as being passive. A data infrastructure is thus more than just a series of sociotechnical interactions, but a system comprised of calculating actors, each making moves on a coevolving landscape [77]. Due to the underlying phenomenon of coevolution, problems should be resolved in isolation as they change the context within which the other problems are framed. As such, the environments in which an asset management data infrastructure can be found can have a profound impact on how the data infrastructure evolves. For example, the lack of a legal framework to deal with automation of overloading inspections has had a delaying effect on complete automation, meaning that although WIM can be used to identify overloading, physical inspections still need to be made for legal purposes. Figure 7 depicts the environment class which often influences the boundaries, form, and evolution of data infrastructures.

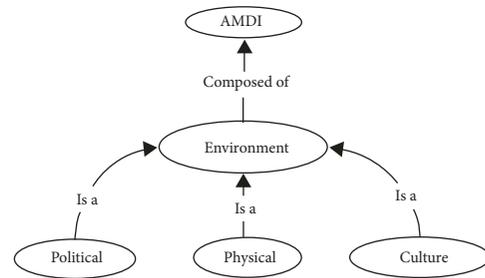


FIGURE 7: The environment class model.

## 6. Discussion

The asset management data infrastructure characteristics found in the literature are often theoretical concepts and it was often not clear how these concepts manifest in practice or how the concepts are interrelated. Furthermore, the literature review reveals that traditional asset management data infrastructure designs often do not take into account the complex, evolutionary nature of data infrastructures. In the literature, theoretical precepts are often discussed, but there are few systematic accounts of the application of IoT data infrastructures in infrastructure management in practice and how these concepts emerge or what the implication of adopting these concepts may be. Systematic analysis of the literature review resulted in the development of propositions which in this research are tested in part by means of an explanatory case study; however, more research in this area is required. These four propositions are discussed below.

The first proposition of the case reads as follows: “(1) IoT data infrastructures are composed of data, agents, and technology.” The case study shows that asset management data infrastructures are complex sociotechnical systems as, for example, understanding sociotechnical complex systems such as the WIM system requires knowledge of both the technical and the social systems; taking only a technical perspective would result in missing important information such as the impact of people on the choice of technology, or the impact of the organization structure on how people respond to adoption of IoT. Modelling the asset management

data infrastructure from either purely an actor perspective or from a technical system approach would therefore either provide too little opportunity for modelling the reflectivity of the actors or not provide enough detail for a complete design of the technical system. This research therefore made use of the “cross-over” modelling technique which forces the modeler to consider problems from the agent perspective, whilst providing insight into known and unknown variables such as the relationship between agents.

The second proposition states that “(2) agents operating in IoT data infrastructures are guided by schema which is defined by data governance”. The case study shows that a formalized data governance structure, which is a fit with the specific organization, does need to be implemented in order to enable IoT adoption in asset management organizations. This is because automating decision-making often incurs business process related changes which can be found in aligning complex data structures. For example, automating the monitoring of overloading means that weigh stations only need to be employed for suspect vehicles and not for every vehicle, greatly reducing the need for inspections and changing the process of vehicle inspection. This is reflected in the data governance alignment class. As such, as seen in the data governance organization class, it is important to ensure that data provenance is well organized so that it is clear where responsibilities and accountabilities lie throughout the data lifecycle. Within the context of the case study, instituting strong data governance procedures which align interfunctional teams behind a common goal has a positive influence on IoT adoption in asset management organizations.

The third proposition states that “IoT data infrastructures develop within environments that are often in a constant state of flux”. Environmental characteristics may refer to the sector within which the organization operates or may represent cultural, societal, political, or geographical conditions. The results of the case study show that asset management organizations with a high level of environmental complexity that also have access to high levels of financial and other resources are more enabled to adopt innovations such as IoT. For example, although the cultural, political, and physical environments in which WIM is managed presents unique challenges, RWS continues to manage it to an exceptional level of quality. RWS is reported to have access to sufficient financial resources and has a broad knowledge base and a strong political lobby. Within the context of the case study, greater environmental complexity in combination with access to sufficient financial resources may stimulate higher rates of IoT adoption in asset management organizations.

The fourth proposition reads as follows: “IoT data infrastructures emerge, evolve and adapt over time”. The case shows that behavior resulting from the interactions between agents is more complicated than a simple summation of the simple components. For example, it is insufficient to simply introduce the new system within the old processes. New processes need to be implemented and even new legal frameworks need to be developed. Thus, the system cannot be predicted by simply understanding how each component works and behaves. As such, within the context of the case

study, asset management data infrastructures, as CASs, are dynamic, changing constantly and discontinuously. Constant change in an asset management data infrastructure is driven by the number of agents, data governance, and the interdependence between the system and its environments. Formal order is not externally imposed from outside of the asset management data infrastructure, but rather *emerges* from interactions between agents.

## 7. Conclusions

IoT may provide a variety of benefits for asset managers such as reduced need for physical inspections due to the automation of real-time data which is of sufficient quality to automate operational, reactive decision-making, as well as allowing asset management organizations to develop a history and view of infrastructure assets for tactical planning and strategic trend analysis [9]. However, current data models of asset management data infrastructures are often static, a characteristic which poses constraints for extensibility and adaptability in the face of adoption of IoT. The first research question therefore asks what a model of asset management data infrastructures looks like. The model requirements which answer this question were derived from a systematic literature review and a case study in the asset management domain. The research shows that modelling the asset management data infrastructure from either purely an actor perspective or from a technical system approach would not sufficiently cover the dynamic nature of the data infrastructure. Our model breaks up the asset management data infrastructure into reusable, logical parts but does not pose a limitation to the extensibility of an element. As such we argue that the asset management data infrastructure model presented in this research also takes into account the typical characteristics of complex system design.

The second research question asks what the benefits of the model for the implementation of data infrastructures are. The model helps us understand the consequences for data infrastructure development and maintenance, particularly when there is a dependence on interactions between the elements of the data infrastructures such as when the development of a new dataset is announced, often the cause of major changes in the behavior and composition of the system, even when the anticipated situation does not arise. For example, new connections between data systems may be introduced in anticipation of a master data management project, greatly increasing the complexity and dependencies of the systems, even if the master data management project is eventually discontinued.

The third research question asks what the benefits of taking a CAS view are. This paper argues that designing asset management data infrastructures as CAS helps the designer meet new design challenges by helping to acquire a better understanding of the elements and behavior of asset management data infrastructures as sociotechnical, complex systems. For example, a CAS lens helps us to identify and better understand the key elements of asset management data infrastructures and mechanisms for their functioning and dealing with change.

## Data Availability

The case study data used to support the findings of this study is included within the article. Interview data is restricted due to privacy considerations.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work was supported by Rijkswaterstaat.

## References

- [1] P. M. Herder, J. De Joode, A. Ligtoet, S. Schenk, and P. Taneja, "Buying real options - Valuing uncertainty in infrastructure planning," *Futures*, vol. 43, no. 9, pp. 961–969, 2011.
- [2] A. Pagano, I. Pluchinotta, R. Giordano, and U. Fratino, "Integrating "hard" and "soft" infrastructural resilience assessment for water distribution systems," *Complexity*, vol. 2018, Article ID 3074791, 16 pages, 2018.
- [3] F. Shrouf, J. Ordieres, and G. Miragliotta, "Smart factories in Industry 4.0: A review of the concept and of energy management approached in production based on the Internet of Things paradigm," in *Proceedings of the 2014 IEEE International Conference on Industrial Engineering and Engineering Management, IEEM 2014*, pp. 697–701, December 2014.
- [4] S. Wang, J. Wan, D. Zhang, D. Li, and C. Zhang, "Towards smart factory for Industry 4.0: a self-organized multi-agent system with big data based feedback and coordination," *Computer Networks, Industrial Technologies and Applications for the Internet of Things*, vol. 101, pp. 158–168, 2016.
- [5] M. Janssen, *Designing Electronic Intermediaries: An Agent-Based Approach for Designing Interorganizational Coordination Mechanisms [Doctoral, thesis]*, Delft University of Technology, Delft, The Netherlands, 2001.
- [6] M. Mohseni, "What does asset management mean to you?" in *Proceedings of the 2003 IEEE PES Transmission and Distribution Conference and Exposition*, vol. 3, pp. 962–964, Greenwood Village, CO, USA, 2003.
- [7] P. M. Herder, I. Bouwmans, G. P. Dijkema, R. M. Stikkelman, and M. P. C. Weijnen, "Designing infrastructures using a complex systems perspective," *Journal of Design Research*, vol. 7, no. 1, pp. 17–34, 2008.
- [8] M. Barnes, E. Matka, and H. Sullivan, "Evidence, understanding and complexity: evaluation in non-linear systems," *Evaluation*, vol. 9, no. 3, pp. 265–284, 2003.
- [9] P. Brous, M. Janssen, and P. Herder, "Internet of Things adoption for reconfiguring decision-making processes in asset management," *Business Process Management Journal*, 2018.
- [10] J. Webster and R. T. Watson, "Analyzing the past to prepare for the future: writing a literature review," *MIS Quarterly*, vol. 26, no. 2, pp. 13–23, 2002.
- [11] Y. Levy and T. J. Ellis, "A systems approach to conduct an effective literature review in support of information systems research," *Informing Science: The International Journal of an Emerging Transdiscipline*, vol. 9, pp. 181–212, 2006.
- [12] P. Brous, I. Overtoom, P. Herder, A. Versluis, and M. Janssen, "Data infrastructures for asset management viewed as complex adaptive systems," *Procedia Computer Science*, vol. 36, pp. 124–130, 2014.
- [13] D. Denyer and D. Tranfield, "Producing a systematic review," in *The Sage Handbook of Organizational Research Methods*, pp. 671–689, Sage Publications Ltd., Thousand Oaks, Calif, USA, 1st edition, 2009.
- [14] M. Wallace and W. Alison, *Critical Reading and Writing for Postgraduates*, SAGE Publications Ltd., Thousand Oaks, CA, USA, 3rd edition, 2016.
- [15] R. K. Yin, *Case Study Research: Design and Methods*, SAGE Publications, 2003.
- [16] M. Janssen and A. Verbraeck, "An agent-based simulation testbed for evaluating internet-based matching mechanisms," *Simulation Modelling Practice and Theory*, vol. 13, no. 5, pp. 371–388, 2005.
- [17] B. Curtis, M. I. Kellner, and J. Over, "Process Modeling," *Communications of the ACM*, vol. 35, no. 9, pp. 75–90, 1992.
- [18] M. P. C. Weijnen, P. M. Herder, and I. Bouwmans, "Designing complex systems: a contradiction in terms," in *Research in Design Series*, IOS Press, pp. 235–254, Delft, The Netherlands, 2008.
- [19] J. Hayes and C. Gutierrez, "Bipartite graphs as intermediate model for RDF," in *Proceedings of the International Semantic Web Conference*, pp. 47–61, Springer, 2004.
- [20] B. McBride, "Jena: implementing the rdf model and syntax specification," in *Proceedings of the Second International Conference on Semantic Web*, vol. 40, pp. 23–28, CEUR-WS, 2001.
- [21] A. Harth and S. Decker, "Optimized index structures for querying RDF from the Web," in *Proceedings of the 3rd Latin American Web Congress, LA-WEB 2005*, p. 10, IEEE, November 2005.
- [22] M. Janssen and G. Kuk, "A complex adaptive system perspective of enterprise architecture in electronic government," in *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*, pp. 71b–71b, IEEE, Kauia, HI, USA, January 2006.
- [23] D. Wollmann and M. T. A. Steiner, "The strategic decision-making as a complex adaptive system: a conceptual scientific model," *Complexity*, vol. 2017, Article ID 7954289, 13 pages, 2017.
- [24] J. H. Holland, "Complex adaptive systems," *Daedalus*, vol. 121, no. 1, pp. 17–30, 1992.
- [25] S. Chan, "Complex adaptive systems," in *Proceedings of the Research Seminar in Engineering Systems*, October 2001.
- [26] S. Y. Auyang, *Foundations of Complex-System Theories: In Economics, Evolutionary Biology, and Statistical Physics*, Cambridge University Press, 1999.
- [27] L. Grus, J. Crompvoets, and A. K. Bregt, "Spatial data infrastructures as complex adaptive systems," *International Journal of Geographical Information Science*, vol. 24, no. 3, pp. 439–463, 2010.
- [28] M. Haghnevis and R. G. Askin, "A modeling framework for engineered complex adaptive systems," *IEEE Systems Journal*, vol. 6, no. 3, pp. 520–530, 2012.
- [29] M. Rupert, A. Rattrout, and S. Hassas, "The web from a complex adaptive systems perspective," *Journal of Computer and System Sciences*, vol. 74, no. 2, pp. 133–145, 2008.
- [30] J. Sutherland and W. van den Heuvel, "Enterprise application integration and complex adaptive systems," *Communications of the ACM*, vol. 45, no. 10, pp. 59–64, 2002.

- [31] J. H. Miller and E. Page, *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*, Princeton University Press, Princeton, NJ, USA, 2009.
- [32] Y. Gong, *Engineering Flexible and Agile Services: A Reference Architecture For Administrative Processes*, Delft University of Technology, Delft, The Netherlands, 2012.
- [33] P. Brous and M. Janssen, "A systematic review of impediments blocking internet of things adoption by governments," in *Open and Big Data Management and Innovation*, M. Janssen, M. Mäntymäki, J. Hidders et al., Eds., vol. 9373 of *Lecture Notes in Computer Science*, pp. 81–94, Springer International Publishing, 2015.
- [34] T. Y. Choi, K. J. Dooley, and M. Rungtusanatham, "Supply networks and complex adaptive systems: control versus emergence," *Journal of Operations Management*, vol. 19, no. 3, pp. 351–366, 2001.
- [35] P. Checkland and S. Holwell, *Information, Systems and Information Systems: Making Sense of the Field*, 1997, <http://www.citeulike.org/group/1702/article/986681>.
- [36] W. J. Kettinger and Y. Li, "The infological equation extended: Towards conceptual clarity in the relationship between data, information and knowledge," *European Journal of Information Systems*, vol. 19, no. 4, pp. 409–421, 2010.
- [37] V. Khatri and C. V. Brown, "Designing data governance," *Communications of the ACM*, vol. 53, no. 1, pp. 148–152, 2010.
- [38] S. Lin, J. Gao, A. Koronios, and V. Chanana, "Developing a data quality framework for asset management in engineering organisations," *International Journal of Information Quality*, vol. 1, no. 1, pp. 100–126, 2007.
- [39] W. Kirstin and B. Otto, *A Contingency Approach to Data Governance*, Cambridge, USA, 2007.
- [40] K.-T. Huang, W. Yang, and Y. Richard, *Quality Information and Knowledge*, Prentice Hall PTR, Upper Saddle River, NJ, USA, 1999.
- [41] R. Price and G. Shanks, "A semiotic information quality framework: development and comparative analysis," *Journal of Information Technology*, vol. 20, no. 2, pp. 88–102, 2005.
- [42] B. Otto, "How to design the master data architecture: Findings from a case study at Bosch," *International Journal of Information Management*, vol. 32, no. 4, pp. 337–346, 2012.
- [43] R. Vilminko-Heikkinen, P. Brous, and S. Pekkola, *Paradoxes, Conflicts and Tensions in Establishing Master Data Management Function*, 2016.
- [44] J. Bester, *Financial Cost Implications of Inaccurate Extraction of Transactional Data in Large African Power Distribution Utility*, 2016, <http://businessperspectives.org/component/option,com-journals/task,view/jid,3/id,435/>.
- [45] B. E. Bargmeyer and D. W. Gillman, "Metadata standards and metadata registries: an overview," in *Proceedings of the International Conference on Establishment Surveys II*, Buffalo, New York, NY, USA, 2000.
- [46] J. H. Holmes, T. E. Elliott, J. S. Brown et al., "Clinical research data warehouse governance for distributed research networks in the USA: a systematic review of the literature," *Journal of the American Medical Informatics Association*, vol. 21, no. 4, pp. 730–736, 2014.
- [47] Y.-C. Lin, "Construction 3D BIM-based knowledge management system: a case study," *Journal of Civil Engineering and Management*, vol. 20, no. 2, pp. 186–200, 2014.
- [48] J. Thomas, S. Bohn, J. Brown, K. Pennock, A. Schur, and J. Wise, "Information visualization: data infrastructure architectures," in *Proceedings of the Seventh International Working Conference on Scientific and Statistical Database Management*, vol. 7, pp. 2–9, Charlottesville, VA, USA, 1994.
- [49] L. Roberts, L. J. Blanshard, K. K. Van Dam, S. L. Price, L. S. Price, and I. Brown, "Providing an effective data infrastructure for the simulation of complex materials," in *Proceedings of the 5th UK e-Science All Hands Meeting (AHM 2006)*, S. J. Cox, Ed., p. 101+, Nottingham, England, 2006, <http://discovery.ucl.ac.uk/100424/>.
- [50] D. D. Nebert, *Developing Spatial Data Infrastructures: The SDI Cookbook*, GSDI, 2004.
- [51] Z. Yue, W. Sun, P. Li, M. U. Rehman, and X. Yang, "Internet of things: architecture, technology and key problems in implementation," in *Proceedings of the 8th International Congress on Image and Signal Processing, CISP 2015*, pp. 1298–1302, IEEE, October 2015.
- [52] A. Botta, W. de Donato, V. Persico, and A. Pescapé, "Integration of cloud computing and internet of things: a survey," *Future Generation Computer Systems*, vol. 56, pp. 684–700, 2016.
- [53] I. Yaqoob, E. Ahmed, I. A. T. Hashem et al., "Internet of things architecture: recent advances, taxonomy, requirements, and open challenges," *IEEE Wireless Communications Magazine*, vol. 24, no. 3, pp. 10–16, 2017.
- [54] L. M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, "A break in the clouds: towards a cloud definition," *ACM SIGCOMM Computer Communication Review*, vol. 39, no. 1, pp. 50–55, 2009.
- [55] P. Mell and T. Grance, *The NIST definition of cloud computing*, 2011, <http://faculty.winthrop.edu/domanm/csci411/handouts/nist.pdf>.
- [56] A. Rajabifard, M.-E. F. Feeney, and I. P. Williamson, "Future directions for SDI development," *International Journal of Applied Earth Observation and Geoinformation*, vol. 4, no. 1, pp. 11–22, 2002.
- [57] J. M. Anderies, M. A. Janssen, and E. Ostrom, "A Framework to analyze the robustness of social-ecological systems from an institutional perspective," *Ecology and Society*, vol. 9, no. 1, 2004.
- [58] J. Ure, R. Procter, Y.-W. Lin et al., "The development of data infrastructures for eHealth: A socio-technical perspective," *Journal of the Association for Information Systems*, vol. 10, no. 5, pp. 415–429, 2009.
- [59] W. H. E. De Man, "Understanding SDI: complexity and institutionalization," *International Journal of Geographical Information Science*, vol. 20, no. 3, pp. 329–343, 2006.
- [60] S. Gianvecchio, M. Xie, Z. Wu, and H. Wang, "Humans and bots in internet chat: Measurement, analysis, and automated classification," *IEEE/ACM Transactions on Networking*, vol. 19, no. 5, pp. 1557–1571, 2011.
- [61] C. W. Furneaux, K. A. Brown, and A. Gudmundsson, "Engineering asset procurement: operationalising complex adaptive system theory," in *Proceedings of the 2008 First International Conference on Infrastructure Systems and Services: Building Networks for a Brighter Future (INFRA)*, pp. 1–6, Rotterdam, Netherlands, November 2008.
- [62] R. Cherinka, R. Miller, and C. Smith, "Beyond web services: towards on-demand complex adaptive environments," in *Proceedings of the 2005 IEEE International Conference on Web Services, ICWS 2005*, pp. 815–816, USA, July 2005.
- [63] P. Brous, P. Herder, and M. Janssen, "Governing asset management data infrastructures," *Procedia Computer Science*, vol. 95, pp. 303–310, 2016.

- [64] P. Malik, "Governing big data: principles and practices," *IBM Journal of Research and Development*, vol. 57, no. 3/4, pp. 1:1-1:13, 2013.
- [65] S. S. Dawes, "Stewardship and usefulness: policy principles for information-based transparency," *Government Information Quarterly*, vol. 27, no. 4, pp. 377-383, 2010.
- [66] P. Brous, M. Janssen, and R. Vilminko-Heikkinen, "Coordinating decision-making in data management activities: a systematic review of data governance principles," in *Proceedings of the International Conference on Electronic Government and the Information Systems Perspective*, pp. 115-125, Springer, 2016.
- [67] A. D. Mathew, L. Ma, and D. J. Hargreaves, "A Conceptual Data Modelling Methodology for Asset Management Data Warehousing," in *Proceedings of the World Congress for Engineering Asset Management*, vol. 3, pp. 1086-1095, Beijing, China, 2008.
- [68] D. M. Strong, Y. W. Lee, and R. Y. Wang, "Data quality in context," *Communications of the ACM*, vol. 40, no. 5, pp. 103-110, 1997.
- [69] R. Y. Wang, Y. W. Lee, L. L. Pipino, and D. M. Strong, "Manage your information as a product," *MIT Sloan Management Review*, 1998.
- [70] D. Ballou, S. Madnick, and R. Wang, "Special section: assuring information quality," *Journal of Management Information Systems*, vol. 20, no. 3, pp. 9-11, 2003.
- [71] K. Sherif and B. Xing, "Adaptive processes for knowledge creation in complex systems: the case of a global IT consulting firm," *Information & Management*, vol. 43, no. 4, pp. 530-540, 2006.
- [72] R. L. Ackoff, "Towards a system of systems concepts," *Management Science*, vol. 17, no. 11, pp. 661-671, 1971.
- [73] A. Gorod, B. Sausser, and J. Boardman, "System-of-systems engineering management: a review of modern history and a path forward," *IEEE Systems Journal*, vol. 2, no. 4, pp. 484-499, 2008.
- [74] L. A. Bollinger and G. P. J. Dijkema, *Resilience and Adaptability of Infrastructures: A Complex Adaptive Systems Perspective*, 2012, <http://resolver.tudelft.nl/uuid:73e6acf7-33d3-4447-bb9f-cbc27-3d2f90f>.
- [75] K. Brown, C. Furneaux, and A. Gudmundsson, "Infrastructure transitions towards sustainability: a complex adaptive systems perspective," *International Journal of Sustainable Development*, vol. 15, no. 1-2, pp. 54-71, 2012.
- [76] M. Khouja, M. Hadzikadic, H. K. Rajagopalan, and L.-S. Tsay, "Application of complex adaptive systems to pricing of reproducible information goods," *Decision Support Systems*, vol. 44, no. 3, pp. 725-739, 2008.
- [77] R. M. Kim and S. M. Kaplan, "Interpreting socio-technical co-evolution: applying complex adaptive systems to is engagement," *Information Technology & People*, vol. 19, no. 1, pp. 35-54, 2006.
- [78] J. Choi, D. L. Nazareth, and H. K. Jain, "Implementing service-oriented architecture in organizations," *Journal of Management Information Systems*, vol. 26, no. 4, pp. 253-286, 2010.
- [79] M. Janssen, S. A. Chun, and J. R. Gil-Garcia, "Building the next generation of digital government infrastructures," *Government Information Quarterly, Building the Next-Generation Digital Government Infrastructures*, vol. 26, no. 2, pp. 233-237, 2009.
- [80] Y. Merali, "Complexity and information systems: The emergent domain," *Journal of Information Technology*, vol. 21, no. 4, pp. 216-228, 2006.
- [81] G. Eoyang and T. Berkas, "Evaluating performance in a complex adaptive system (CAS)," in *Managing Complexity in Organizations*, pp. 313-335, Quorum Books, Westport, CT, USA, 1999.
- [82] T. W. Malone and K. Crowston, "What is coordination theory and how can it help design cooperative work systems?" in *Proceedings of the Conference on Computer Supported Cooperative Work CSCW'90*, pp. 357-370, ACM, New York, NY, USA, October 1990.
- [83] J. G. March and H. A. Simon, *Organizations*, American Psychological Association, 1958.
- [84] H. Mintzberg, *Structure in Fives: Designing Effective Organizations*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.
- [85] H. S. Nwana, L. C. Lee, and N. R. Jennings, "Coordination in software agent systems," *The British Telecom Technical Journal*, vol. 14, no. 4, pp. 79-88, 1996.
- [86] A. Bosman, *Werkverdeling En Organiseren*, Universiteitsdrukkerij, Groningen, The Netherlands, 1st edition, 1995.
- [87] P. Cilliers, *Complexity and Postmodernism: Understanding Complex Systems*, Routledge, 2002.
- [88] J. Rotmans and D. Loorbach, "Complexity and transition management," *Journal of Industrial Ecology*, vol. 13, no. 2, pp. 184-196, 2009.
- [89] O. Hanseth and K. Lyytinen, "Design theory for dynamic complexity in information infrastructures: the case of building internet," *Journal of Information Technology*, vol. 25, no. 1, pp. 1-19, 2010.
- [90] V. Albino, N. Carbonara, and I. Giannoccaro, "Industrial districts as complex adaptive systems: agent-based models of emergent phenomena," *Industrial Clusters and Inter-Firm Networks*, pp. 58-82, 2005.
- [91] R. D. Stacey, "The science of complexity: an alternative perspective for strategic change processes," *Strategic Management Journal*, vol. 16, no. 6, pp. 477-495, 1995.
- [92] P. Anderson, "Complexity theory and organization science," *Organization Science*, vol. 10, no. 3, pp. 216-232, 1999.
- [93] Rijkswaterstaat, *Assetmanagement Binnen Rijkswaterstaat*, Rijkswaterstaat, 2010.
- [94] S. Bagui, A. Das, and C. Bapanapalli, "Controlling vehicle overloading in BOT projects," in *Proceedings of the 2nd Conference of Transportation Research Group of India (2nd CTRG)*, *Procedia - Social and Behavioral Sciences*, vol. 104, pp. 962-971, December, 2013.
- [95] A. Mulyun, D. Parikesit, M. Antameng, and R. Rahim, "Analysis of loss cost of road pavement distress due to overloading freight transportation," *Journal of the Eastern Asia Society for Transportation Studies*, vol. 8, pp. 1020-1035, 2010.
- [96] L. Shien, J. Gao, and A. Koronios, *The Need for a Data Quality Framework in Asset Management*, vol. 1, Adelaide, Australia, 2006.
- [97] M. Broadbent and P. Weill, "Management by maxim: how business and IT managers can create IT infrastructures," *Sloan Management Review*, vol. 38, no. 3, pp. 77-92, 1997.
- [98] E. Lesser, M. Fontaine, and J. Slusher, *Knowledge and Communities*, Routledge, 2009.
- [99] P. Brous, P. Herder, and M. Janssen, "Towards modelling data infrastructures in the asset management domain," *Procedia Computer Science, Complex Adaptive Systems San Jose, CA*, vol. 61, pp. 274-280, 2015.
- [100] S. Russell and P. Norvig, *A Modern Approach*, vol. 25 of *Artificial Intelligence*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1995.

- [101] F. Zambonelli, N. R. Jennings, and M. Wooldridge, "Organisational rules as an abstraction for the analysis and design of multi-agent systems," *International Journal of Software Engineering and Knowledge Engineering*, vol. 11, no. 3, pp. 303–328, 2001.

