Coordinating Data-Driven Decision-Making in Public Asset Management Organizations: A Quasi-Experiment for Assessing the Impact of Data Governance on Asset Management Decision Making

Paul Brous, Marijn Janssen, Paulien Herder

Coordinating Data-Driven Decision-Making in Public Asset Management Organizations: A Quasi-experiment for Assessing the Impact of Data Governance on Asset Management Decision Making

Paul Brous[1*], Marijn Janssen[1] , and Paulien Herder[1]

[1]Delft University of Technology, Delft, The Netherlands
{P.A.Brous, M.F.W.H.A.Janssen, P.M.Herder@tudelft.nl}

Abstract. Public organizations are facing increasing challenges to the management of their infrastructure assets. New sources of data, such as social media and IoT, can provide new insights for organizations to help them deal with these challenges. Yet data must be of sufficient quality in order to be acted upon. The objective of this study is to develop and approach to evaluate how data governance improves decision-making in asset management organizations. This paper describes a quasi-experiment which identifies and quantifies relationships between data governance and improvements in asset management decision-making. The quasi-experiment focusses on data requirements for determining current and future asset conditions, which is critical for assessing remaining service life and risk of failure. The quasi-experiment utilizes a pre-test post-test control group design. We expect that the inclusion of data governance improves the quality of data which allows for improved decision-making in asset management organizations.

Keywords: data, data governance, data quality, data management, experiment

1 Introduction

Public asset management organizations are facing increasing challenges to the management of their infrastructure assets, technological advances, political shifts, changing stakeholders, or economic fluctuations. Many public asset management (AM) organizations routinely store large volumes of data in an attempt to find ways to improve efficiency and effectiveness of their AM processes through data-driven decision-making [8, 15]. Increasing the complexity is the development of techniques which utilise other data sources such IoT and Social Media data to provide information which may provide more timely information than more traditional methods.We follow Mohseni's [24] definition of AM as being a discipline for optimizing and applying strategies related

adfa, p. 1, 2011.

to work planning decisions in order to effectively and efficiently meet the desired objective [17, 22, 24]. 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[24].

This study is centered on the AM process of determining current and future asset conditions, which is critical for assessing the remaining service life of assets and to prevent the risk of failure of assets. This knowledge has a direct impact on decisions regarding the provision of logistic and maintenance support for assets and disposing of, or renewing assets. The objective of this study is to evaluate how data governance supports data-driven decision-making in asset management organizations. Data governance specifies the framework for decision rights and accountabilities to encourage desirable behavior in the use of data [19], ensures that data is aligned to the needs of the organization [13], monitors and enforces compliancy to policy [36], and ensures a common understanding of the data throughout the organization [26].

According to Brous et al. [6], data infrastructures can be seen as a shared, evolving, heterogeneous, set of resources (including human resources, or agents), which are capable of providing facts required to fulfil a social need. Data infrastructures often have a unique character and behave differently. This makes it difficult to implement data governance in different environments and achieve similar outcomes [16]. It is difficult to attribute the contribution of data governance to asset management decision-making to one or more specific factors [4]. This article starts with the introduction and identification of the problem in sections one and two. Subsequently, the design propositions of the quasi-experiment are derived in sections three and four, which encompass an overview of data-driven decision-making in asset management organizations and the potential functional elements of data governance in asset management organizations respectively. In section five we describe the design of the quasi-experiment and the gaming approach whose purpose is to identify the effect that the design propositions have on data quality. In section six we discuss the possible limitations of the quasi-experiment. The paper concludes with a summary of the theory and approach in section seven.

2 The Need for Data Governance

Data quality can be affected by a broad range of outside influences at indiscriminate moments in time [37]. It is because of this that it is exceptionally difficult for asset management organizations to effectively manage their data. Asset management organizations may thus not always be well equipped to handle data [23]. The reasons for this often do not lie in the technology, but rather originate in a wide variety of areas such as organization, or culture [12, 24]. Because data infrastructures are complex [6], there is an interrelationship between their social and technical dimensions.

New sources of data, originating from sources such as social media and IoT, can provide new insights to help organizations face these challenges. But data must be of sufficient quality in order to be acted upon [25, 39] and too much data can create "noise" which detracts van the quality of the information. A widely adopted definition of high quality data is data that is "fit-for-use" [35, 40]. Using the definition provided by Strong et al. [35], the characteristics of high-quality data have intrinsic, accessibility, contextual, and representational aspects. This also means that usefulness and usability are important aspects of quality [13, 35]. Having data infrastructures which produce data of a quality that is aligned to the needs of the organization is therefore essential for asset management organizations which rely on data-driven decision-making processes [2].

According to Panian [28], enforcing policies and processes around the management data is the foundation of an effective data governance practice. The enforcement of data management policies and processes requires coordination. Coordination is the management of dependencies between activities [20]. Coordination mechanisms, such as hierarchies and networks, denote the way interdependent activities and decisions are managed [21]. Coordination mechanisms need to be established to ensure accountability for data quality through a combination of incentives and penalties [2], as accountability can unlock further potential by addressing relevant issues related to data stewardship. Governing data appropriately is only possible if it is properly understood what the data to be managed means, and why it is important to the organization [34]. Attention to business areas and enterprise entities is the responsibility of data stewards [38] who have the entity-level knowledge necessary for development of data for which they are responsible [34].

3 Data-driven decision making in AM

In more and more AM organizations, managerial decisions rely on data-based analytics [8]. Many AM organizations gather extremely detailed data from and propagate knowledge to their consumers, suppliers, alliance partners, and competitors. Also, there are many more opportunities for data collection outside of operational systems. According to Brynjolfsson et al. [8], mobile phones, vehicles, factory automation systems, and other devices are routinely instrumented to generate streams of data on their activities. AM organizations can use sensors to track the performance of their assets, and they can use the data these sensors provide to improve the management of their assets. Similarly, data collected from social media may make the user experience visible and may provide insights into the real-time condition of the assets. However, the use of data for decision-making in processes such as prognostics [18] is still relatively undeveloped, and there are still serious ethical [4] and technical [12] issues which need to be addressed.

More precise and accurate information should facilitate decision making [1, 33]. In this paper, we develop an experiment for measuring the effect of data governance on data-driven decision-making within the context of determining current and future asset conditions. In this research, the assumption is made that all asset management decision-making is data-driven and that better quality data results in better decision-making.

4 Functional Elements of Data Governance in Asset Management Data Infrastructures

According to Brous et al. [6], data infrastructures can be viewed as complex adaptive systems (CAS). In this research we model the elements of data infrastructures viewed from a CAS perspective. In this way, this research builds on previous work published by Brous et al. [6, 7]. According to Brous et al., data infrastructures can be conceptualized as consisting of data and technology, which are stable and simple building blocks and are the basic parts of the system. These building blocks are manipulated by agents who interact with one another, operating within a certain schema. Schema refers to the shared rules which are embodied by norms, values, beliefs, and assumptions [11]. Agents use rules to make decisions within frames of reference or schemata by which they interpret and evaluate information. In this regard, the schema of data infrastructure is defined and maintained by data governance processes which provide coordination for data management activities [6].

As discussed in the previous sections, data governance specifies the framework for decision rights and accountabilities to encourage desirable behavior in the use of data [19], ensures that data is aligned to the needs of the organization [13], monitors and enforces compliancy to policy [36], and ensures a common understanding of the data throughout the organization [26].

A common metric used to measure the effectiveness of data governance is data quality [26, 31, 39]. Data governance, data quality and data (quality) management are closely linked and are often handled by the same individuals in organizations [26, 29]. In this regards, data governance is important for decision making with regard to data quality management [19, 26, 27]. According to Strong et al., data quality is typically determined by the data's fitness for use, which is the capability of data to meet the requirements of the user in order to accomplish a certain goal in a given context. A user can only decide whether or not data is fit for use if the quality of the data is known and reported. This makes it important for organizations to define data quality metrics, which can be used to measure and report the quality of data based on well-defined data quality dimensions. Wang and Strong [37] identify four dimensions of data quality and one hundred and eighteen aspects of data quality. This research follows Otto [26] and Wang & Strong [37] and addresses only the commonly used quality aspects of completeness, consistency, accuracy, relevancy, and timeliness [26, 37]. In this research we follow the definitions of these data quality aspects provided by Pipino et al. [30] pp 212 (see table 1 below).

Table 1. Definitions of data quality aspects (adapted from [30] pp 212)

Data quality aspect	Definition
Completeness	"The extent to which data is not missing and is of sufficient breadth and depth for the task at hand"
Consistency	"The extent to which data is presented in the same format"
Accuracy	"The extent to which data is correct and reliable"
Relevancy	"The extent to which data is applicable and helpful to the task at hand"
Timeliness	"The data to which data is sufficiently up-to-date for the task at hand"

Propositions for the design of a data governance prototype were created based on the elements of data governance discussed above. According to Denyer et al. [14], a design proposition is a general template for the creation of solutions for a particular class of field problems. The design propositions suggest on a high level which functional infrastructure elements may be used to improve data governance in asset management data infrastructures. We propose four key elements to improve data governance: 1. coordination mechanisms; 2. definition of data quality requirements; 3. monitoring of data quality; 4. shared data commons. Although there may be other ways to improve data governance, these infrastructure elements were found to be critical. Based on these key elements, the following design propositions were generated:

1. Coordination mechanisms positively influence data quality in asset management organizations

2. Defining data quality requirements positively influences data quality in asset management organizations
3. Monitoring data quality positively influences data quality in asset management organizations
4. Creating a shared data commons positively influences data quality in asset management organizations.

5 Evaluation method: the Quasi-Experiment and the Use of Serious Games

In section five we discuss the structure of the quasi-experiment and the use of serious gaming to explore system behavior and to simulate data governance in an asset management setting.

5.1 Gaming Approach

According to Shadish et al. [32], an experiment is a study in which an intervention is deliberately introduced to observe its effects. Experiments have factorial designs where independent variables are systematically varied, and the dependent variable(s) are quantitative, objective measures of system performance [1]. A quasi-experiment [9] is an empirical study used to estimate the causal impact of an intervention on its target population [1]. According to Adelman [1], quasi-experiments share similarities with experimental design, but they lack the element of random assignment to treatment or control. Instead, the researcher controls the assignment of the treatment condition to the quasi-experiment using criterion other than random assignment such as an eligibility cutoff mark. In this study the choice was made for a quasi-experiment as opposed to a true experiment as full control over the scheduling of experimental stimuli that make a true experiment possible is lacking [9] and we wish to retain control over selecting and scheduling measures and how the treatment will be organized [32].

The quasi-experiment detailed in this paper uses gaming as a tool to simulate data governance in data-driven decision making in an asset management setting. According to Bekebrede [3], serious gaming can be a useful tool to simulate complex socio-technical infrastructure systems and supports policy makers and designers in understanding the complexity of the planning and design of these systems from the observer perspective [3]. With a support tool is meant a tool or instrument which can contribute to the planning, design, implementation and management of data infrastructures in different ways. Gaming can thus be used as a support tool for understanding the complexity of asset management data infrastructures and the impact of these on asset management decision-making. At the same time, gaming is an experience space in which participants can experience the complexity themselves and increase their understanding of the system, from the player perspective. We aim to evaluate data governance in a game setting in which participants use a prototype application to specify the coordination mechanisms for decision rights and accountabilities, to ensure that data is aligned to the needs of the organization, to monitor and enforce compliancy, and to ensure a common understanding of the data. At the same time we aim to control the variables to test our propositions and to ensure that the effects can be contributed to data governance. Figure 1 shows the variables involved in the quasi-experiments.

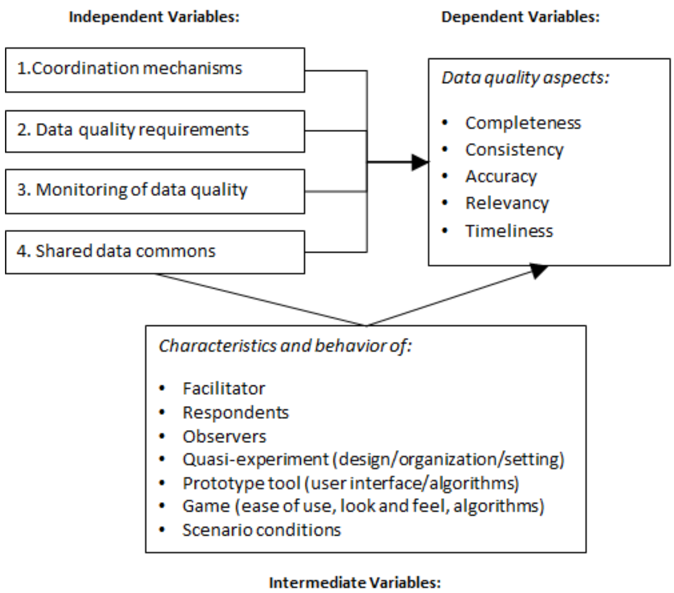

Fig. 1. Variables involved in the quasi-experiments

In this quasi-experiment, participants will be required to maintain assets in Minecraft, a virtual world, using data provided to them by the “game-master”. Virtual worlds, such as Minecraft, allow researchers to explore existing theory and develop new theory in a variety of fields, including information and social sciences [23]. Minecraft is a multiplayer sandbox-building game focused on creativity, building and survival in which players can acquire resources and must maintain their health and hunger at acceptable levels. The core gameplay revolves around construction [10]. In this quasi-experiment, players play as a team, but operate as individuals. The team consists of 5 players.

Within their virtual world, each team will be allocated “assets” which they will be required to manage and maintain based on the data provided to them. The state of the assets will degrade during the course of the game, and will need to be maintained. In a second application, players will be able to govern their data using the functional elements described in the design propositions. Depending on their allocated group, players will have access to varying degrees of functionality. This allows the researcher to manipulate the variables within the game setting in order to test the four design propositions. For example, at the start of the game, teams may be given the opportunity to define the required quality of the data provided to them, and, depending on the game settings, define who is responsible for maintaining the quality of each dataset. During the course of the game, the game master will degrade the quality of the data unless appropriate action is taken. The control group will not have any access to the second application, but will be granted access to the same data. The researchers will be able to monitor the fluctuating quality of all the datasets throughout the game.

5.2 Structure of the quasi-experiments

The quasi-experiments will be conducted as follows (see figure 2 below). Firstly, the quasi-experiment will be introduced to the participants and instructions will be given. Secondly, the pre-test, a participant survey, will be conducted to measure various background characteristics of the participants, as well as their experience with asset management, data governance, and with serious gaming. Thirdly, participants will be asked to complete scenario tasks within the game environment.

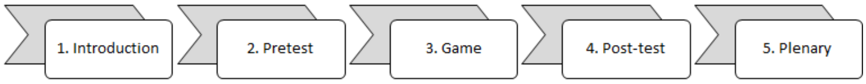

Fig. 2. Structure of the quasi-experiments

While the participants are playing the game, time measures and observations will be made to obtain additional information. Time measures will be made to examine how long it takes to conduct the scenario tasks and to investigate whether there are significant differences between the time used to conduct the scenarios by the treatment group and the time used by the control group. Fourthly, a post-test in the form of a second participant survey will be used to measure the extent to which data governance influences the completion of the scenario tasks within the game. Finally, in a plenary discussion the participants will be questioned as to the levels of difficulty of the tasks and if they have any suggestions to improve the game or the prototype tooling.

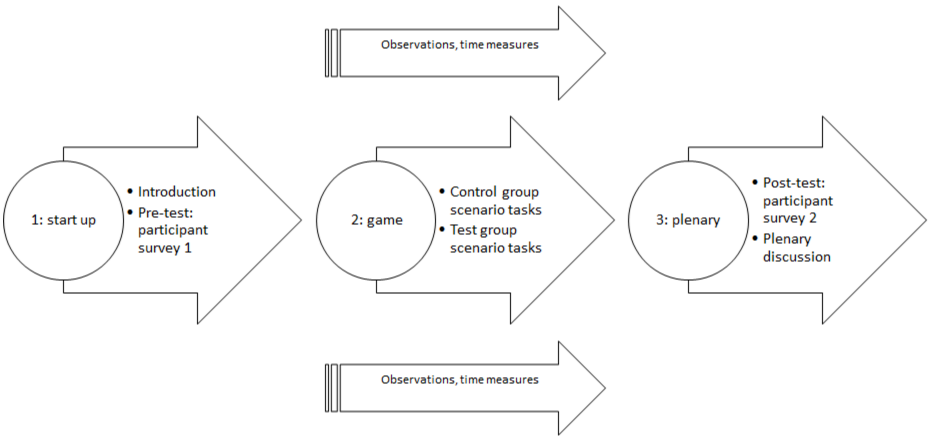

Fig. 3. Approach of the quasi-experiment

5.3 Treatment versus control condition

Five groups will be tasked with completing the scenario tasks within the game to test the effect of the introduction of coordination mechanisms for decision rights and accountabilities, the alignment of data to the needs of the user, options to monitor and enforce compliancy, and options to ensure a common understanding of the data. The fifth group is a control group to establish a base-line. Quasi-experiments will be conducted with at least five groups of 5 people to ensure that sufficient participants are involved and that the responses to the questionnaire can be analysed with statistical tests. Ideally, at least 10 experiments will be conducted to ensure statistical validity. The conditions for the treatment groups and the control group should remain as equal as possible.

5.4 Limitations

Quasi-experiments are subject to concerns regarding internal validity, because the treatment and control groups may not be comparable at baseline [5]. With quasi-experimental studies, it may not be possible to convincingly demonstrate a causal link between the treatment condition and observed outcomes. This is particularly true if there are confounding variables that cannot be controlled or accounted for, such as if the design of the experiment does not control for the effect of other plausible hypotheses that could have improved performance between the pretest and the posttest [1]. For example, external influences may occur between the pretest and posttest that could explain the results. If the selected group represent either the very best or very worst performers, then it is possible that pretest-posttest differences could be affected by statistical regression to the mean. In this experiment, the evaluations focus on a limited number of specific tasks related to the coordination framework, data quality definitions, data quality monitoring and the shared data commons which need to be conducted within a limited time frame. Due to time limitations it may not be possible to conduct additional tasks or to conduct scenario tasks longer than 50 minutes. Participants may not be able to complete the scenarios in this time frame. Also, three types of measures are used in the evaluations, namely time measures, observations and questionnaires. In addition to these three measures, other measures, such as other data quality aspects, of the performance of the participants may be used. By using additional measures, more information may be obtained regarding the contribution of data governance to decision-making in asset management organizations. Moreover, other factors may influence the outcomes, such as the user interface, quality of the tooling, experience with gaming, and experience with information management in general. The final results may therefore not only be attributed to the coordination framework, the definition of quality requirement, the monitoring of data quality or the shared data commons.

6 Summary

Public organizations are facing increasing challenges to the management of their infrastructure assets and many AM organizations are looking for ways to improve the efficiency and effectiveness of their AM processes through data-driven decision-making. New sources of data such as IoT and social media data may provide more timely information than more traditional techniques. In this paper, we develop a measure of data governance for data-driven decision-making within the context of determining current and future asset conditions. The assumption is made that asset management decision-making is data-driven and that better quality data results in better decision-making. It is important to realize that there are still ethical, organizational and technical barriers to the adoption of data driven decision making. In this paper we describe a quasi-experiment to assess how aspects of data governance - a coordination framework, data quality definitions, data quality monitoring and a shared data commons - affect the commonly used quality aspects of completeness, consistency, accuracy, relevancy, and timeliness. The quasi-experiment detailed in this paper uses gaming as a tool to simulate the implementation of data governance in data-driven decision making in an asset management setting. This experiment does have limitations as quasi-experiments are subject to concerns regarding internal validity, because the treatment and control groups may not be comparable at baseline and it may not be possible to convincingly demonstrate a causal link between the treatment condition and observed outcomes.

7 References

1. Adelman, L.: Experiments, quasi-experiments, and case studies: A review of empirical methods for evaluating decision support systems. IEEE Trans. Syst. Man Cybern. 21, 2, 293–301 (1991).
2. Al-Khouri, A.M.: Data ownership: who owns “my data.” Int J Manag Inf Technol. 2, 1–8 (2012).
3. Bekebrede, G.: Experiencing complexity: a gaming approach for understanding infrastructure systems. TU Delft, Delft University of Technology (2010).
4. Bhidé, A.: The Big Idea: The Judgment Deficit. Harv. Bus. Rev. September 2010, (2010).
5. Blackwell, D.: Equivalent Comparisons of Experiments. Ann. Math. Stat. 24, 2, 265–272 (1953).
6. Brous, P. et al.: Data Infrastructures for Asset Management Viewed as Complex Adaptive Systems. Procedia Comput. Sci. 36, 124–130 (2014).
7. Brous, P. et al.: Towards Modelling Data Infrastructures in the Asset Management Domain. Procedia Comput. Sci. 61, 274–280 (2015).
8. Brynjolfsson, E. et al.: Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance? Social Science Research Network, Rochester, NY (2011).
9. Campbell, D.T., Stanley, J.: Experimental and Quasi-Experimental Designs for Research. Wadsworth Publishing (1963).
10. Canossa, A.: Give Me a Reason to Dig: Qualitative Associations Between Player Behavior in Minecraft and Life Motives. In: Proceedings of the International Conference on the Foundations of Digital Games. pp. 282–283 ACM, New York, NY, USA (2012).
11. Choi, T.Y. et al.: Supply networks and complex adaptive systems: control versus emergence. J. Oper. Manag. 19, 3, 351–366 (2001).
12. Coff, R.W.: Human Assets and Management Dilemmas: Coping With Hazards On the Road to Resource-Based Theory. Acad. Manage. Rev. 22, 2, 374–402 (1997).
13. Dawes, S.S.: Stewardship and Usefulness: Policy Principles for Information-based Transparency. Gov. Inf. Q. 27, 4, 377–383 (2010).
14. Denyer, D. et al.: Developing Design Propositions through Research Synthesis. Organ. Stud. 29, 3, 393–413 (2008).
15. Dimitris Bertsimas, Aurélie Thiele: Robust and Data-Driven Optimization: Modern Decision Making Under Uncertainty. In: Models, Methods, and Applications for Innovative Decision Making. pp. 95–122 INFORMS (2006).
16. Grus, L. et al.: Spatial data infrastructures as complex adaptive systems. Int. J. Geogr. Inf. Sci. 24, 3, 439–463 (2010).
17. Hastings, N.A.J.: Physical Asset Management. Springer, London, UK (2010).
18. Heng, A. et al.: Rotating machinery prognostics: State of the art, challenges and opportunities. Mech. Syst. Signal Process. 23, 3, 724–739 (2009).
19. Khatri, V., Brown, C.V.: Designing data governance. Commun ACM. 53, 1, 148–152 (2010).
20. Malone, T.W., Crowston, K.: The Interdisciplinary Study of Coordination. ACM Comput. Surv. 26, 1, 87–119 (1994).
21. Malone, T.W., Crowston, K.: What is Coordination Theory and How Can It Help Design Cooperative Work Systems? Presented at the ACM Conference on Computer-supported Cooperative Work , New York, USA (1990).
22. Mathew, A.D. et al.: Understanding data management in asset management: A survey. Presented at the World Congress for Engineering Asset Management , Beijing, China October 27 (2008).

23. Mehdi, K.-P.: Inventive Approaches for Technology Integration and Information Resources Management. IGI Global (2014).
24. Mohseni, M.: What does asset management mean to you? Presented at the Transmission and Distribution Conference and Exposition , Greenwood Village, USA September 7 (2003).
25. Otto, B.: On the Evolution of Data Governance in Firms: The Case of Johnson & Johnson Consumer Products North America. In: Handbook of Data Quality. pp. 93–118 Springer (2013).
26. Otto, B.: Organizing data governance: Findings from the telecommunications industry and consequences for large service providers. Commun. Assoc. Inf. Syst. 29, 1, 45–66 (2011).
27. Otto, B. et al.: Towards a framework for corporate data quality management. ACIS 2007 Proc. 109 (2007).
28. Panian, Z.: Some practical experiences in data governance. World Acad. Sci. Eng. Technol. 38, 150–157 (2010).
29. Pierce, E. et al.: Industry Report: The State of Information and Data Governance. IADQ (2008).
30. Pipino, L.L. et al.: Data Quality Assessment. Commun ACM. 45, 4, 211–218 (2002).
31. Redman, T.C.: Data driven: profiting from your most important business asset. Harvard Business Press, Boston, USA (2008).
32. Shadish, W.R. et al.: Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Wadsworth Publishing, Boston (2001).
33. Shannon, C.E.: A Mathematical Theory of Communication. SIGMOBILE Mob Comput Commun Rev. 5, 1, 3–55 (2001).
34. Smith, A.: Data Governance and Enterprise Data Modeling – Don't Do One Without the Other! — Enterprise Information Management Institute, http://www.eiminstitute.org/library/eimi-archives/volume-1-issue-2-april-2007-edition/data-governance-and-enterprise-data-modeling-dont-do-one-without-the-other.
35. Strong, D.M. et al.: Data quality in context. Commun ACM. 40, 5, 103–110 (1997).
36. Thompson, N. et al.: Government data does not mean data governance: Lessons learned from a public sector application audit. Gov. Inf. Q. 32, 3, 316–322 (2015).
37. Wang, R.Y., Strong, D.M.: Beyond Accuracy: What Data Quality Means to Data Consumers. J. Manag. Inf. Syst. 12, 4, 5–34 (1996).
38. Weber, K. et al.: One Size Does Not Fit All—A Contingency Approach to Data Governance. J. Data Inf. Qual. 1, 1, 4:1–4:27 (2009).
39. Wende, K.: A Model for Data Governance – Organising Accountabilities for Data Quality Management. Presented at the Australasian Conference on Information Systems , Toowoomba, Australia December 5 (2007).
40. Wende, K., Otto, B.: A Contingency Approach to Data Governance. Presented at the International Conference on Information Quality , Cambridge, USA October 11 (2007).