
A FRAMEWORK FOR THE CORPORATE GOVERNANCE OF DATA – THEORETICAL BACKGROUND AND EMPIRICAL EVIDENCE

Tomi DAHLBERG¹, Tiina NOKKALA²

Turku Business School at University of Turku, Rehtorinpellonkatu 3,

FI-20014 University of Turku, Finland

E-mails: ¹tomi.dahlberg@utu.fi (corresponding author); ²tiina.nokkala@utu.fi

Received 03 December 2014; accepted 17 January 2015

Abstract. In a modern organization, IT and digital data have transformed from being functional resources to integral elements of business strategy. Against this background, our article addresses corporate governance of digital data in general and that of aging societies in particular. To describe the role of executives and managers in data governance, we first review the corporate and IT governance literature. We then propose a theoretical framework for the governance of data: a novel construct. We apply the framework to the governance of aging societies related data, that is, to answer the question of how best to manage the provision of services to citizens with digital data enablement and support. We also disclose the results from two recent surveys, with 212 and 68 respondents respectively, on the business significance of data governance. The survey results reveal that good governance of data is considered critical to organizations. As concluding remarks, we discuss the significance of our results, our contributions to research, the limitations of our study and its managerial implications.

Keywords: governance of data, governance of IT, corporate governance, data management, aging, aging societies, managers and data, data assets.

JEL Classification: M15, G34, H51, H53, J14, L86.

1. Introduction

The motivation for this article comes from early-phase research on how to govern aging societies with the help of information technology (IT) and digital data. These are considered some of the key means of providing solutions to the global problems of population aging (Obi *et al.* 2013) – a promise yet to be fulfilled. Our article focuses on the use and especially the governance of digital data, rather than the full range of IT-enabled services. Although our interest lies in the governance of data in a specific context, we believe that the governance of data for aging societies is fundamentally similar to the governance of data in any other context. Consequently, we propose a generic framework for the governance of data. We approach our topic from a corporate

governance (managerial) perspective as we consider the governance of data a business managerial issue as opposed to an IT or data modelling issue, Technology, IT and data modelling centric approaches use the data governance construct, which actually often describes data management. Therefore, we pay special attention to the role of business management in the governance of data.

The aging of populations affects both developed and developing economies. A recent United Nations study disclosed that the proportion of persons older than 60 years was 11% of the world's population in 2012 (United Nations 2014). The same report estimates that this proportion will double to 22% by the year 2050. According to the OECD, Japan had the highest proportion of citizens older than 65 years at 24.2% in 2012, followed by Italy and Germany with 20.6% (OECD 2014a). The country of the authors, Finland, had the seventh highest proportion of citizens over 65 years among the OECD countries at 18.1%, whilst the average for OECD countries was 15.1% and for the EU28 17.8%. The OECD estimates that by 2050 the proportion of elderly citizens will grow to 35.6% in Japan, 35.3% in Italy and 29.5% in Germany (OECD 2014b). The estimate for Finland is 27.1%, for OECD countries 25.2% and for the EU25 29.5%. The aging of populations is also strong in developing economies. For example, in China it is estimated that the number of citizens over 65 years will have grown from 25 million in 1953 to 349 million in 2050 (Caiwei 2013).

Demographic changes, such as longer life expectancies and lower birth rates, are creating serious economic and social challenges for economies. Peter Wintley-Jensen, the head of the DG-Information Society within the European Commission, has listed critical challenges for the EU by 2025: the dependency ratio (proportion of citizens not in the labour force) will drop from 1:4 to 1:2; the proportion of health and social care to GDP will increase by 4–8%; the headcount of the work force will shrink by 20 million (Wintley-Jensen 2013). The concept of the super-aged society has been introduced to describe aging in Japan (Obi *et al.* 2013). This also describes an approach that aims to turn challenges into opportunities for economic growth. For example, responding to the increasing demand for healthcare and social welfare services requires the promotion of healthier lifestyles, encouraging greater activity in the monitoring of wellbeing, increasing pre-emptive care and self-services, and developing services so that citizens can stay at home; also, among the solutions proposed are empowerment through active aging, new working arrangements and longer inclusion and participation in society.

Digital data are viewed as being able to support service development, innovations and reforms in three ways. At national, regional and local levels, data compiled and analysed concerning citizens and their service needs can be used to design and operate national, regional and local healthcare, social welfare and other service systems and policies. Aged care services are only one element in such systems and policies. Digital data can also be used to develop national, regional and local services by automating data routines, integrating data storages and transferring data electronically. This requires

data interoperability, inter-organizational coordination hubs (Markus, Quang 2012) and standards. Cooperation between organizations and stakeholders is also needed. The governance of data and the daily management of data are fundamental to the coordination of national, regional and local data, as well as data-driven services.

Second, healthcare, social welfare, finance and other professions providing services to (elderly) citizens need good quality data in their work. The requirements for data are similar to those outlined above. Data interoperability and transferability within and between tasks and organizations are required to ensure that professionals have useful information. There is also a need for better coordinated service activities with improved data sharing between professionals (Hovenga 2013; Obi *et al.* 2013). For example, an elderly citizen may need home care (e.g. food delivery), social welfare (e.g. financial support to pay a part of expenses), basic healthcare (e.g. control in the timeliness of drug delivery) and specialized medical care (e.g. periods of care at a hospital); these are services which need to be provided as a coordinated whole. Citizen-authorized access to data on the socio-medical-economic status of the citizen is expected to improve the efficiency and the impact of services.

Finally, (elderly) citizens use data-enabled services and are – or still too often should be – the owners of the data concerning them. Data need to be made available to facilitate its use. As parts of these data are sensitive, such as health, financial and social status data, adequate privacy and authorization mechanisms are needed. On the other hand, mobile technologies, cloud services, games and gamification, social media, robotics, etc., have already produced many services that help in the daily lives of (elderly) citizens. These include services that promote healthier living and self-diagnosis. An (elderly) citizen could even store data using cloud services, for example data series on blood pressure, nutrition, exercise, diabetes and heartbeat. Cooperation between healthcare, service providers and citizens and mechanisms to ensure the (medically required) reliability of this data are needed to achieve data interoperability and integration to the medical and social welfare databases created by professionals.

Currently, data interoperability and transferability are far from reality. Several reasons prevent the electronic transfer and consolidation of data: data creation and handling processes vary, leading to dissimilarities in data coding and content; data concepts, formats and structures differ; data are fragmented and duplicated. There is also a lack of any accepted and widely used international, national and/or local data model or data message standards. Health and medical data are one of the most standardized areas. Nonetheless, out of the 194 Member States of the World Health Organization (WHO), only 34 members were able to provide reliable health data in 2012 (Hovenga 2013). A major reason for the current situation is the fact that each organization typically develops/purchases and runs its own databases and information systems (ISs) without considering data interoperability, transferability and usability (Dahlberg 2010; Dahlberg *et al.* 2011; Hovenga 2013). As the number of ISs and data storage systems continues

to increase, there are data held on the same citizens, services, professionals, etc., in an ever-increasing number of data storages and information systems.

We claim that this is a managerial issue. Why? Because only business professionals know what the content of data should be and what data are needed to perform specific tasks. If the governance of data is unclear, no one in an organization is responsible for the content quality or the availability of data in specific tasks. Currently, even within a single organization, business executives may allow or even demand the purchase/development of ISs and data storage systems that do not share or transfer data.

The situation is similar in general. For example, most large companies have limited visibility for their customers, products, vendors, etc. Data are fragmented into hundreds of incompatible ISs and dozens of data storages. Sales, procurement and manufacturing professionals, as well as customers, vendors, partners, etc., face similar data challenges to those confronting healthcare and social welfare service professionals and citizens. We conclude that the governance of data framework should be generic and have a corporate governance (managerial) focus.

At the same time, the governance of data on elderly citizens is ideal for the testing and validation of the proposed framework. Elderly citizens have lived long and thus data concerning them are stored in multiple storages, may cover several decades and include rich sets of events and changes. Second, when an elderly citizen retires, the collection of occupational healthcare, HR and other work-related data discontinues and salary payments are transformed into pension payments – at least partially. There are also changes in taxation, benefits, social status, etc. Third, many elderly citizens make a gradual transformation from active, “young” elderly citizens to “old” elderly citizens. The young elderly travel and consume cultural services and they may continue to work in some way, participate in societal activities and spend time with their grandchildren, friends, etc. As they age, they start to need more support and care service provision.

To summarize, this article proposes a generic governance of data framework from a corporate governance perspective. Our intention is later to test the framework in an on-going research project entitled “*Governance of aging societies*”. We aim to evaluate the ability of the framework to solve data interoperability and transferability problems in the context of data-driven services offered to elderly citizens. Our article addresses two issues (research questions): first, we discuss the proposed framework and its theoretical basis; second, we answer the question of how important the governance of data is using the evaluations of the respondents gathered from two surveys. Section 2 addresses the theoretical aspects of the study. Section 3 covers the methodological facets of the surveys and section 4 presents the results. Finally, a conclusion and discussion section ends the article.

2. Governance of data framework and its theoretical basis

The work of Data Management Association (DAMA) International to establish data management concepts is probably the most acknowledged cumulative endeavour in data management (DAMA 2009). DAMA's DMBOK (Data Management Book of Knowledge) distinguishes between data, information and knowledge. Data are defined as the representation of facts, such as text, numbers, graphics, images, sound and video (DAMA 2009: 2). Data transforms into information when definition, format, timeframe and relevance are added to the data. Information transforms into knowledge when patterns, trends, relationships and assumptions are added to information. Data have seven phases during their lifecycle. The planning, specification and enabling phases predate the existence of data. Data are processed during the creating and acquiring, maintaining and using, archiving and retrieval, and purging phases (DAMA 2009: 4). The DMBOK framework identifies ten data management functions (DAMA 2009: 7) and seven environmental elements (DAMA 2009: 13). The data governance function is used to steer the other nine functions. Due to nature of the DMBOK data governance tasks and the lack of a governance body (ISO/IEC 2008) this is not governance of data. The mapping of data management functions and environmental elements results in the DAMA-DMBOK functional framework (DAMA 2009: 15). A "context diagram" is also crafted (DAMA 2009: 18). The DMBOK manual then devotes one chapter to each data management function and its context diagram (DAMA 2009: 37–334).

We attempted to apply the DMBOK framework and later the Method for an Integrated Knowledge Environment (MIKE 2014). We chose these frameworks because they are – as far as we know – globally used holistic data management frameworks. We found both frameworks insufficient for three reasons. First, both the DMBOK and the MIKE frameworks use the word "governance". Yet, governance here primarily describes daily management and the use of data, rather than the governance of data management and especially corporate governance of data. Second, governance of data is looked at from the inside, i.e. from an IT perspective. In contrast, corporate governance of data looks at data from the outside, namely how investments in data help to achieve the business objectives of the "corporate" and ensure a financial return on investments. Finally, the DMBOK framework assumes that the meaning of data can be defined universally. Based on the assumption of universality, the meaning of patient, citizen and service event concepts, inter alia, and related data attributes could be defined so that there is a single version of "the truth" applicable to all situations and contexts. We do not share this assumption. Instead, we propose that the meaning of data is defined in its creation and/or use context. For example, the meaning of the patient concept and patient concept-related data attributes differs for authorities, doctors, nurses, relatives of the patient and the patients themselves, depending on what each of them does with the data in a particular situation, or context. Due to these limitations, we decided to propose a new framework by using three building blocks.

2.1. Corporate governance and the governance of IT

According to the classic, widely used definition of Shleifer and Vishny (1997), corporate governance refers to the ways in which suppliers of finance assure returns on their investments. They approach the issue from an agency perspective, by separating ownership and control. Shleifer and Vishny (1997) state that: *“In most general terms, the financiers and the manager sign a contract that specifies what the manager does with the funds, and how the returns are divided between him and the financiers. Ideally, they would sign a complete contract, that specifies exactly what the manager does in all states of the world, and how the profits are allocated”* (p. 741). However, this is impossible in practice due to various uncertainties. Therefore, investors and the manager agree on control rights, which are used to respond to uncertainties as they occur. The activities of setting objectives, agreeing accountability and putting controls in place to secure the achievement of objectives have become descriptive characteristics of corporate governance.

Bebchuk and Weisbach (2010: 940) note that *“[b]ecause returns to suppliers of finance depend on myriad legal and contractual arrangements, the operation of various markets, and the behavior of different types of players, corporate governance has evolved into various subliterations”*. Corporate governance of IT (e.g. ISO/IEC 2008), or simply governance of IT, is one such subliteration, although corporate governance research may not acknowledge governance of IT in this role.

Governance of IT applies corporate governance concepts and principles for IT and digital data. For example, the family of international ISO/IEC 38500 standards states that the governance body of an organization (corporate) should evaluate IT, i.e. it should set objectives for the use of IT and digital data in line with corporate objectives. The governance body should also direct IT, agreeing areas of accountability in activities for the use of IT and digital data to achieve the objectives. Finally, the governance body should monitor IT, ensuring through the use of control mechanisms that the objectives established for the use of IT and digital data are really achieved. In the governance of IT literature, this is called the Evaluate–Direct–Monitor (EDM) cycle. The EDM cycle governs and interacts with the daily management of IT and digital data, both in business projects with IT investments and business operations with IT dependencies. The governance of IT emphasizes the role of business executives and managers (“investors”) in two ways: i) the governance body of the organization, typically the board, CEO, executive committee, etc., is responsible for implementing IT governance; ii) the evaluation of IT builds on the principle that business-driven objectives are established for the governance of IT and data – and their management.

We apply the EDM model of the ISO/IEC 38500 standard to the governance of data, as this is the model that is widely employed, as well as the established international standard for the governance of IT. The results are shown in Table 1.

Table 1. Evaluate-Direct-Monitor (EDM) model applied to governance of data (source: compiled by the authors, applying the format of ISO/IEC 2008)

Governance body task	Description of governance body tasks
Evaluate	Determine for what current and future purposes data are used in an enterprise to execute business strategy and to achieve business objectives and thus to set business objectives for data usage. Consider how changes in business needs and the environment impact the use of data. Set in place a management system with objectives, areas of accountability, controls and behaviours which ensure efficient and effective use of data.
Direct	Implement the management system to govern data creation, maintenance and usage by assigning responsibilities and by directing the preparation and implementation of policies and plans. Ensure that the implementation of projects and other data quality improvement activities, as well as the translation of results into operational practice, are properly planned and managed, including stakeholder involvement and consideration of related business and IT systems. Encourage a sound data and information management culture.
Monitor	Monitor through appropriate measurement systems the performance of data processes, data quality and data usage, including the relation of data to the achievement of business objectives. Ensure that external obligations in data management (e.g., laws and regulations on privacy, continuity, etc.) are met.

Previous research has proposed that structures, processes and cooperation mechanisms be used to design IT governance arrangements (Van Grembergen *et al.* 2004; Van Grembergen, De Haes 2008). Data-related decision right matrices and organizational and control structures for the creation and use of data are examples of data governance structures. The EDM cycle and control processes used to ensure the quality of data are examples of processes. Contracts, discussions and interactions between owners/investors and data managers/experts are examples of cooperation mechanisms.

Corporate governance encourages governance bodies to prepare against uncertainties. Within the context of data management, a typical means of doing so is through enterprise architecture, especially information architecture and data/information risk management. Enterprise/information architecture is used to manage and reduce the complexity of data and thus reduce the possible negative consequences of uncertainties. Architecture also relates data to the business, ISs and infrastructure. Furthermore, information architecture promotes learning and understanding by defining data-related concepts, models, etc., which are then used to promote a common language and learning within an organization. The purpose of data/information risk management is to secure the continuity of the organization's business activities by mitigating data/information-related risks. In addition to continuity management, means of doing so include data

security and data quality management. The purpose of data security – including cyber security – and data quality management is to ensure secure access to data and also the availability, integrity, timeliness, etc., of the data.

2.2. Digitalization of data and digital strategy

The digitalization of data started when computers were first put into use some 60 years ago. In organizations, business data started to accumulate to their ISs and databases and have done so at an increasing pace over the years (Davenport 2007). For a long time, the majority of data were non-digital. According to Hilbert and Lopez (2011), a major transformation started around the year 2000, when audio and video data increasingly became digital, fuelled by innovations in data storage, data transformation, data compression and other technological advances. The cost of processing and storage per unit of data is now negligible. As a consequence, mankind had generated 276 exabytes (billions of gigabytes or 2^{60}) of data by 2007 according to Hilbert and Lopez (2011) and the International Data Corporation (IDC 2011): of this, approximately 95% was digital. In four years the amount of data created grew to 1.8 zettabytes (or 1800 exabytes, Zettabyte = 2^{70}) and the proportion of digital data increased to over 99%. For this year (2014), the estimate is 7.5–9.0 zettabytes. If these estimates are even roughly accurate, they mean that over the last two year mankind has created more data than during its cumulative history prior to that time. According to the Internet Systems Consortium (ISC) there were over 900 million server computers – excluding PCs, tablets, smart phones, etc. – used to process data in mid-2012 (ISC 2012). Despite this huge computing capacity and the low cost of data storage, it was no longer possible to store all data by 2006–2007 according to the IDC (2011). The IDC estimates that we are currently able to store less than half the amount of newly created data.

The emergence of new data concepts is one aspect in the digitalization of data. In addition to transactional business, data organizations use various types of sensor data, including data created by robots. Organizations also use message, audio and video data and spatio-temporal data, i.e. location/space and/or time-stamped data. Organizations may also use open data, which are typically data made available by public sector organizations free of charge. Figure 1 lists the aforementioned types of data sources. In addition to structured data, organizations have an interest in using unstructured data. Furthermore, data may be internal or external to an organization. Figure 2 shows the combinations of structured vs. unstructured and internal vs. external data. Watts *et al.* (2009) have pointed out that the complexity of data management grows with the increase in data volumes. Currently, as the result of this data spread to an increasing number of ISs and data storages, each holding more or less overlapping data.

Data digitalization, data explosion and their consequences emphasize the need for governance of data and the role of business management in this area. The digitalization of data means that it has become business-critical for an organization to understand what data it creates and processes in its various activities. Decisions on what data/informa-

Transactional business data
Sensor data
Audio and video stream data
Spatial and spatio-temporal data
Messages and other www-data
Open data (not covered above)

Fig. 1. Types of data sources (source: compiled by the authors)

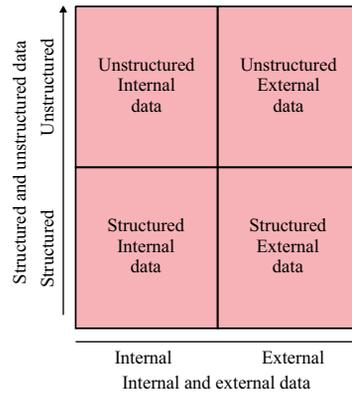


Fig. 2. Structure and internality of data (source: compiled by the authors)

tion to store and what not should be based on a sound business rationale as opposed to technical considerations. The principle of corporate governance, according to which the financiers of investments in digital data need to assure themselves of returns on their investments, is a sound basis for this.

The digitalization of data has also exerted an impact on strategy work. Bharadwaj *et al.* (2013) claim that due to business infrastructure digitalization, IT strategy has become an integral part of business strategy to the extent that the two can no longer be separated (Bharadwaj *et al.* 2013). They use the concept of digital strategy, which is important not only to single organizations, but also business networks (Pagani 2013). Digital strategy represents a fundamental change in business strategy thinking – although exactly what this constitutes is still open for discussion. Looking at the last 30 years, we have first defined business strategy and then considered how IT and digital data should be used to implement business strategy as one of the functional resource strategies. This perspective to business strategy can be found even within the business – IT alignment approach (Henderson, Venkatraman 1993), that is, business strategy is seen to direct IT strategy. The digital strategy approach appears to change the actual business strategy process only slightly. The proposition is that the business significance and the deployment of digital data and IT should to be considered an element of business strategy, rather than a functional strategy following on from business strategy. The digital strategy approach emphasizes the central role/involvement of business executives and managers in the governance of IT and data.

2.3. Data and information assets of organizations and their ontological nature

We found DMBOK to be useful in the classification of the data and information assets that organizations use. Business transactions, documents, content and reports are intuitively evident to most people as data, whereas master data, reference data and metadata

are not. Master data, reference data and metadata are building blocks for the other asset types of data shown in Figure 3. The definitions in Figure 3 are based on DMBOK (DAMA 2009), Cleven and Wortman (2010) and Dahlberg *et al.* (2011).

As discussed in the introduction, our opinion is that one of the key issues in the governance of data is the ontological nature of data. We build on the work of Iivari *et al.* (1998) and Wand and Weber (1993, 2002). Iivari *et al.* (1998) consider an IS – and thus digital data within the IS – to represent real-world, especially human, activities. At the extremes, there are two approaches to interpret what representations mean, universal and contextual. According to the universal approach, an IS and its data describe facts which remain stationary and stable as long as the data exist. Hence, the real world, seen through the lenses of data modelling, is also considered stationary and stable. The contextual approach, on the other hand, proposes that data represent vested interests and dynamic interplay between socially-constructed concepts, especially the representations of human behaviour in the context that the IS serves. As we share this view, we consider that the current challenges in the governance of data, in data management and in the quality of data arise largely from assumptions about the stability, predictability, uniformity and causality of the real world (Dahlberg *et al.* 2011).

The propositions of Wand and Weber (1993, 2002) explain why the increase in digital data has spread into an increasing number of ISs and data storages. According to these authors, any IS represents the ontology of the real world (context) with a minimum set of constructs. Any IS should also map and track the state changes of the real world, such as the diverse data needs of various stakeholders in their anticipated and ad-hoc use/behaviour purposes. Wand and Weber’s propositions explain both the existence of multiple ISs with more or less similar data and the means to govern them. Governance of data builds on understanding the meaning of the various representations, i.e. the contextually defined semantic/descriptive metadata of those representations.

<p>Transactions</p> <p>Operational business data: business transactions within operational applications and data storage systems</p>	<p>Reporting</p> <p>Reporting data: original and processed data within reporting data storage systems used for reporting, analyses and business intelligence</p>	<p>Documents</p> <p>Documents: structured and unstructured drawings, memos, presentations, spreadsheets and other documents</p>	<p>Content</p> <p>Content data: messages, e-mails, images, www-data, audio and video streams; can be combined with other information assets</p>
<p>Master and reference data: Shared non-transactional data used in several applications, e.g. product/item, customer, location, chart of accounts, country</p>			
<p>Metadata – data on data: (1) what data means in a business context, e.g. how users understand item data and how they use them in a specific context (informational); (2) the structure, format and coding of data, how it flows and where it is stored (operational/administrative)</p>			

Fig. 3. Data and information assets used in organizations

Although this discussion appears highly theoretical, it has significant practical consequences as the following example demonstrates. Assume a situation in which we want to compile data from multiple data storages, e.g. data on an elderly citizen. The formats and structure of the data as well as the number and the hierarchy of data attributes (data fields) differ in these data storages. What do we do? The universal approach assumes that they all represent a single truth. This leads to the listing of all data attributes with the objective of finding the universal – also called global – data attributes of the data storage systems. Data vocabulary and synonym tables are then created, as well as transformation tables, to capture the differences in the formats and length of data. Challenges created by situations in which two or more data values have the same meaning, or where one data attribute can have two or more meanings, are solved in the same way. As the last step, the data are harmonized and standardized and duplicates are removed. Although this approach appears logical, it works only for sets of closely related ISs and data storages, i.e. for those, which are contextually close. It does not provide a generic solution because of the increasing complexity of data and the inability of humans to agree on the meaning of the data due to their different needs and interpretations. The first reason is practical and the second is ontological. According to the universal approach, IT professionals are the key persons involved as their expertise is needed to model the data.

The contextual approach acknowledges that the universal approach suits situations in which real-world representations are closely related, such as similar tasks or chains of tasks. For example, applications used in procurement could be related in this way. The contextual approach emphasizes the role of metadata and the use of agreed messages in sharing data between data storages. With metadata repositories and agreed messages, it is possible to create links between data storages so that data may remain in respective storages, including the memory of a computer. Also data ontologies, such as address lists (of elderly citizens) or medical device classifications, offer possibilities to support interoperability and data transfer. They contain descriptions about the meaning of data, which can be used in data queries and analyses. Thus, by knowing the metadata, it is possible to compile data from multiple sources. When data are compiled, the original data (e.g. video and audio clips, or documents) can be attached to the transferred data or the links can be used to provide trusted access to data. This could significantly reduce the need for data transfers as the data can remain in their original destination and still be recognized. Another important difference from the universal approach is the emphasis on business professionals, who know the meaning of data.

2.4. Synopsis – corporate governance of data framework

The proposed governance of data framework is compiled from the theoretical building blocks described above. The full framework is shown in Figure 4. For example, the governance element of the framework includes the EDM cycle as well as structures, processes and cooperation mechanisms. This and all the other elements of the framework have been discussed in detail in the three subsections above.

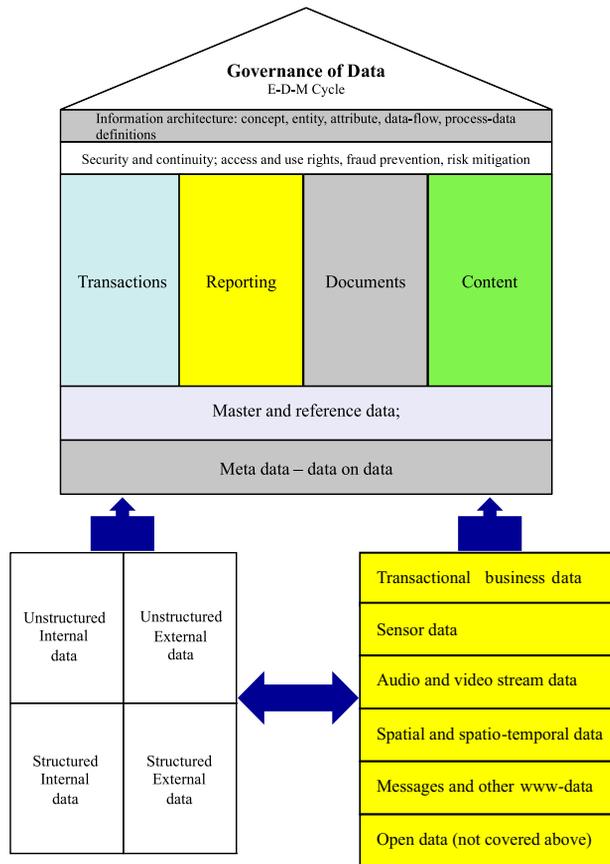


Fig. 4. Corporate governance of data framework (source: compiled by the authors)

We intend to evaluate the framework as follows. The use of robotics to support elderly citizens living at home could be the first empirical contexts for the evaluation of the framework. Robots have sensors, which collect data about the environment, the citizen served, the use of the functionalities in the robots, etc. We evaluate the framework by analysing whether or not it helps to provide better identification of data useful to home care, other social security providers, basic medical care, special medical care and other professionals serving the citizen, as well as the relatives of the citizen and the citizen himself or herself. The framework should help to determine what aspects of the data collected should be made interoperable and transferable with data, IS and data storages used by the aforementioned stakeholders and how. Finally, the framework should make it easier to define areas of accountability for the governance of data as a whole. It is noteworthy that data interoperability and transferability are bi-directional: i) from data compiled by robots to data used by professionals; ii) from data used by

professionals to data used to guide robots. For example, if a citizen has been operated on and is released to go home, the framework is valuable if it helps to determine what kind of data are transferred to the robots to support rehabilitation at home. Our intention is to evaluate the framework also in other contexts, such as electronic commerce.

3. Survey methodology

As we have not yet validated the proposed framework empirically, we investigated the perceived significance of data governance through two survey studies. In the first study, we used a relatively large, existing dataset called IT Barometer 2013, which includes data collected by a National Data Processing Association in mid-2013. One of the authors collected the survey data from CIOs and business executives, mainly from organizations with over 500 employees. We used only the part of the available data that concentrated on the perceived significance in the governance of data. The survey included the following governance of data questions expressed as statements: (1) In my organization, we have agreed clearly the ownership and the decision rights for data; (2) In my organization, we govern data comprehensively by developing the governance of data on the basis of a holistic roadmap.

Invitations to participate in the survey and one reminder were sent to 2,128 people. The response rate was 10%, which we regard as slightly higher than normal for IT management surveys sent to executives and managers. Of the respondents, 53% ($n = 115$) were CIOs and other IT managers, 19% were business executives and 28% were senior business experts. Of the participants, 27% worked in industry, 12% in commerce, 46% in services and 14% in public sector organizations. In terms of reporting to superiors, 49% of the survey respondents indicated that the CIO in their organization reported to CEO, 26% to CFO and 25% to other CxO.

The second survey was addressed to well-recognized healthcare and social welfare experts selected by the Association of Finnish Local and Regional Authorities. Respondents were asked to evaluate the governance principles and the benefits of an inter-organizational IT governance arrangement established between over 100 organizations in Finland. The IT governance arrangement is related to the reform of social welfare and healthcare services with related laws and decrees in the country. Also, in this case, one of the authors collected the survey data, of which we used the part that dealt with aspects of perceived significance for the governance of data. Dahlberg (2014) has described the case in detail. The survey included the following questions on the benefits of IT governance expressed as statements: (1) Increase the interoperability of patient/customer information systems and data storages; (2) Create jointly agreed data models and stick to them.

Invitations to participate in the survey were emailed to 260 persons. The Association mentioned above sent invitations to experts who were deemed to influence decisions concerning the establishment of IT governance arrangements within their regional areas

in the country. After one reminder, 68 responses were received. We regard the 26% response rate to be high for this type of study. Of the respondents, slightly over half (53%) worked in healthcare districts, 37% in cities, towns or municipalities and 10% in other organizations. Executives and managers accounted for 66% and experts the remaining 34%; 55% worked in social welfare or healthcare professions and 45% in social welfare or healthcare IT professions. Healthcare and social welfare executives or managers had worked in those positions on average for 15.9 years and experts for 12.7 years. Healthcare and social welfare CIOs and IT managers had worked in those positions on average for 7.2 years and IT experts for 8.8 years.

In both surveys, respondents were asked to evaluate survey items on a seven-point Likert scale from totally disagree (=1) to totally agree (=7). Both surveys also included demographic, situational and behavioural control variables. The first survey included eight behavioural control variables and the second survey seven behavioural control variables presented as statements on a seven-point Likert scale, similar to the other survey items. The demographic, situational and behavioural control variables in the first survey are shown in Table 2. Similarly, the demographic, situational and behavioural control variables in the second survey are shown in Table 3.

The statistical relationships between the dependent variables – that is, the two statements from the first survey and the two statements from the second survey – and the demographic, situational and behavioural control variables were analysed using Pearson's correlation coefficient. For statistical analyses, behavioural control variable responses were also classified into two classes: "agree" (Likert scale values 5–7) and "disagree" (Likert scale values 1–3). Differences in the distributions of the dependent variables were analysed using Fischer's F-test and differences in the means using the two-tailed Student's t-test. These analyses were done separately for each demographic, situational and behavioural control variable classified into the "agree" and "disagree" classes. The distributions of the dependent variables were skewed. As the results, we therefore report only the statistically significant differences with confidence levels above 95% for the "agree" and "disagree" groups. We also limit the reporting of statistical analysis to the results of the t-test for the same reason. However, it is worth mentioning that correlations between the two variables were statistically significant if there were statistically significant differences in the t-test.

In summary, we used guidance provided by Sireci (1998) and Yin (2009) to design both surveys and also the scales of the survey statements in both surveys.

Table 2. Demographic, situational and behavioural control variables in first survey

Construct	SURVEY ITEM
Demographic and situational control variables	Professional status (business manager, IT manager, expert)
	Size of the organization (<100, 101–500, 500 + employees)
	Industry (Industrial, commerce, service, public sector)
	The composition of IT costs (IT function, IT function + business IT)
	The superior of the CIO (CEO, CFO, other CxO)
Behavioral control variables	The organizational status of the CIO (executive committee member, non member)
	My organization manages IT and develops IT management as its strategic capability
	My organizations develops systematically IT competencies needed in the execution of our business including IT management competencies
	Senior, business unit and IT executives share the accountabilities of IT management on the basis of a clearly defined governance arrangement
	In my organization business strategy, business models, operative model and IT constitute a well-integrated whole
	In my organization the selection of IT solutions is based on the alignment of business and IT needs
	My organization defines measurable objectives for IT purchases so that business needs are well taken care of
	My organization knows well how IT impacts our business as evaluated with reliable metrics
After IT purchases we monitor the achievement of the measurable objectives defined for the purchases	

Table 3. Demographic, situational and behavioural control variables in second survey

Construct	SURVEY ITEM
Demographic and situational control variables	Type of organization (municipal healthcare district)
	Geographic area (the area for which governance of IT was established, other parts of the country)
	Organizational status (H/S manager, H/S expert, IT manager, IT expert), where H/S = healthcare/social welfare
	Experience in years in H/S managerial positions
	Experience in years in H/S expert positions
	Experience in years in IT managerial positions
	Experience in years in IT expert positions
	Involvement in the establishment of the inter-organizational IT governance arrangement

End of Table 3

Construct	SURVEY ITEM
Behavioral control variables	As a whole the role of IT is generally regarded as much too important for healthcare and social welfare services
	As few as possible funds should be used for healthcare and social welfare IT services
	In future IT will be much more important to the development and operation of healthcare and social welfare services
	My organization is a highly competent deployer of IT for healthcare and social welfare services
	In my organization IT governance accountabilities are allocated clearly between healthcare/social welfare and IT professionals
	We have clear measurable objectives for healthcare social welfare IT services
	As a whole my organization applies IT so well to healthcare and social welfare services that it would be graded as A or A+ were it considered in terms of educational grading

4. Survey results

The mean and median values and the proportions of high evaluations among the responses to the questions for both surveys are shown in Table 4. The first survey measured the status in the governance of data and the second survey the significance for the governance of data. The respondents in the first survey evaluated the status in the governance of data low on the Likert scale from 1 to 7. The results of the second survey indicate that the respondents consider the governance of data highly significant. Great caution is necessary in comparing the results of the surveys. Respondents were selected on the basis of different sampling frames, the contexts of the surveys were different and also the formulations of survey items were dissimilar. It is perhaps safe to conclude tentatively that the governance of data is perceived to be more important than its current status suggests. No other empirical data were collected in the first survey. The other survey, however, comprises one set of data among other sets of empirical evidence (Dahlberg 2014). When the governance of IT principles and the benefits of IT governance were presented to the executives, managers and experts within the area of established inter-organizational IT governance, the lack of social and medical data interoperability and transferability was mentioned as the most burning challenge related to IT governance (Dahlberg 2014).

In the first survey, only the position of the CIO (the CIO is or is not a member of the organization's executive committee) produced statistically significant differences in the means of one dependent variable. The average of responses was 4.45 for respondents who indicated that the CIO was a member of their organization's executive committee.

Table 4. Results of the two surveys

Survey Item	Mean	Median	Proportion of strongly agree
<i>Survey 1: (IT barometer data): Business and IT executive evaluations about the IT and its role. (n = 207 for the first item and n = 206 for the second survey item. Strongly agree is composed of values 6 and 7 on the Likert scale)</i>			
In my organization, we have agreed clearly the ownership and the decision rights for data	3.9	4.0	16.4%
In my organization, we govern data comprehensively by developing the governance of data on the basis of a holistic road-map	4.0	4.0	14.1%
<i>Survey 2: Social welfare and healthcare professionals' evaluations about the significance of IT governance principles and IT governance benefits (n = 68)</i>			
Increase the interoperability of patient/customer information systems and data storages	6.3	6.5	86.8%
Create jointly agreed data models and stick to them	6.3	7.0	85.3%

The corresponding average was 3.50 for respondents who indicated that the CIO was not a member of the executive committee. The statistical reliability of this result was greater than 99.999%. All the eight behavioural control variables produced statistically significant differences in the means of both dependent variables. Furthermore, the statistical reliability of all results except one was greater than 99.999%; the statistical reliability of the exception was 99.992%. Figure 5 illustrates the comparison of responses between the groups created on the basis of behavioural control variables.

In our opinion, the results of the two surveys provide clear support for the development of governance of data in general and also for proposing a governance of data framework such as the one provided in this paper. The results shown in the upper part of Table 4 provide an indication that the status of the governance of data is not satisfactory. Furthermore, the results shown in the lower part of Table 4 suggest that data interoperability and the governance of data were considered highly necessary by the respondents in the second survey.

The results visualized in Figure 5 support the use of the building process selected in section 2 for the design of the corporate governance of data framework. Good governance of IT practices includes systematic development of IT competencies, clear accountability, business–IT alignment, the setting of clear objectives for IT purchases and the monitoring of IT usage and purchases using reliable metrics. These are not only in line with the governance of IT principles (e.g. ISO/IEC 2008), but also with corporate governance. Furthermore, the results shown in Figure 5 also demonstrate that the deployment of good IT governance practices is positively related to good corporate governance of data.

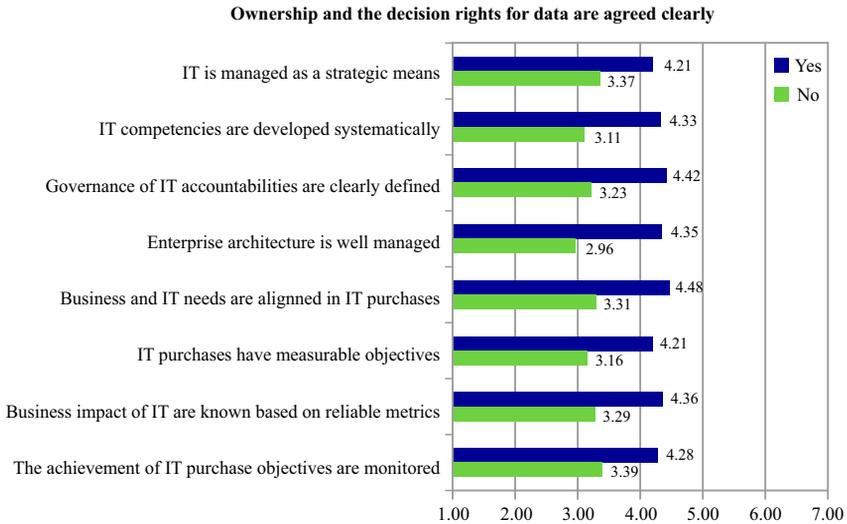


Fig. 5. Differences in the means of response classes formed on the basis of responses for behavioural control variables (“no” values 1–3, “yes” values 5–7, n = 207)

5. Discussion and conclusions

This paper has proposed a framework for the corporate governance of digital data. We have not been able to identify data governance frameworks within prior research, addressing the issue from the perspective of corporate governance and the related governance of IT. Corporate governance and the governance of IT consider investments in IT/data from the outside. The key question is how investors, as the providers of funds, are able to secure a return on their investments. The existing DBBOK and MIKE frameworks approach the issue from technical and data management perspectives, i.e. from the inside. Thus, they are unable to answer the key concern of the “investors”, how to assure returns. The proposed framework emphasizes the central role of business executives in the governance of data. Second, our framework considers data to be defined contextually as opposed to universally. This underlines the role of business executives, managers and professionals even further, as they know for what purposes the data are used and what they mean in specific contexts. For these reasons, the proposed framework was built on the idea that the governance of data is a managerial issue, not a technical or a modelling issue.

The framework was designed by combining three theoretical building blocks. These were corporate governance and governance of IT, the digitalization of data and digital strategy, and data and information assets used by organizations, taking into consideration the ontological nature of these assets. We used established definitions for corporate governance (Shleifer, Vishny 1997) and for the governance of IT (ISO/IEC 2008) in the design of the framework. The framework contributes to knowledge concerning both cor-

porate governance and the governance of IT in addition to establishing a new knowledge construct – governance of data. As IT and digital data have become elements of most investments, the question of how investors are able to secure returns from such investments is increasingly important. Within the IT governance research, the governance of data is a new topic. For example, within the ISO/IEC standardization, it is likely to emerge formally as a new work item in 2015.

The results of the two surveys reported in this paper suggest that there is a clear need for the corporate governance of data. The two surveys indicate that the current status in the governance of data is unsatisfactory and that there is a strong need to improve data interoperability. The surveys also provide empirical evidence supporting the use of the corporate governance approach.

Our article also has some limitations. We once again point out that the proposed framework has not been tested or verified empirically. This may mean that the building blocks of the framework do not fit together and require modifications. Second, the empirical evidence provided concerning the status in the governance of data and its significance was collected only in one country. Multivariate statistical testing could shed more light on the empirical results.

Despite these limitations, our paper offers several scientific contributions. The most significant contribution is the proposed framework. The framework builds on the idea that digital data are contextually defined and central to business management. Thus, the proposed framework fills a research gap and offers an alternative to the DMBOK and other IT and data modelling based frameworks, which assume the universality of data and which address data management rather than the governance of data. Parts of our proposed framework, namely the typification of data sources and the matrix combining the structure and the internality of data shown in Figures 2 and 3, could even be used to extend the DMBOK framework – to better govern and manage data.

The proposed framework opens up several meaningful venues for future research. We have already mentioned the verification of the framework. Alternatively, it would be possible to compare alternative data management frameworks, their ontologies and theoretical bases from a governance perspective. Such research appears to be sorely needed as the volume and complexity of digital data continue to grow at a fast rate. Our motive for proposing the framework is to facilitate better management of the complexity of data and to involve business executives, managers and experts in the governance and management of data with a clear mandate. The proposed framework may, for example, also help to cope with and provide direction in the data management challenges raised under the “big data” concept. The concept of “big data” appears to be far from fulfilling its promise.

Finally, our advice to practitioners is to ensure that business executives and managers are genuinely and actively involved in the governance and management of data. Without their involvement, advances in the governance of digital data, its management and deployment are likely to be slow.

References

- Bharadwaj, A.; El Sawy, O. A.; Pavlou, P. A.; Venkatraman, N. 2013. Digital business strategy: toward a next generation of insights, *MIS Quarterly* 37(2): 471–482.
- Bebchuk, L. A.; Weisbach, M. A. 2010. The state of corporate governance research, *The Review of Financial Studies* 23(3): 939–961. <http://dx.doi.org/10.1093/rfs/hhp121>
- Caiwei, X. 2013. Innovation and change for the elderly in China, Chapter 7 in T. Obi, J.-P. Auffret, N. Iwasaki (Eds.). *Aging society and ICT: global silver innovation*. Amsterdam, Netherlands: IOS Press.
- Cleven, A.; Wortmann, F. 2010. Uncovering four strategies to approach master data management, in *Proceedings of the 43rd Hawaii International Conference on System Sciences*, 5–8 January 2010, Koloa, Kauai, Hawaii. New Jersey: IEEE press.
- Dahlberg, T. 2010. *The final report of Master Data Management Benchmark Best Practices research project*. Aalto University School of Business, Helsinki 2010.
- Dahlberg, T. 2014. Perceived need to cooperate in the creation of inter-organizational it governance for social welfare and health care it services – a case study, Chapter 2 in K. Saranto, M. Castren, T. Kuusela, S. Hysynsalmi, S. Ojala (Eds.). *Communications in computer and information science*. Heidelberg, Germany: Springer Verlag. http://dx.doi.org/10.1007/978-3-319-10211-5_2
- Dahlberg, T.; Heikkilä, J.; Heikkilä, M. 2011. Framework and research agenda for master data management in distributed environments, *Scandinavian Information Systems Research Conference (IRIS2011)*, 16–19 August 2011, Turku, Finland.
- The Data Management Association (DAMA). 2009. *The DAMA guide to the Data Management Body of Knowledge (DAMA-DMBOK guide)*. 1st ed. Bradley Beach, New Jersey: Technics Publications.
- Davenport, T. 2007. *Competing on analytics: the new science of winning*. Boston, MA: Harvard Business School Press.
- Henderson, J. C.; Venkatraman, N. 1993. Strategic alignment: leveraging information technology for transforming organizations, *IBM Systems Journal* 32(1): 4–16. <http://dx.doi.org/10.1147/sj.382.0472>
- Hilbert, M.; Lopez, P. 2011. The world's technological capacity to store, communicate, and compute information, *Science* 332(6025): 60–65. <http://dx.doi.org/10.1126/science.1200970>
- Hovenga, E. J. S. 2013. Impact of data governance on a nation's healthcare system building blocks, Chapter 2 in E. J. S. Hovenga, H. Grain (Eds.). *Health information governance in a digital environment: studies in health technology and informatics*, vol. 193. Amsterdam, Netherlands: IOS Press.
- IDC. 2011. *Overload: Global Information Crated and Available Storage* [online], [cited 11 November 2014]. Available from Internet: <http://www.idc.com>
- Iivari, J.; Hirscheim R.; Klein, H. K. 1998. A paradigmatic analysis contrasting information systems development approaches and methodologies, *Information Systems Research* 9(2): 164–193. <http://dx.doi.org/10.1287/isre.9.2.164>
- ISC. 2012. *Internet Systems Consortium Status Report 2012* [online], [cited 11 November 2014]. Available from Internet: <http://www.isc.org>.
- ISO/IEC. 2008. *International Organization for Standardization and the International Electrotechnical Commission, ISO/IEC 38500:2008 Corporate Governance of Information Technology* [online], [cited 9 November 2014]. Available from Internet: <http://www.iso.org>
- Markus, L. M.; Quang, B. 2012. Going concerns: the governance of interorganizational coordination hubs, *Journal of Management Information Systems* 28(4): 163–197. <http://dx.doi.org/10.2753/MIS0742-1222280407>
- MIKE. 2014. *Method for an Integrated Knowledge Environment* [online]. Continuously updated open source framework and method for data, information and knowledge management [cited 11 November 2014]. Available from Internet: <http://mike2.openmethodology.org>

Obi, T.; Auffret, J.-P.; Iwasaki, N. 2013. *Aging society and ICT: global silver innovation*. Amsterdam, Netherlands: IOS Press.

OECD. 2014a. Elderly population by region, in *OECD Factbook 2014: economic, environmental and social statistics*. OECD Publishing. <http://dx.doi.org/10.1787/factbook-2014-5-en>

OECD. 2014b. *Ageing of OECD countries* [online]. Excel file downloadable from the OECD statistics page [cited 9 November 2014]. Available from Internet: <http://www.oecd.org/statistics/>

Pagani, M. 2013. Digital business strategy and value creation: framing the dynamic cycle of control points, *MIS Quarterly* 37(2): 617–632.

Shleifer, A.; Vishny, R. W. 1997. A survey of corporate governance, *Journal of Finance* 52(2): 737–783. <http://dx.doi.org/10.1111/j.1540-6261.1997.tb04820.x>

Sireci, S. G. 1998. The construct of content validity, *Social Indicator Research* 45: 83–117. <http://dx.doi.org/10.1023/A:1006985528729>

United Nations. 2014. *World population prospects: the 2012 revision* [online]. United Nations, Department of Economic and Social Affairs. Downloadable open data excel spreadsheet [cited 7 November 2014]. Available from Internet: http://esa.un.org/wpp/WPP2012_POP_F13_A_OLD_AGE_DEPENDENCY_RATIO_1564.XLS

Van Grembergen, W.; De Haes, S.; Guldentops, E. 2004. Structures, processes and relational mechanisms for IT governance, Chapter 1, in W. Van Grembergen (Ed.). *Strategies for information technology governance*. Hershey: Idea Group Global.

Van Grembergen, W.; De Haes, S. 2008. *Implementing information technology governance: models, practices and cases*. Hershey: Idea Group Global. <http://dx.doi.org/10.4018/978-1-59904-924-3>

Wand, Y.; Weber, R. 1993. On the ontological expressiveness of information systems analysis and design grammars, *Information Systems Journal* 3(4): 217–237. <http://dx.doi.org/10.1111/j.1365-2575.1993.tb00127.x>

Wand, Y.; Weber, R. 2002. Research commentary: information systems and conceptual modeling – a research agenda, *Information Systems Research* 13(4): 364–376. <http://dx.doi.org/10.1287/isre.13.4.363.69>

Watts, S.; Shankaranarayanan, G.; Even, A. 2009. Data quality assessment in context: a cognitive perspective, *Decision Support Systems* 48(1): 202–211. <http://dx.doi.org/10.1016/j.dss.2009.07.012>

Wintley-Jensen, P. 2013. EU activities on aging, Chapter 27 in T. Obi, J.-P. Auffret, N. Iwasaki (Eds.). *Aging society and ICT: global silver innovation*. Amsterdam, Netherlands: IOS Press.

Yin, R. K. 2009. *Case study research, design and methods*. 4th ed. Thousand Oaks, CA: SAGE Publications.

Tomi DAHLBERG. PhD, professor, research director, senior researcher fellow, executive in residence since 2000 at Aalto University Business School, University of Jyväskylä, Turku School of Economics at the University of Turku, Åbo Akademi University. Director, Senior Vice President, Executive Vice President, CIO, CTO, CEO, Board member in software industry, finance and banking, telecoms, nanotechnology and management consulting since 1984. Member of the ISO/IEC JTC1 SC40 and its WG1 (governance of IT) and wg2 (IT service management) and the chair of the Finnish SFS shadow group SR-08. Research interests: governance of information, governance of data, inter-organizational governance of IT, business models, master and metadata management, governance of aging societies, CIO profession, payment instruments, innovation management processes.

Tiina NOKKALA. MSc (Econ.), PhD student and researcher, PhD research is supervised by Tomi Dahlberg. Previously worked in the finance and insurance industry. Research interests: governance of data, master data governance, governance of aging societies.