Evaluating the Configurations of Data Value Chain Elements for Information-Intensive Services in the Insurance Industry: An fsQCA Approach

Author: Savuth Pen

SNR: 2029156

S.pen@tilburguniversity.edu

Company: Allianz Benelux N.V.

January 12, 2023

Master of Information Management (MSc IM)

Faculty of Economics and Business Administration

Tilburg University

Master's Thesis supervisors:

Dr. Mylène Struijk (Tilburg University)

Dr. Carol Ou (Co-reader Tilburg University)

Coen Brinkman, Félix Vandervorst and Solon Pavlioglou (Allianz Benelux)

Management Summary

In the modern data-rich economy, value creation in services is increasingly driven by the utilization of big data. Insurance companies are incorporating more information-intensive services (IISs) to their business models, where value is primarily created through information exchanges during the process of transforming raw data into valuable information, referred to as the data value chain (DVC). Due to the variety in IISs, there are various DVC factors and elements that can compose the DVC for IIS (IIS-DVC). As a result, insurance providers encounter challenges in evaluating alternative DVC configurations and its impact on the value of IISs. This requires a DVC evaluation framework that assesses which configurations of IIS-DVC factors are resulting in a high value and therefore enables organizations to design a more effective IIS-DVC.

The purpose of this study is to address the research question: *which configurations of IIS-DVC factors lead to high value in information use in the insurance industry?* Value in information use is the degree to which the IIS is operationalized. An archetype of IIS-DVC is constructed that represents the key IIS-DVC factors and elements common in IISs. This framework serves as the foundation for the empirical study, which uses 13 IIS-DVC configurations from the insurance firm Allianz Benelux. The fuzzy-set QCA method, which combines qualitative and quantitative analysis, is used to discover patterns among these configurations and to identify the core IIS-DVC elements that play a significant role in determining the degree of operationalization.

As a result, five IIS-DVC configurations, A1-A5, are found to lead to high degree of operationalization, whereas four configurations, B1-B4, results in low degree of operationalization (figure 3). In addition, the findings indicate type of information as a core element as it is key in determining whether the configurations lead to low or high outcomes. Follow-up research may examine why the type of information used to operate the IIS in insurance businesses has a significant impact on the IIS's performance.

Contents

1 Introduction

1.1 Problem Indication

The services sector is the fastest-growing and most dominant sector of the European economy (Rubalcaba, 2012). As the total volume of data grows exponentially, information continues to play a significant role in this sector and is the driving force behind many business operations (Apte et al., 2010). Information is especially important for organizations that provide mainly information-intensive services (IISs), such as financial services, insurance and engineering services. An IIS is a service where value is mainly co-created via information interactions between the service provider, its partners and information users instead of through tangible engagements (Apte et al., 2010; Lim et al., 2018).

In the modern data-rich economy, with continuous digital transformation, organizations are trying to better understand 'data-based' or 'data-driven' business models to enhance the engagement of IISs and hence extract more value (Morabito et al., 2015; Schüritz et al., 2017). These data-driven business models create value for information users through the data value chain (DVC), which represent a set of factors necessary to convert raw data into useful information and valuable assets (i.e., value in information use) (Curry, 2016; Faroukhi et al., 2020; Miller & Mork, 2013).

However, these factors are composed of elements that vary depending on the service's nature and purpose, and therefore complicating the evaluation process of data-based services (Elragal & Klischewski, 2017; Faroukhi et al., 2020). For example, cloud-services commonly employ push data collection mode, which is an element of data collection (DVC factor), while smart grid service, a service to optimize the production, distribution and use of electricity, uses pull data collection mode to gather raw data from sources (Daki et al., 2017; Schmidt & Möhring, 2013). This shows the diversity of sets of elements that can comprise the DVCs.

Studies on the use of DVC factors and its elements for IISs are mostly focused on defining and explaining these components, but they seldom evaluate whether these selected factors and elements are effective or not (Daki et al., 2017; Lim et al., 2018; Miller & Mork, 2013; Schmidt & Möhring, 2013). By understanding and evaluating these elements and the combinations in which they occur, it enables organizations to configure their DVC elements in such a manner to achieve high benefits from IISs (Lim et al., 2018). However, the impact that a combination of DVC elements has on the value of IIS remains unexplored in the literature.

1.2 Problem Statement

Organizations frequently encounter challenges in designing an effective DVC and are therefore more likely not to achieve the intended value from IISs. These failures are primarily caused by the deficiency of an evaluation framework for DVCs that examines the relationship between a set of elements and the value gained through the use of IIS (Serhani et al., 2016). Moreover, because of the diversity in DVC concepts, there are various elements that can configure the set of factors needed to transform the data, complicating the development of an evaluation framework.

These shortcomings and obstacles are common in the insurance industry since the digital transformation has an impact on all parts of the insurer's value chain, including underwriting, pricing, claims and data services (Albrecher et al., 2019; Puschmann, 2017; Stoeckli et al., 2018). Consequently, insurance products and services are becoming more customized and usage-based, expanding the possibilities for the configurations of DVC elements (Albrecher et al., 2019). For instance, because the services are more usage-based, the DVC requires a different data processing method, such as stream processing, to ensure that services are available when needed (Carbone et al., 2015).

Due to the growth of alternative configurations, insurance companies struggle to comprehend which IIS-DVC configurations are effective and which are not (Albrecher et al., 2019; Stoeckli et al., 2018). As a consequence, reaching the maximum potential of big data in the insurance business is constrained (Soldatos et al., 2022). Evaluating the IIS-DVC can help with this complexity by providing insurance providers insight into the DVC elements that are sufficient to explain the outcome of interest and which ones are not, resulting in more value extracted from big data (Albrecher et al., 2019; Janssen & Sol, 2000; Soldatos et al., 2022) By identifying which elements are key for a high service value, firms are able to modify or improve their DVC for obtaining more value from IIS. Additionally, an evaluation framework supports the design and development of new IISs by assisting in the identification of key IIS-DVC elements and giving insight into the relationship with the IIS value (Janssen & Sol, 2000; Lim et al., 2018).

However, despite the insurance industry being considered to be very data-intensive, it has received relatively little attention in big data and DVC publications (Sadovskyi et al., 2014). The complexities of its operations and legal restrictions, which restrict research and limit the potential for new implementations, may account for the lack of developments (Sadovskyi et al., 2014). Hence, research on the IIS-DVC and their impact on the value of IISs in the insurance industry is needed (Kühne & Böhmann, 2019; Lim et al., 2018).

Therefore, this study aims to identify DVC factors and elements that are commonly used in the IIS-DVC in order to evaluate various IIS-DVC configurations for insurance IISs. It contributes to a more complete understanding of the IIS-DVC and its influence, and helps insurance providers design a more effective IIS-DVC. For identifying the IIS-DVC factors and elements, a literature review combined with interviews will be conducted to establish a complete DVC for IIS. Then, a combination of qualitative and quantitative method, the fuzzy-set qualitative comparative analysis (fsQCA) method, will be applied on several IISs operating at the insurance company, Allianz Benelux. With this, a set of IIS-DVC elements can be identified that reflects a certain outcome of interest.

To summarize, the objective of this research is to evaluate alternative IIS-DVC configurations and determine which IIS-DVC elements are key in achieving a certain value in information use. As a result, insurance companies are provided with knowledge about their current IIS-DVC designs as well as opportunities for improvement in order to extract more value from IISs. The following research question captures the objective of this research:

- *Which configurations of IIS-DVC factors lead to high value in information use in the insurance industry?*

1.3 Research Questions

To answer the research question, the concept of DVC needs to be explained in order to identify the key element of the IIS-DVC. Following that, a framework is built to allow for the examination of the relationship between the set of IIS-DVC elements and the IIS value. Based on the course of this research, the following sub questions have been formulated to answer the research question:

Sub questions:

1. What is the DVC?
2. Which elements of the factors are key in the DVC for IISs?
3. How can the IIS-DVC elements and outcomes be measured?
4. What are the configurations of IIS-DVC elements for high and low value in information use for insurance IISs?

1.4 Research Method

The purpose of this study is to identify various configurations of IIS-DVC elements and explore the relationship to the value of information use. Hence, the research is divided into two parts: the first part identifies the IIS-DVC elements and how to measure them through a literature review, and the second part performs an empirical study to examine the relationship.

The literature review is conducted to find the key IIS-DVC elements. While there are few studies on the DVC for IIS, this study builds on the nine-factor IIS-DVC framework developed by Lim et al. (2018). They have determined nine factors: data source (1), data collection (2), data (3), data analysis (4), information on the data source (5), information delivery (6), information user (7), value in information use (8), and provider network (9) that are essential in the value creation in IISs (Lim et al., 2018). Since this study has thoroughly explored which factors characterize the IIS-DVC, the framework is used as the foundation for identifying additional IIS-DVC key elements.

For that, literature from the fields of Information Management (IM), Information Systems (IS), big data and services is mainly extracted to comprehend the DVC concept as well as its factors and elements. In addition, Allianz Benelux data product managers, data engineers and data scientists are interviewed to obtain additional information on which DVC elements are essential for an IIS. The first three sub-questions can be addressed based on the literature study and interviews with experts.

The second component of the research aims to answer the fourth sub-question, which is about the discovery of the optimal configuration of IIS-DVC elements for a high value in information use. Based on the configuration approach of the study, a theoretical framework will be constructed using the contingency fit theory. This framework serves as the foundation for analysis performed with the fuzzy-set QCA method. Fuzzy-set QCA is designed to explain causality in complex situations with an intermediate number of cases. It is suitable for situations where the number of cases is too low for statistical analysis (Pappas & Woodside, 2021). Therefore, the method uses both qualitative and quantitative approaches and hence suitable for this research. The QCA method has three different variants: crisp-set QCA (csQCA), multi-value QCA (mvQCA), and fsQCA (Zhang et al., 2022). The fsQCA distinguishes itself from other methods by enabling to score factors between 0 and 1, where the csQCA method is limited to binary variables. For that reason, the fsQCA method will be applied in this research. As input for the analysis, 13 IISs operating at Allianz Benelux will be used.

1.5 Academic and Practical Relevance

Academically, this research contributes by providing an evaluation framework that examines the relationships between IIS-DVC and the value in information use. Current research is mainly focused on defining DVC concepts and their use in different contexts but not evaluating the performance of the value chain (Curry, 2016; Daki et al., 2017; Miller & Mork, 2013; Schmidt & Möhring, 2013). Although data-based value creation and its performance have been on the agendas of academics and practitioners, there are still limited studies conducted on (Ekbia et al., 2015; Ostrom et al., 2015; ur Rehman et al., 2016). Consequently, there is ambiguity and incomprehension about why certain DVC configurations are selected in a given context, without understanding what the alternative configuration can lead to. Since there is no evaluation approach for comparing configurations on their performances, this study contributes by developing an archetype of IIS-DVC that can be used to compare IIS-DVC configurations with other DVC configurations from different contexts.

From the practical point of view, this study is mainly useful for data product managers, data engineers, service designers and data scientists working at Allianz Benelux but also relevant for other IIS-oriented organizations such as financial institutions. In general, the examination on the IIS-DVC operating at Allianz Benelux provides data product managers with insight into the high value set of IIS-DVC elements. The findings of the analysis indicate which IIS-DVC elements are key in producing a certain value in information use. This enables insurance providers to incorporate this knowledge into the design of future IIS-DVC in order to improve the service value or to modify their current IIS-DVC configurations.

1.6 Structure of the Thesis

The thesis is organized in seven sections and starts off with the literature review related to the DVC, IIS and value creation in section 2. In section 3, the theoretical framework is constructed by developing the archetypes of IIS-DVC based on the contingency fit theory. Moreover, the section provides the selection of key elements that is included in this research. Subsequently, section 4 describes the research method, which includes determining metrics for all IIS-DVC elements and employing the fsQCA method. Section 4, the results, examines the different configurations of IIS-DVC elements. In section 5, the findings will be discussed and explained how the findings could affect the theories and practices. Then, section 7 covers the limitations and recommendations for future research. Lastly, the research is concluded in section 8.

2 Literature Review

This chapter will elaborate on the DVC concepts, DVC factors and their elements, all of which are necessary for comprehending the study. It starts with explaining the DVC concept, followed by an overview of relevant DVC concepts and then a comprehension of its challenges and limitations. The second part of this section focuses on the IISs and the factors of the IIS-DVC derived from the framework of Lim et al. (2018). Moreover, this review explains what is meant by value in information use in this context. Finally, based on the literature review, the key elements of each IIS-DVC factor are identified and explained.

2.1 The DVC Concept

Data-driven value creation necessitates a comprehension of the DVC concept which is derived from Porter's value chain concept (Porter's, 1985). A value chain defines the activities that contribute to and create value from the goods and services being produced. The concept was first mainly adopted in the manufacturing industry that enables organizations to discover which of their primary and supporting activities creates value for the customers (Simatupang et al., 2017).

Nevertheless, despite the value chain concept's effectiveness in achieving this goal, products and services nowadays are more data-driven and intangible. Therefore, several IS researchers have presented revised conceptions of the value chain that take the digital aspect into account, referred to as the DVC (Curry, 2016; Faroukhi et al., 2020; Miller & Mork, 2013). The DVC is defined as a set of factors that are required to transform raw data into useful insights and valuable assets (Curry, 2016). Another definition is a framework that guides the structured integration of big data to create valuable information for corporate decision-making (Miller & Mork, 2013). Hence, in this study, the DVC is defined as a configuration of factors necessary to obtain value from the usage of useful information or data-based service. Factors refers to both components and activities that characterize data-based value creation.

2.1.1 Current Research on the DVCs

On the contrary, conducted research focuses mostly on the advantages of data-based services, but there are limited studies that explore the performance of the DVC in a configuration approach. Table 1 summarizes the limited research on DVC evaluation. By contrast, scholars have mainly been concerned with defining and explaining the DVC factors in different contexts. During the review of the literature, it was noticed that there are variances in both the set of DVC and the terminology of some factors due to the different purposes for which the DVCs are used. However, the definitions and explanations of the DVC are quite consistent across studies.

Table 1: Overview of different DVC concepts.

References	Series of DVC factors	Context
(Miller & Mork, 2013)	Data discovery Data integration Data exploitation	Structured integration of big data to create valuable information for corporate decision-making.
(Curry, 2016)	Data acquisition Data analysis Data curation Data storage Data usage	Big Data Value Chain model to generate value by analyzing the information flow in information systems, and the integration into the ecosystem.
(Schmidt & Möhring, 2013)	Data acquisition Data transformation Data storage Data processing Data analysis Data visualization	Examining the potential for deploying a DVC using cloud-services.
(Hu et al., 2014)	Data generation Data acquisition Data storage Data analytics	From a system perspective, develop an understanding of diverse approaches and mechanisms used in the BDVC activities.
(Lim et al., 2018)*	Data collection Information creation Value creation	Presented a nine-factor framework that defines the data-based value creation in IISs.
(Daki et al., 2017)	Data source Data integration Data storage Data analytics Data visualization	BDVC for smart grid; optimizing the production, distribution and use of electricity by information technologies.

* The following factors are grouped under the headings, data collection, information creation and value creation: data source, data collection, data, data analysis, information, information delivery, information user, value in information use and provider network (figure 1).

Faroukhi et al. (2020) examined these differences and similarities in factors between fifteen DVC adopted concepts. According to their findings, the following factors are common in use: data generation, data acquisition, data storage, data analysis and data visualization or usage (Faroukhi et al., 2020). In this paper, the focus is on value-created factors for IISs and thus the IIS-DVC factors from Lim et al. (2018) are applicable. Due to the lack of research on the DVC for IIS, the framework of Lim et al. (2018) is only available for use. These IIS-DVC factors are compared to the commonly used set of DVC factors in order to check its completeness, illustrated in figure 1, and to discover additional elements that characterize the IIS-DVC factors, examined in section 2.3. Beyond that, the figure indicates that the DVC factors are essentially the same.

2.2 Information-Intensive Services (IISs)

IISs are defined as services that are primarily co-created through information interaction between the service provider, its partner and the information user that occurs throughout the IIS-DVC (Lim et al., 2018). An example of an IIS is the motor claims handling service in the insurance industry. The service operates by collecting, processing and distributing information from and to several parties such as the information user, claim inspectors, mechanics, motor experts and the insurance company employees (Geoghegan et al., 2021). Other studies across a range of discipline have demonstrated the popularity of IISs, for example, the emergency situation monitoring service (ESMS) and the building energy management services (BEMS), all of which are mainly built on the exchange of information between parties (Dounis & Caraiscos, 2009; Kim & Chung, 2015).

The occurrence of IISs in various industries demonstrates differences in the configuration of the IIS-DVC elements. Whereas humans (an element of data collection) play a key role in data collection in the ESMS, pre-programmed queries are applied in the case of BEMS. Despite these variances, Lim et al. (2018) studied 149 IISs in six different contexts, including health care and insurance service design, data platform service design and student wellness enhancement service design, to identify the IIS-DVC factors and elements that were commonly used in the cases. Hence, these factors and elements are considered to be universal for IIS and thus appropriate for this research. However, the next paragraph will elaborate on this research to investigate its completeness in order to construct a comprehensive evaluation framework.

2.3 Data Value Chain for IIS

The nine factors that define data-based value creation in IIS are parts of three main activities, namely data collection, information creation and value creation (figure 1) (Lim et al., 2018). For the first activity, data source and data collection were identified as the factors. Relating it back to the commonly used activities of the DVC ascertained by Faroukhi et al. (2020), data source can be thought of as the data generation while data collection reflects the data acquisition. The second activity, information creation, is divided into data, data analysis and information (on the source) where data mirrors the data storage. The third activity, value creation, consists of information delivery and information user, which all are related to data visualization and usage, respectively.

All of these IIS-DVC factors directly contribute to value co-creation by supporting the transformation of data to information. In addition, the IIS-DVC concept of Lim et al. (2018) also takes into account the value in information use and the provider network as factors. The value in information use is the dependent variable that results from the execution of all DVC activities. Although it does not directly contribute to the value co-creation process, the factor is necessary to evaluate the performance of the IIS-DVC. Furthermore, the provider network refers to the network of the service provider, their partners. In some cases, the service provider relies on outsourced assets such as sensors for data collection or consultants for the data analysis. This factor acts as a moderator and is thus excluded because the focus of this study is on the configuration of the IIS-DVC instead of examining the moderating effects.

As a result, the following factors for IISs are considered: data source (1), data collection (2), data (3), data analysis (4), information (5), information delivery (6), and customer (7). The value in information use (8) functions as the dependent variable (Faroukhi et al., 2020; Lim et al., 2018). The following paragraphs will examine these factors and their role in value co-creation.

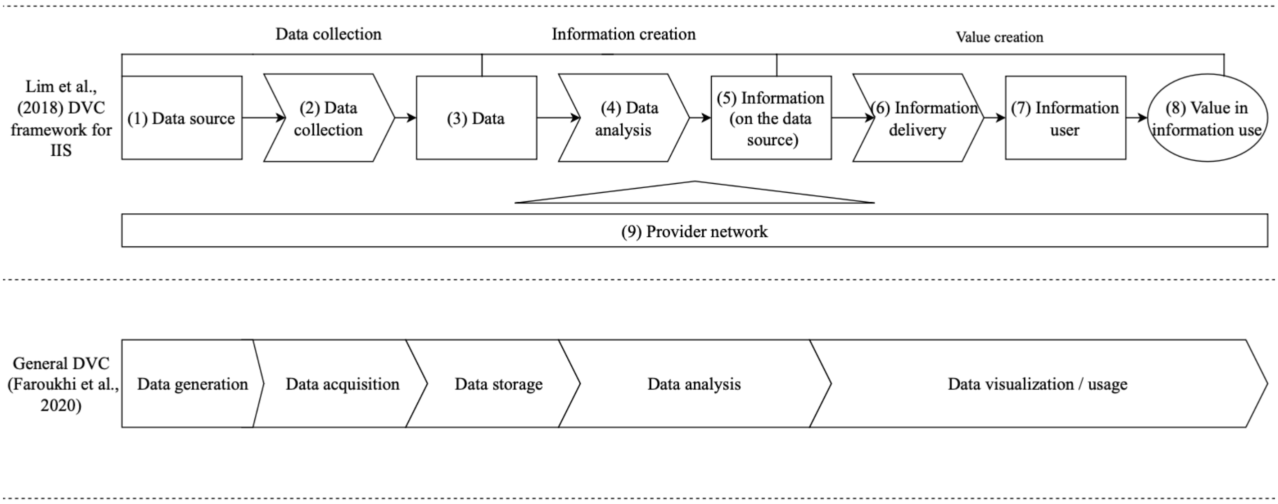

Figure 1: Nine factor-based DVC for IIS of Lim et al. (2018) compared to the general DVC ascertained by Faroukhi et al. (2020).

2.3.1 Data Source

For creating value via data, it has to be generated first. There is a wide variety of sources that generate data in different formats, known as one of the hallmarks of big data (Buhl et al., 2013; Sagiroglu & Sinanc, 2013). Nevertheless, the sources can be classified into those where data is generated by humans and sources where data is created by equipment, systems or sensors (Faroukhi et al., 2020; Lim et al., 2018). In terms of the former, humans are considered as a data source when their actions or behaviors generate data for the IIS. Additionally, the human body is in itself a source of data, which is prevalent in medical services. For example, customers' behavioral data and bio signals were employed in fitness monitoring services (Takacs et al., 2014).

On the contrary, data generation by operating systems, servers, applications and sensors, referred to as engineering systems, happen with the use of computational or electronic devices (Lim et al., 2018). Besides the type of source, data can be categorized into unstructured, semi-structured or structured formats (Faroukhi et al., 2020; Lim et al., 2018). In this factor, data are in their rawest form and offer no useful insights (Hu et al., 2014).

2.3.2 Data Collection

Data collection takes place after the data generation process in the data sources. In the domain of IM, data collection is considered as a subprocess of data acquisition. Data acquisition is an aggregate term for the sub-processes of data collection and data pre-processing, which is, respectively, the order of execution (Hu et al., 2014). The act of gathering data from sources (human or engineering systems) summarizes the data collection process. Designing an appropriate data collection method that depends on the source's characteristics and the purpose for which the data will be used, is crucial in this stage since it affects the ability and result of the data analysis process (Curry, 2016; Hu et al., 2014).

Sensors, log files and web crawlers are commonly utilized data collection methods (Hu et al., 2014; Silva et al., 2019). Sensors are suitable for transforming physical properties, such as temperature, humidity and pressure, into understandable digital data. Another widely used collection method is the log files, for instance, the web server log file that records user actions on a website (Ingram, 1999). The technique documents all the activities in an engineering system and stores them in a specific file format depending on the requirements from the analytic stage. At last, the web crawler is a software-based method that collects all kinds of data from webpages for a search engine (Kausar et al., 2013). This form of collecting is fully automated as the software operates on its own (Dhenakaran & Sambanthan, 2011).

To conclude, these three main data collection methods and others can be subdivided into two kinds of modes, the pull mode and the push mode (Hu et al., 2014). Pull-based systems collect data based on the interval and frequency set by the user, making this approach dependable on user's queries (Sathe et al., 2013). In contradiction, the push-based systems autonomously gather data at the command of the central server or base station. There is an underlying model between the central server and engineering system that sets the conditions for when the system has to collect data (Sathe et al., 2013). Additionally, due to the large variety of techniques, the methods can be categorized into the degree of human involvement in order to simplify the analysis of the IIS design (Lim et al., 2018).

Subsequent to data collection, data pre-processing starts as collected data has arrived at the data center. It aims to integrate raw data from different sources and deliver complete and high quality data for the analysis stage. This process incorporates cleansing the data by eliminating errors (e.g., inaccurate, incomplete and unreliable data); data redundancy elimination for consistency in data; data reduction to optimize storage capacity; and transforming data into an appropriate format, all of which contribute to a more meaningful dataset (Faroukhi et al., 2020; Ridzuan & Zainon, 2019; Zhang & Ansari, 2013). This data pre-processing process has two typical designs: batch processing and stream processing. Batch processing is a method of processing data in batches that can be determined by data size or the period over which the data is collected. Stream processing, on the other hand, continually processes data close to the moment it was generated, also referred to as real time processing (Carbone et al., 2015; Faroukhi et al., 2020).

2.3.3 Data

After going through the first two factors, the data will be stored for further processing. The previous activities, especially the data transmission and pre-processing, may significantly affect the quality of data that will be used during the analysis stage (Faroukhi et al., 2020; Lim et al., 2018). In addition, the design of the storage infrastructure and system also affects the degree of data quality (Padgavankar, Gupta, et al., 2014). The infrastructure is responsible for a reliable storage capacity while storage systems must enable scalability to ensure the accessibility of data for querying and analyzing (Hu et al., 2014; Padgavankar, Gupta, et al., 2014). Hence, assessing the quality of data can be imposed as an overarching evaluation method to evaluate this factor.

2.3.4 Data Analysis

Data analysis is often regarded as the most important and influential DVC factor since its goal is to extract meaningful value from data (Hu et al., 2014). Although there are many analytical methods to achieve different objectives, the methods are classified into three broad areas: descriptive, predictive, and prescriptive analysis (Maimon & Rokach, 2005; Roy et al., 2022).

Descriptive analysis focuses on comprehending the present and past circumstances by analyzing the data (Saggi & Jain, 2018). Therefore, the approach is mainly suitable for modelling the data for efficiency, regressing certain trends, and visualizing insights. In contrast, predictive analysis relies more on identifying patterns in historical data in order to predict future events or trends (Saggi & Jain, 2018)). In regards to trends, statistical techniques such as regression are typically employed for this type of modelling. Data mining techniques, including machine learning and classification, are more appropriate for learning patterns and discovering insight from them (Maimon & Rokach, 2005). Prescriptive analysis expands the range of possibilities. A relevant difference from predictive analysis is that prescriptive models' computations are not necessitated to be time series-related. For instance, prescriptive models seek for changes in climate in order to discover the best outcome possible given specific constraints, whereas predictive models do not aim to modify the climate but rather offer insight into the developments over time to moderate its influence.

Regardless of the analysis model employed, Lim et al. (2018) characterize the data analysis for IISs as either a fully automatic process performed by a pre-programmed algorithm or a process that involves humans since it requires their professional expertise. The IISs associated with healthcare are an example of a sector that is well-suited to adopting human-involved methods. Other IISs that simply need algorithms require humans to develop and maintain but the actual analysis is carried out entirely mechanically.

Thus, the alternatives for designing the analysis model largely rely on the objectives and requirements of the IIS. Because of the diverse objectives and requirements, the design of these models is a challenging aspect of data modelling (Nasser & Tariq, 2015). Additionally, the nature of big data and its complexity, difficulties in scaling queries and the complexity of integrating analytics tools are examples that also contribute to the design's challenges (Fan et al., 2014; Nasser & Tariq, 2015). Therefore, the level of robustness is an effective evaluation tool for assessing the model's resilience to these challenges (Tallon, 2011).

2.3.5 Value in Information Use

Data analysis activities produce information that conveys knowledge about the data source. The type of information resulting from the data analysis can be descriptive, predictive, or prescriptive based on the purpose of use (Lim et al., 2018). The delivery of this information to end users can be accomplished by various methods such as reporting, applications, or devices, whereby the extent of human involvement is also a significant aspect of information delivery in IIS (Lim et al., 2018). Likewise, the choice of delivery depends on the information user's objectives. The set of IIS-DVC factors has only created value if the information or assets, obtained from that IIS-DVC, are actually used. Therefore, it is important that the IIS has at least one user (information user). Moreover, there are variances in how IIS information or assets can be used, such as reporting, for dashboards, decision makings or serving as a data product for subsequent operations inside an organization, resulting in differences in value in information use. Because not all IISs lead directly to financial benefits, the IISs that have a supporting function, value in information use is defined in this study as the degree of how well the IIS operates. Thus, the current level of engagement compared to the intended usage level also known as the degree of operationalization (Vuorinen et al., 1998).

3 Theoretical Framework

As the purpose of the study is to evaluate various configurations in the IIS-DVC, an archetype of IIS-DVC will be constructed in this section using the contingency fit theory. The theory, which asserts that internal and external contingency factors impact the design and effectiveness of organizations, is therefore used as the underlying conceptual framework (Donaldson, 2006; Drazin & Van de Ven, 1985). Following the example set by J.-N. Lee et al. (2019), a theoretical framework is constructed and shown in figure 2 by incorporating the IIS-DVC elements congruent with the value in information use and configuring the fit between this congruence and the contingency factor. Moreover, these factors and their key elements are explained in this section. So, this section serves as the foundation for the analysis.

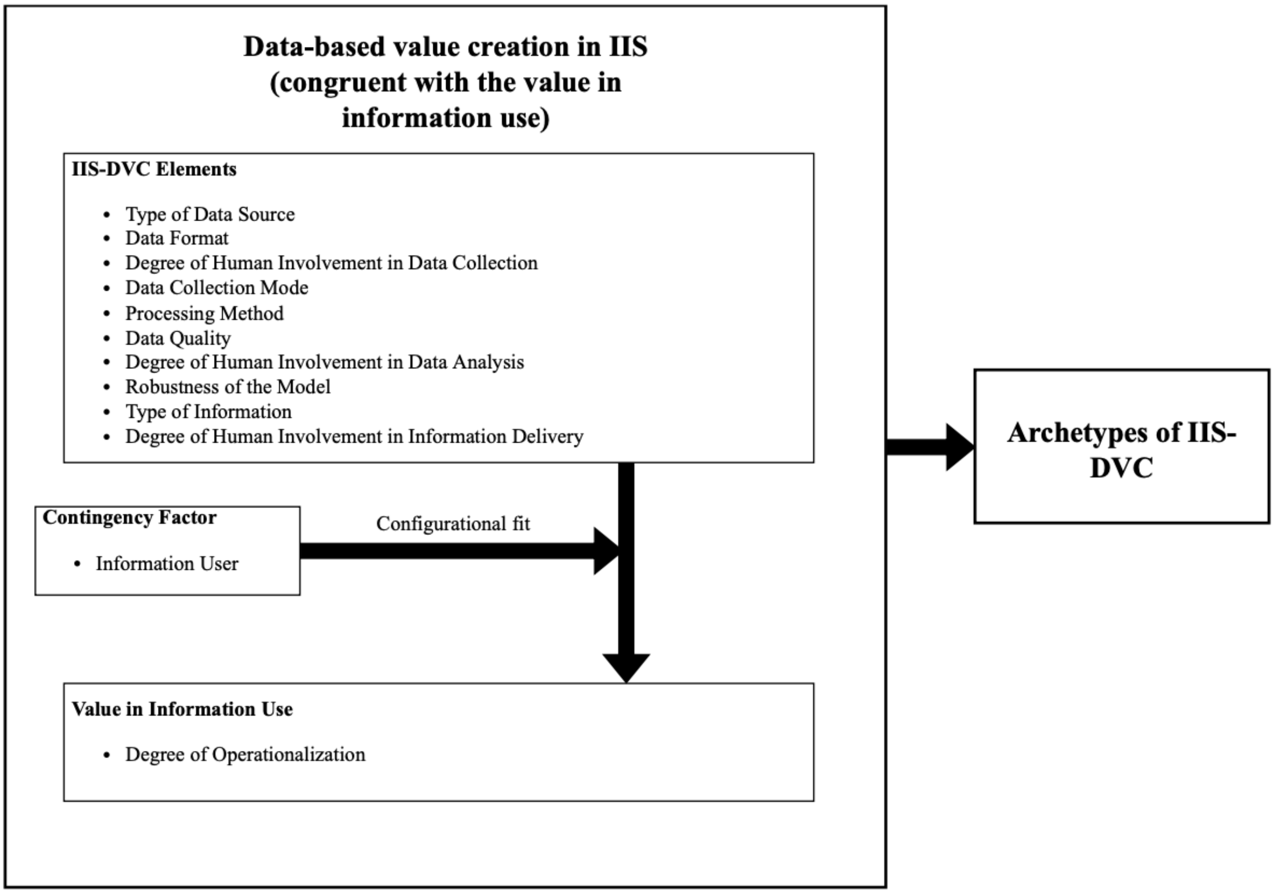

Figure 2: Theoretical Framework of the Archetypes of IIS-DVC

3.1 Archetypes of IIS-DVC

Archetypes are used in analysis to examine how a phenomena's attributes are configured and to find alternative models that can account for the phenomenon (Eisenack et al., 2019). Thus, a configuration approach that examines different sets of attributes, yet supportive of one another and frequently taking place together, in relation with the phenomenon (Eisenack et al., 2019; J.-N. Lee et al., 2019). The approach allows scholars to study on how these attributes interact with one another and how it can be employed in conjunction to achieve the desired output (Fiss, 2011; J.-N. Lee et al., 2019).

Archetype of data-based services refers to conceptual models or modular services that offer a systematic view of different opportunities for creating customer value through data and data analysis (Hunke et al., 2020; Schilling et al., 2017). Consequently, organizations are able to configure the DVC elements in a manner that best fits the desired value for the customer. Hence, in this research, the archetypes of IIS-DVC are defined as modular IIS-DVC models where IIS-DVC elements are logically coherent with the value in information use (or degree of operationalization) and fitting the contingency factors, information user. Based on the archetype theory, the optimal approach to explain the value in information use is to analyze different configurations of elements rather than just one of them individually.

Table 2: Configuration elements of the Archetypes of IIS-DVC

Independent factors	**Elements**	**Measures**	**References**
(1) Data Source	Type of data source	Human system or engineering system	(Lim et al., 2018)
	Data format	Unstructured, semi-structured or structured data	(Faroukhi et al., 2020; Hu et al., 2014)
(2) Data Collection	Degree of human involvement	Fully human involved, semi-human involved or fully automated	(Dhenakaran & Sambanthan, 2011; Ingram, 1999; Lim et al., 2018)
	Data collection mode	Pull mode or push mode	(Hu et al., 2014)
	Processing method	Batch processing or stream processing	(Carbone et al., 2015; Faroukhi et al., 2020)
(3) Data	Data quality	Accuracy, completeness, timeliness and consistency	(Ballou & Pazer, 1985; Y. W. Lee et al., 2002)
(4) Data Analysis	Degree of human involvement	Human-involved analysis or automated analysis	(Lim et al., 2018)
	Robustness of the model	Likert scale of 5: strongly disagree - strongly agree	(Fan et al., 2014; Nasser & Tariq, 2015; Tallon, 2011)
(5) Information (on the source)	Type of information	Descriptive, predictive or prescriptive information	(Lim et al., 2018; Maimon & Rokach, 2005; Saggi & Jain, 2018)
(6) Information Delivery	Degree of human involvement	Fully human involved, semi-human involved or fully automated	(Hu et al., 2014; Lim et al., 2018)
Contingency Factor	**Elements**	**Measures**	**References**
(7) Information User	Type of information user	Internal or internal and external users	(Lim et al., 2018)
	Size of information user	Number of information user	(Curry, 2016)
Dependent Factor	**Elements**	**Measures**	**References**
(8) Value in Information Use	Degree of operationalization	Level of engagement/usage compared to the targeted level of usage	(Vuorinen et al., 1998)

3.1.1 The Contingency Fit

The theoretical framework of the IIS-DVC archetype was developed using the contingency fit theory (figure 2). The theory claims that there is no ideal design to operate an organization or service, rather, the optimum course of action depends on how effective the fit is between contingency factors and others (Donaldson, 1996; Jesmin & Hui, 2012). In line with the theory, no single IIS-DVC configuration is wildly successful for all IISs. Thus, the optimum configuration changes depending on the fit with contingency factors such as the size and type of information user. Hence, the approach allows finding patterns in multiple IIS-DVC configurations and observing the relation with the value in information use.

3.2 Configurations of the IIS-DVC Archetype

To find patterns in the archetypes of IIS-DVC, the factors and its key elements are defined and listed in table 2. First, the independent factors (1-6), all of which directly contribute to value creation and are characterized by elements that, according to literature, constitute the significant determinant of the factor. These IIS-DVC factors and most of their elements are derived from the DVC framework of Lim et al. (2018). They studied value creation in IIS in a wide range of contexts, making these factors and elements universal for IIS and thus useful for this research. However, with regards to the IIS-DVC elements, it is a limited representation of the factors. Based on existing research in the domain of DVC, the IIS-DVC elements are expanded to provide a more complete description of each factor's key elements (table 2).

3.2.1 Independent factors

As a result, the type of source and data format are defined as the key elements of data source (1). The source type, in particular, is an essential part in the IIS-DVC since understanding where data is generated is closely related to the objective of the service (Lim et al., 2018). Moreover, as the variety in big data expands (data format), data sources become more diverse, requiring the design of systems that generate, collect and analyze data to be format-dependent (Faroukhi et al., 2020; Hu et al., 2014). Thus, these elements affect how the subsequent IIS-DVC elements are designed and therefore identified as key elements in this study. Contrary to other data source elements, such as response time and accuracy, are not included since they are concerned with the performance of the data source (i.e., dependent on data format and source type) rather than the design on which this research focuses.

For the data collection (2), three main elements are defined, namely the degree of human involvement, data collection mode and processing method. All elements are closely related to the efficiency and productivity of the service (i.e. degree of operationalization). The degree of human involvement indicates how automated the data collection process is, whereas the data collection mode influences the efficiency of querying data from the data source, and the processing method determines the processing time, data preparation and data flow to storage (Carbone et al., 2015; Dhenakaran & Sambanthan, 2011; Faroukhi et al., 2020; Hu et al., 2014; Ingram, 1999; Lim et al., 2018).

Furthermore, data (3) reflects the scope and possibilities of an IIS and consequently influences the degree of operationalization (Lim et al., 2018). Because of this, data quality is considered as a key element of data since its ability to assess the scope and possibilities by evaluating the dimensions:

accuracy, completeness, timeliness and consistency (Y. W. Lee et al., 2002; Padgavankar, Gupta, et al., 2014). These four dimensions are used by the majority of scholars (Cappiello et al., 2004; Y. W. Lee et al., 2002). Additionally, in this study, data is used as input for analytical models, whose outcomes are partly affected by the accuracy, completeness, timeliness and consistency of data (Padgavankar, Gupta, et al., 2014).

Subsequently, the elements characterizing the data analysis (4) is considered to be the degree of human involvement and the robustness of the model. Similar to factor 2, these elements affect the effectiveness of the analysis model (Fan et al., 2014; Lim et al., 2018; Nasser & Tariq, 2015). For the information on the source (5), the information type is defined as a key element (Hu et al., 2014; Lim et al., 2018). Lastly, the main element of information delivery (6) is as well the degree of human involvement, which differentiates between fully human involved, semi-human involved or fully automated (e.g., dashboards and reports). The review of the available literature led to the identification of all six factors and their elements.

3.2.2 Dependent factor

Congruent with the independent factors, the dependent factor, value in information use (8), is the degree of operationalization, which determines the level of usage compared to its targeted level. Although different metrics can be used to determine the value in information use, the operationalization is favored over other measures due to the nature of the IISs researched on. Affected IISs do not always directly lead to financial benefits but can instead initiate other business operations or decision making. In this case, comparing economic benefits is challenging since certain IISs do not result in benefits, which can lead to misleading findings. By contrast, the degree of operationalization is commonly used as measurement for the productivity of services, and more suitable for comparing the value in information use among the IISs (Vuorinen et al., 1998).

3.2.3 Contingency factor

Lastly, the information user (7) is designated as the contingency factor because of its relatedness to the IIS-DVC elements and their influence on the value in information use and it does not directly contribute to value creation. For instance, the type of information user is associated with the information delivery since internal users may require low human involvement whereas external users may need additional assistance due to knowledge gaps. Furthermore, designing DVC for new services necessitates a thorough understanding of the user's requirements, expectations and intended uses that vary between internal and external users (J. H. Lee & Waterman, 2012). Studies on service design emphasize the significance of user co-creation in service design (Mager, 2009; Teixeira et al., 2012). According to this, the information user is incorporated in the IIS-DVC configuration and functions as a contingency factor instead of an independent factor. The size and type of information user (internal, external or both) are identified as key elements (Curry, 2016; Lim et al., 2018).

4 Research Methods

This section builds on the theoretical framework that constructs the examination of various configurations in IISs (unit of analysis), to evaluate their impact on the value in information use. An empirical study of 13 IISs at Allianz Benelux was used as a research method to capture the design of the IISs. The data are collected by a semi-structured interview, with questions based on the IIS-DVC elements. For the analysis of the ten IISs, the fuzzy-set qualitative comparative analysis (fsQCA) was employed after calibrating all elements.

4.1 Analysis with fsQCA

In accordance with the research's configuration approach, the IISs of Allianz Benelux were analyzed using the fsQCA methodology, a variant of QCA. QCA is an analysis technique that examines the interdependence of variables instead of studying each independently, enabling the study of the link between a set of causal conditions (configuration) and their output (Pappas & Woodside, 2021; Rihoux & Ragin, 2008; Schneider & Wagemann, 2010). A set membership in a factor is referred to as the condition, in this study, the elements of the independent, dependent and contingency factors. An element is completely out of the set if its value is 0, whereas 1 implies that it is completely in a set (Mattke et al., 2022). Through these conditions and their configurations, a logically simpler explanation of the relationship with the outcome can be discovered, which is consistent with the theoretical purpose of this research. Additionally, the analysis method defines which conditions are core conditions, implying a strong causal link with the outcome, and which are contributing conditions, indicating a less causal relationship but still having an impact on the outcome (Fiss, 2011; Ragin, 2008).

In the base, the method is intended to integrate qualitative and quantitative analytical techniques that need an in-depth understanding of cases in order to score conditions for obtaining findings that are generalizable to larger populations (Pappas & Woodside, 2021; Ragin, 1994; Rihoux & Ragin, 2008). Because of its capabilities, the QCA methodology has grown in popularity as a monitoring and evaluation tool, especially in organizational, sociological and strategic studies (Ho et al., 2016; Rihoux, 2006; Soda & Furnari, 2012). However, the older QCA variants, the crips-set (csQCA) and multi-value (mvQCA), limited the adoption in other fields including IS, IM and marketing research due to several limitations. A significant limitation is that conditions in csQCA and mvQCA can only allow binary data (0 or 1) as input. To overcome this restriction, the fsQCA can score conditions ranging from 0 to 1, allowing it to capture complexity in cases that inherently differ in degree (Pappas & Woodside, 2021; Rihoux & Ragin, 2008). Hence, the fsQCA provides a more practical method as conditions can obtain all the values. Consequently, the fsQCA variant is recently being used more frequently in IS and IM studies, as well as other disciplines (Dawson et al., 2016; J.-N. Lee et al., 2019;

Ordanini et al., 2014; Park et al., 2020).

Compared to conventional analysis techniques, such as the variance-based approaches, fsQCA is more appropriate for this research because it explains the relationship between the outcome of interest and its conditions, whereas conventional analysis techniques are restricted to analyzing the relationships between the variables in a model (Pappas & Woodside, 2021). Thus, fsQCA can identify the combination of IIS-DVC elements that are sufficient to attain a certain value in information use, and therefore useful for this study.

Although the fsQCA method is well suited to this study, there is a drawback that needs to be considered. The method is sensitive to subjectivity since the choice of cases, factors, elements and measures has a significant influence on the findings of the analysis (Sehring et al., 2013). These decisions must be justified by valid arguments to prevent subjectivity. Therefore, the factor 1 to 6, IIS-DVC elements and measures were carefully selected from several IM, IS and big data research and cross-checked to determine whether they were relevant in multiple DVCs. The selection of IISs was based on predetermined criteria and the importance of services operating at Allianz Benelux. The service must first adhere to the definition and criteria of an IIS and complete all of the factors (1-6) with at least one information user (7) using the service. Additionally, no distinction was made in the degree of operationalization of the services as this is part of the study, therefore both low and high operationalized IISs were included in the analysis. The following part will go into further depth about the selected IISs, the data sample and collection, and how data were processed.

4.2 Sample and Data Collection

Data on IISs were collected from the insurance provider Allianz Benelux. Table 3 gives a summary of 13 IISs cases that were used in the analysis, all of which exclusively operate in the Benelux. The IISs are services that provide information for insights (dashboards or reports), decision making or subsequent operations. A semi-structured interview with data product managers, data engineers and data scientists, who specialize in a particular IIS, was conducted in order to collect the data. Therefore, an interview protocol was formulated based on the IIS-DVC elements, the contingency and dependent factor (see Appendix A). During the interviews, the prepared questions were presented with context about the IIS-DVC factors and elements, and processed immediately as most of the questions consisted of closed questions from which the respondent had to choose between a number of options applicable to the IIS. Moreover, all interviews were recorded but not transcribed because the type of questions were relatively simple.

Table 3: Description of the selected IISs from Allianz Benelux.

IISs at Allianz Benelux	**Description**
Sales Funnel	Sales Funnel service provides a complete overview of the active Property & Casualty (P&C) portfolio on registration date and effective date, showing the policy in and outflow.
P&C Portfolio Lapse Dashboard	Lapse dashboard evaluates the P&C portfolio (excluding Marine insurance) based on renewal data, technical price and its variables available to assess portfolio quality for Retail Motor.
Claims Frequency and Severity Report (CRS)	The CRS report provides insights on the claim frequency & severity (refers to the cost of a claim) and their development over time. Based on historical data, the observed frequency and severity are projected to their ultimate values. The report provides details on wide long different dimensions such as feeder systems, customer segment and line of business.
Customer Dashboard	The service gives insights about Allianz customers. The most important KPI is the number of active customers, displayed along various dimensions such as time, geography, and products. Other KPIs are customer satisfaction, complaints and number of customers with a digital ID.
Operations Steering Report (OSR)	ORS tracks operational work item, in and outflow data with related productivity KPI for call and non-call data.
Provider Management Dashboard	The main purpose of the service is to help information users form Claims department to analyze various payments that were made to providers for various cost types, including medical costs and legal costs.
Steering Dashboard	Steering the P&C portfolio by evaluating the premium KPIs based on effective date and loss KPIs based on reporting data.
Allianz Customer Model	The ACM dashboard service provides information on customer characteristics along with the ACM principles, which set the standards for simplicity, digitalization and scalability as core enablers to transform Allianz Benelux.
Fraud Underwriting	The service aims to manage UW fraud effectively by deter fraudsters from becoming customers and active customers from committing fraud. As a result, honest customers will be protected against paying the costs of UW fraud and this will improve Allianz overall performance and competitiveness.
Fraud Claims	Aims to manage claims fraud effectively by deter fraudsters from successfully extracting payments for either misrepresented claims or completely fake or fraudulent claims.
Pricing 2.0 motor	The service relates to the risk assessment and implementation of pricing (including scenario testing) for motor products.
Pricing 2.0 non-motor	The service relates to the risk assessment and implementation of pricing (including scenario testing) for property (non-motor) products.
MIRA Marine & Property	MIRA supports the underwriting process by analyzing the Marine & Property insurances. Insuring a ship is an example of Marine insurance, whereas Properties refer to factories, office buildings, etc.

Table 4: Descriptive statistics of the IIS-DVC elements.

IIS-DVC elements	Mean	S.D.	Min	Max
Type of data source	Human system 61.5%, Engineering system 38.5%	NA	NA	NA
Data format	Unstructured data 53.8%, Semi-structured data 15.4%, Structured data 30.8	NA	NA	NA
Degree of human involvement in Data Collection	Fully human involved 0.0%, Semi-human involved 92.3%, Fully automated 7.7%	NA	NA	NA
Data collection mode	Pull mode 76.9%, Push mode 23.1%	NA	NA	NA
Processing method	Batch processing 92.3%, Stream processing 7.7%	NA	NA	NA
Data quality	Accuracy 3.731	0.768	3	5
	Completeness 4.269	0.519	3	5
	Timeliness 4.138	0.376	3	5
	Consistency 4.769	0.599	3	5
Degree of human involvement in Data Analysis	Human-involved analysis 53.8%, Automated analysis 46.2%	NA	NA	NA
Robustness of the model	3.461	0.746	2	4
Type of information	Descriptive 61.5%, Predictive 38.5% and Prescriptive 0.0%	NA	NA	NA
Degree of human involvement in Information Delivery	Fully human involved 0.0%, Semi-human involved 92.3% Fully automated 7.7%	NA	NA	NA
Type of information user	Internal 84.6% Internal and external 15.4%	NA	NA	NA
Size of information user	14.769	9.689	2	30

4.3 Measuring and Calibrating IIS-DVC elements

In this study, the configuration is a set of IIS-DVC and contingency elements that result in the degree of operationalization. Several of these elements, including the degree of operationalization and the size of information user, are non-binary, necessitating the employment of the fsQCA. Using the fsQCA requires the calibration of all conditions, a procedure that converts the elements and outcomes value into a score between 0 and 1. The direct calibration approach was employed with elements and outcomes calibrated based on three anchors: full membership, full non-membership and the cross-over anchor (Ragin, 2008). A score of 0 represents a full non-membership, whereas 1 shows full membership. The cross-over anchor was used to distinguish between these memberships (Ragin, 2008).

There are numerous methods for calibrating the condition into these anchors, the majority of which are reliant on the researcher's conceptualization of the condition (Mendel & Korjani, 2018).

Therefore, conditions can be calibrated using the 75th, 50th, and 25th percentiles, indicating that full membership contains the top 25 percent of condition values, whereas full non-membership contains values up to 25 percent (Fiss, 2011). Another method is to calibrate using the min, max, and mean values. With the max representing full membership, the mean representing the cross-over point, and the min indicating full non-membership (Ragin, 2008; Schneider & Wagemann, 2010). Thus, the methods for determining these three anchors vary and are decided by the researcher.

In this study, the calibration was only applied on the elements: data quality, the robustness of the model, size of information user and on the outcome, degree of operationalization. Calibration was not necessary for the other elements as the value was either 0 or 1 (dummy variables).

First, as demonstrated in table 4, the IIS-DVC and contingency elements, as well as the outcome of interest, the degree of operationalization, were measured and analyzed. The elements of data source (1), type of data source and data format, were measured by classifying into human system or engineering system and unstructured, semi-structured or structured data, respectively. Therefore, 0 refers to the human system and 1 represents the engineering system. With regards to data format, 0 indicates unstructured data and 1 when data is either semi-structured or structured. These formats were merged as the data processing and analysis methods were quite similar (Faroukhi et al., 2020; Hu et al., 2014).

For the data collection (2), the degree of human involvement was measured by categorizing into fully human involved, semi-human involved or fully automated, where 0 represents both fully human and semi-human involvement and 1 indicates fully automated. Moreover, the data collection mode was categorized in pull, coded to 0, or push mode, coded to 1. The processing method was also measured using two values: batch or stream processing, where 0 indicates batch processing and 1 indicates stream processing.

The data (3) was assessed for quality on the dimensions: accuracy, completeness, timeliness and consistency. Statements concerning these dimensions were extracted from the information quality (IQ) questionnaire of Y. W. Lee et al. (2002) and presented to the respondents in a Likert scale of 5 (strongly disagree – strongly agree), having a Cronbach's alpha coefficient of 0.671, below the threshold of 0.70 (Taber, 2018). However, because data quality was the only measure for data, it was included in the analysis despite failing to reach the threshold. The IQ measurements are of high quality, as is attested by its widespread use in Information System (IS) and Information Management (IM) research, and fits the criteria of this study (Nelson et al., 2005; Zheng et al., 2013).

Additionally, the data quality was calibrated using the Likert scale, where the value 5 (strongly agree) denotes full membership, 1 (strongly disagree) indicates fully non-membership and 3 (neither agree nor disagree) was selected as the cross-over point (Fiss, 2011; Ragin, 2008). The data quality was assessed based on 19 statements and were enumerated so that they could be included in the analysis because incorporating each statement independently resulted in an excessive number of input variables that the fsQCA program cannot process.

Subsequently, the degree of human involvement in data analysis (4) was measured by classifying into human-involved, coded into 0, or automated analysis, coded into 1. The robustness of the model was measured by a statement about the level of resilience against difficulties and challenges, on a 5-point Likert scale and calibrated in the same way as data quality. The single statement was created to collect simple data from the respondents, and hence not necessary to assess their construct validity.

Moreover, the received information (5) from the analysis was simply classified by its type which can be descriptive, predictive or prescriptive information. The element was coded into 0 if the type was descriptive and otherwise 1. Predictive and prescriptive information was combined because it is primarily concerned with the future, while descriptive information explains the past (El Morr & Ali-Hassan, 2019). The degree of human involvement in information delivery (6) was coded into either fully human involvement or semi-human involvement (0) or fully automated (1). Factor (6), the information delivery was also measured by the degree of human involvement where 0 indicates fully human or semi-human involvement and 1 indicates fully automated.

Regarding the contingency factor (7), the size was measured by the number of users while the type of information user was categorized into internal users and internal as well as external users (both), coded to only internal (0) and internal as well as external (1). For the size of information users, the three thresholds for full membership, cross-over point and full non-membership were determined to be 22.5, 15 and 7.5, respectively, based on the 75, 50 and 25 quantiles. This distribution indicates large information users. Lastly, the outcome, degree of operationalization was measured by the current level of engagement or usage compared to the targeted level of usage. The 75, 50 and 25 percent made the three anchors, full non-membership, cross-over point and full membership, where an operationalization level higher than 75% indicates a high engagement of the IISs.

Table 5 provides an overview of the calibrations. For all calibrations, the thresholds of 25th, 50th and 75th percentile were used to distribute the full non-membership, cross-over point, and full membership. However, the membership matched with the cross-over anchor in some cases, which indicates that the element was neither in nor out of a set. These calibrated elements were adjusted by adding 0.001 to guarantee that the analysis covered all cases (Fiss, 2011). Following the calibration process, the fsQCA software was used to run the truth-table algorithm, which determines the sufficient configurations of elements for generating a high degree of operationalization (Ragin, 2008). Thus, the truth-table algorithm examined whether IISs with the same elements configuration constantly led to a high degree of operationalization. Appendix B contains the results of the truth-table.

Table 5: Calibration of Causal Condition and Outcome

Conditions and Outcome	**Coding and Calibration**
Degree of operationalization	**For high operationalization level** - Full membership anchor = more than 75% - Cross-over anchor = 50% - Full non-membership anchor = less than 25%
Type of data source	0 = Human systems, 1 = Engineering systems
Data format	0 = Unstructured data, 1 = either Semi-structured or Structured data
Degree of human involvement in Data Collection	0 = either Fully human involved or Semi-human involved, 1 = Fully automated
Data collection mode	0 = Pull mode, 1 = Push mode
Processing method	0 = Batch processing, 1 = Stream processing
Data quality	**Data quality (sum of 19 items). Each item was measured by 5-point Likert scale (5 = strongly agree, 3 = neither agree nor disagree, 1 = strongly disagree)** - Full membership anchor = 95 - Cross-over anchor = 57 - Full non-membership anchor = 19
Degree of human involvement in Data Analysis	0 = Human-involved analysis, 1 = Automated analysis
Robustness of the model	**Each item was measured by 5-point Likert scale (5 = strongly agree, 3 = neither agree nor disagree, 1 = strongly disagree)** - Full membership anchor = 5 - Cross-over anchor = 3.001 - Full non-membership anchor = 1
Type of information	0 = Descriptive information, 1 = either predictive or prescriptive information
Degree of human involvement in Information Delivery	0 = either Fully human involved or Semi-human involved, 1 = Fully automated
Type of information user	0 = Internal users, 1 = Internal and external users
Size of information user	**For large users based on the 75, 50, and 25 percentiles** - Full membership anchor = 22.5 - Cross-over anchor = 15.001 - Full non-membership anchor = 7.5

5 Results

Using the fsQCA method, different configurations of the IIS-DVC elements congruent with the contingency factor and outcome were analyzed. The calibrated and coded data were used as input for the analysis. As a result, the truth-table algorithm provides five different configurations for high level of operationalization, and four configurations that generate low level of operationalization. This section explains the findings, how to read the graphical expression of the results and what the different configurations means.

Table 6: Configurations of IIS-DVC elements sufficient for high degree of operationalization.

Outcome	**Parsimonious solutions**	**Intermediate solutions**
High degree of operationalization	Predictive + Robust&∼SemiAndStructured&∼AutoAnalysis + ∼LargeUsers&AutoAnalysis + ∼LargeUsers&Robust&∼SemiAndStructured ->High operationalization of IIS	LargeUsers&**Robust**&DQ&∼InAndExternal&∼Engineering&∼**SemiAndStructured**&∼AutomatedDC&∼Push&∼Stream&∼**AutoAnalysis**&∼Predictive&∼AutomatedID (1 case) + LargeUsers&Robust&DQ&∼InAndExternal&Engineering&SemiAndStructured&∼AutomatedDC&∼Push&∼Stream&AutoAnalysis&**Predictive**&∼AutomatedID (2 case) + ∼**LargeUsers**&**Robust**&DQ&InAndExternal&∼Engineering&∼**SemiAndStructured**&∼AutomatedDC&∼Push&∼Stream&**AutoAnalysis**&∼Predictive&∼AutomatedID (1 case) + ∼**LargeUsers**&**Robust**&DQ&∼InAndExternal&∼Engineering&∼**SemiAndStructured**&∼AutomatedDC&Push&∼Stream&∼**AutoAnalysis**&**Predictive**&∼AutomatedID (2 case) + ∼**LargeUsers**&∼Robust&DQ&∼InAndExternal&∼Engineering&∼SemiAndStructured&AutomatedDC&Push&Stream&**AutoAnalysis**&**Predictive**&AutomatedID (1 case) + ->High operationalization of IIS

*IIS-DVC elements in the intermediate solution that also appear in the parsimonious solution are in **bold** and considered as **core** elements (Fiss, 2011).

** LargeUsers = User size (large), Robust = Robustness of the model, DQ = Data quality (high), InAndExternal = Internal and external users, Engineering = Data Source type (engineering), SemiAndStructured = Semi-structured/structured data, AutomatedDC = Human involvement in data collection (fully automated), Push = Data collection mode (push), Stream = Processing method (Stream), AutoAnalysis = Data analysis (automated), Predictive = Information type (predictive) and

AutomatedID = Information delivery (automated).

Table 7: Configurations of IIS-DVC elements sufficient for low degree of operationalization.

Outcome	Parsimonious solutions	Intermediate solutions
Low degree of operationalization	LargeUsers&AutoAnalysis&∼Predictive + ∼LargeUsers&∼AutoAnalysis&∼Predictive + ∼LargeUsers&∼Push&∼AutoAnalysis + ∼LargeUsers&∼InAndExternal&∼Predictive + ∼LargeUsers&∼InAndExternal&∼Push ->Low operationalization of IIS	**LargeUsers**&Robust&DQ&∼InAndExternal&∼Engineering&∼SemiAndStructured&∼AutomatedDC&∼Push&∼Stream&**AutoAnalysis**&∼**Predictive**&∼AutomatedID (1 case) + **LargeUsers**&Robust&DQ&InAndExternal&Engineering&SemiAndStructured&∼AutomatedDC&∼Push&∼Stream&**AutoAnalysis**&∼**Predictive**&∼AutomatedID (1 case) + ∼**LargeUsers**&∼Robust&DQ&∼**InAndExternal**&∼Engineering&∼SemiAndStructured&∼AutomatedDC&∼**Push**&∼Stream&∼**AutoAnalysis**&∼**Predictive**&∼AutomatedID (1 case) + ∼**LargeUsers**&Robust&DQ&∼**InAndExternal**&∼Engineering&SemiAndStructured&∼AutomatedDC&∼**Push**&∼Stream&∼**AutoAnalysis**&∼**Predictive**&∼AutomatedID (1 case) + ->Low operationalization of IIS

5.1 Parsimonious and Intermediate Solution

The fsQCA offers three solutions, namely complex, parsimonious and intermediate solutions. Only the parsimonious and intermediate solutions were employed since the complex solution is more helpful in more complex cases and also provides little insight into the causal relationships between conditions, making it less relevant for this study (Fiss, 2011). Table 6 and 7 show the solutions using the Boolean notation, where '+' indicates a logical disjunction (i.e. OR), the expression '&' denotes 'AND', and the tilde '∼' indicates the negation of the condition, which evaluates to absent or 0 when the condition evaluates to present or 1.

In case of the parsimonious solution, there are several logically disjunct sets of IIS-DVC elements, implying that at least one set is sufficient for realizing the operationalization of an IIS. Thus, for the first part of the parsimonious solution, 'predictive + Robust&∼SemiAndStructured&∼AutoAnalysis' for high operationalization, the type of information that has the form of a predictive feature, or the high robustness of the analytical model along with the unstructured data format (i.e. neither semi-structured nor structured) and in combination with human involvement in data analysis (i.e. not automated analysis), leads to the high operationalization of the IIS. Moreover, if the element is included in the parsimonious solution, it is identified as the core element, meaning that it has a strong causal relationship with the outcome (Fiss, 2011; Ragin, 2008). To indicate whether a core element is included in the configuration, it must be present in both solutions, as shown in Table 6 and 7 where the elements are in bold.

The intermediate solution provides, in this case, five configurations that produce a high degree of operationalization and four for low operationalization, by determining the core elements and others that have to be present in conjunction. Therefore, the first configuration for high outcome indicates that the robustness of the analysis model, unstructured data format and human involved data analysis, are core to the outcome of interest. The other elements that are present in the configuration, thus, occur in the intermediate solution but not in the parsimonious, are considered as peripheral or contributing conditions (Fiss, 2011; Ragin, 2008). These conditions are less causally connected to a specific degree of operationalization.

However, for interpretation convenience, the results of table 6 and 7 are graphically presented, using the format developed by Fiss (2011) and Ragin (2008), in figure 3. The columns A1 through A5 illustrate each a single configuration of the IIS-DVC elements that leads to a high level of operationalization based on the intermediate solution. In addition, the columns B1 to B4 represent the configurations for low level of operationalization. A graphical representation of the configurations makes it feasible to illustrate the complexity of causal conditions in a simplified manner. Consequently, it can be determined which conditions are necessary, sufficient or not sufficient.

5.2 Validity of the Solutions

In order to justify this interpretation and others, the validity of both solutions was assessed based on consistency and coverage (Ragin, 2008). In terms of consistency, it measures the extent to which cases with comparable sets of conditions agree on recognizing the same outcome (Fiss, 2011; Ragin, 2008). Thus, a high consistency score indicates that a certain combination of IIS-DVC elements consistently leads to the matching result. In this research, the consistency for each of the five high level configurations is 1.00, which is considerably above the accepted standard of 0.80 (Fiss, 2011; Misangyi & Acharya, 2014; Ragin, 2008). Because all configurations individually scored full consistency, the overall solution consistency is also 1.00, which evaluates all configurations together to determine whether they collectively result in a high degree of operationalization. A value above 0.80 is likewise acceptable here (Ragin, 2008). Each of the low level configurations has a consistency level of 1.00 except B3, which has the value of 0.95. As a result, the overall solution consistency of 0.99.

Regarding the coverage measures, consisting of raw coverage, unique coverage and overall solution coverage, help to figure out what proportion of the output is covered by the solution (Ragin, 2008; Schneider & Wagemann, 2010). The raw coverage indicates the percentage of the outcome that is explained by a given configuration, whereas the unique coverage determines the degree to which a configuration exclusively covers the outcome (Ragin, 2008; Schneider & Wagemann, 2010). A high score on the coverage metrics indicates a high level of empirical relevance and demonstrates how well the configuration performs to achieve the outcome in question (J.-N. Lee et al., 2019; Schneider & Wagemann, 2010). In this research, the configurations A2 and B3 have the widest coverage , and hence empirically relevant and most efficient in obtaining the intended degree of operationalization (Schneider & Wagemann, 2010).

IIS-DVC Elements	**Configuration for High Operationalization**					**Configuration for Low Operationalization**			
	A1	**A2**	**A3**	**A4**	**A5**	**B1**	**B2**	**B3**	**B4**
Data Source									
Engineering systems	⊗	●	⊗	⊗	⊗	⊗	●	⊗	⊗
Semi-structured/structured data	⨂	●	⨂	⨂	⊗	⊗	●	⊗	●
Data Collection									
Fully automated	⊗	⊗	⊗	⊗	●	⊗	⊗	⊗	⊗
Push mode	⊗	⊗	⊗	●	●	⊗	⊗	⨂	⨂
Stream processing	⊗	⊗	⊗	⊗	●	⊗	⊗	⊗	⊗
Data									
Data quality	●	●	●	●	●	●	●	●	●
Data Analysis									
Automated analysis	⨂	●	⬤	⨂	⬤	⬤	⬤	⨂	⨂
Robustness	⬤	●	⬤	⬤	⊗	●	●	⊗	●
Information									
Predictive information	⊗	⬤	⊗	⬤	⬤	⨂	⨂	⨂	⨂
Information Delivery									
Fully automated	⊗	⊗	⊗	⊗	●	⊗	⊗	⊗	⊗
Information user									
Internal and external users	⊗	⊗	●	⊗	⊗	⊗	●	⨂	⨂
Large information users	●	●	⨂	⨂	⨂	⬤	⬤	⨂	⨂
Consistency	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.95	1.00
Raw coverage	0.09	0.14	0.09	0.13	0.09	0.14	0.14	0.15	0.14
Unique coverage	0.09	0.14	0.09	0.13	0.09	0.14	0.14	0.15	0.14
Overall solution Consistency			**1.00**				**0.99**		
Overall solution Coverage			**0.54**				**0.57**		

Figure 3: Different configurations of IIS-DVC elements for achieving Low and High Degree of Operationalization.

* The presence of the IIS-DVC element is represented by a solid circle, whereas the absence of the element is indicated by a circle with a cross. In addition, large circles are core elements and small circles are contributing elements.

** If an IIS-DVC element (circle with a cross) is absent, it implies that the elements coded to 0 are present.

5.3 Configuration for High Degree of Operationalization

According to the findings, there are five configurations (A1-A5) that result in high IIS operationalization (figure 3). First, the solution shows that all IIS-DVC elements are required to create value in IISs since there are no blank spaces in the graphic. A blank space in fsQCA indicates that it does not matter if the element is present or not (Ragin, 2008). Thus, all IIS-DVC elements are present in the configuration, however, there are differences in measures of the elements. Elements that are identified as absent, shown with a crossed circle, the 0 coded or negation of the element is present (table 5).

Second, for all configurations, including for low operationalization, high data quality is present and therefore designated as a trivial condition, meaning that its presence contributes to the outcome as well as its absence (Schneider & Wagemann, 2010). Furthermore, the solution highlights the IIS-DVC factor, data analysis, because in most configurations its elements are essential (core) for producing a high degree of operationalization. This is also true for data format, type of information and the user size. Moreover, the elements of data collection (2) are highly correlated because they mostly occur in the same set of measures, except for A4. Lastly, the raw and unique coverage's of each of the five configurations are identical. It signifies that the set of conditions is distinct and does not intersect with others, however, there is some overlapping which is indicated by a low unique coverage (J.-N. Lee et al., 2019; Schneider & Wagemann, 2010).

Starting with the first configuration, A1 represents an IIS-DVC in which the presence of structured data format, human-involved analysis and high robustness as core elements is sufficient to achieve high operationalization. However, the other elements are needed to accomplish this objective, but have a weaker causal link with the output. The coverage of this particular configuration is relatively low compared to others, which indicates that A1 is least relevant for empirical purposes. In contrast to A1, the configuration of A2 is less dependent on core elements. Only the presence of predictive information as a core element with the support of peripheral elements, is sufficient to achieve the outcome. However, this IIS-DVC design provides the largest raw and unique coverage, making it the most empirically relevant.

For A3, A4 and A5, these configurations do not require a high number of information users unlike the previous designs. However, they are reliant on more core elements for achieving the same result. A3 differentiates itself from others as the configuration requires both internal and external users, whereas A5 stands out due to its unique approach to deliver information in IIS, namely fully automated, and as the design of data collection (2) differs completely from others. A full automated collection approach based on the push mode, and the stream processing method to pre-prepare the data shape the data collection for A5. In regards to A4, it needs the most core elements to produce a high degree of operationalization.

5.4 Configuration for Low Degree of Operationalization

Additionally to the configurations for high level of operationalization, figure 3 also demonstrates four configurations that lead to a low operationalization degree (B1-B4). The core elements defined in the sets for low operationalization are quite similar to A1 to A5. This means that these elements are mainly decisive for the success of the IIS at Allianz Benelux. Moreover, the first two solutions, B1 and B2, significantly overlap each other, indicated by a relatively high unique coverage. It is identical for the other configurations, B3 and B4. Since the configuration turns around between B2 and B3, the size of information users seems to be responsible for this change in design. To summarize, the configurations B1 to B4 are not the inverses of the configurations for high operationalization, meaning that Allianz Benelux does not necessarily need to change the entire IIS-DVC but rather modify to improve the operationalization degree.

6 Discussion and Implications

The research findings have offered insight into various configurations of IIS-DVC elements and their impact on the value in information use, which is classified as high or low operationalization. Nevertheless, given the limitations of this study, these findings need to be interpreted and discussed with caution. Therefore, this section examines the limitations along with implications of the findings.

Due to variations in data-based service types, there are various DVC configurations applicable (table 1). However, organizations are experiencing challenges and difficulties in designing an effective DVC in general, but this study focuses specifically on DVC for IISs operational in the insurance industry (Albrecher et al., 2019; Puschmann, 2017; Serhani et al., 2016). Evaluating its design helps organizations in determining which configurations of IIS-DVC elements lead to a certain value in information use. The archetype of IIS-DVC is developed that constructs the relationship between factors and elements, and therefore enables identifying distinct IIS-DVC configurations and its core elements. As a result, the configurations A1 to A5 lead to high degree of operationalization, while B1 to B4 result in low operationalization level (figure 3).

Table 8: Configurations for High and Low Degree of Operationalization

IIS-DVC elements	A1	A2	A3	A4	A5	B1/B2	B3/B4
Information user size	Large	Large	Not large	Not large	Not large	Large	Not large
Robustness of the model	High robustness	High robustness	High robustness	High robustness	Low robustness	High robustness	High/low robustness
Data quality	High quality	High quality	High quality	High quality	High quality	High quality	High quality
Type of information user	Internal	Internal	Internal and external	Internal	Internal	Internal/internal and external	Internal
Type of data source	Human systems	Engineering systems	Human systems	Human systems	Human systems	Human/engineering systems	Human systems
Data format	Unstructured	Semi-structured/ structured	Unstructured	Unstructured	Unstructured	Unstructured/semi-structured/structured	Unstructured/semi-structured/structured
Degree of human involvement in Data Collection	Semi-human involved	Semi-human involved	Semi-human involved	Semi-human involved	Fully automated	Semi-human involved	Semi-human involved
Data collection mode	Pull mode	Pull mode	Pull mode	Push mode	Push mode	Pull mode	Pull mode
Processing method	Batch processing	Batch processing	Batch processing	Batch processing	Stream processing	Batch processing	Batch processing
Degree of human involvement in Data Analysis	Human-involved analysis	Automated analysis	Automated analysis	Human-involved analysis	Automated analysis	Automated analysis	Human-involved analysis
Type of information	Descriptive	Predictive	Descriptive	Predictive	Predictive	Descriptive	Descriptive
Degree of human involvement in Information Delivery	Semi-human involved	Semi-human involved	Semi-human involved	Semi-human involved	Fully automated	Semi-human involved	Semi-human involved

6.1 Core IIS-DVC Elements of The Configurations

Table 8 summarizes these results consisting of the necessary elements needed in the configurations. According to the empirical results, the core elements for both low and high level of operationalization are the size of the information user, the degree of human involvement in data analysis, and the type of information derived from this. Thus, the second part of the IIS-DVC, beginning with information creation and ending with the use of information, is dominant in creating high or low value of the

IIS, which meets the expectation and is aligned with the theories about DVCs (Faroukhi et al., 2020; Hu et al., 2014). As theories observes that the actual value from data are mainly created during the analysis process and subsequent stages, the results confirm this observation (Curry, 2016; Faroukhi et al., 2020; Lim et al., 2018; Miller & Mork, 2013; Moreno-Mateos & Carou, 2022).

Additionally, regardless of whether the data analysis process requires human involvement or is fully automated, or whether the IIS is aimed at a large or low number of information users, both sets of elements can lead to both outcomes, except for the information type. The absence of predictive information has a large share in determining the IIS to be less operationalized (B1 and B4). When Allianz Benelux uses descriptive information for creating value in information use, it is important to have unstructured data format and high robustness of the analysis model as core elements to cover the absence of predictive information, and therefore to prevent the IIS becoming low operationalized (A1 and A3). Thus, to make the IIS-DVC more valuable when it uses descriptive information, the configuration requires high robustness to properly transform unstructured data into more useful descriptive information and therefore resulting in high operationalization instead of low operationalization.

The correlation between robustness and unstructured data is confirmed by studies on challenges and difficulties in the analysis process related to the nature of big data (Fan et al., 2014; Nasser & Tariq, 2015; Tallon, 2011). As unstructured data is more complex by nature than semi-structured or structured data, integrating these data into the analysis model along with extracting value from it, is consequently complicated (Fan et al., 2014; Nasser & Tariq, 2015). To conclude, when unstructured data is a core element in the configuration, high robustness must also function as a core element in order to handle the complexity. It is illustrated by the configurations A1, A3 and A4.

6.2 Contributing IIS-DVC Elements

Besides the core elements, all configurations demand the presence of contributing elements, which is reasonable considering that the DVC needs all factors to be completed in order to transform raw data into useful information or assets ((Curry, 2016; Hu et al., 2014)). The DVC factors must be performed in the sequential form, which has been concluded from studies on various DVC concepts (Curry, 2016; Faroukhi et al., 2020; Miller & Mork, 2013). However, in practice, it occurs that part of the chain is executed repeatedly, creating a loop of several DVC factors. During the interpretation of the findings, this research assumes that the IIS-DVC factors are run correctly once, which is a limitation of the research design. Thus, it is assumed that the subsequent factor is dependent or related to the preceding factor.

Building on this assumption, it can be concluded that the contributing elements of data collection (2) are highly correlated. In other words, if the pull data collection mode is present in the configuration for both low and high operationalization, it is likely that the batch processing method is applied for cleaning and preparing the data and requires human involvement in both processes (table 8). This observation is made based on all configurations except A4. Most configurations show that the formation of data collection is the same, confirmed by A5 showing the inversion of this, suggesting that the other set of data collection elements consists of push mode, stream processing and fully automated data collection.

Nevertheless, this coherence of elements has not been confirmed in literature. Although Hu et al. (2014) observe that stream processing, i.e. processing collected data in near real-time, which implies that shortly after the data is collected it is processed immediately, requires a corresponding real-time data collection mode. Hence, the push collection mode is most suitable because the mechanism is made to automatically collect data, while pull mode requires queries that often need human input and therefore cannot meet the criteria of stream processing (Curry, 2016; Faroukhi et al., 2020; Hu et al., 2014). A5 confirms this observation, however, the sample size and scope of IISs included in this study are not sufficient to convincingly confirm all observations regarding data collection (2).

Moreover, studies on DVC have emphasized the importance of data collection design's reliability on data source characteristics and the purpose of data use (i.e. descriptive, predictive or prescriptive) (Hu et al., 2014; Miller & Mork, 2013). Figure 3 indicates no evident cohesion between data collection, data source and information type in case of Allianz Benelux. For future research, this cohesion and other relationships within the DVC can provide more understanding in why certain configurations of IISs deviate from theory.

Lastly, the contributing element, data quality, is consistently high in all configurations and thus insufficient in determining the outcome. In regards to information users, it was expected to observe some correlation with other IIS-DVC elements as studies emphasize the importance of user co-creation in service design (Mager, 2009; Teixeira et al., 2012). Because there might be differences in requirements between internal and external users, the element was predicted to be more influential. However, the dataset contains few IISs that serve external users and thus insufficient to justify the theory. This also applies to the degree of human involvement in information delivery. Aside from A5, where information is delivered automatically, all configurations include humans in delivering information to users. It is interesting to note that the conclusion regarding the cohesiveness of data collection elements is further strengthened by the presence of automated information delivery in A5, which is in line with push mode, stream processing and automated data collection (Curry, 2016; Hu et al., 2014).

6.3 Limitations

In general, because the IISs included in this study are derived from a single organization (Allianz Benelux) and fairly particular, it is difficult to relate a configuration as a whole to literature. Consequently, it is less generalizable, however, the purpose of this research is to emphasize the importance of evaluating DVCs. Moreover, the research is aimed to provide insight into the design of IISs within Allianz Benelux, but demonstrates relatability with existing studies.

The limitations have been addressed during the findings discussion, yet there remain some limitations regarding the research design unmentioned. First, the items used to measure data quality were validated for internal consistency, yielding a Cronbach's Alpha of 0.671, which is below the recommended threshold of 0.70 (Taber, 2018). Nevertheless, the data quality was included as it was the only key element defined for data (3). In addition, the robustness was measured by a single statement and therefore cannot be tested on consistency. The statements can be extended for future research in evaluating DVC configurations to better test robustness of the analysis model.

6.4 Theoretical and Practical Contributions

This study builds on the work of Lim et al. (2018), who investigated value creation in IIS by finding factors that characterize the DVC for IIS. In absence of a comprehensive analysis of the proposed nine-factors DVC framework, this study contributes by identifying additional key elements and constructing archetypes of IIS-DVC to determine configurations that lead to a certain value in information use. Although various configurations of the DVC have been covered in the literature, few studies have been conducted regarding IIS as well as what influence a configuration has on the value of IIS. An empirical study was conducted to acquire information on the various IIS-DVC configurations, allowing elements of the configuration to be compared with what has been acknowledged in literature.

In addition, although existing studies on DVC primarily describe and explain the factors, the findings of this study expand knowledge by determining which factors play a key role in the value creation process and provide more details on the contribution of individual IIS-DVC elements. A key discovery is that whether or not the configuration succeeds depends on the type of information derived from the analysis stage. This provides a new perspective on the DVC, and further research is needed to clarify the reasoning behind its influence on value in information use.

Lastly, this configuration approach research contributes methodologically to DVC and IIS literature by demonstrating how to employ the fuzzy-set QCA method to find patterns in DVC configurations, providing an alternative method to existing pattern discovery methods such as classification and clustering. Thus, this study gives guidance for applying fuzzy-set QCA in IM studies.

When compared to theoretical contributions, this study mostly benefits Allianz Benelux in practice as the primary contribution is an evaluation of the current state of operational IISs at Allianz Benelux. The findings cover the differences and overlaps between configurations that result in low operationalization and those that lead to high outcomes. First, it can be stated that configurations A1 and B1 largely overlap in terms of IIS-DVC elements, but the difference in outcome is primarily due to a shift in the degree of human involvement in Data analysis. Such information informs Allianz Benelux on potential areas for improvement in IISs that are under-performing.

Second, configuration A2 and B2 are also similar yet produce different outcomes. There are minor variances that make differences in this case as well, where an absence of predictive information mostly leads cases in B2 to attain a low level of operationalization. Other comparisons, such as A3 with B3 or A3 with A4, can be analyzed to gain insight into where the points for improvement are. In addition, comparisons of configurations can be made within the two defined outcomes. It helps in the design of new IISs by providing insights into which combination of elements occur together to produce a high degree of operationalization.

Hence, it can be concluded that the effectiveness of IISs that employ descriptive information for service use is lower. While Allianz Benelux is in the process of modifying more IISs to use predictive information, the findings can assist in deciding what alterations can be done to obtain better results. Thus, slight variations can create different outcomes and therefore necessitate specific modifications to change the outcome without having to alter the entire DVC.

7 Conclusion

This research aimed to evaluate configurations of the DVC for IISs in the insurance industry by identifying which configurations lead to a certain value in information use, where value is categorized as high or low degree of operationalization. Hence, the research question were formulated as follows:

- *Which configurations of IIS-DVC factors lead to high value in information use in the insurance industry?*

The theoretical framework of the archetype of IIS-DVC was constructed based on the configurational fit between IIS-DVC factors. Subsequently, the configurations of IIS-DVC factors were examined using 13 IISs operating at Allianz Benelux as input for analysis that used the fuzzy-set QCA method. The empirical study discovered patterns in the configurations, as well as sets of core and contributing elements that produce either a high or low degree of operationalization. First, the findings indicate that all IIS-DVC elements were required to be present in a configuration in which the IIS-DVC elements, degree of human involvement in data analysis, size of information user and type of information are identified as core elements in most configurations. The results then allow for pattern discovery by comparing low-performance configurations with the high-performance configurations.

As a result, two archetypes of IIS-DVC were developed. The first archetype, for high degree of operationalization, contains predictive information, unstructured data format and high robustness of the analysis model as core elements. The presence of these three core elements differentiates the archetype from the archetype for low degree of operationalization. All configurations that lead to low operationalization contain descriptive information as a core element. As a result, only the descriptive information element forms the second archetype for low degree of operationalization. Thus, the type of information has a significant role in influencing the outcome.

Although the configuration requires all elements, the remaining contributing elements show inconsistency across the configurations for both outcomes. Consequently, it was difficult to identify a complete configuration as an archetype for a certain value in information use. Yet, there are five configurations discovered for producing high degree of operationalization, namely A1 to A5, and four configurations that result in low operationalization, B1 to B4 (figure 3)

As inconsistency of contributing elements prevented its inclusion in the archetypes of IIS-DVC, future research on configurations of the IIS-DVC, using more cases from across the insurance industry, is needed to understand better which elements are frequently occurring together to produce a particular value in information use. Moreover, evaluating DVC for IIS in other industries is recommended in order to expand on these findings. In doing so, the limitations of this research can be considered to improve the assessment.

References

Albrecher, H., Bommier, A., Filipović, D., Koch-Medina, P., Loisel, S., & Schmeiser, H. (2019). Insurance: Models, digitalization, and data science. *European Actuarial Journal*, *9*(2), 349–360.

Apte, U. M., Cavaliere, R. A., & Kulkarni, S. S. (2010). Analysis and improvement of information-intensive services: Evidence from insurance claims handling operations. *Production and Operations Management*, *19*(6), 665–678.

Ballou, D. P., & Pazer, H. L. (1985). Modeling data and process quality in multi-input, multi-output information systems. *Management science*, *31*(2), 150–162.

Buhl, H. U., Röglinger, M., Moser, F., & Heidemann, J. (2013). Big data.

Cappiello, C., Francalanci, C., & Pernici, B. (2004). Data quality assessment from the user's perspective. *Proceedings of the 2004 international workshop on Information quality in information systems*, 68–73.

Carbone, P., Katsifodimos, A., Ewen, S., Markl, V., Haridi, S., & Tzoumas, K. (2015). Apache flink: Stream and batch processing in a single engine. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, *36*(4).

Curry, E. (2016). The big data value chain: Definitions, concepts, and theoretical approaches. In *New horizons for a data-driven economy* (pp. 29–37). Springer, Cham.

Daki, H., El Hannani, A., Aqqal, A., Haidine, A., & Dahbi, A. (2017). Big data management in smart grid: Concepts, requirements and implementation. *Journal of Big Data*, *4*(1), 1–19.

Dawson, G. S., Denford, J. S., Williams, C. K., Preston, D., & Desouza, K. C. (2016). An examination of effective it governance in the public sector using the legal view of agency theory. *Journal of Management Information Systems*, *33*(4), 1180–1208.

Dhenakaran, S., & Sambanthan, K. T. (2011). Web crawler-an overview. *International Journal of Computer Science and Communication*, *2*(1), 265–267.

Donaldson, L. (1996). The normal science of structural contingency theory. *Handbook of organization studies*, *44*(20), 57–76.

Donaldson, L. (2006). The contingency theory of organizational design: Challenges and opportunities. *Organization design*, 19–40.

Dounis, A. I., & Caraiscos, C. (2009). Advanced control systems engineering for energy and comfort management in a building environment—a review. *Renewable and Sustainable Energy Reviews*, *13*(6-7), 1246–1261.

Drazin, R., & Van de Ven, A. H. (1985). Alternative forms of fit in contingency theory. *Administrative science quarterly*, 514–539.

Eisenack, K., Villamayor-Tomas, S., Epstein, G., Kimmich, C., Magliocca, N., Manuel-Navarrete, D., Oberlack, C., Roggero, M., & Sietz, D. (2019). Design and quality criteria for archetype analysis. *Ecology and Society*, *24*(3).

Ekbia, H., Mattioli, M., Kouper, I., Arave, G., Ghazinejad, A., Bowman, T., Suri, V. R., Tsou, A., Weingart, S., & Sugimoto, C. R. (2015). Big data, bigger dilemmas: A critical review. *Journal of the Association for Information Science and Technology*, *66*(8), 1523–1545.

El Morr, C., & Ali-Hassan, H. (2019). Descriptive, predictive, and prescriptive analytics. In *Analytics in healthcare* (pp. 31–55). Springer.

Elragal, A., & Klischewski, R. (2017). Theory-driven or process-driven prediction? epistemological challenges of big data analytics. *Journal of Big Data*, *4*(1), 1–20.

Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. *National science review*, *1*(2), 293–314.

Faroukhi, A. Z., El Alaoui, I., Gahi, Y., & Amine, A. (2020). Big data monetization throughout big data value chain: A comprehensive review. *Journal of Big Data*, *7*(1), 1–22.

Fiss, P. C. (2011). Building better causal theories: A fuzzy set approach to typologies in organization research. *Academy of management journal*, *54*(2), 393–420.

Geoghegan, M., Cormican, K., & Wu, Q. (2021). Sustainable team design: A challenge to traditional beliefs in information-intensive service industries. *Sustainability*, *13*(24), 13552.

Ho, J., Plewa, C., & Lu, V. N. (2016). Examining strategic orientation complementarity using multiple regression analysis and fuzzy set qca. *Journal of Business Research*, *69*(6), 2199–2205.

Hu, H., Wen, Y., Chua, T.-S., & Li, X. (2014). Toward scalable systems for big data analytics: A technology tutorial. *IEEE access*, *2*, 652–687.

Hunke, F., Seebacher, S., Schüritz, R., & Satzger, G. (2020). Pathways from data to value: Identifying strategic archetypes of analytics-based services. *Wirtschaftsinformatik (Zentrale Tracks)*, 1035–1050.

Ingram, A. L. (1999). Using web server logs in evaluating instructional web sites. *Journal of educational technology systems*, *28*(2), 137–157.

Janssen, M., & Sol, H. G. (2000). Evaluating the role of intermediaries in the electronic value chain. *Internet Research*.

Jesmin, I., & Hui, H. (2012). A review of literature on contingency theory in managerial accounting. *African journal of business management*, *6*(15), 5159–5164.

Kausar, M. A., Dhaka, V., & Singh, S. K. (2013). Web crawler: A review. *International Journal of Computer Applications*, *63*(2).

Kim, S.-H., & Chung, K. (2015). Emergency situation monitoring service using context motion tracking of chronic disease patients. *Cluster Computing*, *18*(2), 747–759.

Kühne, B., & Böhmann, T. (2019). Data-driven business models-building the bridge between data and value. *ECIS*.

Lee, J.-N., Park, Y., Straub, D. W., & Koo, Y. (2019). Holistic archetypes of it outsourcing strategy: A contingency fit and configurational approach. *Mis Quarterly*, *43*(4).

Lee, J. H., & Waterman, N. M. (2012). Understanding user requirements for music information services. *ISMIR*, 253–258.

Lee, Y. W., Strong, D. M., Kahn, B. K., & Wang, R. Y. (2002). Aimq: A methodology for information quality assessment. *Information & management*, *40*(2), 133–146.

Lim, C., Kim, K.-H., Kim, M.-J., Heo, J.-Y., Kim, K.-J., & Maglio, P. P. (2018). From data to value: A nine-factor framework for data-based value creation in information-intensive services. *International Journal of Information Management*, *39*, 121–135.

Mager, B. (2009). Service design as an emerging field. *Designing services with innovative methods*, *1*, 27–43.

Maimon, O., & Rokach, L. (2005). Data mining and knowledge discovery handbook.

Mattke, J., Maier, C., Weitzel, T., Gerow, J. E., & Thatcher, J. B. (2022). Qualitative comparative analysis (qca) in information systems research: Status quo, guidelines, and future directions. *Communications of the Association for Information Systems*, *50*(1), 8.

Mendel, J. M., & Korjani, M. M. (2018). A new method for calibrating the fuzzy sets used in fsqca. *Information Sciences*, *468*, 155–171.

Miller, H. G., & Mork, P. (2013). From data to decisions: A value chain for big data. *It Professional*, *15*(1), 57–59.

Misangyi, V. F., & Acharya, A. G. (2014). Substitutes or complements? a configurational examination of corporate governance mechanisms. *Academy of Management Journal*, *57*(6), 1681–1705.

Morabito, V., et al. (2015). Big data and analytics. *Strategic and organisational impacts*.

Moreno-Mateos, M. A., & Carou, D. (2022). A note on big data and value creation. In *Machine learning and artificial intelligence with industrial applications* (pp. 1–18). Springer.

Nasser, T., & Tariq, R. (2015). Big data challenges. *J Comput Eng Inf Technol 4: 3. doi: http://dx. doi. org/10.4172/2324*, *9307*(2).

Nelson, R. R., Todd, P. A., & Wixom, B. H. (2005). Antecedents of information and system quality: An empirical examination within the context of data warehousing. *Journal of management information systems*, *21*(4), 199–235.

Ordanini, A., Parasuraman, A., & Rubera, G. (2014). When the recipe is more important than the ingredients: A qualitative comparative analysis (qca) of service innovation configurations. *Journal of service research*, *17*(2), 134–149.

Ostrom, A. L., Parasuraman, A., Bowen, D. E., Patrıcio, L., & Voss, C. A. (2015). Service research priorities in a rapidly changing context. *Journal of service research*, *18*(2), 127–159.

Padgavankar, M., Gupta, S., et al. (2014). Big data storage and challenges. *International Journal of Computer Science and Information Technologies*, *5*(2), 2218–2223.

Pappas, I. O., & Woodside, A. G. (2021). Fuzzy-set qualitative comparative analysis (fsqca): Guidelines for research practice in information systems and marketing. *International Journal of Information Management*, *58*, 102310.

Park, Y., Fiss, P. C., & El Sawy, O. A. (2020). Theorizing the multiplicity of digital phenomena: The ecology of configurations, causal recipes, and guidelines for applying qca. *Management of Information Systems Quarterly*, *44*, 1493–1520.

Porter's, V. C. M. (1985). What is value chain. *E-Commer.*, 1–13.

Puschmann, T. (2017). Fintech. *Business & Information Systems Engineering*, *59*(1), 69–76.

Ragin, C. C. (1994). Introduction to qualitative. *The comparative political economy of the welfare state*, 299.

Ragin, C. C. (2008). Measurement versus calibration: A set-theoretic approach.

Ridzuan, F., & Zainon, W. M. N. W. (2019). A review on data cleansing methods for big data. *Procedia Computer Science*, *161*, 731–738.

Rihoux, B. (2006). Qualitative comparative analysis (qca) and related systematic comparative methods: Recent advances and remaining challenges for social science research. *International Sociology*, *21*(5), 679–706.

Rihoux, B., & Ragin, C. C. (2008). *Configurational comparative methods: Qualitative comparative analysis (qca) and related techniques*. Sage Publications.

Roy, D., Srivastava, R., Jat, M., & Karaca, M. S. (2022). A complete overview of analytics techniques: Descriptive, predictive, and prescriptive. *Decision intelligence analytics and the implementation of strategic business management*, 15–30.

Rubalcaba, L. (2012). Business services in european economic growth. *Strategic Direction*.

Sadovskyi, O., Engel, T., Heininger, R., Böhm, M., & Krcmar, H. (2014). Analysis of big data enabled business models using a value chain perspective. *Multikonferenz Wirtschaftsinformatik (MKWI 2014)*, 1126–1137.

Saggi, M. K., & Jain, S. (2018). A survey towards an integration of big data analytics to big insights for value-creation. *Information Processing & Management*, *54*(5), 758–790.

Sagiroglu, S., & Sinanc, D. (2013). Big data: A review. *2013 international conference on collaboration technologies and systems (CTS)*, 42–47.

Sathe, S., Papaioannou, T. G., Jeung, H., & Aberer, K. (2013). A survey of model-based sensor data acquisition and management. In *Managing and mining sensor data* (pp. 9–50). Springer.

Schilling, R., Haki, K., & Aier, S. (2017). Introducing archetype theory to information systems research: A literature review and call for future research.

Schmidt, R., & Möhring, M. (2013). Strategic alignment of cloud-based architectures for big data. *2013 17th IEEE International Enterprise Distributed Object Computing Conference Workshops*, 136–143.

Schneider, C. Q., & Wagemann, C. (2010). Standards of good practice in qualitative comparative analysis (qca) and fuzzy-sets. *Comparative sociology*, *9*(3), 397–418.

Schüritz, R., Seebacher, S., & Dorner, R. (2017). Capturing value from data: Revenue models for data-driven services. *Proceedings of the 50th Hawaii international conference on system sciences*.

Sehring, J., Korhonen-Kurki, K., & Brockhaus, M. (2013). *Qualitative comparative analysis (qca): An application to compare national redd+ policy processes* (Vol. 121). CIFOR.

Serhani, M. A., El Kassabi, H. T., Taleb, I., & Nujum, A. (2016). An hybrid approach to quality evaluation across big data value chain. *2016 IEEE International Congress on Big Data (BigData Congress)*, 418–425.

Silva, B. N., Diyan, M., & Han, K. (2019). Big data analytics. In *Deep learning: Convergence to big data analytics* (pp. 13–30). Springer.

Simatupang, T. M., Piboonrungroj, P., & Williams, S. J. (2017). The emergence of value chain thinking. *International Journal of value chain management*, *8*(1), 40–57.

Soda, G., & Furnari, S. (2012). Exploring the topology of the plausible: Fs/qca counterfactual analysis and the plausible fit of unobserved organizational configurations. *Strategic Organization*, *10*(3), 285–296.

Soldatos, J., Troiano, E., Kranas, P., & Mamelli, A. (2022). A reference architecture model for big data systems in the finance sector. In *Big data and artificial intelligence in digital finance* (pp. 3–28). Springer, Cham.

Stoeckli, E., Dremel, C., & Uebernickel, F. (2018). Exploring characteristics and transformational capabilities of insurtech innovations to understand insurance value creation in a digital world. *Electronic markets*, *28*(3), 287–305.

Taber, K. S. (2018). The use of cronbach's alpha when developing and reporting research instruments in science education. *Research in science education*, *48*(6), 1273–1296.

Takacs, J., Pollock, C. L., Guenther, J. R., Bahar, M., Napier, C., & Hunt, M. A. (2014). Validation of the fitbit one activity monitor device during treadmill walking. *Journal of Science and Medicine in Sport*, *17*(5), 496–500.

Tallon, P. P. (2011). Value chain linkages and the spillover effects of strategic information technology alignment: A process-level view. *Journal of Management Information Systems*, *28*(3), 9–44.

Teixeira, J., Patrıcio, L., Nunes, N. J., Nóbrega, L., Fisk, R. P., & Constantine, L. (2012). Customer experience modeling: From customer experience to service design. *Journal of Service management*.

ur Rehman, M. H., Chang, V., Batool, A., & Wah, T. Y. (2016). Big data reduction framework for value creation in sustainable enterprises. *International journal of information management*, *36*(6), 917–928.

Vuorinen, I., Järvinen, R., & Lehtinen, U. (1998). Content and measurement of productivity in the service sector: A conceptual analysis with an illustrative case from the insurance business. *International Journal of Service Industry Management*.

Zhang, Y., & Ansari, N. (2013). On protocol-independent data redundancy elimination. *IEEE Communications Surveys & Tutorials*, *16*(1), 455–472.

Zhang, Y., Sun, Z., Sun, M., & Zhou, Y. (2022). The effective path of green transformation of heavily polluting enterprises promoted by green merger and acquisition—qualitative comparative analysis based on fuzzy sets. *Environmental Science and Pollution Research*, 1–17.

Zheng, Y., Zhao, K., & Stylianou, A. (2013). The impacts of information quality and system quality on users' continuance intention in information-exchange virtual communities: An empirical investigation. *Decision support systems*, *56*, 513–524.

A Interview Protocol

Interview protocol for collecting data on the 13 IISs operating at Allianz Benelux.

Dependent factor	**Elements**	**Items**	**Outcome**
Value in information use	Degree of operationalization	What is the current level of engagement/usage compared to the targeted/intended level of usage?	precent
Contingency factor	**Elements**	**Items**	**Outcome**
Information user	Type of information user	For what kind of information user is the information-intensive service intended?	a) internal users b) external users
	Size of information user	What is the number of total information users?	number
Independent factors	**Elements**	**Items**	**Outcome**
Data source	Type of data source	What kind of data source is used to generate data for the IIS?	a) human system b) engineering system
	Data format	What is the data format obtained from the data source?	a) unstructured data b) semi-structured data c) structured data
Data collection	Degree of human involvement	What is the degree of human involvement in the data collection process?	a) fully human involved b) semi-human involved c) fully automated
	Data collection mode	Which mode of data collecting is employed?	a) pull mode b) push mode
	Processing method	Which processing method is used to collect, transmit, preprocessing data (i.e. data cleansing, redundancy and reduction)?	a) batch processing method b) stream processing method

Independent factor	Elements	Items	Outcome
Data	Data quality assessment	All items are assessed on a scale of 1 to 5, with 1 being strongly disagree and 5 being strongly agree. Items containing the symbol '(R)' are reverse coded. Accuracy - This information is correct. - This information is incorrect (R). - This information is accurate.- This information is reliable. Completeness - This information includes all necessary values. - This information is incomplete (R). - This information is complete. - This information is sufficiently complete for our needs. - This information covers the needs of our tasks. - This information has sufficient breadth and depth for our task. Timeliness - This information is sufficiently current for our work. - This information is not sufficiently timely (R). - This information is not sufficiently current for our work (R). - This information is sufficiently timely. - This information is sufficiently up-to-date for our work. Consistency - This information is consistently presented in the same format. - This information is not presented consistently (R). - This information is presented consistently. - This information is represented in a consistent format.	Likert scale 1-5 1 = strongly disagree 2 = disagree 3 = neither agree nor disagree 4 = agree 5 = strongly agree
Data analysis	Degree of human involvement	What is the degree of human involvement in the data analysis process?	a) human-involved analysis b) automated analysis
	Robustness of the model	The analysis model is robust against difficulties and challenges.	Likert scale 1-5 1 = strongly disagree 2 = disagree 3 = neither agree nor disagree 4 = agree 5 = strongly agree
Information	Type of information	What type of information about the data source is created from the data analysis?	a) descriptive information b) predictive information c) prescriptive information
Information delivery	Degree of human involvement	To what extent are humans involved in the information delivery process?	a) fully human involved b) semi-human involved c) fully automated

B Truth-table fsQCA

Truth-table for low and high degree of operationalization. After calibration of the IIS-DVC elements, the truth-table algorithm was performed in order to find the different configurations for both outcomes.

User size	Robust-ness	Data quality	User type	Source type	Data format	Human in Data collection	Mode	Proces-sing	Human in Data analysis	Informat-ion type	Human in infor-mation de-livery	Number	High degree of operatio-nalization	Raw consist-ency	PRI consist-ency
0	1	1	0	0	0	0	1	0	0	1	0	2	1	1	1
1	1	1	0	1	1	0	0	0	1	1	0	2	1	1	1
1	1	1	0	0	0	0	0	0	0	0	0	1	1	1	1
0	1	1	1	0	0	0	0	0	1	0	0	1	1	1	1
0	0	1	0	0	0	1	1	1	1	1	1	1	1	1	1
User size	**Robust-ness**	**Data quality**	**User type**	**Source type**	**Data format**	**Human in Data collection**	**Mode**	**Proces-sing**	**Human in Data analysis**	**Informat-ion type**	**Human in infor-mation de-livery**	**Number**	**High degree of operatio-nalization**	**Raw consist-ency**	**PRI consist-ency**
0	1	1	0	0	1	0	0	0	0	0	0	1	1	1	1
1	1	1	0	0	0	0	0	0	1	0	0	1	1	1	1
1	1	1	1	1	1	0	0	0	1	0	0	1	1	1	1
0	0	1	0	0	0	0	0	0	0	0	0	1	1	0.95	0.94