DIGITAL DATA: STEPPING-STONES FOR DIGITAL INNOVATION!

By

Christian Ringstad Schultz, PhD Candidate, University of Oslo, Chrirsc@ifi.uio.no
Lene Pettersen, Associate Professor, University of Oslo, Lenpe@ifi.uio.no

Abstract

Digital data is argued to be a part of "the new oil" and a key driver for digital innovation, digitalization, business development, and the data-driven economy, but the term "digital data" is not well defined in the literature and numerous definitions exist. Without a more unified understanding of "digital data", its assumed potential may be difficult to unlock. This study conducted a comprehensive review of the data construct in the information systems (IS) literature from a meso-perspective and gained insights from a qualitative case study of a platform as a service (PaaS) company. Building on these insights, we developed a conceptual model we labeled the RIQA framework to illustrate fundamental characteristics of "digital data" that were (1) considered to be important for future definitions of digital data, and (2) needed for the enablement of digital data. These insights also led to one of our main discoveries: the importance of data accessibility in enabling digital data to be effectively utilized, a consideration of which was lacking in the literature. We contributed to the IS literature by examining, in-depth, how studies have approached digital data and organizations have used it to deliver data services to clients. We also contributed by proposing a conceptual framework to facilitate a better understanding of what constitutes digital data.

Keywords: digital data; data economy; information systems; literature review; case study

1. Introduction

"The new oil" is an expression used in Norway to suggest something that can be capitalized on when Norway, as a country, run out of oil—the main source of income. Data is often considered to be part of this "new oil" and, thus, a key driver for digital innovation, digitalization, business development, and a data-driven economy (Lukic, 2014). For example, by the end of 2020, the Norwegian government will present its policy on a data-driven economy and innovation. A data-driven economy is a digital ecosystem for which data is a strategic asset that is gathered, organized, and exchanged by a network of actors for the purpose of deriving value from the asset (Opher, Chou, Onda, & Sounderrajan, 2016). The accumulated information—large datasets—is argued to be leading to a new *data economy*, in which big data infrastructures, products, and services will generate massive economic growth (Flyverbom & Madsen, 2015). *Digital data*, as a component of data-driven business models, is important in a data-driven economy; however, most studies have pointed out what digital data can do (its potential) and focused less on what digital data is (the key characteristics of digital data). When approaching the literature, we found that the concept of digital data and its explanation were not well defined. Several definitions are used interchangeably in the information systems (IS) community and between different academic silos. If data is a driving force for digital innovation, we need a unified understanding of what digital data is, how it is used, and the importance of data accessibility. This might assist organizations in unlocking the assumed potential that data represents and increase the amount of data shared between organizations and companies. More specifically, we addressed the following research question (RQ) in this study:

- Which of digital data's key characteristics are needed in order to exploit its potential for digital innovation?

In order to answer the RQ, we first reviewed the IS literature, and literature from related disciplines (e.g., computer science, algorithms, strategy, etc.), to explore definitions of digital data from an organizational, meso-perspective. This was followed by an analysis of how these definitions corresponded with the definitions of employees in, and clients of, a platform as a service (PaaS) company. The company offers a data integration platform that collects, connects, and shares digital data among its customers (public and private). Their specialist expertise with digital data shed light on how digital data was defined by people working directly with data.

The paper is organized as follows: First we present a review of how the data construct was defined in the IS literature, then describe a qualitative case-study that provided insights from the PaaS-company. This is followed by our findings on the extent to which the literature matched the practitioners' understandings of the data construct. We close the article with a discussion and present a conceptual model of the attributes or pillars that should be used when defining and understanding digital data".

2. Literature review

As a starting point, digital technology is characterized by (1) being reprogrammable, (2) having a self-referential nature, and (3) its homogenization of data (Yoo, 2010). The latter is the process of bringing data into a common framework to create a structure (Damen, 2015). Nambisan, Lyytinen, Majchrzak, and Song (2017) defined digital innovation as outcomes made possible by digital technologies and digitized processes; hence, digital innovation was viewed through the lens of value creation, going beyond its current value proposition to examine its creation of future value. In order to define digital data and understand the value it represents to organizations; it was vital to investigate the usage of digital data.

Data usage for value-creating purposes can be approached on several different levels. Some contemporary discussions in the IS community have approach "digital data" from a societal and macro perspective, drawing inspiration from media science and directing their attention to the like- or attention-economy (Gerlitz & Helmond, 2013), surveillance capitalism (Zuboff, 2019), algorithmic power (Gillespie, 2014; Bucher, 2018), value creation in ecosystems and platforms (Alaimo & Kallinikos, 2018; Alaimo, Kallinikos, & Valderrama-Venegas, 2020), and the digital data value chain (Alaimo, Kallinikos, & Aaltonen, 2020), among other issues. These studies focused on the creation of digital data and how companies (e.g., Facebook and Spotify) used data from users in their business models, in order to derive value from the accumulated information.

Another approach to data is on a meso organizational level, with digital data deriving from organizations' IT systems, rather than from users or citizens. Although it is well known that data and algorithms are biased by the values and perspectives of those who establish the rules for structuring IT systems (Monteiro & Parmiggiani, 2019; Pettersen, 2018), the goal of this paper was to focus our lens on the fundamental aspects and characteristics of digital data, in order to achieve a clearer understanding of what digital data is, and what attributes data needs to have to facilitate value creation. As a first step in this direction, we approached the literature to explore how digital data was defined and approached. This review revealed that the data construct in the IS literature fell into at least four non-exclusive categories (see 3.1 Literature Review for details of how the review was conducted). The articles focused on digital data from the perspectives of: (1) big data, (2) representations of data, (3) data as value, and (4) data as an enabler.

	Definition	Author	Context
Big Data	Big data is used to refer to data that includes information acquired externally together with data gathered internally.	(Baesens, Bapna, Marsden, Vanthienen,	Paper on how to adopt and innovate applications of data sciences in business.
	The term big data describe the massive volumes of data analyzed and the point is about finding new value in and outside conventional data sources.	(Demirkan, et al., 2015)	Paper on how big data can be important in-service innovation. Data gathered from a debate between five industry leaders.
	Big data can be described as the "new oil" to fuel innovation in the new economy	(Sahay, 2016)	Paper on how big data can be, and is, used to strengthen health systems.
Representation	Data is the raw material for analogical reasoning	(Almklov, Østerlie, &	Paper on how petroleum engineers work with digital sensor data.
	Data is what is stored in a database and processed by an IS.	(Tilly, Posegga, Fischbach, & Schoder,	Literature review on how data and quality insurance can be explained in social infra structures.
	Open data can be defined as "publicly available data structured in a way to be fully accessible and usable".	(Link, et al., 2017)	Paper on how attendees at a Pre-ICIS 2016 workshop (academics) think we should one should embrace open data in research.
	Data can be considered as old.	(Sen, Nelson, & Bumaniam, 2015)	Paper on how an open software can be successful and survive over a longer period of time.
	Data are carriers of facts, sign tokens of actual or potential meanings, media or lenses through which people construct and share the realities they confront.	(Alaimo, Kallinikos, & Aaltonen, 2020)	Paper on data and value and this is their try on a more generic definition of data.
	Data is a spatial representation organized from an information flow.	(Sanders, 2016)	Paper on how data can be viewed and how we can interpret terms in order to preserve a model of adaptive behavior.
	High-quality data can be defined as data that is fit for use by data consumers—a widely adopted criteria.	(Strong, Lee, & Wang, 1997)	Paper on how an organization can use data to decrease the cost by working towards better data.
	Data is a generic term and is equivalent to information	(Wald, 2002)	Paper on clarifying the concept of data fusion and information fusion to provide a better understanding for both terms.
	Data is a rhetorical concept. Data means — and has meant for a very long time — that which is given prior to argument.	(Rosenberg, 2013)	Paper on defining data and what it is, but from a philosophical stand.
Value	Data is the foundation of information systems and data collection is a crucial and costly activity	(Liang, 2010)	Introduction by the chief-editor to a journal focusing on innovation on RFID-
	Data is more than oil of the future economy.	(Demchenko & Los, 2018)	Mixed literature review focusing on the value of data and how it has evolved during the last decade.
Enabler	Data is a primary business asset and in organizational settings, the information technology (IT) function is tasked with managing and integrating data as an "enabler" of data-driven business processes and decision making	(Abbasi, Sarker, & Chiang, 2016)	Paper presenting various perspectives on promising big data research topics and highlight some of the challenges that big data poses.
	Physiological data is, in most cased, used for offline processing, analyzed at a later time.	(Sherer, 2014)	Paper focusing on healthcare value and tries to identify gaps in patient-centered e-health applications.
	Digital data can enable users to participate economically while acting favorably from a collective perspective	(Hildebrandt, Hanelt, & Firk, 2018)	Paper investigating smart cities and how connectivity and digital data can enable users to participate economically.
	Data generated during underwriting process are then made available to securitizes in the secondary markets who use them to make decisions on which loans to select for securitization (i.e., pooling)	(Kaniadakis & Constantinides, 2014)	Paper on the role of financial information infrastructure innovation as a process and how it enabled the transition to mortgage securitization for UK banks.
	Data may allow a company (or partnering companies) to create a personalized experience.	(Wyatt & Piggott, 2019)	Paper with examples on how emerging technologies and experiences challenge previous theories in discussion on new data ecosystems and interaction models.
	Data can be harnessed and transformed into actionable knowledge.	(Perdana, Robb, & Rohde, 2019)	Paper on interactive data and information visualization and how it enhances information presentations by providing users with multiple visual representations, active controls, and analytics.

Table 1. Literature review with definitions, authors and context

2.1 Big data

Articles that framed the data construct from the perspective of big data tended to define it from value-driven theoretical perspectives. They considered big data's value potential (Sahay, 2016) and how big data may be used to create value. Big data was recognized in the literature as massive amounts of data (Baesens, Bapna, Marsden, Vanthienen, & Zhao, 2014) that can be used to create value (Demirkan, et al., 2015).

2.2. Representation

Another batch of articles approached the construct of digital data from the perspective of "representation", because theoretical approaches to data representation vary from philosophical approaches (Rosenberg, 2013) to more technical approaches (Tilly, Posegga, Fischbach, & Schoder, 2017). What united these articles was that they provided context to data so it could be described as information. In this batch of articles, "data characteristics, such as "old" (Sen, Nelson, & Bumaniam, 2015), "raw material" (Almklov, Østerlie, & Haavik, 2014) or "carrier of facts" (Alaimo et al., 2020), were identified.

2.3. Value

A third category of definitions of digital data revealed by our review took an economic theoretical perspective on data. Studies approaching data as value ranged from seeing data as cost-reductive (Abbasi, Sarker, & Chiang, 2016) to perceiving it as an economic enabler (Hildebrandt, Hanelt, & Firk, 2018). The literature shared an understanding of data as potentially being of great value (Sahay, 2016; Liang, 2010; Demchenko & Los, 2018). However, in order to create value from data, an organization needs to collect it, which can be costly (Liang, 2010). The ultimate end goal of data-usage was finding new value outside conventional data sources (Demirkan, et al., 2015)

2.4. Enabler

The last category of definitions revealed in the literature understood data as something that can enable strategic activities; for example, organizational transformation (Perdana, Robb, & Rohde, 2019), user participation (Hildebrandt et al., 2018), or decision-making (Abbasi et al., 2016; Kaniadakis & Constantinides, 2014). The enablement approaches varied, but their theorizations were generally efforts to create value from data. This was also supported by the more practical approach the papers took regarding cases and how to reach customers, partners, and other stakeholders.

In summary, our review of the literature found that definitions of the digital data construct in the IS literature mainly approached data from a value-creating perspective, paying less attention to what data is and the different attributes data needs to have in order to represent or create value. The exceptions were articles in the "representation" category, with scholars and work providing insight into what needs to be in place for digital data to be usable by organizations, which provided a fundamental reason for representation to be highlighted as an individual category. The three other categories focused primarily on value creation but offered guidance on the expected outcomes of data. Moreover, most of the literature concerned case studies of organizations, which accorded with our meso perspective on digital data.

3. Methodology

The methods chosen for this study were (1) a systematic review of the literature and (2) a case study of a PaaS company.

3.1. Literature review

The literature review was conducted in June 2020. It was initiated by a search to determine how digital data was defined and identify the attributes used to define it. The goal was to obtain a broader view of the IS literature in order to create a structure for the literature review. We used the AIS Electronic Library and employed search criteria as follows: (1) "data" and "information systems" included in the *Abstract*, (2) "definition" OR "define" included in *All Fields*, and (3) "data center" excluded in the *Abstract,* for (4) peer-reviewed articles only and (5) sorted by relevance according to AIS Electronic Library search hits. This led to 226 results and we selected the 100 most relevant articles. Through analysis and coding of how data was described in the papers, a pattern emerged, which described data in terms of "big data", "representation of data", "value of data", and "data as an enabler".

	Big data	Representation of data	Value of data	Data as an Enabler
Number of articles	**9**	**12**	**13**	**16**

Table 2. Number of articles according to the categories revealed by the literature review. N = 50.

Of the 100 articles. 50 articles aligned with the emerging categories of definitions, as shown in Table 2. The 50 that were discarded lacked relevance, since they focused on cases (e.g., "Overview of Data Security Issues In Hospital Information Systems" by Masrom and Rahimly [2015] or "Motion Picture Industry Pension Plan" by Sankaran and Wedel [2020]), on data collection (e.g., "Data Model Development for Fire Related Extreme Events" by Chen, Sharman, Rao, and Upadhyaya [2013]), or on the risks of using data (e.g., "Managing Security Requirements" by Ullah and Lai [2011]).

3.2. Data analysis

The four categories of definitions were applied to the main search. We selected general papers we already knew of, which discussed definitions of data or similar. In addition, we conducted a new search of the AIS Electronic Library, using search criteria as follows: (1) '"data" included in the *Abstract*, (2) at least one of the four categories of definitions included in the *Abstract*, (3) "definition" or "define" in *All Fields,* (4) date range from 2010 to 2020, for (5) peer-reviewed journals only. This led to 92 results. Google Scholar was searched to provide a broader scope, using the following search criteria: (1) "defining data" or "data definitions" included in the abstract, (2) date range set to "none", (3) all texts included, and (4) texts not necessarily peer-reviewed. This search identified no further papers, because either (1) papers had already been found or (2) papers did not fit our scope. The review process is illustrated in Figure 1.

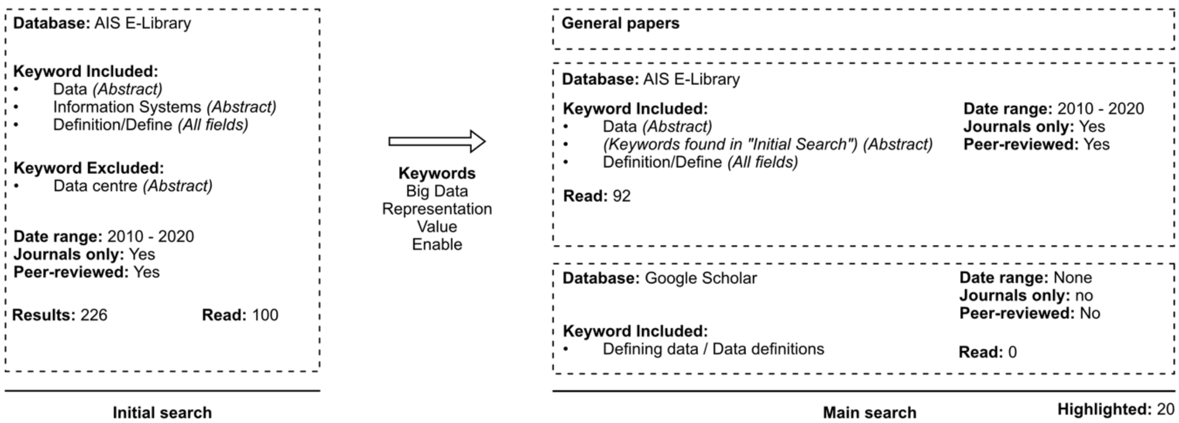

Figure 1. The literature review process

After reading the 92 articles from the main search, 20 papers were selected, as highlighted in Table 1. These 20 articles provided clear definitions of data in light of one, or more, of the four non-exclusive categories of definitions. The other 72 papers offered broader perspectives on digital data and its implications in relation to "big data", "representation", "value", and "enabler".

3.3. Case study

To analyze how digital data was approached and understood in the literature, we conducted a case study of a platform as a service (PaaS) company in November and December 2019. Because we wanted to gain in-depth insights into how employees and clients understood the data construct, an exploratory case study and qualitative approach were chosen (Yin, 2012). An exploratory case study is both descriptive and inductive and is applicable when knowledge and theory about a given field are lacking (Yin, 2012). The research method involved the collection of personal and confidential information and was approved by the Privacy Issues Unit of the Norwegian Social Science Data Service (NSD Project No. anonymized). This approval ensured that the collection, protection, storage, and reuse of personal data complied with ethical standards and legal requirements.

3.4. Choice of company

The organization chosen for studying the data construct from a qualitative perspective is a Norwegian company that, in this paper, is anonymized as Magic (M). M provides a data integration platform for its customers, which was originally developed in the late 2000s by the parent company (a Norwegian IT consulting company). The platform has undergone several major changes and is currently in its third version. The platform represents a new way of thinking about data, since it integrates an organization's data, rather than its IT systems. It collects raw data from source systems and stores it in a log-based data hub. Simple extension points are provided to allow developers to connect systems that lack out-of-the-box adaptors in order to create value. This is in line with the digital innovation lens provided by Nambisan and colleagues (2017), who stated that, when *problem–solution pairs* are identified, there is a need for methodologies that focus on matching specific conditions to specific outcomes, which this paper addresses.

M plays an important role in assisting several companies with data integration and was thus a strong candidate for providing practical insights about digital data. Furthermore, M has a wide variety of Norwegian and foreign employees with different disciplinary profiles (e.g., statisticians, mathematicians, sales representatives, data analysts, and lawyers), which ensured diverse views on the data construct. M's characteristics helped us to achieve in-depth insight into how 'digital data' is perceived and defined in practice; hence, the inclusion of clients and partners, as well as employees, in this study.

3.5. Ethnographic field work and unstructured interviews

Ethnography is the close study of groups' and people's everyday lives in their social settings (Emerson, Fretz, & Shaw, 2011) and typically involves the development of close connections between the ethnographer and the subject or situations being studied (Hammersley & Atkinson, 1995). An anthropologist's most important research tool is fieldwork (Eriksen, 2001). The second author of this paper is an anthropologist and conducted an ethnographic field study in M in November and December 2019. The discipline of the second author was important for determining the value of the identified theoretical perspectives (Nambisan et al., 2017). An interdisciplinary effort was likely to enhance the determination of the value of the theoretical perspectives identified, leading to higher research quality. The company informed its employees about the second author and her research agenda. During the fieldwork, the second author spent her days in the workplace, joining meetings, workshops, and discussions with employees. This researcher took on a learning role in addition to using her background as a consultant—a field access strategy recommended by (De Jong, Kamsteeg, & Ybema, 2013).

Ten out of sixteen informants were, during the in-depth and open-ended interviews, specifically asked to explain how they understood the concept of data, in order to understand how working with data affected their working lives. The informants had different roles, seniorities (six had worked within M for a number of years), hierarchical ranks, and disciplinary backgrounds (e.g., programming, marketing, sales, customer service, etc.). The interview sample is listed in table 3.

	Who	Role	Organizational location
1	Male, 40s	Consultancy manager*	Sales and consultancy
2	Male, 40s	Consultant*	Sales and consultancy
3	Male, 40s	Managing director*	Management
4	Male, 40s	Head of sales and marketing*	Management
5	Male, 50s	CEO of a small IT-company	M Client**
6	Male, 40s	Programmer*	Product development
7	Male, 30s	R & D architect*	R&D
8	Female, 30s	Lawyer*	Administration
9	Male, 30s	Product manager*	Product development
10	Female, 30s	Partner Manager	Sales and consultancy
11	Male, 40s	Business developer*	Sales and consultancy
12	Male, 30s	Support	Support
13	Male, 50s	Key role in a large multinational energy company (16,000 employees in 20 countries) *	M client**
14	Male, 30s	Sales Manager	Sales and consultancy
15	Male, 50s	Project Manager	M client**
16	Male, 40s	Head of product development	Management

*Table 3. Table of informants from company M (*Asked what data is, **Informant not employed by company M)*

As shown in Table 3, the informants differed in age, work role, and hierarchical level. Only two of the informants were female, which was a lower number than the Norwegian IT industry average of one out of every four employees being women (Kampevoll, 2019).

3.6 Data analysis

The analysis of the qualitative data was conducted in four phases. First, the second author analyzed the collected data, which involved moving back and forth between the interviews and the field notes, including new dimensions along the way. Second, the first author read the transcripts of the interviews and the field diary notes numerous times, searching for themes, taking notes, and developing analytical categories and constructs. Third, the two authors individually reviewed the interviews several times to identify overall themes and findings, coding and analyzing them according to their patterns, similarities, and differences. Having two pairs of analytical eyes examining the data minimized the risk of biased interpretations (Emerson et al., 2011; Gallenga, 2013). This step was followed by several workshops, meetings, and discussions between the two authors in June 2020, during which they analyzed and discussed the different categories revealed by the interview data. Finally, the findings from the literature review and the case study were compared and analyzed in order to look for similarities and differences. This led to the findings which we will now present.

4. Findings

Based on the literature review, we determined that the construct digital data had no common definition in the IS literature. The review revealed four non-exclusive categories of definitions: (1) big data, (2) representations of data, (3) value of data, and (4) data as an enabler. We then investigated how these categories corresponded with the data construct as defined and perceived by employees and clients in company M.

Independently of our literature review, the analysis of the informants' explanations of what data *is* revealed one main element and three cornerstones, or attributes, that were closely interrelated. We labelled the main element **R**epresentation and the three cornerstones **I**nterpretation, **Q**uality, and **A**ccessibility—shortened for the *RIQA* framework. While the first category closely related to what data is (object), the three other categories focused more on data's opportunities or hindrances for exploitation (actions). The following section will elaborate on these categories.

4.1. Representation

Some of the employees referred, during the interviews, to data as a representation or description of an object: "It is a digital representation of what someone thinks are facts about something", one programmer in M explained. This perception was also shared by one of M's clients, who stated: "Data is information that we have chosen a representation for". Data as representation refers to objects or signs that denote something else, in the way Magritte's famous painting of a pipe illustrated that the pipe was not a pipe, but an image or model of a pipe. According to this approach, data can vary from the abstract and conceptual to hard facts. Data as an object or representation can also be explained in terms of traces left behind, or as something being stored for later use, as the CEO of M pointed out:

> It is a digital representation of an object or an incident. It's what we have managed to digitally collect from traces that are left behind, that someone has either recorded or collected and saved a digital copy of. It's difficult to answer actually, and alone it has little value.

Consequently, data needs to have a history and a timeframe to denote that something has occurred. This "something" could be anything, from weather details to employees' sick leave, recorded in an IT system; however, data located in IT systems or silos exists in a closed space and is thus of little use to an organization. All the informants stressed that data has no meaning or value in isolation. They explained that, in order for data's potential to be unlocked, three closely related sub-aspects of the 'representation' category are required: meaning and interpretation, data quality, and data accessibility.

4.2. Interpretation

First of all, data needs to inhabit a form or have a characteristic that provides it with a meaning that can be interpreted; otherwise, data is merely bits and pieces of information that have no meaning in isolation. The analysis found that, in addition to history as mentioned above, time and context were two key aspects necessary for providing data with meaning—and thus transforming data into a meaningful form or state that IT systems can interpret. M's CEO explained how time and place (context) play significant roles in such a "data transformation":

> [Data] is only valuable if it can be dated and related to a context. Data does not represent any value unless it has meaning, and that's why semantics is so important. The goal with semantic technologies is to provide the data with context.

Language and semantics are conceptual tools in the process of transforming data into meaningful entities. Having a shared understanding, or language, is equally important for an organization, as the R&D architect stated:

> If you look at the hard disk at your computer, it's just a string of ones and zeros, and that's data, but it doesn't mean anything … You [need to] have a way of understanding and interpreting it, or it is completely meaningless.

Moreover, with interpretation, data changes its "state" from an object to a subjective interpretation, because people add layers of meaning to the data: humans provide data with meaningful layers. Interpretation is not universal or generic. As the R&D architect explained: "You have to interpret it [the data], and you can interpret it in endless ways"; for example, the concept 'apple' can denote both a fruit and a US company.

Creating rules for what something should mean is therefore important. Having such a shared conceptual framework is vital for making data meaningful outside the context in which it was created, but creating a shared conceptual framework is not straightforward: “Who has ownership to define what the data should mean and represent?”, as one of M’s clients asked. He illustrated this point with an example from his own company and industry:

> If you state “Tørrvekt” in Norwegian, it can sometimes be translated as “dry weight”, but at other times it may mean “net weight”. So what are we actually saying? What’s the meaning of it and how can we understand each other and feel secure enough to share our data? What we’re working on is a definition model that tells us what we’re talking about and what is being represented in the different systems.

M’s clients showed that different names for objects are a consequence of different practices of naming within an organization, posing a risk that an organization’s digital data may not be fully exploited. M’s client also pointed out the difficulties of creating a standard, universal rule for what a certain piece of data or construct should mean and represent. This challenge or dilemma, in the business world and IT industry, typically relates to standardization. Standardization is a framework of agreements, which all relevant parties in an industry or organization must adhere to ensure that all processes are performed within a set of similar guidelines (one kilo or meter is one kilo or meter regardless context and time). However, very complex processes might be difficult to define as referring to one particular meaning, or one best practice or standard, and the creation and definition is therefore also related to the power to define what a certain piece of data should mean.

M’s clients also indicated the importance of having a shared understanding of what something should mean, from a meso-perspective. Thus, creating a conceptual framework to assist in the interpretation of data is not only a technical matter relating to the standardization of meaning, but is also closely related to people and organizational practices. Establishing organizational practices is important for providing digital data with meaning and obtaining the greatest value out of using a data-integration platform like M’s. It is important for building trust and establishing new kinds of collaborative practices and it also requires new ways of thinking and collaboration for individuals and departments in an organization and between organizations; hence, establishing a shared understanding of what digital data should denote and mean relates to both technical aspects and organizational elements, such as practices and power.

4.3. Quality

The category of data concerning interpretation also closely related to data quality. Depending on who you ask, data quality can mean different things. M is a company that interprets digital data on a meso level. For a company, data quality is achieved when the data is fit for its intended use in operations, decision-making, and planning. The data used must satisfy customers (Fleckenstein & Fellows, 2018) and be useable in operations, decision-making, and planning. This is in line with the literature approaching data as value, meaning that it should be useful, accurate, and correct for its application.

Quality typically denotes how well data is connected to its context and how strongly and accurately the data represents an object. M’s CEO used a table to explain this point:

> If you only notice the data, without knowing its surroundings, it won’t mean much until you get context in a system, for example, that table number X1092 in SAP means that X equals 1. Data doesn’t make much sense without being put into context.

Data can be interpreted and transformed into something meaningful in several ways, but quality plays a vital role in this transformation. Data is typically said to be of high quality when it correctly represents the real-world construct to which it refers; however, data quality can be a key barrier for organizations in fulfilling the potential of data to lead to a digital transformation (Haug et al., 2013). If the data is not of good quality, it is irrelevant whether you have large amounts of it. M’s product manager used a picture to illustrate how data quality can provide a big picture and help organizations to make strategic and vital decisions and predictions:

> Data is zeros and ones. Data is a component of information that can show history, against the present time, and use this to predict the future … When obtaining a picture of the history, the more data you have, the more concrete the picture is. Because strong data is fundamental in a decision-making phase … the more and correct data you have, or good data quality as we say, the better the foundation laid for making decisions.

In order to increase data quality, data quantity is important. The more total data we have access to, the better are our chances of obtaining data that fits certain criteria, leading in turn to better chances of executing certain tasks within an application (Fürber, 2016).

4.4. Accessibility

Finally, when they were asked to define data, the informants stressed the importance of data accessibility. Data accessibility refers to data being available for use in two respects: IT-related and organizational. IT-related issues relate to 'digital data' being difficult to extract from IT systems, because it is typically locked up in silos; organizational concerns relate to data and IT systems being controlled by specific units or people in an organization who claim ownership of the data, as stated by the project manager at M: "To be able to access the project data, there are several contracts with the customer stating that 'the only ones to access the data are the ones who can prove that they need to use it'."

Because data represents power, it can be controlled. This typically concerns the attitudes of people or departments in organizations who claim ownership of their data systems. The product manager at M illustrated how some people tried to control the data by sharing only small portions:

> We had a discussion about one of our source systems, about those who deliver data. They deliver only small portions, rather than offering access to the database, where you can get all the data you need and do what you want with it.

Another organizational aspect besides power and control, which some informants mentioned when discussing the data construct, was established practices and organizational routines. Some employees are reluctant to accept new ways of doing things and are blindfolded by established practices and ways of thinking, as client M explained:

> When discussing Application Programming Interfaces (APIs), to serve the data, that's not something they know how to use. They are able to export Portable Document Formats (PDFs) and are happy with that. So that's the challenge: the access and how to export documents … If you are going to create new value and knowledge out of [data], you can't just see it as it has always been defined.

Established practices, routines, and organizational "myopia" may hinder organizations from exploiting their digital data and consequently hinder innovation. Myopia denotes an organization's inability to see new strategic opportunities for the business (Levinthal & March, 1993). Not only established practices and myopia hinder organizations from exploiting their data's potential. Our findings showed that digital data needs to have some key characteristics in order for organizations to exploit its potential. We will now proceed to discuss these characteristics and conclude.

5. Discussion

When returning to our RQ regarding which of digital data's key characteristics are needed in order to exploit its potential for digital innovation, we found in the IS literature that digital data was defined from four different angles: big data, representation, value, and enablement. These understandings of what data is emerged in the interviews with employees and clients of a PaaS company. In addition to the aspects revealed by the literature, when the informants were asked to define digital data, they commented on aspects relating to power and the control of data, as well as to organizational practices and myopia that could hinder innovation. Moreover, the informants provided more details of data's characteristics than were described in the IS 'representation' articles. Such details concerned the importance of "accessibility in order to enable data and create quality outcomes"; "how subjective interpretation of data is, and how easy it is to manipulate its interpretation"; and "how data is viewed differently in an organization".

The different categories revealed in the literature and in the interviews were merged into a conceptual figure we labelled the RIQA framework (Figure 2). The framework consists of one main element (representation) and three cornerstones or attributes (interpretation, quality, and accessibility) that need to be in place when defining digital data for value creating purposes. The three attributes are connected and, consequently, one cannot be removed without affecting the others.

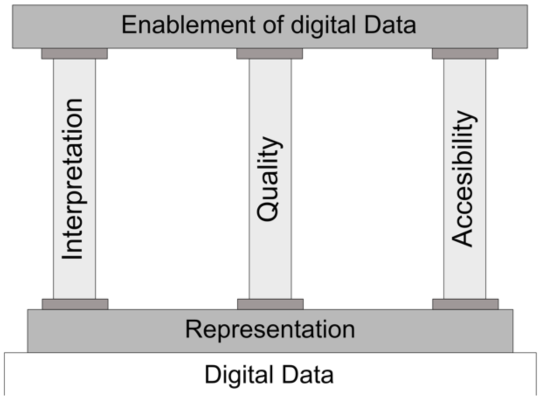

Figure 2: *The RIQA-Framework*

5.1 Representation

Representation is the main element, because representation of digital data means that it needs to be something in order to exist and be represented. Rosenberg (2013) stated that data is a rhetorical concept, which aligned with the informants' statements asserting that data, in isolation, has no meaning. What both the literature and the interview subjects agreed on is that data needs be given meaning in order for it to be enacted (i.e., to be enabled). The IS literature discussing data from a representation perspective did not provide a clear definition of digital data; instead, different characteristics of data were listed (e.g., "raw" or "old"). This was also the case with the informants, who provided a wide variety of definitions. This may have been due to each informant having their own experience of data and, hence, their own interpretation of data, which aligned with Almklov et al., (2014) claim that data is raw material for analogical reasoning.

5.2 Interpretation

For data to be used, it needs to be interpreted, and time and context are essential for accomplishing this, as mentioned by M's CEO. Sen et al., (2015) stated that data can be considered "old", but the literature did not mention time as particularly important for interpretation. The literature focused on the context, which was similar to M's general approach. It is arguable whether time is part of a context, but it is not clear whether time is crucial for interpretation. Context, however, is crucial for meaning to be interpreted. Context is what gives data the potential to be valuable, which both the literature (Demirkan et al., 2015; Hildebrandt et al., 2018; Perdana et al., 2019; Wyatt & Piggott, 2019) and the interview subjects agreed on. Context challenges the distinction between objective and subjective data, as mentioned by the R&D architect at M. These different interpretations can be used for the greater good or misused as a form of power.

5.3 Quality

Context is important when discussing the quality of data, due to data not making much sense without being placed in a context, as mentioned by M's CEO. This context provides the criteria for what constitutes "quality data", defined as data that is fit for use by data consumers (Strong, Lee, & Wang, 1997). M's product manager described "the fit" as data that can help organizations make vital strategic decisions and predictions. The chance to do so will determine whether the data coming from different sources is well-connected and open. Opportunities will be enhanced if the data actually fits the data consumers' criteria (Fürber, 2016).

5.4 Accessibility

When discussing how data can be represented and interpreted, it is important to understand that it needs to be accessible. If an organization cannot access its own data, the data provides few opportunities for value creation and innovation, but if data is accessible, the organization can define and interpret it. Larger amounts of data (e.g., from other sources and IT systems) can provide a bigger picture, and thus higher quality, as was stated by the product manager at M. Even when data can be easily accessed, technically speaking, a challenge is the reluctance of employees to see its potential value, as an M client stated. This point was not covered by the IS literature but was present in the strategic management and organizational studies literature. Reluctance to accept new ways of doing things has often been explained in terms of myopia (Levinthal & March, 1993), which reduces organizational innovation ability. This accorded well with Nambisan and colleagues (2017), who argued that successful digital innovation depends on how actors come to understand, share, and modify their understandings of innovation outcomes, processes, and related markets.

6. Conclusion

To conclude, based on our findings, and from an organizational perspective, we argue that digital data needs to strongly represent the real-world constructs or phenomena to which it refers, be underpinned by a meaningful and shared language for interpretation, have high quality, and be accessible. The main element and the cornerstones in our RIQA framework are stepping-stones for enablement and innovation. The three main attributes are connected, and one cannot be removed without affecting the others.

Our research is not without limitations. Since we mainly focused on published journal articles in the IS literature, and set a time frame, some articles discussing the data construct might have been overlooked; however, due to our systematic and in-depth search of the IS literature, we are confident that the main IS approaches to defining data have been covered. Empirically, we only interviewed individuals in one company. This may have limited the insights, but this company, with its interdisciplinary informants, was a good example of a data company with employees who work directly with data, and it also provided technical insights from a PaaS. Although this company and its employees may have colored our perspective of what 'digital data' is, it provided us with a broader perspective due to the informants' interdisciplinary backgrounds. Future research should employ and develop our RIQA framework further and should also aim to understand how digital data can enable value and what kind of value can be created.

Digital data is a complex subject, and a challenge for IS research is to better understand what digital data actually is. With a better understanding of the data construct, we can learn more about enabling organizations to transform their digital data into value and further enhance digital innovation. Hopefully, our RIQA framework is a step in this direction.

References

Abbasi, A., Sarker, S., & Chiang, R. H. (2016). Big Data Research in Information Systems: Toward an Inclusive Research Agenda. *Journal of the assiociation for information systems*(17 (2)).

Alaimo, C., & Kallinikos, J. (2018). Objects, metrics and practices: an inquiry into the programmatic advertising ecosystem. *Conference on Information Systems and Organizations*, (pp. 110-123). Springer, Cham.

Alaimo, C., Kallinikos, J., & Aaltonen, A. (2020). Data and Value. In S. Nambisan, K. Lyytinen, & Y. Yoo (Eds.), *The Handbook of Digital Innovation.* Cheltenham: Edward Elgar Publishing.

Alaimo, C., Kallinikos, J., & Valderrama-Venegas, E. (2020). Platform Evolution: A study of TripAdvisor. *53rd Hawaii International Conference on System Sciences.* Hawaii.

Almklov, P. G., Østerlie, T., & Haavik, T. K. (2014). Situated with Infrastructures: Interactivity and Entanglement in Sensor Data Interpretation. *Journal of the Association for Information Systems:*(15).

Baesens, B., Bapna, R., Marsden, J. R., Vanthienen, J., & Zhao, J. (2014). Transformational issues of big data and analytics in networked business. *MIS Quarterly: Management Information Systems, 38*(2), pp. 629-631.

Bucher, T. (2018). *If...Then: Algorrithmic Power and Politics.* Oxford University Press.

Chen, R., Sharman, R., Rao, R., & Upadhyaya, S. J. (2013). Data Model Celopment for Fire Related Extreme Events: An Activity Theory Approach. *Management Information Systems Quarterly, 37*(1), pp. 125-147.

Damen, M. (2015). *Charim.net*. Retrieved july 2, 2020, from http://www.charim.net/datamanagement/63#:~:text=Data%20homogenization%20is%20the%20process,analysis%20and%20validity%20of%20results.

De Jong, M., Kamsteeg, F., & Ybema, S. (2013). Ethnographic strategies for making the familiar strange: struggling with 'distance' and 'immersion' among Moroccan-Dutch students. *Journal of Business Anthropology*(2), pp. 168-186.

Demchenko, Y., & Los, W. (2018). DATA AS ECONOMIC GOODS: DEFINITIONS, PROPERTIES, CHALLENGES, ENABLING TECHNOLOGIES FOR FUTURE DATA MARKETS . *ITU Journal: ICT Discoveries*(2).

Demirkan, H., Bess, C., Spohrer, J., Rayes, A., Allen, D., & Moghaddam, Y. (2015). Innovations with Smart Service Systems: Analytics, Big Data, Cognitive Assistance, and the Internet of Everything. *Communications of the Association for Information Systems, 37*.

Emerson, R. M., Fretz, R. L., & Shaw, L. L. (2011). *Writing Ethnographic Fieldnotes.* Chicago: University of Chicago Press.

Eriksen, T. H. (2001). *Small places, large issues.* London: Pluto.

Fürber, C. (2016). *Data Quality Management with Semantic Technologies.* Wiesbaden: Springer Fachmedien .

Fleckenstein , M., & Fellows, L. (2018). *Modern data strategy.* Springer International Publishing AG.

Flyverbom, M., & Madsen, A. K. (2015). Sorting data out – unpacking big data value chains and algorithmic knowledge production. *Die Gesellschaft der Daten : Über die digitale Transformation der sozialen Ordnung*.

Gallenga, G. (2013). Elements of reflexive Anthropology in three fieldwork studies of the workplace. *Journal of Business Anthropology*(2(2)), pp. 187-208.

Gerlitz, C., & Helmond, A. (2013). The Like Economy: Social Buttons and the Data-Intensive Web. *New Media & Society, 15*(8), pp. 1348-1365.

Gillespie, T. (2014). The relevance of algorithms. *Media Technologies: Essays on communication, materiality and society*(167).

Hammersley, M., & Atkinson, P. (1995). *Etnography: Principles in Practice* (Vol. 2). London: Routledge.

Haug, A., Arlbjørn, J. S., Zachariassen, F., & Schlichter, J. (2013). Master data quality barriers: an empirical investigation. *Industrial management & data systems*(113(2)), pp. 243-249.

Hildebrandt, B., Hanelt, A., & Firk, S. (2018). Mitigating Moral Hazard in Access-Based Consumption through IS-Enabled Value CoCapturing with Consumers. *Business & Information Systems Engineering: The International Journal of WIRTSCHAFTSINFORMATIK, 60*(3), pp. 227-241.

Kampevoll, F. (2019). *E24*. Retrieved August 7, 2020, from https://e24.no/karriere-og-ledelse/i/WLQ6PK/vil-bedre-kvinneandelen-i-it-med-aa-hente-fra-finans-handel-og-offentlig-sektor

Kaniadakis, A., & Constantinides, P. (2014). Innovating Financial Information Infrastructures: The Transition of Legacy Assets to the Securitization Market. *Journal of the Assiocation for Information Systems, 15*, pp. 244-262.

Levinthal, D. A., & March, J. G. (1993). The myopia of learning. *Strategic Management Journal, 14*(2), pp. 95-112.

Liang, T. P. (2010). RFID Technology and Automated Data Collection. *Pacific Asia Journal of the Association for Information Systems, 2*.

Link, G. J., Lumbar, K., Conboy, K., Feldman, M., Feller, J., George, J., . . . Willis, M. (2017). Contemporary Issues of Open Data in Information Systems Research: Considerations and Recommendations. *Contemporary Issues of Open Data in Information Systems Research: Considerations and Recommendations*(41).

Lukic, J. (2014). Role of Big Data in Open Innovation Practices: The case of Serbian ICT industry. *Ekonomika Preduzeca*(62), pp. 294-304.
Masrom, M., & Rahimly, A. (2015). Overview of Data Security Issues in Hospital Information Systems. *Pacific Asia Journal of the Assiociation for Information Systems, 7*(4).
Monteiro, E., & Parmiggiani, E. (2019). Synthetic Knowing: The Politics of the Internet of Things. *MIS Quarterly, 43*(1), pp. 167-184.
Nambisan, S., Lyytinen, K., Majchrzak, A., & Song, M. (2017). Digital Innovation Management: Reinventing Innovation Management Research In A Digital World. *MIS Quarterly, 41*(1).
Opher, A., Chou, A., Onda, A., & Sounderrajan, K. (2016). *The Rise of the Data Economy: Driving Valye through Internet of Things Data Monetization.* New York: IBM.
Perdana, A., Robb, A., & Rohde, F. (2019). Interactive Data and Information Visualization: Unpacking its Characteristics and Influencing Aspects on Decision-making. *Pacific Asia Journal of the Association for Information Systems, 11*(4).
Pettersen, L. (2018). Why Artifical Intelligence Will Not Outsmart Complex Knowledge Work. *Work, Employment and Society, 33*(6), pp. 1058-1067.
Rosenberg, D. (2013). Data Before The Fact. In L. Gitelman (Ed.), *"Raw data" is an Oxymoron.* Cambridge Massachusetts: The MIT Press.
Sahay, S. (2016). Big Data and Public Health: Challenges and Opportunities for Low and Middle Income Countries. *Communications of the Association for Information Systems*(39).
Sanders, J. D. (2016). Defining Terms: Data, Information and Knowledge. *SAI Computing Conference.* London.
Sankaran, S., & Wedel, t. L. (2020). Motion Picture Industry Pension Plan: A Database Design Case. *Journal of Informatiomn Systems Education, 31*(2), pp. 124-130.
Sen, R., Nelson, M. L., & Bumaniam, C. (2015). Application of Survival Model to Understand Open Source Software Release. *Pacific Asia Journal of the Association for Information Systems*(7 (2)), pp. 1-24.
Sherer, S. A. (2014). Patients Are Not Simply Health IT Users or Consumers: The Case for "e Healthicant" Applications. *Communications of the Association for Information Systems, 34*(17).
Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data Quality in Context. *Communications of The ACM, 40 (5).*
Tilly, R., Posegga, O., Fischbach, K., & Schoder, D. (2017). Towards a conceptualization of Data and Information Quality in Social Information Systems. *Business & Information Systems Engineering: The International Journal of WIRTSCHAFTSINFORMATIK, 59*(2(1)), pp. 3-21.
Ullah, A., & Lai, r. (2011). Managing Security Requirements: Towards Better Alignment Between Information Systems And Business. *Pacis 2011 Proceedings*.
Wald, L. (2002). *Data Fusion, Definitions and architectures.* Paris: les Presses de l'École des mines.
Wyatt, J., & Piggott, A. (2019). The Design of Not-so-e The Design of Not-so-everyday Things: Designing for Emer y Things: Designing for Emerging Experiences. *AIS Transactions on Human-Computer Interaction*(11 (4)), pp. 244-252.
Yin, R. K. (2012). *Applications of Case Study Research.* Washington DC: SAGE.
Yoo, Y. (2010). Computing in Everyday life: A Call for research on Experiential Computing. *MIS Quarterly, 34*(2), pp. 213-231.
Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power.* Profile Books.