UNDERSTANDING THE RELATION BETWEEN INTEROPERABILITY AND DATA QUALITY:

A STUDY OF DATA HUB DEVELOPMENT IN SWEDISH ELECTRICITY MARKET

JWAN KHISRO
Dept. of Information System and Technology
Mid Sweden University
Sweden
jwan.khisro@miun.se

Abstract

What makes this digital age so interesting is the prevalence of data and how it could be utilised, processed and made interoperable to create new businesses or to revolutionise old ones. Data quality, or lack thereof, is widely considered one of the most critical problems for achieving interoperability. To achieve high interoperability, such as the Swedish electricity market data hub, data hub development needs to better comprehend the relation between interoperability and data quality. Thus, this study inquired how the relation between interoperability and data quality can be understood. A qualitative study with a deductive approach was chosen for this purpose, as this enabled deeper understanding. The chosen theory formed the basis for the analytical framework. Seven deep interviews with actors in the Swedish electricity market provided empirical data. The results demonstrated that interoperability and data quality possess a make-or-break relationship. Consequently, the understanding is that high data quality is capable of decreasing complexity in a development process and increasing its reliability.

Keywords: Data quality, data hub, development interoperability, relation.

1. Introduction

Today's digital age is witnessing increasing collaboration among enterprises during the entire product or service life cycle. As such, they have to cope with internal changes from both a technical and organisational point of view (Chen and Doumeingts, 2003; Camarinha-Matos, 2016; Sullivan and Skelcher, 2017). Large enterprises need to exchange data among numerous separately developed systems. In order for this exchange to be useful, the individual systems must agree on the meaning of their exchanged data i.e. the enterprise must ensure interoperability (Sciore et al.,1994). What makes this digital age so interesting is the prevalence of data and how it could be utilised, processed and made interoperable to create new businesses or to revolutionise old ones. As articulated by Grey (2015): "*You are only as good as your data*". Data collection has thus grown considerably, with data quality becoming increasingly important for business transformation, as interoperability includes an exchange of data elements via digital artefacts.

Panetto et. al. (2016) further explained that data is meaningful only as data interoperates with other data and context can be provided to it to transform the data into information. For instance, data consistency has been identified as a critical need to reach an interoperable solution (Tolk and Muguira, 2003; Pandit et al., 2018). The unambiguous interpretation of the meaning of the data to be interchanged between the two systems is crucial to gain interoperability (Tolk and Muguira, 2003). Extensive research has accordingly been performed in the areas of data and information quality (Wang et al., 1995; Shankaranarayanan and Cai, 2006; Madnick et al., 2009; Storey et al., 2012; Du and Zhou, 2012). Despite these works, however, data quality or lack thereof, is widely considered as one of the most critical problems for achieving interoperability as it comes back to the point regarding the exchange of data between systems, which resides at the core of interoperability (Umar et al., 1999; Zhu and Wu 2014; Zhao and Xia, 2014; Daraio et al., 2016; Zhu et al., 2016). Data quality is considered a success factor for achieving interoperability (Khisro and Sundberg, 2018). Fenton et al., (2013) even went as far as to say, '*Without data standards and data quality, the future of interoperability is bleak*'. In line with Panetto et al., (2016) and Fenton et al., (2013), I argue, that data functions as a springboard for interoperability. To achieve high interoperability and avoid collaboration difficulties. such as from multiple sources of data being exchanged across boundaries, a deeper understanding of the relation between interoperability and data quality is thus required. As such, this study inquired as to how the relation between interoperability and data quality can be understood.

This paper is structured as follows: In the Theory section, the concepts of interoperability and data quality are introduced, followed by an analytical framework for understanding the relation between the them. The Methodology section presents the chosen method and context, data collection procedures and method of analysis. The Result and analysis section present the empirical results together with the analysis. The relation between the interoperability and data quality is then discussed and further clarified in the Discussion section. The Conclusion section provides an answer to the research question. Finally, suggestions for future study are presented.

2. Theory

Understanding of the relation between the two concepts of interoperability and data quality are central in this study, as this affects, for instance, data hub development. This is in agreement with Lindblad-Gidlund (2005), who argued that development process requires a relational perspective. Relation is the way that two things relate, or the way in which one relates to another, indicates the sort of connection that exists between them. If something relates to a particular subject, it concerns that subject. Furthermore, relation represents the state or condition of being related or the manner in which things are related. This comprise the connection or similarity between two or more things as being, belonging, or working together (Collins dictionary 2018, Merriam-Webster, 2018). In information systems, human-to-human relations consist of interactions between humans, which can be carried by digital communication; meanwhile, human-to-machine relations describe interactions between humans and machines through digital artefacts that gather necessary data based on partners' roles, thus supporting communication with external partners. Beyond this, machine-to-machine relations indicate consistent digital processes and communication among enterprises linking different applications and information systems (Reimers, 2001; McAfee, 2005). This section thus, provides an understanding of the interoperability and data quality concepts. The analytical framework then presents a summary of how the relation between interoperability and data quality can be understood.

2.1 Interoperability

The concept of interoperability involves the ability for a system, service or product to communicate with other systems, services or products effectively without user interference (IEEE, 1991; Ford et al., 2007).

Gürdür and Asplund (2017) conceptualised data interoperability as the capability of data involving documents, multimedia content, and digital resources to be accessible, reusable, and comprehensible by all transaction parties, such as in a human-to-machine and machine-to-machine basis. The use of different representations, purposes, contexts, and syntax-dependent approaches will result in a lack of common understanding. However, EC (2006) pointed out that the interaction between enterprises occurs not only on the level of information technology, but also on organisational and semantic levels. Such an interaction needs to be flexible and can be developed without much expense. Meanwhile, Van Sinderen et al. (2013) focussed on enterprises' ability to actively contribute to their own and each other's organisational goals through a defined effect on each other's operations. This is based on the IEEE definition of interoperability, however, wherein enterprise interoperability emphasises meaningfulness and utility for the enterprises involved. Interoperability is believed to be more adaptable due to its decreased cost and faster implementation. Generally, interoperability refers to coexistence, autonomy and federated environments, whereas integration refers to the concepts of coordination and coherency (Panetto, 2007). Therefore, two integrated systems are inevitably interoperable, but two interoperable systems are not necessarily integrated Panetto and Cecil (2013).

Tolk and Migura (2003) proposed the levels of conceptual interoperability model (LCIM) to address the need for levels of conceptual interoperability. This model considers interoperability as a conceptual rather than technical problem. It aims to act as a go-between for conceptual and technical design by focussing on the data to be interchanged between systems. The focus of LCIM rests on the data to be interchanged and the available interface documentation. The LCIM can aid in better understanding the data goals and fulfilling them effectively and efficiently. The LCIM defines levels that signal increasing interoperability with regard to design data. These levels are not isolated, and the content of each level (i.e. its value) is complex (Tolk and Muguira, 2003). Wang et al. (2009) further developed and deepened the LCIM. The levels are defined based on Tolk and Muguira (2003) and Wang et al. (2009), as follows:

Level 0: No interoperability, with no connection between systems and no interoperability.

Level 1: Technical interoperability, where one can observe symbols of communication through a common established communication protocol. This level possesses technical connections through which systems can exchange data in bits and bytes.

Level 2: Syntactic interoperability, wherein the data structure is defined, but not the meaning of the data elements. The systems share a common data format and agree on a common syntax.

Level 3: Semantic interoperability, in which a common reference model facilitates interacting systems to exchange terms to be semantically analysed. This level implies agreement on the definition of terms through a process of disambiguation. Not only data, but also its contexts (i.e. information) can be exchanged.

Level 4: Pragmatic interoperability, characterised by a common workflow model where the context of information can be exchanged. This level implies awareness and sharing of a common reference logical model. This model could constitute a subset of reality and the search for solutions.

Level 5: Dynamic interoperability, wherein one common execution model allows data changes to propagate. The systems understand the processes that will use the symbols they exchange. At this level, the assumptions and constraints of processes are described unambiguously, and the systems' behaviour is predictable during interoperation.

Level 6: Conceptual interoperability, featuring a common conceptual model that allows interacting systems to understand each other's information, processes, contexts, and modelling assumptions. The underlying concepts represented by the symbols are described unambiguously. This level implies alignment of the models represented in systems. The systems share a common reference conceptual model that captures the assumptions and constraints of the corresponding real or imaginary object.

2.2 Data quality

Data quality, a multidimensional concept defined by Wang and Strong (1996), describes data that is fit for use by data consumers. It should be intrinsically good, contextually appropriate for the task, clearly represented, and accessible. Turner (2004) and Herzog et al. (2007) characterised data as being of high quality if it is fit for use in their intended operational, decision-making and other roles, or if it conforms to standards. Redman (2013) and Duvier et al. (2018) identified that successful organisations define data quality as its ability to fulfil important customer needs. Wang and Strong (1996) defined a set of attributes or dimensions that represent a single aspect of data quality to allow it to be measured, analysed, and improved in a valid manner. Ozmen-Ertekin and Ozbay (2012) focussed on data quality dimensions as the measurable forms of data characteristics. Wang et al. (2006) mentioned dimensions as a set of data relevant to specific industries and which most consumers react to in a fairly consistent manner. Although data quality dimensions have frequently been mentioned, it has been widely accepted and noted that what are most important to information consumers are accuracy, consistency, timeliness, and completeness (Parssian 2006; Blake and Mangiameli 2011; Song et al., 2016).

Accuracy refers to the degree to which data are equivalent to their corresponding real values (Ballou and Pazer, 1995). This can be assessed by comparing values with external values that are either known or considered to be correct (Redman, 1996).

Timeliness refers to the degree to which data are up-to-date. This is associated with three occurrences: first, when there is a change in the real world; second, when change is recorded as data in an information system; and third, on the use of that data (Blake and Mangiameli, 2011).

Completeness refers to the degree to which data are full and complete in content, featuring no missing data. This dimension can be described as a data record that captures the minimum required amount of information (Wand and Wang, 1996), or data for which all values have been captured (Hazen et al., 2017). Every field in the data record is needed in order to paint a complete picture of what the record strives to represent in the 'real world' (Blake and Mangiameli, 2011).

Consistency refers to the degree to which related data records match in terms of format and structure (Wang and Strong, 1996). Ballou and Pazer (1995) defined consistency as the data value's representation remaining the same in all cases. Batini et al. (2009) developed the notion of both intra-relation and inter-relation constraints on data consistency. Intra-relation consistency assesses the data's adherence to a range of possible values (Coronel and Morris, 2016), whereas inter-relation assesses how well data are presented using the same structure.

2.3 Analytical framework

Based on the presented interoperability and data quality theories, this section presents further explanations and a framework for deeper understanding the relation between interoperability and data quality. Thus, Table 1 gives an account of how each part of this complex relation can be understood according to theory.

The technical interoperability and data quality relation was created through a combination of understanding at the technical interoperability level and the data quality dimensions of accuracy, timeliness, completeness and consistency. This part of the relation was understood as a common communication protocol (Wang et al., 2009) that accurately describes data (Redman, 1996), encourages data to be up-to-date (Blake and Mangiameli, 2011), completely captures all necessary values (Wand and Wang, 1996) and consistently allows all data to possess the same format and structure (Wand and Wang, 1996).

The syntactic interoperability and data quality relation was created through a combination of understanding at the syntactic interoperability level and the data quality dimensions of accuracy, timeliness, completeness and consistency. This part of the relation was understood as a common communication protocol (Wang et al., 2009) that was understood as common language rules (Wang et al., 2009) that accurately provide correct format and structure of all data (Ballou and Pazer, 1995), apply to all updated data (Blake and Mangiameli, 2011), completely define all data elements (Hazen et al., 2017) and consistently present all data to be exchanged in the same format and structure (Wand and Wang, 1996).

The semantic interoperability and data quality relation was created through a combination of understanding at the semantic interoperability level and the data quality dimensions of accuracy, timeliness, completeness and consistency. This part of the relation was understood as a common communication protocol (Wang et al., 2009) that was understood as a common reference model (Wang et al., 2009) that accurately reflects correct descriptions of data and its real values (Ballou and Pazer, 1995), contributes to a common understanding of relevant data that are dependent on updated data (Blake and Mangiameli, 2011), contains only necessary data and information (Wand and Wang, 1996) and consistently possesses the same data value in all cases (Ballou and Pazer, 1995).

The pragmatic interoperability and data quality relation was created through a combination of understanding at the pragmatic interoperability level and the data quality dimensions of accuracy, timeliness, completeness and consistency. This part of the relation was understood as a common communication protocol (Wang et al., 2009) that was understood as a common workflow model (Wang et al., 2009) that is accurately significant for the usability of correct data and information exchange (Redman, 1996), features sufficiently updated data (Blake and Mangiameli, 2011), uses only necessary data and information (Wand and Wang, 1996) and employs consistent common data value representation for all contexts (Ballou and Pazer, 1995).

The dynamic interoperability and data quality relation was created through a combination of understanding at the dynamic interoperability level and the data quality dimensions of accuracy, timeliness, completeness and consistency. This part of the relation was understood as common process understanding (Wang et al., 2009) that is accurately achieved through clearly described data (Ballou and Pazer, 1995), regularly updated data (Blake and Mangiameli, 2011), complete access to all required data elements (Wand and Wang, 1996) and consistently applicable data structure and format (Wand and Wang, 1996).

The conceptual interoperability and data quality relation was created through a combination of understanding at the conceptual interoperability level and the data quality dimensions of accuracy, timeliness, completeness and consistency. This part of the relation was understood as a common understanding of assumptions regarding information, processes, context and constraints (Wang et al., 2009) affecting the accurate description of data (Redman, 1996), provides proper data update (Blake and Mangiameli, 2011), features complete relation between data elements (Wand and Wang, 1996) and is consistently aligned with the representation of data value (Ballou and Pazer, 1995).

Table 1. How the relation between interoperability and data quality can be understood.

Data quality dimensions **Interoperability levels**	**Accuracy**	**Timeliness**	**Completeness**	**Consistency**
Technical interoperability	The common communication protocol is based on accurately described data.	The common communication protocol encourages data to be up to date.	The common communication protocol captures all needed values.	The common communication protocol allows all data to share the same format and structure.
Syntactic interoperability	Correct format and structure of all data elements follows common language rules.	Correct format and structure of all data elements follows common language rules.	Data elements are completely defined in common data structure and formats following common language rules.	All data to be exchanged are presented in the same format and structure.

Table 1. How the relation between interoperability and data quality can be understood – continued.

Data quality dimensions / Interoperability levels	Accuracy	Timeliness	Completeness	Consistency
Semantic interoperability	Correct descriptions of data and its real values are reflected in the common reference model.	Common understanding of relevant data depends on updated data.	The common reference model contains only necessary data.	Common data understanding is based on the same data value in all cases.
Pragmatic interoperability	Common workflow model is significant for usability of correct data and information exchange.	Data are sufficiently updated for use in the corresponding workflow context.	Only data and information needed are used in the common workflow model.	Common data value representation is used for all workflow model contexts.
Dynamic interoperability	Process understanding is achieved through clearly described data.	Process understanding is accomplished by regularly updated data	All data elements are required for common process understanding.	Data structure and format are applicable in a common process understanding.
Conceptual interoperability	Common understanding of assumptions about information, processes and context influence accurate data description.	Common understanding of constraints regarding information, processes and context influence proper data update.	Relation between data elements is captured in common understanding of information, process and context assumptions and constraints.	Representation of data value is aligned with information, process and context assumptions and constraints.

3. Methodology

In order to coordinate and manage the processes and complex data transactions between electricity suppliers and grid companies, data hubs are being developed. The Nordic countries have achieved different development steps ranging from requirements to implementation of the data hub, with interoperability and data quality at the centre of interest (Swedish Energy Markets Inspectorate, 2017; Svenska kraftnät, 2018). This study sought to deepen our understanding of how the relation between interoperability and data quality can be understood. A qualitative study with a deductive approach was chosen for the purpose as this enabled understanding of practitioners' way of expression employed during seven semi-structured interviews (Elo and Kyngäs, 2008; Creswell 2014, Yin, 2017). Interoperability and data quality theories were utilised as themes for the interview guide (Bengtsson, 2016), with the exception of Level 0 'No interoperability' and Level 1 'Technical interoperability' from the LCIM. Respondents possessed different roles and responsibilities in the different companies and in the development of the data hub for the Swedish electricity market, see Table 2. They were selected based on their interest and active engagement in the development process of interoperability and data quality in the data hub. The interviews were conducted in Sweden through two physical meetings in the workplace and through five online meetings. Each interview lasted approximately two hours in length. They were recorded on the author's computer, as accepted by all respondents upon asking. The recorded material was transcribed in Swedish and then translated into English. Content analysis was employed for analysing the result, as this contributed to understanding the human contribution within the context (Kohlbacher 2006; Bengtsson, 2016; Erlingsson and Brysiewicz, 2017). In conformance with Seale et al. (2004), the results were written in a narrative form together with citations from the interview material to illustrate the respondents' understanding of interoperability and data quality, see section 4 Empirical result and analysis. The result was colour coded based on the interoperability levels and data quality dimensions themes. The content of each individual answer was analysed according to theory to identify similarities and differences between theory and practice.

As the questions in the interview guide were based on the available knowledge of each concept, it was not possible to arrive at the respondents' particular understanding of the relation between them. It was therefore necessary to create an analytical framework, see section 2.3 Analytical framework, that would assist the further analysis aimed at deepening our understanding of the relation between interoperability and data quality. Therefore, citations from 4.1 Interoperability and 4.2 Data quality were analysed and structured according to the analytical framework, see Table 3, in order to further analyse how the relation between interoperability and data quality was understood, see section 4.3 The relation between interoperability and data quality. Citations as well as No citations in Table 3, i.e. the result were then discussed in accordance with the theory to identify strong and weak understanding of the relation between the interoperability and data quality. Thus, a more detailed comprehension of the relation between interoperability and data quality were possible when they were read together. In this way, the analysis served as a basis to discuss and conclude how the relation between interoperability and data quality were understood.

Table 2. Respondent role and responsibility in the company and in the data hub development.

Role	Company	Tasks and responsibilities in the company.	Role and responsibility related to developing the data hub.
Operations officer	Energy company	Responsible for the Measuring Systems Group.	Provides billing data for staff in electricity companies through the hub.
Operations administrator	Energy company	Works in the department of business support and development. Has contact with both grid and supplier companies.	No specific role at the moment. Takes special interest in grid and supplier companies' situation in the hub.
Member of expert group	Electricity supplier	Works at the IT department. Ensures systems run properly. Involved in electricity market issues. Attends meetings for the hub.	Involved in business processes and technology issues for the hub.

Table 2. Respondent role and responsibility in the company and in the data hub development – continued.

Development engineer	Grid company	Has worked with issues regarding the upcoming hub.	Grid companies are responsible for collecting and reporting measuring data to the hub.
Customer service coordinator	Energy company	Work tasks largely consist of changes in supplier, billing, in and outgoing payments, requirement management and customer service.	Takes part in the analysis of information, data and of the hub.
Data hub owner 1	Energy authority	The company develops and manages the hub on order from the government.	Works with change management issues for the hub project and is responsible for dialogue and communication between actors in the electricity market to ensure they understand what to do, how to do it and when.
Data hub owner 2	Energy authority	The company develops and manages the hub on order from the government.	Hub product owner. Responsible for systems requirements.

4. Empirical result and analysis

This section presents the empirical result and analysis of the study concepts, interoperability and data quality.

4.1 Interoperability

The Operations officer and the Development engineer both defined interoperability as the level of integration between systems. The Customer service coordinator also confirmed:

"We have never come across the phrase before. Integration is the common term for us. By that we understand when our systems are integrated or can easily talk to each other. "

According to Panetto and Cecil (2013), confusion is common in understanding the concepts of integration and interoperability. However, two integrated systems are inevitably interoperable, but two interoperable systems are not necessarily integrated. Furthermore, the Operations administrator answered that interoperability means an effective data exchange. The Member of an expert group answered that interoperability refers to cooperation between different actors. The respondents from the data hub owner responded that the interoperability means there is capacity between two systems. This is compatible with IEEE (1991) and Ford et al. (2007), who determined that interoperability describes the ability for a system, service or product to communicate with other systems, services or products effectively and without user interference.

Syntactic interoperability

I asked how one can ensure that no data in the data hub will contradict each other in communication with multiple organisations. The Member of an expert group answered that this could be achieved by having as little data as possible in the hub, and only data that is truly necessary. Data must also be controlled against available official records.

The Operations administrator described a lack of understanding of data in the hub, but also that there is not yet much interest in understanding how the system would work for them. The Development engineer engineer stated that be controls must be present in the hub to ensure that, for example, a customer cannot possess active data with two suppliers. The Data hub owner explained that none of the partners can upload data that is inconsistent with the others, as each partner possesses their own responsibility. The Operations officer answered as follows:

> *"By minimising unnecessary data that can cause conflicts. It must have a good structure and it must be decided in advance what data should be in the hub."*

This is in harmony with Tolk and Muguira (2003) and Wang et al. (2009), who described syntactic interoperability as the data structure defined.

Semantic interoperability

I next asked how a common understanding of the data that one reads can be ensured when communicating with multiple organisation. The Operations officer insisted on the importance of meetings, stated that they needed to talk more when they did not understand each other. This occurred because different departments possessed different perspectives. The Operations administrator stated that there is currently a lack of understanding regarding who is responsible for what. No one bears total responsibility when it comes to this. The initiative has to come from top management, who has to choose who will be responsible for the process of creating a shared understanding of data. The respondents from the Data hub owner emphasized that it must be clear why certain information, formats or attributes must be chosen for the hub. They also added:

> *"It needs to be explained, defined and exemplified for everyone to understand. This is how we will reach consensus"*

This correspond with Wang et al. (2009), who explained that semantic interoperability facilitates the exchange of not only data, but also its contexts (i.e. information) while interacting systems can also exchange terms that can be semantically analysed. Furthermore, the Member of an expert group stated that, by ensuring all partners understand the hub information model, the process will verify that information is correct as well as contribute to understanding among the businesses involved. The Development facing no such problem of misinterpretation or differing understandings of data. The Customer service coordinator stated that a regulated list of words should be available explaining meaning and interpretations for everyone to share.

Pragmatic interoperability

I then asked how it can be ensured that a common logical reference model is in place. The Operations officer answered that it was quite early in the process of implementing a common reference model. However, this comes back to decision-makers at the managerial level who did not know who was responsible for what data in the hub. While the Operations administrator believed this could be accomplished through process mapping. Every manager involved possesses a responsibility to develop a common logical reference model. They must also take an interest in the different processes, how they relate and their consequences for each other.

The Member of an expert group stated that:

> *"There are processes developed in the BRS guide that tell how to exchange information. We need a handbook explaining how to retrieve and control data. Everything should refer to information model. It is the hub's information model that all actors will stick to. "*

This goes with Wang et al. (2009), who described a common workflow model where the context of information can be exchanged and implied the awareness and sharing of a common reference logical model. From the Development engineer perspective, to guarantee measurement values, logic must be present in the hub. The hub must be able to handle data, to calculate, summaries and distribute data among everyone. The Customer service coordinator believed in discussions with colleagues and the data hub owner. The data hub owner must play a central role here. The respondents from the Data hub owner, meanwhile, stated that the concept should remain the same as today, with few new concepts added. Furthermore, there should be continuous publication lists of concepts, processes should be defined in collaboration with expert groups and discussions about the project, and documentation should be provided. It is a machine-to-machine issue. Therefore, grid owners need to consider the API and the message as such, and not the information model. This is partly harmonised with Gürdür and Asplund (2017), who discussed data interoperability as the capability of data involving documents, multimedia content, and digital resources to be accessible, reusable, and comprehensible by all transaction parties, such as in a human-to-machine and machine-to-machine basis. Most of the respondents described this as a machine-to-machine communication and missed the human-to-machine perspective.

Dynamic interoperability

I next asked how the importance of information can be maintained when changes occur in the data hub. The Operations officer stated this could be achieved by initially leaving space in the hub for future information and functions, because as soon as they write a new field in the system, matters become difficult and complicated. The Operations administrator clearly expressed that the data in the hub must be right. That is what is assumed, and if it is wrong, then they have to act. Therefore, it is important that every partner in the electricity market knows their responsibility he/she has in the hub. The respondents from the Data hub owner replied, through version management, and also mentioned:

> *"The partners are responsible for data and information; we could send them a list of wrong data and they would have to correct them. There will be a migration support. It is a system that takes in information from the partners and validate it according to rules. If there is anything stuck then it will be sent back. "*

This is in line with Wang et al. (2009), who explained dynamic interoperability as featuring one common execution model, which allows changes to data to propagate. The systems understand the processes that will use the symbols they exchange. Further, The Member of an expert group emphasized that every change in the hub must be traceable and possess a history. Each new change should contain time information so the actor can go back and see what it looked like before. The Development engineer stated that, as soon as there was a change in any system, this must be communicated to or from the hub; if the change is not confirmed, then the old data remains valid. The Customer service coordinator talked about no difference from today. Through a track list of changes, so they can follow the history. At this level, the assumptions and constraints of processes are described unambiguously, and the systems' behaviour is predictable during interoperation Wang et al. (2009).

Conceptual interoperability

I then asked what role the respondents held in the process of creating a common conceptual model for the data hub. The Operations officer reflected on a common conceptual model for the data hub that does nothing more than inform each other, while those responsible for the data sign up for information from the data hub owner. The Member of an expert group stated that they developed their process diagram and follow up how it should work forward. What they thought two years ago has changed today, resulting in the conceptual model being further developed. The respondents from the Data hub owner expressed that there was an existing conceptual model being used today needs alterations to serve the expert groups, but nothing new needs to be added. Together with the expert groups, they were responsible for the developing a conceptual model two years ago. The Operations administrator also said:

> *"Today we have three different systems. They are like silos; you only collect specific information needed from each system. They do not talk to each other."*

This is what Wang et al., (2009) focussed on when developing a conceptual model that allows interacting systems to understand each other's information, processes, contexts, and modelling assumptions. The Development engineer mentioned that they had just started a project to go through everything about the hub. This concerned what grid companies and suppliers do today versus what they will do in the new model, and to obtain an effective understanding of each other's role. Moreover, the Customer service coordinator said they were active and work together as partners in developing the hub, the conceptual model and their own systems. They had to find the right way to do so together. According to the researchers, conceptual interoperability implies the alignment of the models represented in systems (Wang et al., 2009).

4.2 Data quality

The Development engineer explained explained that they currently collect customer data in dialogue with the customer, with the customer turning to them first. In the hub, suppliers collect and ensure that customer data are correct. This is in line with Redman (2013) and Duvier et al. (2018), who stated that successful organisations define data quality as fulfilling important customers' needs. The Customer service coordinator further declared that low data quality creates a great deal of extra work. This is in line with Wang and Strong (1996), who explained that data quality should be intrinsically good, contextually appropriate for the task, clearly represented, and accessible. Moreover, the Operations officer stated that data quality means that data specifications meet system requirements and expectations, while the Operations administrator, The Data hub owner and the Member of an expert group a most agreed that data quality means correctness, format, validation and completeness. The Development engineer added that data quality means the data is accurate and available at the right time.

Customer service coordinator also confirmed that:

> *"Data quality means information is right, correct, updated or rather up-to-date and available for the right staff."*

This agrees with Herzog et al. (2007), who described data quality as a multidimensional concept. Moreover, Blake and Mangiameli (2011) and Song et al. (2016) mentioned accuracy, consistency, timeliness, and completeness as the most important dimensions for information consumers.

Accuracy

The Development engineer explained that it is a considerable challenge to ensure that customer data is correct, such as the customer's name, address, and customers moving in and out. The Member of an expert group said correct data is created through immediate access and update of for instance address data from official sources. Each partner is responsible for their data, for instance customer data. The hub bears no responsibility in this respect. The Operations administrator and the Customer service coordinator focused on the continuous daily job of data cleansing and on improving the cleansing procedure. The respondents from the Data hub owner pointed out that each partner is responsible for the correctness of their data stored in the hub. The hub itself bears no such responsibility, but can validate the data. For example, the hub cannot tell whether it is the right customer moving in, but they can report if a personal code number is invalid or if a surname should be lacking. In order to avoid incorrect data, the Data hub owner engages in validation and technical format control This is accuracy problem that refers to the degree to which data are equivalent to their corresponding real values (Ballou and Pazer, 1995). Moreover, the Operations officer mentioned:

> *"It must be possible for register administrators to compare different periods, for instance last year with the present."*

This is in line with Redman (1996), who claimed that accuracy can be assessed by comparing values with external values that are known to be or considered correct.

Timeliness

All the respondents focused solely on the direct and immediate update of the data and on obtaining confirmation from the hub. In particular, the Member of an expert group said an update of, for instance, a customer moving in or out of a dwelling, should be made immediately in the hub. Customer service coordinator explained that if the data owner discovers data is wrong then it must be corrected. The Development engineer highlighted the importance of very carefully defining which actors should update which data and who is responsible for correct data. The Customer service coordinator also mentioned:

> *"If the data owner discovers data is wrong then it must be corrected. There must be rules stating who owns the data; the grid company or the electricity supplier."*

This, to a certain degree, aligns with Blake and Mangiameli (2011), who pointed to three occurrences of change: in the real world, in an information system and in the use of data. Several respondents mentioned a particular type of change occurrence, such as changes in the real world or in an information system, whereas others did not mention any particular type of change occurrence at all. However, none of the respondents mentioned all three occurrences of change, which disagrees with Blake and Mangiameli (2011). Furthermore, Blake and Mangiameli (2011) did not mention change in data ownership.

Completeness

The Customer service coordinator stressed that the hub, and the data they obtain, must eventually be trusted. The Operations officer mentioned the need for some kind of plausibility assessment of data in or out of the hub. The Operations administrator claimed that the consequences of not possessing all the information in the hub must be clarified and communicated. The Member of an expert group mentioned that the hub will possess a limited amount of data due to the General Data Protection Regulation (GDPR) law, and will therefore not cover all necessary data. For example, it is the suppliers' duty to ensure billing addresses are correct, meaning the hub will not contain billing addresses. The Development engineer said people have put a great deal of effort into visualising and defining every possible situation that could occur in the business. The respondents from the data hub owner clarified:

> *"Expert groups have been working for the last two years with defining what information needs to be in the hub. The hub will then mirror the content of each partners system. "*

Several respondents close to the hub development, such as the Development engineer and Data hub owner, indicated effective understanding of the completeness dimension in line with Wand and Wang (1996), who stated that is can be described as a data record that captures the minimally required amount of information. This result also agrees with the assumption that every field in the data record is necessary in order to paint the complete picture of what the record is attempting to represent in the 'real world', as suggested by Blake and Mangiameli (2011). The other respondents also understood this, but expressed it more generally in ordinary language.

Consistency

The Operations officer clarified that the hub should contain sufficient data, but not too much of it. Numerous unnecessary parameters are present. The Operations officer further added that they have to create a shared understanding among the partners in the hub in order to be able to agree on what data should look like in the hub. The respondents from the Data hub owner insisted that in order for the hub to function properly, correct data structure must be migrated before implementation. This is in accordance with Wang and Strong (1996) who referred to consistency as being the degree to which related data records match in terms of format and structure. By contrast, the Member of an expert group highlighted that the current issue is not data structure, but rather what processes will work in the hub. By using the hub, only one partner needs to be contacted by either phone or website. The hub should also facilitate the customer process of choosing a supplier. The Development engineer also supported the process issue by claiming they must have access to the measurement series and records made available to each supplier in the hub as well as to data in such a way that they can perform their necessary checks and reconciliations. Consistency theory does not support issues of finding the work processes suitable for the hub and the enterprises, though.

4.3 The relation between interoperability and data quality

Regarding syntactic interoperability and data quality relation, the respondents demonstrated effective understanding of the importance of accurate and consistent data format and structure, and of data being updated. However, there was no indication of understanding timeliness and completeness in relation to syntactic interoperability.

Concerning semantic interoperability and data quality relation, the respondents demonstrated effective understanding of the importance of accurate and correct customer data, even though this was considered a considerable challenge. By arguing that information in the hub must be clarified and communicated, they also demonstrated understanding of the importance of data and information being completely contained in order to achieve interoperability. On the other hand, there was no indication of understanding timeliness and consistency in relation to semantic interoperability

Regarding pragmatic interoperability and data quality relation, the respondents demonstrated effective understanding of the importance of consistent common data value representation for all contexts, such as a handbook to explain how to retrieve and control data. However, there was no indication of understanding accuracy, timeliness and completeness in relation to pragmatic interoperability.

Concerning dynamic interoperability and data quality relation, the respondents demonstrated effective understanding of the importance of correct data and consistent communication of system changes. However, there was no indication of understanding timeliness and completeness in relation to dynamic interoperability.

Regarding conceptual interoperability and data quality relation, the respondents demonstrated no understanding of the relation between the two concepts. There was no indication of understanding accuracy, timeliness, completeness and consistency in relation to conceptual interoperability.

Concerning the timeliness dimension and interoperability relation, the respondents demonstrated no understanding of the relation between the two concepts. There was no indication of understanding syntactic, semantic, pragmatic, dynamic and conceptual levels in relation to the timeliness dimension.

Regarding the completeness dimension and interoperability relation, the respondents demonstrated no understanding of the relation between the two concepts except for the semantic interoperability level. There was no indication of understanding syntactic, pragmatic, dynamic and conceptual levels in relation to the completeness dimension.

Tabell 3. Respondents understanding of the relation between interoperability and data quality.

Data quality dimensions Interoperability levels	Accuracy	Timeliness	Completeness	Consistency
Technical interoperability	Not applicable	Not applicable	Not applicable	Not applicable
Syntactic interoperability	*"To avoid incorrect data, the Data hub owner makes validation and technical format control"* (Data hub owner)	No citation.	No citation.	*"There must be a high quality and control of the data structure otherwise data cannot be exchanged properly."* (Data hub owner)
Semantic interoperability	*"To have correct customer data is a big challenge."* (Development engineer)	No citation.	*"Information in the hub must be clarified and communicated."* (Operations officer)	No citation.
Pragmatic interoperability	No citation.	No citation.	No citation.	*"We need a handbook to explain how to retrieve and control data."* (Member of an expert group)
Dynamic interoperability	*"The data in the Hub must be right."* (Operations administrator)	No citation.	No citation.	*"A change in no matter what system must be communicated to or from the hub."* (Development engineer)
Conceptual interoperability	No citation.	No citation.	No citation.	No citation.

5. Discussion

From the analysis, it can be understood that the respondents' knowledge regarding interoperability and data quality varied from little to considerable compared to theory. However, those who demonstrated effective understanding appeared unsure of how to clearly explain interoperability and data quality. The analysis indicated that interoperability was harder for the respondents to define than data quality. For instance, some respondents alternated between interoperability and integration. While the theory clearly distinguishes between the data quality dimensions of accuracy, timeliness, completeness and consistency, the respondents could not clearly differentiate between them, which makes managing data exchange difficult. The respondents touched upon the idea of the relation between interoperability and data quality, but none expressed a clear vision about it, even though data quality and interoperability represent essential concepts in their daily work. The analysis indicated a split in understanding of the relation between interoperability and data quality. As a consequence, misunderstanding created information management silos. Specifically, no respondent expressed understanding of the relation between the two concepts at the conceptual level of interoperability, reflecting their common understanding of data, information, processes and context in relation to all data quality dimensions.

Furthermore, the timeliness dimension of data quality needs to be taken more into consideration at all interoperability levels. However, a particular focus is needed at the semantic and pragmatic interoperability levels in relation to all data quality dimensions, as they are shifting levels between technical and organisational details, making enterprises easier to align. In addition, most respondents described interoperability as a machine-to-machine exchange of data. This represents half of the truth, because the interoperability reaches only the semantic level, which is the technical view of interoperability. However, it is not only systems that exchange data; people in organisations has to inform each other as well in order to exchange data. That represents part of the organisational view of interoperability.

According to Chen and Doumeingts (2003), enterprises have to cope with internal changes from both a technical and organisational point of view. Based on previous theoretical and empirical illustrations, it appears indisputable that interoperability depends on exchanged data. Consequently, successful interoperability requires a common agreement on data quality. Data quality is considered a success factor for interoperability (Khisro and Sundberg, 2018). It is therefore essential for any enterprise to recognise and understand the importance of the relation between data quality interoperability. According to Sciore et al. (1994), for a successful data exchange, the individual systems must agree on the meaning of their exchanged data that is, the organisation must ensure interoperability. For instance, a data hub that collects data from different information systems of network enterprises filters, refines and shares data for the correct participants. Thus, achieving high data quality represents an essential ongoing process for interoperability in the development process. Enterprises that exchange data among numerous different systems require sufficient interoperability. In order for this exchange to be effective, the individual systems must also agree on the quality of their exchanged data.

6. Conclusion

This study sought to contribute to a deeper understanding of the relation between interoperability and data quality. To this end, this study has demonstrated that the relation between interoperability and data quality is mutually imbricated. In other words, data quality constitutes the backbone for interoperability, and so the exchange of quality data between systems, and the ability to use the other system's functionality, represents the core of the interoperability concept. If data lacks the proper attributes or a format that is usable by collaborating enterprises, the data's meaningfulness becomes compromised, and interoperability is not achieved. As such, data quality faced a make-or-break relationship with interoperability.

The relation between interoperability and data quality is thus concluded as follows:

- Interoperability can reveal poor data quality.
- The relation between semantic and pragmatic interoperability levels with all data quality dimensions requires particular focus, as these represent shifting levels between technical and organisational perspectives of interoperability.
- Interoperability can help enterprises prioritise what data is required.
- Data quality is a prerequisite for interoperability.
- Data quality can help identify interoperability obstacles.
- Data quality needs to be high enough to sufficiently meet the interoperability aim.

Consequently, the understanding is that high data quality is capable of decreasing complexity in a development process and increasing its reliability as this pave the way to successful interoperability. It is therefore important to consider the deeper understanding of the relation between interoperability and data quality when creating new businesses or revolutionising old ones in this digital age.

Future studies

This paper has embedded a path towards future research on, for instance, how to improve data quality to overcome interoperability obstacles in collaboration, or how the relation between interoperability and data quality affects digital innovation. Furthermore, this study has opened the door for future investigations into how the relation between interoperability and data quality affects decision-making.

Funding

This work was supported by the European Regional Development Fund, ISERV project [ID 20201030].

References

Ballou and Harold L. Pazer (1995). Designing Information Systems to Optimize the Accuracy-Timeliness Tradeoff. (1995). *Information Systems Research, 6*(1), 51-72.

Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. *ACM Computing Surveys (CSUR), 41*(3), 1-52.

Bengtsson, M. (2016). How to plan and perform a qualitative study using content analysis. *NursingPlus Open, 2*(C), 8-14.

Blake, R., & Mangiameli, P. (2011). The Effects and Interactions of Data Quality and Problem Complexity on Classification. *Journal of Data and Information Quality (JDIQ),2*(2), 1-28.

Camarinha-Matos, L.M., 2016. Collaborative smart grids–a survey on trends. *Renewable and Sustainable Energy Reviews, 65*, pp.283-294.

Chen, D. & Doumeingts, G. (2003). European initiatives to develop interoperability of enterprise applications—basic concepts, framework and roadmap. *Annual reviews in control*, *27*(2), 153-162.

Coronel, C., & Morris, S. (2016). *Database systems: design, implementation, & management*. Cengage Learning.

Collins dictionary https://www.collinsdictionary.com/dictionary/english Accessed december 2018.

Creswell, J. W., 2014. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. 4th ed. Los Angeles: Sage Publications.

Daraio, C., Lenzerini, M., Leporelli, C., Naggar, P., Bonaccorsi, A., & Bartolucci, A. (2016). The advantages of an Ontology-Based Data Management approach: Openness, interoperability and data quality. *Scientometrics*, *108*(1), 441-455.

Du, & Zhou. (2012). Improving financial data quality using ontologies. *Decision Support Systems, 54*(1), 76-86.

Duvier, Neagu, Oltean-Dumbrava, & Dickens. (2018). Data quality challenges in the UK social housing sector. *International Journal of Information Management, 38*(1), 196-200.

EC, Enterprise Interoperability Research Roadmap, Final Version (V. 4.0), 31 July, 2006.

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, *62*(1), 107-115.

Erlingsson, & Brysiewicz. (2017). A hands-on guide to doing content analysis. *African Journal of Emergency Medicine, 7*(3), 93-99.

Fenton, S., Giannangelo, K., Kallem, C. and Scichilone, R. (2013) "Data Standards, Data Quality, and Interoperability (2013 update)"*Journal of AHIMA,* 84 (11): 64-69

Ford, T. C., Colombi, J. M., Graham, S. R., & Jacques, D. R. (2007). *Survey on Interoperability Measurement*. AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH.

Gürdür, & Asplund. (2017). A systematic review to merge discourses: Interoperability, integration and cyber-physical systems. *Journal of Industrial Information Integration,*Journal of Industrial Information Integration.

Hazen, Weigel, Ezell, Boehmke, & Bradley. (2017). Toward understanding outcomes associated with data quality improvement. *International Journal of Production Economics, 193*, 737-747.

Herzog, F.J. Scheuren, & W.E. Winkler (2007). *Data quality and record linkage techniques*. New York: Springer Science+Business Media. 234 pp. US$44.95. ISBN: 978-0-387-69502-0.

IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries. (1991). Piscataway, USA: IEEE.

Khisro Jwan & Sundberg Håkan (2018): Enterprise interoperability development in multi relation collaborations: Success factors from the Danish electricity market, Enterprise Information Systems, DOI: 10.1080/17517575.2018.1528633

Kohlbacher Florian. (2006). The Use of Qualitative Content Analysis in Case Study Research. *Forum: Qualitative Social Research, 7*(1), 1-30.

Lindblad-Gidlund, K. (2005). Techno Therapy: A relation with technology. Research reports in informatics, 2005.

Madnick, S., Wang, R., Lee, Y., & Zhu, H. (2009). Overview and Framework for Data and Information Quality Research. *Journal of Data and Information Quality (JDIQ), 1*(1), 1-22.

Pandit, H.J., O'Sullivan, D. and Lewis, D., 2018. GDPR data interoperability model. In *the23rd EURAS Annual Standardisation Conference, Dublin, Ireland*.

McAfee, A. (2005): Will Web Services Really Transform Collaboration? MIT Sloan Management Review, Vol. 46, No. 2, pp. 78-84.

Panetto, H., 2007. Towards a classification framework for interoperability of enterprise applications. International Journal of CIM, 20 (8), 727–740.

Panetto, H., & J. Cecil (2013) Information systems for enterprise integration, interoperability and networking: theory and applications, Enterprise Information Systems, 7:1, 1-6, DOI: 10.1080/17517575.2012.684802

Panetto, H., Zdravkovic, M., Jardim-Goncalves, R., Romero, D., Cecil, J. & Mezgár, I. (2016). New perspectives for the future interoperable enterprise systems. *Computers in Industry, 79*, 47-63.

Parssian, A. (2006). Managerial decision support with knowledge of accuracy and completeness of the relational aggregate functions. *Decision Support Systems, 42*(3), 1494-1502.

Ozmen-Ertekin, & Ozbay. (2012). Dynamic data maintenance for quality data, quality research. *International Journal of Information Management, 32*(3), 282-293.

Tolk, A., & Muguira, J. A. (2003, September). The levels of conceptual interoperability model. In *Proceedings of the 2003 fall simulation interoperability workshop* (Vol. 7, pp. 1-11). Citeseer.

Turner, S. (2004). Defining and measuring traffic data quality: White paper on recommended approaches. Transportation Research Record No. 1870.

Umar, A., Karabatis, G., Ness, L., Horowitz, B., & Elmagardmid, A. (1999). Enterprise Data Quality: A Pragmatic Approach. *Information Systems Frontiers, 1*(3), 279-301.

Redman, T. (1996) *Data Quality for the Information Age.* Artech House, Boston, MA

Redman, T. (2013). Data Quality Management Past, Present, and Future: Towards a Management System for Data. Handbook of Data Quality. S. Sadiq, Springer: 15-40.

Reimers, K. (2001): Standardizing the new e-business platform: Learning from the EDI experience, Electronic Markets, Vol. 11, No. 4, pp. 231-237.

Sciore, E., Siegel, M., & Rosenthal, A. (1994). Using semantic values to facilitate interoperability among heterogeneous information systems. *ACM Transactions on Database Systems (TODS), 19*(2), 254-290.

Seale Clive, Giampietro Gobo, Jaber F.Gubrium, & Silverman David. (2004). *Qualitative Research Practice*. London: SAGE Publications.

Shankaranarayanan, & Cai. (2006). Supporting data quality management in decision-making. *Decision Support Systems, 42*(1), 302-317.

Song, Sun, Wan, & Liang. (2016). Data quality management for service-oriented manufacturing cyber-physical systems. *Computers and Electrical Engineering,*Computers and Electrical Engineering.

Storey, Dewan, & Freimer. (2012). Data quality: Setting organizational policies. *Decision Support Systems, 54*(1), 434-442.

Sullivan, H. and Skelcher, C., 2017. *Working across boundaries: collaboration in public services*. Macmillan International Higher Education.

Swedish Energy Markets Inspectorate (2017). Ny modell för elmarknaden, R2017:05. https://www.ei.se/Documents/Publikationer/rapporter_och_pm/Rapporter%202017/Ei_R2017_05.pdf Accessed May 07 2018.

Svenska kraftnät (2018). DEN SVENSKA ELMARKNADSHUBBEN - SAMLAD INFORMATION.https://www.svk.se/sok/?searchfield=2018+DEN+SVENSKA+ELMARKNADSHUBBEN+-+SAMLAD+INFORMATION Accessed september 15 2018.

Van Sinderen, Johnson, & Doumeingts. (2013). Computer in Industry Special Issue on "Interoperability and Future Internet for Next-Generation Enterprises" Editorial and state of the art. *Computers in Industry, 64*(8), 881-886.

Wand, Y., & Wang, R. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM, 39*(11), 86-95.

Wang, R., & Strong, D. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems, 12*(4), 5-33.

Wang, Reddy, & Kon. (1995). Toward quality data: An attribute-based approach. *Decision Support Systems*, *13*(3), 349-372.

Wang, R. Y., Ziad, M., & Lee, Y. W. (2006). *Data quality* (Vol. 23). Springer Science & Business Media. New York, Boston, Dordrecht, London, Moscow.

Wang, Wenguang, TOLK, Andreas, & Wang, Weiping. (2009). The Levels of Conceptual Interoperability Model: Applying Systems Engineering Principles to M&S.

Webster's Dictionary http://www.merriam-webster.com/ Accessed december 10 2018.

Yin, R.K., 2017. *Case study research and applications: Design and methods*. Sage publications.

Zhao, Kexin, & Xia, Mu. (2014). Forming Interoperability Through Interorganizational Systems Standards. *Journal of Management Information Systems, 30*(4), 269-298.

Zhu, & Wu. (2014). Assessing the quality of large-scale data standards: A case of XBRL GAAP Taxonomy. *Decision Support Systems, 59*(1), 351-360.

Zhu, H., Lee, Y., & Rosenthal, A. (2016). Data Standards Challenges for Interoperable and Quality Data. *Journal of Data and Information Quality (JDIQ), 7*(1-2), 1-3.