ELSEVIER
CrossMark
Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 61 (2015) 274 – 280
Procedia
Computer Science

Complex Adaptive Systems, Publication 5
Cihan H. Dagli, Editor in Chief
Conference Organized by Missouri University of Science and Technology
2015-San Jose, CA

Towards Modelling Data Infrastructures in the Asset Management Domain

Paul Brous[ab*], Paulien Herder[a], Marijn Janssen[a]

[a]*Delft University of Technology, Jaffalaan 5, 2628 BX, Delft, The Netherlands*
[b]*Rijkswaterstaat, Derde Werelddreef 1, 2622 HA, Delft, The Netherlands*

Abstract

More and more asset management organizations are relying on trusted data to drive their decision making process. Yet little systematic research has been performed regarding the generation of trusted data. Trusted data has a clear, defined and consistent quality which meets the expectations of the user. The aim of the research presented in this paper is to develop a conceptual model to support asset management organizations with regards to their development of data infrastructures which produce trusted data and provide organizations with the capability to make the right decisions at the right time. The autonomous characteristics of agent-based systems and process orientation of discrete-event simulation are combined in our conceptual model. In this way both the autonomous social behavior of organizations and their business processes can be modelled. Several predefined components containing the behavior of agents through various coordination mechanisms have been developed. These components can be further customized to model an empirical situation more closely. The approach is illustrated with a case study in the asset management domain.

Peer-review under responsibility of scientific committee of Missouri University of Science and Technology

Keywords: complex adaptive systems; CAS; data infrastructure; asset management; data quality

1. Introduction

The complex nature of data infrastructures is the reason for the difficulties encountered in trying to understand and assess them[1], and it can be difficult to attribute success or failure of data infrastructure implementations to one or more specific factors. Because data infrastructures are complex, there is an interrelationship between their sociological and technical dimensions, and it is difficult to track cause-and-effect relationships. The dynamic sociological and technological interrelations between agents and the components of data infrastructures are therefore hard to predict and control. All data infrastructures have a unique character and behave differently. This makes it

Peer-review under responsibility of scientific committee of Missouri University of Science and Technology
doi:10.1016/j.procs.2015.09.215

difficult to implement data infrastructures in different environments and achieve similar outcomes[1]. Complex adaptive systems (CASs) are often described as systems of interactive, mutually interdependent, individual elements which merge over time into coherent forms[2]. Data infrastructures have been identified as CASs[1,3], and using a CAS lens can help us to identify and better understand the key elements of data infrastructures and the coordination methods necessary for their functioning and dealing with change[3]. According to Auyang[4] (1999) CASs are systems that are built from individual agents which are capable of adapting as they interact with each other and their environment. Thus, by conceptualizing data infrastructures as CAS, policy-makers and decision-makers can gain a better understanding of the dependencies involved[5]. This conceptualization acknowledges that it is impossible to exert a hierarchical control over complex systems of organizations and projects spanning multiple levels and jurisdictions. Instead, one must take into account the various typical characteristics of complex adaptive systems[6].

This research takes place in the physical asset management domain. Asset management (AM) is important for industry, the argument being that the success of an enterprise often depends on its ability to use and manage its assets efficiently[7]. Asset management is regarded as an essential business process in many organizations[7] and trusted data is regarded as essential to aiding the decision making process[8]. Having data infrastructures which produce trusted data is therefore essential for organizations which have data driven decision making processes. We follow Mohseni's[9] (2003) definition of AM as being a discipline for optimizing and applying strategies related to work planning decisions in order to effectively and efficiently meet the desired objective[9–11]. AM is therefore essentially a matter of understanding risk, followed by developing and applying the correct business strategy, and the right organization, process and technology models to solve the problem[9]. The activities associated with AM are: identifying what assets are needed, identifying funding requirements, acquiring assets, providing logistic and maintenance support for assets and disposing or renewing assets[10]. These activities provide the scope for the conceptual model.

2. Method and Structure

We combine the autonomous characteristics of agent-based systems and the process orientation of discrete-event simulation in this research to develop an agent based conceptual model of data infrastructures. Agent-based systems provide a decentralized solution based on centralized decision making. This gives the system a high degree of flexibility and robustness[12]. Discrete event simulation provides an intuitive and flexible approach to simulating complex systems[13].

Asset managers have struggled over the years to develop systems which produce data they can trust, and asset data is regularly observed to be lacking in quality, to be "noisy" (embedded within significant amounts of meaningless data), or to be missing the required detail[14]. Addressing this issue requires an holistic approach[3] which describes the sociological as well as the technological components. The conceptual model described in this research aims at helping asset managers understand the nature of data infrastructures, the consequences of the choice of coordination mechanism and how this affects data quality. By modelling data infrastructures we can illustrate and simulate the basic components of data infrastructures and their interrelationships. In the following section we discuss on the basis of a review of literature how using a CAS lens can help us to identify and better understand the key elements of data infrastructures and the coordination methods necessary for their functioning and dealing with change as well as describing various kinds of coordination mechanisms. The literature review continues in section four where we discuss the integration of the agent-based and the process oriented discrete-event simulation approaches. In section five we develop an agent-based architecture. The combination of elements and coordination mechanisms make up our agent-based conceptual model. We illustrate the conceptual model by means of a case study in section six. The case under study was that of asset management within the Directorate General of Public Works and Water Management of the Netherlands. The Directorate General of Public Works and Water Management of the Netherlands is commonly known within The Netherlands as "Rijkswaterstaat", often abbreviated to "RWS", and is referred to as such within this research.The case study was explorative in method and descriptive in nature. Unstructured interviews were held with managers, subject matter experts, and internal consultants. Internal documentation concerned with the description of the data infrastructure of the RWS was studied. We draw conclusions and make recommendations for future research in the final section.

3. CAS

Despite the large number of examples used by researchers to describe what a CAS is, there appears to be little agreement as to what exactly the characteristics of a CAS are. In this research our focus is on data infrastructures as CAS. We follow Grus et al.[1] (2010), whose research field is spatial data infrastructures, and we use the definition given by Barnes et al.[15] (2003). With regards to data infrastructures, CASs are defined 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" (Barnes et al. 2003 p.276).

A factor which contributes to the difficulty in defining the characteristics of data infrastructures from a CAS perspective is that characteristics of CASs can be divided into physical *elements* and also into functional and operational *behaviors*. Elements are sets of physicality's that together make CASs different from other systems. Behaviors are the collection of functions and operations that make CAS behavior unique. Few researchers have made this distinction when defining CAS characteristics and there have been a number of calls for attention to this topic[15]. In this research we focus on modelling the elements of data infrastructures as viewed from a CAS perspective. This research builds on previous work published by Brous et al.[3] (2014). According to Brous et al.[3] (2014), being CAS's, data infrastructures consist of data and technological *components*[1,16–18], which are stable and simple building blocks and are the basic parts of the system. These components are manipulated by *agents* who interact with one another in large variety of ways. The agents operate within a certain *schema.* Schema refers to shared rules which are embodied by norms, values, beliefs, and assumptions[19]. Agents use rules to make decisions within frames of reference or schemata by which they interpret and evaluate information[20]. But data infrastructures can also have competing stakeholders[21] and conflicting schemata which requires coordination. As such, *coordination mechanisms* are necessary tools used by coordinating agents. A coordination mechanism is a process which determines how information is obtained and used in decision-making[3]. Coordination mechanisms present within data infrastructures are*:* self-organization, feedback, planning, and contracting[3].

4. Multi-agent systems and discrete-event simulation

Simulation strives to develop a dynamic model of a system, and experiment with this model as well as with possible alternative models of the system in order to attempt to understand a known problem[22]. Simulation is performed prior to implementation and can be used to evaluate process design options[23]. With discrete-event simulation, the temporal aspects of a sequence of discrete events are modeled[23,24]. However, in this research we need to develop a model which describes not only the organizational processes of asset management, but also the different coordination mechanisms used by asset managers. Generally speaking, the most powerful models attempt to minimize the semantic gap between the units of analysis and the constructs present in the modeling approach[23]. But, with modelling, it is also important to reduce complexity by eliminating unnecessary detail[23,25] in order to highlight the essence of the problem.

Multi-agent systems (MAS) can exhibit the characteristics of organizations, and of intentional organization design[23,26]. We follow Janssen & Verbraeck[23] (2005) and take an agent-based approach which represents the decentralized nature of data infrastructures, the multiple loci of control, differing forms of governance and the competing interests of multiple stakeholders[23]. In agent-based simulations, the agents interact in a simulated environment, where modelling reductions have been applied[23]. Once fully developed, the simulator created by this research will provide an artificial time mechanism that allows interactions to take place faster or slower than reality[27]. By placing the agents in an agent-based simulation, it should be possible to study the results of each coordination mechanism, both in detail and over a prolonged period of time, where the experimental conditions can be manipulated[23]. But in order to evaluate the coordination mechanisms in practice, the interactions between the independent organizations, the business processes within the organizations, and the utilization of their resources also need to be modelled.

5. An agent-based conceptual model

We follow Janssen & Verbraeck[23] (2005) as well as the ideas of agent architectures developed for MAS by Jennings[12] (2001) and develop an agent architecture using object orientation[12]. Object oriented environments require communication between objects[23]. Implementing an agent within an object orientation therefore requires developing the objects as agents to enable an agent to comply with the common characteristics of agents, such as autonomy, communication, and behavior to either react on the environment or to deliberately perform an action[23]. According to Janssen & Verbraeck[23] (2005), an agent-based model should ensure that all modeled entities meet the characteristics that make up an autonomous agent. Our model breaks up the data infrastructure into reusable, logical parts but does not pose a limitation to the extensibility of an agent, as different organizations should be modeled and new coordination mechanisms might have to be added[23].

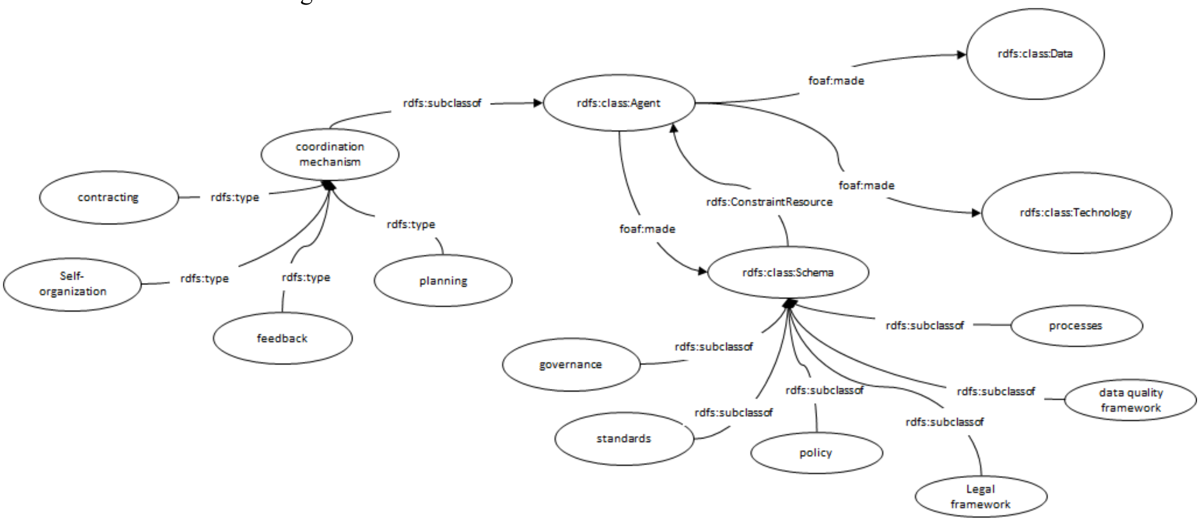

Fig. 1.The agent-based conceptual model.

As seen above in figure 1, within our agent-based model, all independent actors are viewed as agents. Agents are encapsulated within the *agent class,* which can be extended to include a wide variety of types of agents. We adopt Janssen and Verbraeck's[23] (2005) definition of an agents and define 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" (Janssen and Verbraeck, 2005, pp.375)[23]. In many multiple agent architectures, complexity is generally reduced and problems are decomposed, with sub-problems being assigned to specific agents[23]. The greater problem is in this way resolved by including multiple agents. Our model follows this line of thinking as each agent has a particular role to play in the implementation of the data infrastructure, based on their position within the organization and the underlying processes.

We thus further develop the agent-based model by examining the characteristics of agents with regards to the underlying schema. The *schema class* helps to define the nature of agents, as well as shaping the behavior and interactions of agents. Agents are always situated in a particular environment. They receive inputs related to the states of their environment, and they act on the environment. Schema provides the guidelines which guide the actions of the agents.

An agent's behavior is often viewed as a manifest of intelligence[30]. The *coordination mechanism 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[23,31]. According to Janssen and Verbraeck (2005), behavior is dependent on the circumstances, and as a result modeled agents should be able to have various types of behavior. The behavior of the agents dictates which technology is implemented and which data is developed. This behavior also dictates how the data and the technology are maintained.

The *technology class* encapsulates the collection of Information Technology (IT) artifacts, hardware and software, used in the production of data or services or in the accomplishment of objectives, such as data analysis or data management. Creating and managing a business driven IT involves decisions based on a sound understanding of an organization's strategic context[34]. IT has led many organizations to imagine a world of leveraged knowledge, but whilst IT has inspired this vision, it in itself cannot bring it into being[35]. In the model, IT is regarded as an important enabler of data infrastructures.

Data have long been recognized as a core factor in IS and data infrastructures, and have been generally defined as the measure or description of objects or events[1,3,32,33]. Data is usually referred to as a set of interrelated data items that measure the attributes of the objects or events[33]. The data produced by data infrastructures is encapsulated in the *data class*.

6. Case Study: Asset Management at RWS

The core function of AM at RWS is that objects such as roads, viaducts, or bridges are managed so that they fulfill the function they have in the network. To operate effectively, RWS focusses on being able to make the right choices with regards to management and maintenance[36]. Choices must be made in the way the bridge or viaduct is maintained. According to a RWS official, "It is more important that the bridge functions properly than that it looks good with a fresh coat of paint". However, these choices are not always as straightforward, as, for example, during maintenance procedures of bridges or viaducts, roads still need to be accessible. In order to make these choices, RWS requires data that can be trusted to conform to the quality that is required. RWS bases its asset management processes on data driven decision making. A well-functioning data infrastructure is required for the acquisition and maintenance of trusted data[3]. The model defined in the previous section can be used to illustrate the components of data infrastructures and simulate data processes throughout the asset management process.

In order to maintain their asset management data system, RWS has developed a data management organization which implements and enforces uniform data entry and data management protocols and processes. This data management organisation encompasses a wide variety of agents. All of which can be described by extending the *agent class* in the model. Within the data management process, there are many different organisational levels, each level and each link in the information chain acing as an agent in the process. For example, divisions of RWS are organised according to geographic location, and each division is an independent agent. Each independent division implements standardised processes in their own way, and each individual advisor, in his turn, is capable of acting independently. Furthermore, RWS also has electronic "agents" in the form of Internet of Things (IoT) implementations in which the "things" acquire ambient intelligence[37]. As an organisation, RWS contracts out a good deal of the data entry to external contractors, who in turn are also agents. This means that RWS does not always have full control of the data entry process. RWS addresses this issue by implementing various *coordination mechanisms* to influence the behaviour of the agents. The model can be manipulated in order to simulate the effect of using different coordination mechanisms for different agents to monitor the effect on data quality and the relationship of the quality to the advertised level of quality and the expected level of quality.

RWS is looking more and more towards an integral approach to managing the entire network of assets. 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." The Information Delivery Specification (IDS) is a part of the contract between RWS and the contractor in which the data transfer is specified. This contract document guarantees a uniform exchange of information on structures between the different partners. RWS has formally accepted the contract method as the coordination mechanism of choice. Also planning through the implementation of yearly portfolio plans is an important coordination mechanism. But depending on the level of organisation and independence of the agent, other, less formal coordination mechanisms such as self-organisation and feedback become important behavioural tools. But successful coordination also requires making standardised agreements which can be encapsulated in the *schema class*.

Within the schema class, for example, RWS has a policy of using systems engineering in her projects. The COINS standard has been introduced for exchange of data between RWS and the contractor. COINS is a BIM standard. It is complementary to standards such as IFC, IFD Library and IDM. COINS supports the exchange of Systems Engineering information and ensures that an object tree, GIS data, 2D drawings, 3D models, IFC models and the object type library can be stored in association in a database[38]. It also provides a BIM-container interchange format. RWS is also in the process of introducing a state-of-the art data quality management system based on a data quality framework in order to gain much needed insight into the actual quality of the data and to be able to monitor the data quality trends over time.

RWS requires a complete and constantly up-to-date insight into all asset data within the *data class*. The data class can be extended to include all data formats including structured and unstructured data. Asset data has a quantitative and qualitative aspect. RWS has identified data quality as being essential for achieving organisational goals, and is introducing a data quality monitoring system as a strategic objective. Decomposition is an important aspect for quantitative data. Decomposition is the classification of assets according to a number of levels. Decomposition, whilst not a new subject for RWS has been approached differently by RWS in the last few years. RWS is in the process of introducing the Building Information Model (BIM) and is developing a uniform asset management system based on a uniform Object Type Library (OTL). RWS is extremely ambitious with regards to decomposition. The OTL has a complex mapping, with regards to both quantitative and qualitative data. BIM is an important means of exchanging information regarding assets between RWS and her contractors. For the application of the Building Information Model (BIM) RWS has developed a range of products that make it possible to exchange uniform and reliable information between construction partners.

Within the *technology class*, RWS currently has over one hundred and eighty separate applications which have been implemented to provide the information required by asset managers. RWS regards this number of applications as an untenable situation due to the major burden on application and infrastructural management. Furthermore, many of the applications use much the same data, but often the results of the different applications reveal a mixed level of data quality, which does little to generate trust. RWS is currently implementing a new system known at RWS as "Areaal Informatie Rijkswaterstaat" (Area Information RWS), or "AIR". The objective behind AIR is to reduce the number of asset management information systems from one hundred and eighty to four. Furthermore, RWS is also implementing a generic data quality monitoring system which is designed to assure data quality levels and provide insight into potential quality issues within the data production process.

7. Conclusions

In this paper we describe an agent-based conceptual model of a data infrastructure within the asset management domain. Agents in this domain can take the form of organizations, divisions and departments of organizations, individuals, software agents and even a collection of things. Agents can represent the decentralized nature of data management, the multiple loci of control, the multiple perspectives, and the competing interests. Finally, agents have behavior to autonomously decide how to organize their activities based on a chosen coordination mechanism. Future research can focus on more advanced architectures containing other characteristics commonly included in multi-agent systems such as goals, desires and intentions, as well as extending the model to include class types in each (sub) class.

Our case study shows that our conceptual model is suitable for modelling data infrastructures. By explicitly representing organizations by agents and by taking their goals into account it becomes possible to discuss alternative coordination mechanisms and to decide upon which mechanisms are acceptable to the organizations involved in the design process.

The agent-based architecture helps in breaking up a conceptual model into logical parts that can be reused and ensures the behavior of agents is autonomous. In this way the architecture ensures that a model can be built quickly, can be extended where necessary, and has a high degree of correspondence with reality.

Acknowledgements

We acknowledge and thank the people of the Rijkswaterstaat who gave of their time and expertise during the case study research. This research is funded by Rijkswaterstaat.

References

1. Grus, L., Crompvoets, J. & Bregt, A. K. Spatial data infrastructures as complex adaptive systems. *Int. J. Geogr. Inf. Sci.* **24,** 439–463 (2010).
2. Miller, J. H. & Page, S. E. *Complex Adaptive Systems: An Introduction to Computational Models of Social Life: An Introduction to Computational Models of Social Life*. (Princeton University Press, 2009).
3. Brous, P., Overtoom, I., Herder, P., Versluis, A. & Janssen, M. Data Infrastructures for Asset Management Viewed as Complex Adaptive Systems. *Procedia Comput. Sci.* **36,** 124–130 (2014).
4. Auyang, S. Y. *Foundations of Complex-system Theories: In Economics, Evolutionary Biology, and Statistical Physics*. (Cambridge University Press, 1999).
5. Janssen, M. & Kuk, G. A complex adaptive system perspective of enterprise architecture in electronic government. in *System Sciences, 2006. HICSS'06. Proceedings of the 39th Annual Hawaii International Conference on* **4,** 71b–71b (IEEE, 2006).
6. Herder, P. M., Bouwmans, I., Dijkema, G. P. J., Stikkelman, R. M. & Margot P.C. Weijnen. Designing infrastructures using a complex systems perspective. *J. Des. Res.* **7,** 17–34 (2008).
7. Koronios, A., Lin, S. & Gao, J. A Data Quality Model for Asset Management in Engineering Organisations. in **17,** (2005).
8. Haider, A., Koronios, A. & Quirchmayr, G. in *Engineering Asset Management* (eds. Mathew, J., Kennedy, J., Ma, L., Tan, A. & Anderson, D.) 288–300 (Springer, 2006). at <http://link.springer.com/chapter/10.1007/978-1-84628-814-2_32>
9. Mohseni, M. What does asset management mean to you? in **3,** 962–964 (IEEE, 2003).
10. Hastings, N. A. J. *Physical Asset Management*. (Springer, 2010). at <http://www.springer.com/engineering/production+engineering/book/978-1-84882-750-9>
11. Mathew, A. D., Ma, L. & Hargreaves, D. J. Understanding data management in asset management: A survey. in 1096–1107 (Springer London, 2008). at <http://eprints.qut.edu.au/15444>
12. Jennings, N. R. An agent-based approach for building complex software systems. *Commun. ACM* **44,** 35–41 (2001).
13. Karnon, J. *et al.* Modeling using Discrete Event Simulation: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-4. *Value Health* **15,** 821–827 (2012).
14. Lin, S., Gao, J. & Koronios, A. The need for a data quality framework in asset management. in **1,** (2006).
15. Barnes, M., Matka, E. & Sullivan, H. Evidence, Understanding and Complexity Evaluation in Non-Linear Systems. *Evaluation* **9,** 265–284 (2003).
16. Haghnevis, M. & Askin, R. G. A Modeling Framework for Engineered Complex Adaptive Systems. *IEEE Syst. J.* **6,** 520–530 (2012).
17. Rupert, M., Rattrout, A. & Hassas, S. The web from a complex adaptive systems perspective. *J. Comput. Syst. Sci.* **74,** 133–145 (2008).
18. Sutherland, J. & van den Heuvel, W.-J. Enterprise Application Integration and Complex Adaptive Systems. *Commun ACM* **45,** 59–64 (2002).
19. Choi, T. Y., Dooley, K. J. & Rungtusanatham, M. Supply networks and complex adaptive systems: control versus emergence. *J. Oper. Manag.* **19,** 351–366 (2001).
20. Furneaux, C. W., Brown, K. A. & Gudmundsson, A. Engineering asset procurement: Operationalising complex adaptive system theory. in *2008 First International Conference on Infrastructure Systems and Services: Building Networks for a Brighter Future (INFRA)* 1–6 (2008). doi:10.1109/INFRA.2008.5439596
21. Cherinka, R., Miller, R. & Smith, C. Beyond Web services: towards on-demand complex adaptive environments. in *2005 IEEE International Conference on Web Services, 2005. ICWS 2005. Proceedings* -816 (2005). doi:10.1109/ICWS.2005.30
22. Shannon, R. E. Introduction to the Art and Science of Simulation. in *Proceedings of the 30th Conference on Winter Simulation* 7–14 (IEEE Computer Society Press, 1998). at <http://dl.acm.org/citation.cfm?id=293172.293175>
23. Janssen, M. & Verbraeck, A. An agent-based simulation testbed for evaluating internet-based matching mechanisms. *Simul. Model. Pract. Theory* **13,** 371–388 (2005).
24. Pegden, C. D. Introduction to SIMAN. in *Proceedings of the 15th conference on Winter simulation-Volume 1* 231–241 (IEEE Press, 1983). at <http://dl.acm.org/citation.cfm?id=801501>
25. Curtis, B., Kellner, M. I. & Over, J. Process modeling. *Commun. ACM* **35,** 75–90 (1992).
26. Weiss, G. *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*. (MIT Press, 1999).
27. Zeigler, B. P., Praehofer, H. & Kim, T. G. *Theory of Modeling and Simulation, Second Edition*. (Academic Press, 2000).
28. Nwana, H. S., Lee, L. C. & Jennings, N. R. Coordination in Software Agent Systems. *Br. Telecom Tech. J.* **14,** 79–88 (1996).
29. Nwana, H. S. Software agents: An overview. *Knowl. Eng. Rev.* **11,** 205–244 (1996).
30. Russell, S., Norvig, P. & Intelligence, A. A modern approach. *Artif. Intell. Prentice-Hall Egnlewood Cliffs* **25,** (1995).
31. Zambonelli, F., Jennings, N. R. & Wooldridge, M. Organisational rules as an abstraction for the analysis and design of multi-agent systems. *Int. J. Softw. Eng. Knowl. Eng.* **11,** 303–328 (2001).
32. Checkland, P. & Holwell, S. Information, systems and information systems: making sense of the field. (1997). at <http://www.citeulike.org/group/1702/article/986681>
33. Kettinger, W. J. & Li, Y. The infological equation extended: towards conceptual clarity in the relationship between data, information and knowledge. *Eur. J. Inf. Syst.* **19,** 409–421 (2010).
34. Broadbent, M. & Weill, P. Management by maxim: How business and IT managers can create IT infrastructures. *Sloan Manage. Rev.* **38,** 77–92 (1997).
35. Lesser, E., Fontaine, M. & Slusher, J. *Knowledge and Communities*. (Routledge, 2009).
36. Rijkswaterstaat. Assetmanagement binnen Rijkswaterstaat. (2010).
37. Atzori, L., Iera, A. & Morabito, G. The internet of things: A survey. *Comput. Netw.* **54,** 2787–2805 (2010).
38. COINS. *http://www.coinsweb.nl/* (2015). at <http://www.coinsweb.nl/>