Data Governance and Information Governance: Set of Definitions in Relation to Data and Information as Part of DIKW

Jan Merkus[1][a], Remko Helms[1][b] and Rob Kusters[1][c]
[1]*Faculty of Management, Science & Technology, Open Universiteit, Valkenburgerweg 177, Heerlen, The Netherlands*

Keywords: Data Governance, Information Governance, Data Management, Information Management, Data Information Knowledge Wisdom (DIKW), Meaning.

Abstract: Chaos emerges with the ever growing amounts of data and information within organisations. But it is problematic to manage these valuable assets and also remain accountable and compliant for them because there is no agreement about even their definitions. Our objective is to propose a coherent set of definitions for data governance and information governance within and across organisations in relation with data and information as underlying concepts. As a research method, we explore elements from existing definitions in literature about the Data-Information-Knowledge-Wisdom pyramid and about data governance and information governance. Classification of these elements and coding them in concepts during discussions among peers resulted in a new vocabulary. This forms the basis for formulation and design of an original coherent set of definitions for data, information, meaning, data governance and information governance. This research is grounded, goal oriented and uses multiple accepted literature review methods. But it is limited to the literature found and the IS domain.

1 INTRODUCTION

The amount of data and information within and across organisations is growing epidemically worldwide. One of the main causes is that we need data and information for every activity and transaction, but that human capacity of remembering information is limited (Miller, 1956). Computerized systems assist us here. But where data storage costs decreases, storage availability increases and computer capacity grows exponentially, data disposal is neglected (Moore, 1965; Murphy and Chang, 2009). This fuels the collection of an ever growing amount of data and information.

But, first, chaos accumulates when the amounts of data and information become uncontrollable. Therefore, this situation demands adequate data and information management in order to gain control. Second, the situation escalated before. The debacles of e.g. Enron, Parmalat or Ahold urged governments to issue new regulations requiring improved accountability, also for data and information. Third, data and information require accountability, as valuable assets for organisations (Bhansali, 2013; Khatri and Brown, 2010; Kooper et al., 2011).

Establishing the management of data and information is known as data governance (DG) and/or information governance (IG). In general, governance is about controlling, so data governance is about controlling data. Likewise, information governance is about controlling information e.g. (Tallon et al., 2013; Weber et al., 2008). The scope of DG and IG can be an organisation and/or a network or even a complete value chain (Begg and Caira, 2012; Rasouli et al., 2016b). Data and information have been researched extensively as part of the information systems (IS) domain IT governance (De Haes and Van Grembergen, 2004; Luftman, 2003; Weill and Ross, 2004). And DG and IG literature starts growing (Alhassan et al., 2016; Jang and Kim, 2016; Niemi and Laine, 2016). But so far, academics have different opinions about DG and IG and disagree on definitions of data and information. For example, information has been defined by some researchers as being both

[a] https://orcid.org/0000-0003-2216-7816
[b] https://orcid.org/0000-0002-3707-4201
[c] https://orcid.org/0000-0003-4069-5655

Merkus, J., Helms, R. and Kusters, R.
Data Governance and Information Governance: Set of Definitions in Relation to Data and Information as Part of DIKW.
DOI: 10.5220/0007411901430154
In *Proceedings of the 21st International Conference on Enterprise Information Systems (ICEIS 2019)*, pages 143-154
ISBN: 978-989-758-372-8

physical, objective, subjective and sociocultural while others define on one hand data being physical and on the other hand information being subjective (Baškarada and Koronios, 2013; Boell, 2017). Likewise, some do not distinguish DG from IG while others do (Khatri and Brown, 2010; Kooper et al., 2011; Olaitan et al., 2016).

However, coherency among concepts is necessary for establishing adequate management of data and information within and across organisations. A coherent set of definitions therefore contributes to the body of knowledge on DG/IG, data and information. The implication for practice is that this gives better mutual understanding and vocabulary for managing data and information within and across organisations.

The remainder of this paper is structured as follows. First, we describe the theoretical background of DG, IG and DIKW. Then, we justify our literature research method followed by our literature research findings. Next, we present our analyses of the elements from definitions of data and information resulting in a conceptual model which coherently organises the vocabulary of articulated concepts. After that we give our analysis of the elements from definitions of DG and IG resulting in additional vocabulary and definitions. Last, we finish with a discussion whether to distinct DG from IG.

2 THEORETICAL BACKGROUND

Governance definitions are agreed upon. Other research already recognized that governance is establishing management by setting standards, policies and processes to create organizational structures (ARMA International, 2013; Panian, 2010). And corporate governance cascades to subdomains like IT in order to become accountable for corporate business goals cascading to these subdomains (De Haes et al., 2013; Neff et al., 2013; Olaitan et al., 2016).

In the past, DG and IG were seen as IT governance subdomains and observed from three corporate governance perspectives while focusing on accountability of data and information as valuable assets. The first is the COSO controlling perspective. The COSO mission for Corporate Governance is to improve organizational performance and governance since 1992. Cobit5 as part of COSO covers IT controlling and both are management and risk driven. Much research is carried out on Cobit since its start in the nineties and having five versions already (Cheong and Chang, 2007; COSO, 2014; De Haes and Van Grembergen, 2004). The second perspective is a relationship between CG, ITG and IG via compliance, quality and value of information focusing on IT in general and on information and data specifically since 2008. However, there is not much research for CG in relation with ITG (ISO/IEC38500:2015/38505:2017),(De Haes and Van Grembergen, 2004; Lajara and Maçada, 2013; Tallon and Diego, 2013). A third corporate perspective on ITG and DG is governance, risk management & compliance (GRC) mainly focusing on data since 2010 (Gregory, 2010; Hagmann, 2013; Racz et al., 2010b). Not much research is carried out for CG in relation with DG either.

However, DG and IG definitions became less clear. Although not mentioned as such, DG and IG originate from regulations preventing misuse and abuse of data and information within organisations. In the UK, the Data Protection Act was passed in 1998 requiring protection of personal and organizational data. This resulted in the National Health Service (NHS) IG initiatives (Liaw et al., 2014; Lomas, 2010). In the USA, financial crises Enron in 2001 and WorldCom have given rise to legislation as Sarbanes-Oxley in 2002 requiring protection of financial assets under corporate governance by means of proper data and information (Cheong and Chang, 2007; Williams, 2011). In the EU, GDPR legislation of 2018 was proceeded by EU and national privacy legislations. Banking organisations started similar initiatives with the Basel accords I (1988) – IV (2017) being enforced by governments. The earliest mentioning of IG by Peterson had already distinct information from IT in 2002 (Peterson, 2002). DG was first mentioned by Wende in 2007 and positioned as part of ITG (Wende, 2007). In the same year, Cheong and Chang mentioned the business as data drivers (Cheong and Chang, 2007). Furthermore, some did not separate DG from IG (Khatri and Brown, 2010; Olaitan et al., 2016; Tallon et al., 2013). But others did separate IG from DG with the argument that, unlike data, stakeholders outside IT, such as business users, own information because of its subjectivity (Beijer and Kooper, 2010; Kooper et al., 2011). Researchers call for further research on DG and IG (Otto, 2011a; Tallon et al., 2013).

Considering the data and information as underlying concepts of DG and IG does not help much either (Beijer and Kooper, 2010; Weber et al., 2009). Data's singular form is datum which means "a fact or piece of information" according Oxford Learners Dictionary. Handling of both these concepts

was part of IT management and ITG in the IS research domain (De Haes and Van Grembergen, 2004; Weill and Ross, 2004). Moreover, data and information have been discussed extensively in the DIKW-pyramid literature. Here, data and information are defined in relation with knowledge and wisdom. This model suggests that wisdom is abstracted from knowledge, knowledge from information and information from data as raw material (Ackoff, 1989; Bierly III, 2000; Elliot, 1934; Huizing, 2007; Rowley, 2007; Zins, 2007a). But over the years, the DIKW discussions showed many different definitions of data and information. Researchers are still calling for clear definitions of data and information (Aven, 2013; Bawden and Robinson, 2015; Boell, 2017). We conclude from this that researchers heavily disagree on definitions of DG, IG, data and information.

But clear vocabulary and definitions are necessary for adequate management of data and information. What is needed is a coherent set of definitions for DG, IG and its underlying concepts data and information. A set of definitions is coherently built with interrelated definitions. Good theory definitions are constructed by deciding how to code the basic elements into concepts, build conceptual definitions, determine how these are related, limit the domain and do predictions (Carley, 1993; Wacker, 1998).

3 METHOD

To make this research traceable and reproducible, this research is conducted with a predefined objective, is structured according a method elaborated from literature and is assessed for validity and reliability.

3.1 Objective

The objective of our research is to take decisions in the formulation of a coherent set of definitions for DG and IG in relation to data and information in order to enable improvement of the management of data and information.

Consequently, our research question is: **What is a coherent set of definitions for Data Governance and Information Governance in relation to the underlying concepts data and information.**

Decomposing and understanding this problem shows the necessity to first articulate vocabulary and formulate definitions for data and information as underlying concepts. Second, we articulate vocabulary and formulate definitions for DG and IG. Consequently, this research answers two sub-questions. These are: (1) What are elements of existing definitions for data and information in the context of DIKW and how do these interrelate. (2) What are elements of existing definitions for DG and IG in relation with the concepts of data and information.

3.2 Approach

Defining the concepts of DG, IG, data and information can be based on literature research because some research for DG and IG and much research for DIKW already exists. Our literature review approach is based on the availability of this literature. On one hand there is plentiful literature with definitions of data and information which requires inventory and order. On the other hand there is shortage of literature on DG and IG. Consequently our research method consists out of two consecutive phases. The first phase reviews step-by-step literature on the definitions of data and information to answer the first sub-question. Rowley has conducted similar research for a period from 2003 to 2007 which we review and extend to 2018. The same steps are applied in the second phase, but adjusted to the definition of DG and IG. Each phase follows a content analysis method derived from grounded theory (Carley, 1993; Corbin and Strauss, 1990; Finney and Corbett, 2007; Krippendorff, 1989; Rowley, 2007). In addition, the applied LR method is further elaborated with other widely used LR methods (Bandara et al., 2011; Kitchenham and Charters, 2007; Tranfield et al., 2003; vom Brocke and Rosemann, 2010; Webster and Watson, 2002). To improve LR quality, we applied literature selection criteria from literature (Dybå and Dingsøyr, 2008). The LR typology used is theoretical LR and it is structured by design (Paré et al., 2015). Only peer reviewed journal articles, conference articles and text books containing searched definitions have been selected as relevant literature.

The five steps iteratively applied in each two phases are the following. For literature selection: A) finding literature and B) extracting definitions. For analysis: part C) coding and D) inferring. And for synthesis: E) designing (or defining). The outcomes are discussed and elaborated within the research team.

Phase 1 - Data and Information Definition

This phase explores elements of most existing definitions of data and information as part of DIKW within the IS domain. The literature selection contains the iterative steps A and B:

A) **Finding literature** by defining keywords on basis of the research questions which serve formulating search criteria and finding relevant literature with search engines
B) **Extracting definitions** by screening found peer reviewed literature, extract definitions by reading and understanding the texts and tabulation of found definitions (Dybå and Dingsøyr, 2008).

The goal of the analysis is to come to understanding by ordering and classifying similar elements of existing definitions of data and information from literature as in the steps C and D:
C) **Coding** by determining elements (e's) in definitions as concepts, classifying these to deduce meaning and coding the classes into subcategories for exploring new insights in the phenomena resulting in articulating new vocabulary (Carley, 1993; Corbin and Strauss, 1990; Krippendorff, 1989).
D) **Inferring** by expressing referenced meaning, notable concepts, inferences and attendant thoughts.

The goal of the syntheses is to combine elements to a coherent whole in step E:
E) **Designing** a coherent set of definitions for data and information

Phase 2 - Data Governance and Information Governance Definition

In phase 2 a similar approach is applied for defining DG and IG, but adjusted as follows. For LR steps A and B, the goal is to explore elements of existing definitions of data governance and information governance within the IS domain as basis for vocabulary and for the demarcation of DG and IG. For analysis steps C and D, the goal is to analyse elements of DG and IG, classify and code concepts coherently for articulating vocabulary, interrelated with the analysis of data and information. For synthesis step E, the goal is to formulate own definitions for DG and IG and demarcation of these. The vocabulary and definitions of DG and IG are coherently interrelated with the definitions of data and information.

Design Requirements

Designing a coherent set of definitions requires design requirements for structuring and quality reasons. As a 'good theory', a set of definitions has four basic criteria: conceptual definitions, domain limitations, relationship-building and predictions. Next to the good theory criteria we also choose criteria supported by literature: clarity, conciseness, parsimony, granularity or elementary, utility (Wacker, 1998; Zielstorff, 1998).

3.3 Method Assessment

Assessing the rigor of this approach reveals what makes this LR valid and reliable. This research is goal oriented, grounded in literature and replicates earlier research. Additionally, LR is suitable for explorative research (Boell, 2017; Paré et al., 2015). The approach is based on several accepted LR methods and only uses literature which meets accepted selection quality criteria. This makes the method transparent and repeatable for others. Furthermore, analysis of existing definitions seeks correspondence with existing literature. Iteratively searching for text samples and formulating hypotheses allows learning, understanding and reflection over time. Exploring underlying concepts aims at finding basic theory elements. Moreover, the set of definitions is based on an existing model which also improves internal validity and reliability. Seven discussion sessions within the research group promotes intersubjective validity and reduces researcher bias.

A threat to reliability is that this LR is limited to existing definitions of data and information within the DIKW because of the amount of literature on data and information already available for that topic. We think that this limitation to DIKW does not affect validity significantly. A measure to reduce the risk of excluding existing definitions is to keep looking for definitions until the same definitions re-occur. Another threat is the possibility to make wrong choices in articulating vocabulary and formulating definitions. A measure against this is to work in a goal oriented manner, find definitions structurally, reference coding during analysis and use design requirements during design.

4 FINDING DEFINITIONS

Exploring elements of existing definitions in IS literature for data and information as part of DIKW, DG and IG resulted in the following findings. First we describe the our literature search and the definitions found. Then we summarize what we understand from extracted definitions.

4.1 Definitions Found

The LR was first conducted in the period from October until December 2016 simultaneously for phase one and two and repeated in August 2017 and in April 2018. Research passed steps A to E iteratively while gaining more insight. Google scholar is used for searching and Open University online

library for accessing articles, Google books is used for searching and Koninklijke Bibliotheek using Worldcat for accessing textbooks. The search string *allintitle*: *"Data Information Knowledge Wisdom" OR "DIKW" OR "Data Governance" OR "Information Governance"* resulted in 832 hits in total in the period from 2007 until April 2018. Backward snowballing the books mentioned by Rowley in 2007 and selected articles resulted in 82 text books in total. Screening search results with the quality selection criteria from literature resulted in 87 relevant articles and 29 relevant text books containing definitions for DG (52 sources), IG (25) and DIKW (40). The number of different definitions extracted from literature is 109 from articles and 37 from text books; for data 36, for information 38, for DG 47 and IG 25. Some articles contain several definitions. The number of definitions found since 2007 shows that the discussion is still going. It also shows that the number of definitions of DG and IG increases over time which is supported by other research (Jang and Kim, 2016). And it shows the growing importance of DG and IG over time. The list of all definitions is available from the author.

4.2 Data and Information Definitions

Existing definitions for each of the DIKW have roots in ancient history e.g. Solomon's wisdom (Rowley, 2006). This research concentrates on the definitions of data and information by objective and delineates knowledge and wisdom. In Latin data is plural of the noun datum meaning 'given', so data are facts (Oxfords Learners). In Latin the verb *informare*, in Middle French *enformer* and in Middle English *enforme* all mean to give form to, so information has been formed (Pijpers, 2006), (Zins, 2007b). In between then and 2007 many more definitions were formulated (Rowley, 2007), (Zins, 2007a). The literature research resulted in 74 different definitions of data and information since 2007.

Data could be physical or logical data according literature: "Data are symbols that represent the properties of objects and events", (Ackoff, 1989) or: "physical substance...electronically recorded", (Hey, 2004) or: "properties of events or things "out there", (Boisot and Canals, 2004) or: "data is described as a discrete physical entity, external to the individual", (Newell et al., 2009) or: "anything recordable", (Frické and Martin, 2007) or: "physical signs", (Baškarada and Koronios, 2013) or: "raw, alphanumeric characters", (Smallwood, 2014). Data are kind of facts: "raw facts", (Bierly III, 2000) or: "objective facts", (Jennex, 2009). And data have (no) value: have no value (Rowley, 2007), (Bierly III, 2000) or: are assets (Rifaie et al., 2009).

The meaning of information should depend on its usage or context: "information consists of processed data, the processing directed at increasing its usefulness", (Ackoff, 1989) or: "meaningful, useful data", (Bierly III, 2000) or: "information is defined as data in context", (Otto, 2011b) or: "the context and the usage provide a meaning to the data that constitute information", (Al-Khouri, 2012) or: "information is data that have been shaped into a form that is meaningful and useful to human beings", (Laudon and Laudon, 2013). Information should have (no) value and be objective or subjective: "Information has many definitions by itself more or less without value. information has meaning for more than one person and can be objective", (Pijpers, 2006) or: "the value of information is subjective, since it may be more useful in satisfying the wants of one person than another, or of no use to one person and of use to another. Giving meaning to information is a human element and by definition subjective, since objectivism cannot deal with the human sense making", (Huizing, 2007) or: "and having no intrinsic value of its own", (Newell et al., 2009) or: "Information assets (or data) are defined as facts having value potential value", (Khatri and Brown, 2010) or: "information in the literature distinguishing four broad stances on information: physical stance, objective stance, subject-centered stance, and sociocultural stance", (Boell, 2017). And information needs demarcation: if knowledge is a property of people and embodies prior understanding, experience and learning, it is difficult to argue that explicit knowledge, recorded in documents, is any more or less than information (Rowley, 2007) or: information is what is known in other literature as 'weak knowledge' (Frické and Martin, 2007) or: information is processed data (Otto, 2013) or: others have argued that the concept of information is difficult to define due to its multidimensional nature.

4.3 DG and IG Definitions

In the LR we found confirmation that DG and IG are not much discussed. This confirms the several calls for further research. The selected literature shows the following discussion about DG and IG. DG's first discussions focus on how to rule and decide on data in alignment with corporate objectives (Begg and Caira, 2012; Cheong and Chang, 2007; Otto, 2011a; Rainer et al., 2009; Wende, 2007). Later on, continental scientists focus on the value of data as an asset (Khatri and Brown, 2010; Kooper et al., 2011;

Otto, 2011a). The Anglo-Saxons focus on procedural elements in the meantime (Gregory, 2010; Loshin, 2009; Plotkin, 2013; Rosenbaum, 2010). Others merely focus on data elements like quality, availability or reliability (Korhonen et al., 2013; Smallwood, 2014; Tarantino and Cernauskas, 2011). IG discussions focus on how to rule and decide on information over its life cycle in alignment with (corporate) objectives, (Bhansali, 2013; Lajara and Maçada, 2013; Smallwood, 2014; Tallon et al., 2013). Later on, continental scientists focus on the value of data as an asset (Niemi and Laine, 2016; Olaitan et al., 2016; Rasouli et al., 2016a). Although not yet much developed, scientist explore practical application of DG and IG in specific contexts. For DG in health care, asset management and data warehouses (Brous et al., 2016; Rifaie et al., 2009; Rosenbaum, 2010). For IG in defense and between organisations (Kravets and Zimmermann, 2012; Lajara and Maçada, 2013; Rasouli et al., 2017). Analyses of the definitions for data, information, DG and IG found in literature according described method results in growing understanding of the underlying concepts. These concepts form the basis for articulating vocabulary necessary for formulating and designing a coherent set of definitions as realized in syntheses.

5 ANALYSES & SYNTHESIS

Analyses of the definitions for data, information, DG and IG results in underlying concepts which form the basis for articulating vocabulary necessary for formulating own definitions and conceptual model as realized in syntheses.

5.1 Data

Analysis of definitions for data in literature results in our conclusion for vocabulary based on numbers of elementary concepts in existing definitions. The results of the number of elements are shortened in number of e's. The statement that data are representations is substantiated by 82 elements (82e's) identified in found definitions. These representations are recorded (18e's) signals from the real (including virtual) world (76e's). But data has no meaning (29e's) nor structure (23e's). And data are assets (8e's). Notable concepts in found definitions of data are the following for us. Data have intrinsic value and as ore for information, knowledge and wisdom it is recognized as corporate asset (Hovenga and Grain, 2013; Rasouli et al., 2016a; van Helvoirt and Weigand, 2015). Data have no meaning nor structure without metadata, therefore data needs additional ordering to be of value. Metadata are data about data and are needed for turning data into information with the same meaning. Metadata reflects information about information. High quality metadata is a requirement for DG, lack of metadata results in inconsistent data (Khatri and Brown, 2010; Mosley et al., 2009). The inference is that data are signals from the real world which are captured with sensors and recorded by instruments. Real world signals may be physical readings but also verbal answers like survey responses for trust levels or virtual signs like in computer games like scores. Both sensors and instruments are designed with subjective purposes which originate from knowledge needs intentionally collecting data. Therefore, signals are objective but what to record is subjective.

5.2 Information

We conclude that information is formed (71e's) data (58e's), structured according a certain format or logic. Formed data gets meaning (65e's) in human minds by understanding. Information has a goal (27e's) and requires context (13e's) in order to become meaningful. And information is an asset (10e's). Notable concepts of information is when data is structured according a format with knowledge and (subjective) goal, it gets (subjective) meaning for humans and thus value. The meaning of information is discovered by understanding. Differences in understanding by humans of explicit information results in different (subjective) meanings or opinions. So far, information has been qualified as an implicit product only based on meaning given by human understanding (Ackoff, 1989; Baškarada and Koronios, 2013; Bierly III, 2000; Bocij et al., 2008; Liew, 2013; Rowley, 2007). The inference is that information has only value when used, it is difficult to valuate while being subjective. However, information stored in systems may be of great value and is therefore widely recognized as a corporate asset. But information in IS is not only used by humans but also by machines like robots to steer processing. Consequently, when adding goal to explicit data, explicit information can be created and stored in systems.

5.3 Meaning

While analysing the data and information definitions, the concept meaning is found 34 times. Meaning is distinct from data and information because both have

no meaning and therefore meaning demands a definition to itself. We conclude that meaning is understanding (9e's) perceived by humans (8e's) requiring usage (3e's) and context (8e's). Notable concept for meaning is understanding as in cognitive processing, it is typically human and thus subjective. And meaning has usage and is related to context (Baškarada and Koronios, 2013; Huizing, 2007; Liew, 2013; Zins, 2007b). According the DIKW pyramid: the more understanding, the more meaning, the more knowledge, the more wisdom. And knowledge is used to base decisions on, which can result in actions. Therefore, better understanding leads to better actions. And meaning is key between information and knowledge. Meaning governance is unknown so far and deserves further research, but is excluded from this research.

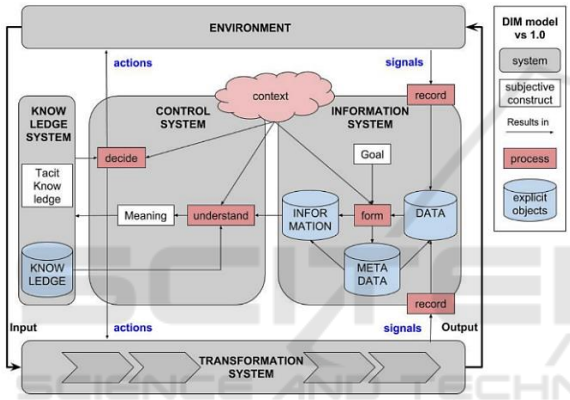

Figure 1: Data-Information-Meaning (DIM) Model.

5.4 DIM Model

We combined the results of our choices in a conceptual model: the Data - Information - Meaning (DIM) model (see figure 1). The information steering model of Bemelmans forms the basis of this conceptual model (Bemelmans, 1982). This model together with the new articulated vocabulary forms our DIM model and is explained as follows. Input from its environment is transformed in transaction systems. Data is recorded from transactions systems and the environment in information systems where information is formed from data with goal and context. Explicit data, metadata and information in (computerized) information systems get subjective meaning by understanding and are influenced in (computerized) control systems. "Information is different from meaning", (Boell, 2017). Meaning created from information with understanding and context can be stored in (computerized) knowledge systems as both explicit and tacit knowledge (Nonaka et al., 1996; Polanyi, 1966; Wognin et al., 2012). Reversely, knowledge can be employed to grow or even create understanding and more and/or deeper meaning. Decisions taken in control systems on basis of meaning in a given context result in actions in the transformation system or systems environment. The DIM model is applicable and scalable to any controlled transformation system just like the steering model of Bemelmans (Bemelmans, 1982).

5.5 Data Governance

Analysis of existing definitions for DG results in our conclusion that DG is establishing (59e's) of management (146e's) in organisations (20e's) assuring aspects as data quality (14e's), access (14e's) and data assets (8e's). Notable concept for DG is the objective of governance is to provide accountability. Governance establishes management, so data governance establishes data management. Quality of data is an important element of data management. Quality as 'fitness for use' mainly concerns accessibility, accuracy, completeness and consistency. Access is based on elements like access (rights), privacy, security. And data is considered an asset being valuable or useful. (Begg and Caira, 2012; Niemi and Laine, 2016; Otto, 2011b; Racz et al., 2010a). Inferences are first that DG is not limited to the legal limits of an organisation. Instead, DG is only limited to the information and transaction systems in the span of control, possibly exceeding the formal organisations boundaries. Second, like data governance establishes data management and information governance establishes information management, X governance establishes X management. And third, because of accountability is X governance not only a matter of IT but also a matter of the business and it needs business alignment. Fourth, DG addresses data quality over its entire lifecycle to safeguard data assets. The data lifecycle includes: definition, production, retention and retirement of data (Alhassan et al., 2016; Khatri and Brown, 2010; Korhonen et al., 2013). So, DG's aim is to manage data.

5.6 Information Governance

In analysis of IG definitions we conclude that IG is establishing (77e's) of management (55e's) in organisations (21e's) assuring aspects as information quality (7e's), access (26e's) and data assets (16e's), but also over its life-cycle (22e's) and in order to be accountable (19e's). Notable concepts for information governance are analogous to DG, IG

establishes information management in organisations. Access is important in the use of information with elements like access security, privacy and control. Life cycle itself is not much mentioned but its discerning stages are. These stages are important when operationalizing information. And as well as for DG, the concept of being accountable is proven. Furthermore, information is widely recognized as a valuable asset. Moreover, quality as 'fitness for use' for users and user groups is also proven (Hagmann, 2013; Kooper et al., 2011; Tallon and Diego, 2013). Inference is first that IG addresses information quality as part of information management over its lifecycle to safeguard information as an asset. Secondly, information value depends on its use. Therefore, IG's aim is to manage information.

5.7 Concepts

Four significant concepts were already identified in theoretical background: governance, establishing, management and organisation. This research identifies four other significant concepts in DG and IG literature: quality, access, asset and life-cycle as elaborated below.

Quality is the first significant concept in both DG and IG (Khatri and Brown, 2010; Otto, 2011a). The notable concept quality is based on the found elements (32e's) accuracy, timeliness, relevance, completeness, credibility and consistency (Korhonen et al., 2013; Weber et al., 2009). Inference is that literature shows relevant quality elements which are valid both for data and for information.

Access to data and information is the second significant concept affecting both data and information (Olaitan et al., 2016). The notable concept access is based on the elements (72e's) access, security, rights, privacy and protect (Ackoff, 1989; Hey, 2004; Loshin, 2009). Inference is that relevant elements of access are valid for data and information.

Asset is the third significant concept in both DG and IG (Khatri and Brown, 2010; Kooper et al., 2011; Otto, 2011c). Notable concept asset is based on the elements (67e's) value, valuable, valuation and asset(s) (Begg and Caira, 2012; Gregory, 2010). Inference is that data and information are valuable or useful for organisations as assets. Similar to data and information, knowledge as related concept in the DIM model is also qualified as valuable asset requiring knowledge management (Helms and Buijsrogge, 2006; Helms et al., 2009).

Life-cycle is the fourth significant concept in both DG and IG (Alhassan et al., 2016; Khatri and Brown, 2010; Korhonen et al., 2013). Notable concepts are define (6e's), acquire (6e's), store (3e's), use (6e's), manage (3e's), archive (3e's) and delete (6e's). Inference is that data are ore for the creation of information. Therefore, data and DG are distinguished from information and IG. However, both follow the same life cycle requiring similar management and similar governance activities. And the life cycles of data and information have similar production processes. E.g. both life cycles may be related according the causality as in figure 2. Knowledge identifies information which defines the required data, although data is multi-interpretational (Dahlberg and Nokkala, 2015; Mosley et al., 2009). Acquiring required data becomes acquired information by predefined formatting with metadata (van Helvoirt and Weigand, 2015). Thereafter, acquired information can only be stored electronically as binary data. When stored, the data can be used multiple times as information (Boell, 2017; Pearlson and Saunders, 2013). And a lot of information asks to be managed, but tools only handle data e.g. archiving. To decide what to archive or delete needs informational insights (Begg and Caira, 2012). And deletion is only effected by electronical removal of data. Alternative relationships are possible but a strong relation is evident. Therefore, DG and IG can share the same management roles, tasks, accountabilities and other efforts. Synthesis leads to the following life cycle both for data and for information: *define, acquire, store, use, manage, archive and delete.*

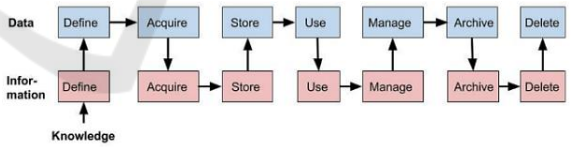

Figure 2: Related data and information life cycles.

5.8 Coherent Set of Definitions

Synthesis of the concepts and formulation of definitions results in a coherent set of definitions as given in table 1. We start with the definitions for data, information and meaning as part of the DIM model. In the processing chain, data is the ore for forming information. Recorded signals are being formed intro information by adding goal. This information gets meaning by understanding the usage and in context. And meaningful information is the ground for taking decisions, which are followed by actions in either the information systems or the real world. Meaning can

also be stored as knowledge in order to be used for decisions or to gather or create new information or data. In literature we found earlier that governance is establishing management within organisations to be accountable. DG's aim is to manage data only, IG's aim is to manage information and it is usage oriented. In DG and IG definitions we found important concepts valid for both DG and IG: quality, access, assets, life-cycle. And both life-cycles are intertwined. DG focusses on data management, the scope of DG is limited to data. Likewise, IG focusses on information management with information as scope. This vocabulary facilitates defining DG and IG in relation with data and information.

Table 1: Set of definitions for data, information, meaning, data governance and information governance.

Data *are recorded representations of signals from the real world*	***Information*** *is data formed with a goal*
Meaning *is human understanding having usage and context*	
Data governance is *establishing management of data in an organisation assuring quality and access during its life-cycle to be accountable for data assets*	***Information governance is*** *establishing management of information in an organisation assuring quality and access during its life-cycle to be accountable for information assets*

5.9 Design Evaluation

Evaluating the design of the set of definitions against the design criteria and requirements results in the following assessment outcomes. The basic criteria of 'good theory' are assessed as follows. Conceptual definitions; the set of definitions is abstracted from elements of existing definitions. Domain limitations; the set of definitions delimits the concepts against other related concepts. Relationship-building; the set of definitions shows the coherency between the concepts. Predictions; the definitions framework answers the predictive questions why, what, where, who and how.

Assessment with design requirements give these results. Clear; definitions and models are unambiguous, without doubt, precise, exact and/or accurate for higher internal validity. Conciseness; definitions and models have brief descriptions that contribute to compact, understandable explanations for higher reliability. Parsimony; the definitions formulations are highly economized and are not redundant. Elementary; ideas are broken up in elementary concepts to find the basic elements for constructing vocabulary for higher internal validity. Utility; concepts, categories and definitions are made practicable in order to use it in practice or for further research. We believe to have met the design requirements from literature sufficiently, although further improvement might be possible with testing in practice.

5.10 Discussion: is DG IG

The outcomes of analyses and syntheses were discussed with peers in iterative discussions (*step m)* resulting in the following discussion. Earliest research mentioning IG was published in 2002, earliest DG articles appeared in 2007 (Cheong and Chang, 2007; Peterson, 2002; Wende, 2007). At first, DG was not distinguished from IG in literature but later on scientist disagreed because of the following reasons. One, information is different from data when subjectively understood by the different business stakeholder, needing goal and context to make sense, but that does not matter for IT which deals with objectively stored facts (Kooper et al., 2009; Wende, 2007). Two, also in DIKW information is often seen in between data and knowledge as formed data or representations of knowledge, having subjective meaning and depending on sociocultural context. Three, lately information is also recognized as objective, observer independent signals or as explicit physical entities like structure, both relevant for resp. IT and IS research (Boell, 2017). Four, information was not seen as an asset before but only as a product when receiving meaning. Now that we found that information can be stored explicitly without meaning it is worth finding out what is the value of information as an asset like data being an asset.

We regard DG as the same as IG like some others do for the following reasons. One, the underlying concept information is formed out of data and both can explicitly be stored in systems. Two, both matters are seen as valuable and can be classified as assets and objects of accounting, an objective of governance. Three, both lifecycle processes are similar and share the same management activities. Four, the governance of both objects is similar: establishing management in an organisation in order to be accountable. Besides, there is no research indicating that DG of data in practice is different from IG of information or from X Governance of X.

6 CONCLUSIONS & RECOMMENDATIONS

The objective of this research is to explore elements of existing definitions of data, information and data governance and information governance, based on literature about Data-Information-Knowledge-Wisdom and data governance and information governance in the IS domain in order to articulate vocabulary and formulate a set of definitions for data and information as underlying concepts of DG and IG within and across organisations.

For method we applied an approach containing (a) a literature review method grounded on accepted LR methods, (b) an analysis of concepts in definitions, (c) a synthesis of coded concepts based on common methods for concept analysis, (d) discussions amongst peers.

The results of this research are: (a) an exhaustive list of definitions for data and information and descriptions of DG and IG from relevant IS literature of the last decade, (b) new vocabulary articulated in analysis of existing definitions from literature and (c) a coherent set of related definitions for better managing data, information with DG and IG. We assessed the articulated vocabulary and formulated a set of definitions against design requirements from literature and believe to have met all. Research limitations are that although this research is grounded, goal oriented and uses an approach based on multiple other accepted methods, it is limited to literature found and formulating definitions only aims towards a related set of concepts. We suggest to consider DG and IG the same because both impose the same, although the underlying matters differ.

Future research is recommended for validating the findings of this literature research in practice. Practical implication is that officers responsible for managing data and/or information can use the vocabulary and set of definitions to better govern data or information by improved management. This assists data and/or information governance officers in creating order in the chaos of ever emerging amounts of data and information.

REFERENCES

Ackoff, R.L., 1989. From data to wisdom. J. Appl. Syst. Anal. 16, 3–9.

Al-Khouri, A.M., 2012. Data Ownership: Who Owns 'My Data'? IJMIT 2, 1–8.

Alhassan, I., Sammon, D., Daly, M., 2016. Data governance activities: an analysis of the literature. J. Decis. Syst. 25, 64–75.

ARMA International, 2013. Information Governance Maturity Model 1–8.

Aven, T., 2013. A conceptual framework for linking risk and the elements of the data–information–knowledge–wisdom (DIKW) hierarchy. RE&SS 111, 30–36.

Bandara, W., Miskon, S., Fielt, E., 2011. A Systematic, Tool-Supported Method for Conducting Literature Reviews in IS. ISJ 1–14.

Baškarada, S., Koronios, A., 2013. Data, information, knowledge, wisdom (DIKW): A semiotic theoretical and empirical exploration of the hierarchy and its quality dimension. AJIS 18, 5–24.

Bawden, D., Robinson, L., 2015. Information and the gaining of understanding. JIS 42.

Begg, C., Caira, T., 2012. Exploring the SME Quandary: data governance in practise in the small to medium-sized enterprise sector. In: EJISE. pp. 3–13.

Beijer, P., Kooper, M., 2010. Information governance: beyond risk and compliance. In: Proceedings of 6th ECMLG. Academic Publishing Inter, Wroklaw: College of Management Edukacja, pp. 10–34.

Bemelmans, prof. dr. T.M.A., 1982. Bestuurlijke informatiesystemen en automatisering. Stenfert Kroese, Leiden.

Bhansali, N., 2013. Data governance : creating value from information assets. CRC Press.

Bierly III, P.E., 2000. Organizational learning, knowledge and wisdom. J. Organ. Chang. Manag. 13, 595–618.

Bocij, P., Greasley, A., Hickie, S., 2008. Business information systems : technology, development and management. FT Prentice Hall.

Boell, S.K., 2017. Information: Fundamental positions and their implications for information systems research, education and practice. Inf. Organ. 27, 1–16.

Boisot, M., Canals, A., 2004. Data , information and knowledge : have we got it right? - annotated. J. Evol. Econ. 14, 43–67.

Brous, P., Janssen, M., Vilminko-Heikkinen, R., Herder, P., 2016. 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. IFIP 2016 9844, 573–583.

Carley, K., 1993. Coding Choices for Textual Analysis: A Comparison of Content Analysis and Map Analysis. Sociol. Methodol. 23, 75.

Cheong, L.K., Chang, V., 2007. The Need for Data Governance : A Case Study. ACIS 2007 Proc. 18, 999–1008.

Corbin, J.M., Strauss, A., 1990. Grounded theory research: Procedures, canons, and evaluative criteria. Qual. Sociol. 13, 3–21.

COSO, 2014. Governance and Operational Performance: Improving Organizational Performance and Governance. Committe Spons. Organ. Treadw. Comm. 32.

Dahlberg, T., Nokkala, T., 2015. A Framework for the Corporate Governance of Data - Theoretical Background and Empirical Evidence. Business, Manag. {&} Educ. 13, 25–45.

De Haes, S., Van Grembergen, W., 2004. IT Governance and its Mechanisms. Inf. Syst. Control J. 1, 27–33.

De Haes, S., Van Grembergen, W., Debreceny, R.S., 2013. COBIT 5 and Enterprise Governance of Information Technology: Building Blocks and Research Opportunities. JIS 27, 307–324.

Dybå, T., Dingsøyr, T., 2008. Empirical studies of agile software development: A systematic review. Inf. Softw. Technol.

Elliot, T.S., 1934. The Rock. London.

Finney, S., Corbett, M., 2007. ERP implementation: a compilation and analysis of critical success factors. BPM J. 13, 329–347.

Frické, Martin, 2007. The Knowledge Pyramid: A Critique of the DIKW Hierarchy. JIS 35, 1–13.

Gregory, A., 2010. Data governance — Protecting and unleashing the value of your customer data assets Stage 1. JDDMP 12, 230–248.

Hagmann, J., 2013. Information governance – beyond the buzz. Rec. Manag. J. 23.

Helms, R., Buijsrogge, K., 2006. Application of knowledge network analysis to identify knowledge sharing bottlenecks at an engineering firm. ECIS 1–13.

Helms, R.W., Bosua, R., Ignacio, R., 2009. Impact assessment of knowledge sharing bottlenecks: the Knowledge Sharing Environment Model (KSEM). Proc. ACIS 132–143.

Hey, J., 2004. The data, information, knowledge, wisdom chain: the metaphorical link. Intergovernmental Oceanographic Commission., 2004.

Hovenga, E.J.S., Grain, H., 2013. Health Data and Data Governance. Stud. Health Technol. Inform. 193, 67–92.

Huizing, A., 2007. The Value of a Rose: Rising Above Objectivism and Subjectivism. Inf. Manag. Setting Scene 91–110.

Jang, K., Kim, W.-J., 2016. Trend of Research in Data Governance Using Time Series Analysis. In: Advanced Science and Technology Letters Vol.139.

Jennex, M.E., 2009. Re-visiting the knowledge pyramid. Proceeedings 42nd Annu. HICSS.

Khatri, V., Brown, C. V., 2010. Designing data governance. Commun. ACM 53.

Kitchenham, B., Charters, S., 2007. Guidelines for performing Systematic Literature Reviews in Software Engineering. Engineering 2, 1051.

Kooper, M., Maes, R., Roos Lindgreen, E., 2009. Information Governance : In search of the Forgotten Grail. Management 1–22.

Kooper, M.N., Maes, R., Lindgreen, R., 2011. On the governance of information: introducing a new concept of governance to support the management of information. IJIM 31, 195–200.

Korhonen, J.J., Melleri, I., Hiekkanen, K., Helenius, M., 2013. Designing Data Governance Structure: An Organizational Perspective. GSTF J. Comput. 2, 11–17.

Kravets, J., Zimmermann, K., 2012. Inter-organizational information alignment: A conceptual model of structure and governance for cooperations. 18th AMCIS 6, 1127–1136.

Krippendorff, K., 1989. Content Analysis. Int. Encycl. Commun. 6.

Lajara, T.T., Maçada, a C.G., 2013. Information governance framework: The defense manufacturing case study. In: 19th AMCIS. Chicago, Illinois, pp. 1984–1993.

Laudon, K.C., Laudon, J.P., 2013. Management Information Systems, 13th ed. Pearson College, Harlow.

Liaw, S.-T., Pearce, C., Liyanage, H., Liaw, G.S.S., de Lusignan, S., Lusignan de, S., 2014. An integrated organisation-wide data quality management and information governance framework: theoretical underpinnings. Inform. Prim. Care 21, 199–206.

Liew, A., 2013. DIKIW : Data , Information , Knowledge , Intelligence , Wisdom and their Interrelationships. Bus. Manag. Dyn. 2, 49–62.

Lomas, E., 2010. Information governance: information security and access within a UK context. Rec. Manag. J. 20, 182–198.

Loshin, D., 2009. Master data management. Elsevier/Morgan Kaufmann.

Luftman, J., 2003. Assessing IT business alignment. Inf. Syst. Manag.

Miller, G.A., 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol. Rev. 63, 81–97.

Moore, G.E., 1965. Creaming more components onto integrated circuits. Electronics 38, 114–117.

Mosley, M., Brackett, M., Earley, S., Henderson, D., 2009. The DAMA Guide to The Data Management Body of Knowledge, LLC Post.

Murphy, G., Chang, A., 2009. A capability maturity model for data acquisition and utilisation. In: ICOMS Proceedings. Sydney, p. 8.

Neff, A.A., Schosser, M., Zelt, S., Uebernickel, F., Brenner, W., 2013. Explicating Performance Impacts of IT Governance and Data Governance in Multi-Business Organisations. In: 24th ACIS. Melbourne.

Newell, S., Robertson, M., Scarbrough, H., Swan, J., 2009. Managing Knowledge Work and Innovation, 2nd ed. Palgrave Macmillan.

Niemi, E., Laine, S., 2016. Designing Information Governance with a Focus on Competence Management in a Knowledge- Intensive Project Organization. In: 21th ICIQ. Ciudad Real, Spain.

Nonaka, I., Umemoto, K., Senoo, D., 1996. From information processing to knowledge creation: A paradigm shift in business management. Technol. Soc. 18, 203–218.

Olaitan, O., Herselman, M., Wayi, N., 2016. Taxonomy of literature to justify data governance as a prerequisite for information governance. In: Proceedings of 28th SAIMS. Pretoria, South Africa, pp. 586–605.

Otto, B., 2011a. Data governance. BISE 3.

Otto, B., 2011b. A Morphology of the Organisation of Data Governance. In: ECIS 2011. Helsinki, Finland, p. 272.

Otto, B., 2011c. Organizing Data Governance: findings from the telecommunications industry and consequences for large service providers. Commun. AIS 29, 45–66.

Otto, B., 2013. On the Evolution of Data Governance in Firms: The Case of Johnson & Johnson Consumer Products North America. In: Shazia Sadiq (Ed.), Handbook of Data Quality. Springer-Verlag Berlin, pp. 93–118.

Panian, Z., 2010. Some Practical Experiences in Data Governance. WASEC 939–946.

Paré, G., Trudel, M.-C., Jaana, M., Kitsiou, S., 2015. Synthesizing information systems knowledge: A typology of literature reviews. Inf. Manag. 52, 183–199.

Pearlson, K.E., Saunders, C.S., 2013. Managing and Using Information Systems, 5th ed. Wiley & Sons.

Peterson, R.R., 2002. Information Governance, PhD Thesis. Tilburg University.

Pijpers, G.G.M., 2006. Information Identified. In: Information Usage Behavior. Boom Publishers, Amsterdam, p. 53.

Plotkin, D., 2013. Data stewardship : an actionable guide to effective data management and data governance. Newnes.

Polanyi, M., 1966. The Tacit Dimensions, 1st ed. Doubleday & Co, inc, New York.

Racz, N., Panitz, J.C., Amberg, M., Weippl, E., Seufert, A., 2010a. Governance, Risk & Compliance (GRC) Status Quo and Software Use : Results from a Survey among Large Enterprises. In: ACIS. Brisbane, p. 11.

Racz, N., Weippl, E., Seufert, A., 2010b. A process model for integrated IT governance, risk, and compliance management. DB&IS'10 155–170.

Rainer, R.K., Turban, E., Potter, R.E., 2009. Introduction to information systems. John Wiley & Sons (Asia) Pte Ltd.

Rasouli, Eshuis, R., Trienekens, J.J.M., Kusters, R.J., 2016a. Information governance in service-oriented business networking, IFIP. Eindhoven.

Rasouli, M.R., Eshuis, R., Grefen, P.W.P.J., 2017. Information Governance in Dynamic Networked Business Process Management. IJCIS 26, 37.

Rasouli, Trienekens, J.J.M.M., Kusters, R.J., Grefen, P.W.P.J., 2016b. Information governance requirements in dynamic business networking. Ind. Manag. Data Syst. 116, 1356.

Rifaie, M., Alhajj, R., Ridley, M., 2009. Data governance strategy: A key issue in building enterprise data warehouse. iiWAS2009 587–591.

Rosenbaum, S., 2010. Data governance and stewardship: Designing data stewardship entities and advancing data access. Health Serv. Res. 45, 1442.

Rowley, J., 2006. Where is the wisdom that we have lost in knowledge? JOD 62, 251–270.

Rowley, J., 2007. The wisdom hierarchy: representations of the DIKW hierarchy. J. Inf. Sci. 33, 163–180.

Smallwood, R.F., 2014. Information governance: concepts, strategies, and best practices. Wiley & Sons.

Tallon, P.P., Diego, S., 2013. Corporate Governance of Big Data: Perspectives on Value, Risk, and Cost. Computer (Long. Beach. Calif). 46, 32–38.

Tallon, P.P., Ramirez, R. V., Short, J.E., 2013. The Information Artifact in IT Governance: Toward a Theory of Information Governance. JMIS 30, 141.

Tarantino, A., Cernauskas, D., 2011. Essentials of risk management in finance. John Wiley & Sons.

Tranfield, D., Denyer, D., Smart, P., 2003. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. Br. J. Manag. 14, 207–222.

van Helvoirt, S., Weigand, H., 2015. Operationalizing Data Governance via Multi-level Metadata Management. In: Janssen, M., Mäntymäki, M., Hidders, J., Klievink, B., Lamersdorf, W., van Loenen, B., Zuiderwijk, A. (Eds.), Open and Big Data Management and Innovation. Springer International Publishing, pp. 160–172.

vom Brocke, J., Rosemann, M., 2010. Handbook on Business Process Management 2.

Wacker, J.G., 1998. A definition of theory: research guidelines for different theory-building research methods in operations management. J. Oper. Manag. 16, 361–385.

Weber, K., Cheong, L., Otto, B., Chang, V., 2008. Organising Accountabilities for Data Quality Management-A Data Governance Case Study. Data Warehous. 347–362.

Weber, K., Otto, B., Osterle, H., Schneider, B., McManus, J., Kooper, M., Maes, R., Lindgreen, E.R., Iannarelli, J.G., O'Shaughnessy, M., Fortes, V.B., Boff, S.O., 2009. One Size Does Not Fit All — A Contingency Approach to Data Governance. ACM J. Data Inf. Qual. 1, 27.

Webster, J., Watson, R.T., 2002. Analyzing the past to prepare for the future: Writing a literature review. MIS Q. 26, 13--23.

Weill, P., Ross, J.W., 2004. How Top Performers Manage IT Decisions Rights for Superior Results. IT Gov. 1–24.

Wende, K., 2007. A Model for Data Governance – Organising Accountabilities for Data Quality Management. ACIS 2007 Proc. 80.

Williams, P.A.H., 2011. Information governance: a model for security in medical practice. J. Digit. Forensics, Secur. Law 2, 57–73.

Wognin, R., Henri, F., Marino, O., 2012. Data, Information, Knowledge, Wisdom: A Revised Model for Agents-Based Knowledge Management Systems. In: The Next Generation of Distance Education: Unconstrained Learning. Springer US, Boston, MA, pp. 181–189.

Zielstorff, R., 1998. Characteristics of a Good Nursing Nomenclature From an Informatics Perspective. Online J. Issues Nurs. 3, 6.

Zins, C., 2007a. Conceptual approaches for defining data, information, and knowledge. JASIST 58, 479.

Zins, C., 2007b. Conceptions of information science. JASIST 58, 335–350.