Association for Information Systems
AIS Electronic Library (AISeL)

Spring 5-15-2016

UNDERSTANDING THE ANATOMY OF DATA-DRIVEN BUSINESS MODELS – TOWARDS AN EMPIRICAL TAXONOMY

Adrian Engelbrecht
TU Darmstadt, engelbrecht@is.tu-darmstadt.de

Jin Gerlach
TU Darmstadt, gerlach@is.tu-darmstadt.de

Thomas Widjaja
Universitat Passau, thomas.widjaja@uni-passau.de

Follow this and additional works at: http://aisel.aisnet.org/ecis2016_rp

UNDERSTANDING THE ANATOMY OF DATA-DRIVEN BUSINESS MODELS – TOWARDS AN EMPIRICAL TAXONOMY

Research

Engelbrecht, Adrian, TU Darmstadt, Darmstadt, Germany, engelbrecht@is.tu-darmstadt.de

Gerlach, Jin P., TU Darmstadt, Darmstadt, Germany, gerlach@is.tu-darmstadt.de

Widjaja, Thomas, Universität Passau, Passau, Germany, thomas.widjaja@uni-passau.de

Abstract

As a consequence of the increasing digitization, massive amounts of data are created every day. While scholars and practitioners suggest that organizations can use this data to develop new data-driven business models, many organizations struggle to systematically develop such models. A fundamental challenge in this regard is presented by the limited research on data-driven business models. Accordingly, the goal of this research is to better understand data-driven business models by identifying key dimensions that can be used to distinguish them and to develop a taxonomy. As our taxonomy aims to guide future studies in a way that ultimately serves organizations, it is based on dimensions regarded to be most relevant from the practitioners' perspective. To develop this taxonomy, we utilize an established empirical approach based on a combination of multidimensional scaling (MDS), property fitting (ProFit), and qualitative data. Our results reveal that the most important dimensions distinguish data-driven business models based on the data source utilized, the target audience, and the technological effort required. Based on these dimensions, our taxonomy distinguishes eight ideal-typical categories of data-driven business models. By providing an increased understanding regarding the topic, our results form the foundation for subsequent investigations in this new field of research.

1   Introduction

As a consequence of the increasing digitization, massive amounts of data are created every day. Both scholars and practitioners suggest that this trend towards "big data" could be an important source for companies to create new business value and to develop innovative business models (e.g., Chen et al., 2012; Lycett, 2013; Manyika et al., 2011; Woerner and Wixom, 2015). However, since the process of business model innovation is rather unstructured (e.g., Schneider and Spieth, 2013), the development of new business models based on data still remains a challenging endeavor. A fundamental reason for this challenge refers to the limited knowledge regarding data-driven business models. As previous research suggests, such new business models could significantly change the way organizations create value and therefore differ substantially from traditional ones (e.g., Bharadwaj et al., 2013; Veit et al., 2014). Following this perspective, we argue that it is crucial to provide a solid foundation for future research on this topic. Consequently, this research intends to better the understanding of data-driven business models by identifying key dimensions that can be used to develop a taxonomy in this regard.

Taxonomies are "systems for grouping objects of interest […] based on common characteristics" (Nickerson et al., 2013, p. 338). Thus, taxonomies help to structure a phenomenon and create a foundation for further research in this direction. According to the "science of diversity" (McKelvey, 1978; McKelvey, 1982), a solid understanding regarding the similarities and distinctions of objects is a crucial requirement to enable further research on these objects. Likewise, we argue that it is equally problematic for our research community trying to engage in collective research efforts and make generalized statements about data-driven business models, if there is a lack of common understanding. A central task for developing taxonomies is the selection of dimensions that are used to distinguish the considered objects. Usually, there are many possible dimensions that can be chosen to arrive at such a distinction. In order to identify the most relevant dimensions in a given context, it is crucial to clearly define the purpose of the taxonomy (Nickerson et al., 2013). The purpose of our research is to identify those dimensions considered to be the most relevant from a practitioners' perspective given that our taxonomy intends to guide future studies with the target of ultimately serving organizations.

In order to develop our taxonomy, we utilize an established empirical approach based on a combination of multidimensional scaling (MDS), property fitting (ProFit), and qualitative data (e.g., Padgett and Mulvey, 2007; Posey et al., 2013; Robinson and Bennett, 1995). In accordance with our research goal, we rely on data obtained from business model experts, as this group of people faces the challenge of developing new business models in their companies. In a first step, these experts need to assess the similarities of different data-driven business models. Afterwards, their similarity ratings are used to create a model of their cognitive mind sets. Thereby, we obtain dimensions relevant for distinguishing different business models and then associate meaningful labels to these dimensions. Our results reveal that the three most important dimensions distinguish data-driven business models according to the data source utilized (non-user data vs. user data), the target audience (consumer-focus vs. organization-focus), and the technological effort required (low vs. high). Based on these dimensions, our taxonomy distinguishes eight ideal-typical categories of data-driven business models.

Our research makes important contributions to research and practice. From a theoretical perspective, we provide an increased understanding regarding data-driven business models by identifying relevant dimensions by which these models can be distinguished. In the spirit of the "science of diversity" (McKelvey, 1978; McKelvey, 1982), our results form the foundation for subsequent investigations in this new field of research. For instance, future studies concerning the design of new methods to support the development of data-driven business models could build on our results considering the differentiation dimensions to structure the topic. As these differentiation dimensions relate to both data-related (e.g., big data) and business model research, we also address calls for more integrated investigations of these areas (e.g., Buhl et al., 2013; Loebbecke and Picot, 2015; Veit et al., 2014). From a practitioner's perspective, our research helps to understand data-driven business models by highlighting eight ideal-typical categories and the dimensions differentiating them. Consequently, organizations can develop new business models in a more structured manner as these categories help to inspire the innovation process and to define a target state. By diving deeper into relevant categories, organizations may also learn about category-specific challenges and existing practices to address them.

The remainder of this paper is structured as followed: in the next section, we will provide an overview of the theoretical background relevant for this study. We will then present the methodology used in our research and describe each step of our data collection and analysis in detail. In the subsequent section, we present the results of our study and finally conclude by discussing its results, implications, and limitations and offer suggestions to further investigate this topic.

2 Theoretical Background

In the following subsections, we provide an overview of two streams of research that constitute a relevant background for the presented study: 1) literature concerned with current developments on the

topic of "big data" and 2) research related to business models. As outlined below, both areas are increasingly concerned with the joint topic of data-driven business models.

2.1   How big data may foster business model innovation

The term "big data" refers to the emergence and use of massive amounts of data in nearly every part of our lives. This trend thereby describes both the quantitative raise of existing kinds of data and the availability of new kinds of data like those created by social media use (e.g., Woerner and Wixom, 2015). Big data is commonly conceptualized by the three dimensions volume, velocity, and variety (e.g., Chen et al., 2012; Lycett, 2013). As the mere availability of more data does not necessarily imply better data quality or an improved organizational performance (Buhl et al., 2013), the exact scope of this definition is not crucial. Instead, it is rather important to investigate how data can create new business value (Lycett, 2013) and how this development may lead to the transformation of organizations (Goes, 2014).

Three major areas of research that suggest how big data may create value and have an impact on organizations can be distinguished: first, organizational decision making and strategizing may become even more data-driven as it is today (Chen et al., 2012). This also refers to the challenge of how established management practices can be improved with new kinds of data (e.g., Bhimani, 2015; Constantiou and Kallinikos, 2014) and how companies in general are required to establish a data-driven organizational culture (Davenport, 2006; Sharma et al., 2014). Second, big data allows the improvement of existing products and services as well as the development of new ones (e.g., Davenport et al., 2012). Thereby, a major obstacle refers to the challenge that certain business potentials may not be generalizable across industries. Existing literature highlights several examples along these lines, but struggles to find common mechanisms that can be widely applied in organizations at a more general level (e.g., Davenport, 2006; Chen et al., 2012). Third, organizations may develop new business models on the foundation of their data, which we will discuss in the following.

Big data has the potential to foster business model innovation in organizations (e.g., Buhl et al., 2013; Manyika et al., 2011; Loebbecke and Picot, 2015). Yet, little empirical research has been conducted on this topic. Furthermore, no comprehensive overview of different possibilities and best practices regarding data-driven business model innovation exists. Nevertheless, several suggestions about how big data could be utilized for such innovative business models are put forward. According to Woerner and Wixom (2015), one possibility to innovate business models is to monetize data by selling raw data, enhanced data, or data-driven reports. In this case, traditional companies could establish new types of products and services, where the data itself is a part of the offering. Another aspect in creating data-driven business models refers to new dynamics in "the interplay between the offering and the customer" (Lycett, 2013, p. 382). By collecting and analyzing consumer data, organizations can customize existing products to perfectly match customer demands. A third approach that affects new business models relates to the development of new business ecosystems (Woerner and Wixom, 2015). Based on the data-driven interconnectedness of diverse entities in a market, increasing possibilities exist to integrate and unite business partners or customers in future business models.

(Big) data-driven business models are becoming a topic of increasing interest, as it has been suggested that innovating traditional business models from a data-driven perspective is necessary to be successful in the long-run (Buhl et al., 2013; Loebbecke and Picot, 2015). In this context, Loebbecke and Picot (2015) refer to the high pressure that results from new start-ups entering existing markets due to potentially low barriers for new players. Therefore, research needs to assess the potential of new data-driven business models (Buhl et al., 2013) in order to deepen the understanding of how companies can change their business models to realize the potentials of big data along with gaining

competitive advantage. In order to follow these calls for research, it seems crucial to first arrive at a shared understanding of what is meant by data-driven business models.

2.2   Business model research

In recent years, the business model concept has gained growing attention in academic literature, which is reflected in a rising number of publications (Zott et al., 2011). To date, there is no consistent definition of the term "business model", which has led to discussions regarding the existing definitions (e.g., Al-Debei and Avison, 2010; Burkhart et al., 2011; Zott et al., 2011). In the following, we refer to the definition by Osterwalder and Pigneur (2010, p. 14), who state that "a business model describes the rationale of how an organization creates, delivers and captures value."

Several studies are dedicated to the question which components are parts of a business model (e.g., Al-Debei and Avison, 2010; Hedman and Kalling, 2003; Osterwalder et al., 2005). In general, many possibilities to delineate the components of a business model exist. Osterwalder et al. (2005) describe a business model as a combination of nine elements: value proposition, target customer, distribution channel, customer relationship, value configuration, core competency, partner network, cost structure, and revenue model. In contrast, Hedman and Kalling (2003) choose to identify a business model by the components of customers, competitors, offering, activities and organization, resources, suppliers, and process. To gain a consensus from the diverse components offered in existing research, Burkhart et al. (2011) conducted a literature review to identify four groups of components that are common for the majority of approaches to conceptualize business models at a general level (see Table 1).

Component	Description
Offering factors	"Describe how the company creates value for its stakeholders"
Market factors	"Express for whom the company creates values"
Internal capability factors	"Deal with the internal activities and competences of the company"
Economic factors	"Bundle all economic-related aspects of the company"

Table 1.   Consensus of business model components (Burkhart et al., 2011, p. 10).

With regard to the increasing availability of potentially valuable data, researchers have argued that new technological developments can be an essential trigger affecting established business models (e.g., Veit et al., 2014). This aspect supports the proposition that companies need to adjust their business models in order to be successful in the long-run (Hanelt et al., 2015). This is particularly relevant for digital business models that can build on the large amounts of data, which arise due to the omnipresent digitization (Veit et al., 2014). Hartmann et al. (2014) have undertaken a first step to investigate the peculiarities of data-driven business models. In their working paper, they analyze these business models regarding the role of different data sources and key activities. Despite this first step, further research is required that provides guidance on how these new business models can be developed and how their success can be fostered (George et al., 2014).

Furthermore, extant research has emphasized the challenge of analyzing business models in the context of specific domains. As a consequence, various scholars have developed specific business model taxonomies in order to account for the particular characteristics of the considered domain (e.g., Burkhart et al., 2011; Pateli and Giaglis, 2004). For example, Schief and Buxmann (2012) examined the peculiarities of business models in the software industry considering the specific characteristics of software compared to other products. Likewise, one aim of the present taxonomy is to consider the specific characteristics of data-driven business models and therefore to extend the more generic business model literature in this direction.

Ultimately, both research streams on big data as well as business models are increasingly influenced by each other. This is remarkable as researchers from both fields have suggested that new data-driven business models could be an essential source for organizations to create new value. Consequently, we argue that there is a need for more integrated investigations of data-driven business models. Using experts' perceptions of data-driven business models to develop a taxonomy presents an important step to provide a relevant foundation supporting further research endeavors. As there is currently no precise definition of the term "data-driven business model", we introduce the following definition: "A business model of an organization is data-driven, if its core business necessarily requires digital data.[1]"

3 Methodology

In general, the purpose of a taxonomy is to group "objects of interests […] based on common characteristics" (Nickerson et al., 2013, p. 338). We refer to these characteristics as dimensions by which objects of interest can be differentiated. Thus, a taxonomy supports researchers to differentiate individual objects and to understand their relationships. This is helpful to examine complex topics and to potentially reveal new research areas (Nickerson et al., 2013).

In this research, we develop a taxonomy of data-driven business models. Our methodology is based on an established procedure, which involves a combination of multidimensional scaling (MDS), property fitting (ProFit), and qualitative data. Several previous studies have successfully used this procedure to develop taxonomies in contexts such as the technological influence on service interactions (Padgett and Mulvey, 2007), security behaviors (Posey et al., 2013), or workplace behaviors (Robinson and Bennett, 1995). The main advantage of this pluralistic approach lies in the combination of qualitative and quantitative analyses and therefore a better understanding of the examined research area by not focusing on a single approach (Orlikowski and Baroudi, 1991). Furthermore, in contrast to other approaches for developing taxonomies, this pluralistic approach involves a solid empirical basis used to identify relevant dimensions to group the objects of interest.

MDS is a powerful "set of mathematical techniques that enable researchers to uncover the 'hidden structure' of data bases" (Kruskal and Wish, 1984, p. 5). In our case, we uncover the structure (i.e., dimensionality) within a population of data-driven business models. The foundation to develop a taxonomy using MDS relies on the elicitation of mental perceptions of individuals regarding the similarities among objects of this population (Schiffmann et al., 1981). The obtained similarity ratings are then used to produce a representation of the data that builds the basis for identifying and labelling the dimensions underlying this representation (Kruskal and Wish, 1984; Schiffmann et al., 1981). Simply put, MDS allows researchers to investigate how people differentiate a set of objects (Posey et al., 2013). Relying on external sources for comparing the objects of interest has the additional benefit that a potential bias towards the researchers' subjectivity during the analysis can be reduced (Robinson and Bennett, 1995).

Going forward, we strictly adhere to this established process which involves five subsequent steps: (1) selecting data-driven business models; (2) acquiring similarity ratings; (3) determining the structure and dimensionality of the experts' perceptions using MDS; (4) identifying common characteristics based on qualitative data; (5) mapping of the characteristics to the dimension using a ProFit analysis.

3.1 Selecting data-driven business models

The first step aims at building a record of data-driven business models that represents the population of the objects of interest (i.e., data-driven business models). This record should be as exhaustive as

---

[1] While the trend of "big data" fosters the development of new business models, some data-driven business models might not require fulfilling the definition of a big data-driven business model. "Small data" can be used to create new business models as well and our study seeks to incorporate these business models.

possible in order to cover all existing kinds of data-driven business models. For this purpose, we chose to extract data from a database of start-ups, as innovative business models are often launched in start-ups before they are introduced in established organizations. Specifically, we retrieved data from "CrunchBase". CrunchBase gathers data on innovative companies using a crowdsourcing approach with a strong focus on start-ups as it maintains a large partnership program with more than 2,000 participants from the start-up community (e.g., accelerators, venture funds, and university programs) (CrunchBase, 2016a). According to the crowdsourcing approach, different groups of people (e.g., the user community or partnering organizations) can participate and improve or extend the database (CrunchBase, 2016b). Up to now, CrunchBase counts 10,000 individual contributors per month that have produced more than 500,000 datasets. These datasets are accessed by more than 2 million unique visitors per month (CrunchBase, 2016c). To ensure high data quality, CrunchBase provides several mechanisms: user authentication, algorithmic and personal reviews, and error reporting features (CrunchBase, 2016d).

CrunchBase uses categories to organize registered start-ups. One specific start-up can be assigned to several categories. In order to identify data-driven business models, two IS researchers independently went through the list of existing categories and tagged data-related categories to find a consensus. To validate the relevance of the chosen categories, we analyzed the first 20 start-ups (sorted by relevance for the according category) to decide whether or not a certain category should be included. Based on this procedure, we considered the following six categories of start-ups: big data, big data analytics, business analytics, predictive analytics, analytics, and data mining.

Using this set of categories we extracted the 50 most relevant start-ups of each category, resulting in an initial record of 300 start-ups. We ensured that every start-up included was based on a data-driven business model by comparing available information (collected from CrunchBase as well as the website of the start-up) with our definition of a data-driven business model. To ensure a high amount of objectivity, two IS researchers reviewed the start-ups independently. According to this procedure, 253 business models were dismissed, as the core of their activities obviously did not require data. After a consensus was found, we started to analyze the resulting 47 data-driven start-ups regarding duplicate business models. Thereby, we examined if two or more businesses in this record were based on a similar business model. Subsequently, the two IS researchers first coded outstanding similarities and afterwards discussed, which start-ups could be discarded. We finally ended up with 33 distinct data-driven business models that were examined in the subsequent steps.[2]

3.2   Acquiring similarity ratings

In a second step, we gathered similarity ratings that form the foundation to extract the taxonomy's dimensionality using MDS in step three. As a purpose of our research is to ultimately help practitioners concerned with the development of new business models, we consulted experts from this domain. Considering their perceptions builds the foundation for future research to help this group of interest. Consequently, we searched for business model experts that also provide entrepreneurial experiences as they had to assess business models of start-ups. We addressed both founders of existing start-ups as well as professionals with a strong focus on business models. Since entrepreneurs are frequently confronted with the challenge to improve and innovate their business model, they should provide extensive knowledge in this area. With regard to other business model professionals, we limited the search results to those who had an explicit record of entrepreneurial activities to ensure a high quality of our sample.

---

[2] A list of all data-driven business models with corresponding information (names, CrunchBase profiles, and descriptions) is available from the authors upon request.

We collected similarity ratings for each pair of data-driven business models. As the comparison of each possible pair by a single expert is not appropriate in terms of cognitive load, we followed Posey et al. (2013) and asked each expert to compare one specific data-driven business model with all remaining ones. Accordingly, 33 experts were involved in comparing all possible combinations of business models. Overall, we collected 1,055 similarity ratings from these experts regarding the presented business models. Experts' age ranged from 20 to 50 resulting in an average of 32.07 years of age (SD = 7.75). The business models were presented using a short text-based description adapted from the information available on CrunchBase and the specific website of the start-up. To ensure that each description contained sufficient information about the business model, we followed Burkhart et al. (2011) who proposed offering factors, market factors, internal capability factors, and economic factors as common features of every business model. To avoid a bias in the experts' similarity ratings, we did not explicate which business model components were used. The similarity ratings were provided using nine-point bipolar scales ranging from 1 = "not at all similar" to 9 = "very similar." For the qualitative part of the study, experts were asked to reflect on the criteria they used to compare the business models and enter these reflections into a text field after their similarity assessments.

3.3 Determining the structure and dimensionality of experts' similarity perceptions

In the third step, the data obtained from the previous step was analyzed using MDS. Specifically, we applied the PROXSCAL implementation included in SPSS. To calculate the structure of the experts' perceptions using the similarities, it is first necessary to decide how many dimensions should be used. Research on MDS proposes that interpretability is an important criterion that affects the choice of a reasonable number of dimensions (Kruskal and Wish, 1984; Schiffmann et al., 1981). Therefore, most researchers use a maximum number of three dimensions (e.g., Padgett and Mulvey, 2007; Posey et al., 2013; Robinson and Bennett, 1995) as it is hardly possible to interpret a higher dimensionality. To decide whether three or fewer dimensions should be chosen, the stress index can be used to further guide this decision. The stress index states how well the similarity ratings can be matched to a certain dimensionality (Robinson and Bennett, 1995). Therefore, it is required to minimize this stress index. In our case, we obtained the lowest stress level for a three-dimensional solution ($.06 < .11 < .28$). The second part of this step implies the graphical mapping of the experts' perceptions. Thus, MDS was used to calculate a position for every business model regarding the three dimensional space (coordinates for each dimension x, y, z). Thereby, MDS tries to locate all business models in a way that achieves the best fit with the empirical similarity ratings.

3.4 Identifying common characteristics based on qualitative data

The two subsequent steps were aimed at interpreting and labelling the three dimensions that explain the similarities and differences of the business models in the data. To later obtain meaningful labels, the procedure consults the qualitative data to extract those attributes that were used by the experts to arrive at their similarity ratings (Posey et al., 2013). In order to identify these attributes, all three authors studied the qualitative data in two steps: first, the authors independently coded the data and thereby created separate lists of attributes. The following example (see Table 2) illustrates how an expert's statement was used to extract different attributes:

Exemplary statement	*"[I compared the business models based on the] target group (consumers, companies) to which the business is selling a product or service."*
Resulting attributes	"The offering is not relevant for consumers." vs. "The offering is relevant for consumers."
	"The offering is not relevant for organizations." vs. "The offering is relevant for organizations."

Table 2.   Coding example.

Afterwards, potential differences were discussed in order to arrive at a single list of attributes. If two attributes were highly similar, they were merged. This resulted in 16 attributes. In a second step, the authors independently counted the occurrences of these attributes in the data to determine their relevance. To deal with possible subjectivities, the resulting frequencies of all three authors were added up to arrive at a joint ranking. As the next step associates the attributes extracted from the qualitative data with the dimensions resulting from MDS, the resulting set of attributes needs to be reduced for feasibility reasons for ProFit analysis (Posey et al., 2013; Robinson and Bennett, 1995). We selected the nine most frequently mentioned attributes that resulted from the coding process which are displayed in Table 3.

Attribute	Left Anchor	Right Anchor
REL_C	"The offering is not relevant for consumers."	"The offering is relevant for consumers."
REL_O	"The offering is not relevant for organizations."	"The offering is relevant for organizations."
COST_C	"The offering is available for free for consumers."	"The offering is not available for free for consumers."
COST_O	"The offering is available for free for organizations."	"The offering is not available for free for organizations."
TECH	"The business model requires small technological efforts."	"The business model requires high technological efforts."
U_DATA	"The business model is not based on user data."	"The business model is based on user data."
ADV	"The business model is not based on advertisement."	"The business model is based on advertisement."
SALE	"The business model is not based on the sale of data."	"The business model is based on the sale of data."
SERV	"The business model is not based on offering a service."	"The business model is based on offering a service."

Table 3.   Relevant attributes extracted from the qualitative data provided by the experts.

3.5   Mapping attributes and dimensions using ProFit analysis

ProFit analysis is based on multiple regressions to determine how well an object's location within the n-dimensional space (obtained by MDS) explains its value for each of the relevant attributes. To obtain the business models' values for the attributes extracted in the previous step, it was required to collect an additional round of data. We therefore asked 16 expert raters (i.e., 11 IS researchers and five research assistants trained on the topic) to evaluate all data-driven business models regarding each attribute identified. These raters have a profound background in the field of information technology and intensively cooperate with the local start-up center that fosters entrepreneurial activities. As they are confronted with the development of innovative business models in this regard, the raters are also well educated with respect to this topic.

As all attributes should be rated for every business model, we presented and surveyed each attribute on a 7-point bipolar scale as shown in Table 3. Thus, we collected scores for each attribute for every business model from all raters, which resulted in over 500 ratings for each attribute. These ratings allow subsequent regression analyses for the attributes regarding their relationship with the business model locations. One separate regression was computed for each attribute in relation to its position in the space created by MDS (Padgett and Mulvey, 2007; Posey et al., 2013; Robinson and Bennett, 1995). Particularly, attributes were used as the dependent variables and the coordinates of the business models' positions (x,y,z) were used as independent variables. Table 4 shows the results of these regressions.

Attribute	R²	F	Dim 1	Dim 2	Dim 3	Correlations					
						REL_C	REL_O	COST_C	TECH	U_DATA	ADV
REL_C	.22	4.07*	-.00	-.54**	-.02						
REL_O	.27	4.89**	.05	.48**	.32*	-.50**					
COST_C	.17	3.15*	-.32	-.37*	.04	.28	-.28				
TECH	.28	5.24**	.27	-.19	-.49**	.16	-.24	-.81			
U_DATA	.48	10.97***	.66***	-.25	.19	.26	-.29	-.98	.18		
ADV	.28	5.12**	.14	-.45**	-.36*	.51**	-.49**	-.12	.15	.19	
SERV	.20	3.64*	.31	.26	.33*	-.39*	.48**	-.20	.11	.20	-.37*
***p < .001, **p < .010, *p < .050.											

Table 4.   Results of the property fitting analysis.

Two of the initial nine attributes were excluded in the process (COST_O and SALE) as they had no significant relationship with any dimension resulting from MDS (COST_O: F = 2.24 with p = .11; SALE: F = .74 with p = .54). Regarding the remaining attributes, two were significantly related to more than one dimension (REL_O; ADV). In this case, we analyzed the corresponding regression weights to determine, to which dimension the attribute was related more strongly. Below, we describe how these results were finally used to arrive at meaningful labels for the dimensions obtained from MDS.

3.5.1   First dimension: data source (non-user data vs. user data)

In order to interpret the first dimension, we focused on the attribute U_DATA, because it was the only attribute that had a significant relationship with this dimension. As this attribute explains a high amount of variance and has the highest regression weight in this analysis (ß = .66), it was clear that this attribute appropriately describes the first dimension. Another indicator for the usefulness of this attribute relates to fact that it is not significantly correlated with another attribute. Thus, the first dimension can be clearly distinguished from the others. Since the associated attribute was surveyed using the bipolar scale "The business model is not based on user data" vs. "The business model is based on user data", we label this dimension "data source". The continuum of this dimension is therefore described using the extrema "non-user data" vs. "user data". To illustrate the characteristics of this dimension, we will provide two examples for each extreme point from our population of business models.

Based on the experts' ratings, "Company 3" had the highest score on this first dimension (.825) and therefore provides a good example for a company that is based on a non-user data-driven business model. Specifically, the company sells access to past and current high quality satellite images that can be used for different organizations, for example, to analyze the workload for providers of logistics services as they can assess how many trucks are currently located within a certain part of a harbor area. Obviously, the company is completely independent from user data as it relies entirely on satellite image data. In contrast, "Company 9" that had a score of -.829 on the first dimension runs a localized search engine that is entirely dependent on user data in form of their search queries and profiles. This dependency relates to the fact, that the provided user data builds the foundation to earn money by selling personalized advertisement.

3.5.2   Second dimension: target audience (consumer-focus vs. organization-focus)

Compared to the first dimension, four attributes were significantly related with the second dimension (REL_C; REL_O; COST_C; ADV). According to the highest regression weights, this second dimension was strongly associated with ratings whether business models were relevant for consumers

or organizations (REL_C, ß = -.54; REL_O, ß = .48). Since the dimension's relationships with these attributes are of opposite signs, dimension two indicates whether the business model's audience has a consumer-focus or an organizational-focus. Furthermore, this dimension also distinguishes business models in respect of whether the offering was a paid service for consumers (COST_C) and whether the business model was based on advertisement (ADV). This is reasonable as manifestations on these attributes should strongly depend on whether the business model is targeted at organizational customers or end users. Note that no significant correlation between COST_C and ADV could be observed in the data. This makes sense as businesses models targeted at consumers can be based on a service fee and advertising simultaneously. Likewise, there was no significant correlation between a business model's relevance for consumers (REL_C) and the costs for consumers (COST_C). This accounts for the possibility that consumer-oriented services within our population of business models were both available for free or for a charge. In sum, this dimension can be labelled "target audience", which varies from consumer-focused to organization-focused business models. As in the previous subsection, we will provide two examples of particular business models that are located near the extreme points of the dimension.

"Company 29" with a value of -.705 on this dimension is entirely focused on consumers as it offers personal health assistance that enables users to get an answer to their medical questions. Accordingly, everything this company does is aligned to serve consumers in a certain way. Therefore, they provide adaptive algorithms that are continuously getting better to intelligently support the users. In contrast, "Company 23" that scored .933 on this dimension, sells insights from legislative and regulatory data to foster transparency in the political and legal system. Accordingly, this offering is focused on helping organizations that are working in highly regulated environments like the financial sector. Therefore, the offering can be considered as largely irrelevant for consumers.

3.5.3 Third dimension: technological effort (high vs. low)

Four attributes were significantly related to dimension three (REL_O; TECH; ADV; SERV). Considering the regression weights, the technological effort of a business model is the most important attribute for this dimension (TECH, ß = -.49). We labelled this dimension accordingly referring to the amount of technological effort required to provide the service (e.g., data collection, transformation, integration, analysis). Based on this label, it seems curious at first that a business model's relevance for organizations (REL_O) and the dependency on advertisement (ADV) were also significantly related to this dimension. Looking into the population of our business models, business models targeted at organizations often required less technological effort as they often offered data or data-based insights without complex processing to their organizational customers who might use these data or insights in their own processes. For consumer-focused business models, no such relationship could be observed which could explain why the business model's relevance for consumers was not associated with dimension three. In a similar fashion, advertisement often requires high technological effort as personalization algorithms are based on a rather complex integration and analysis of different kinds of data.

The fourth attribute associated with dimension three refers to the question whether a business models' offering was based on providing services. This attribute seems problematic for the purpose of differentiating the three dimensions as its regression weights were almost equal for all three dimensions (the significance levels of the coefficients for the first two dimensions were slightly above 5% level). Therefore, we excluded this attribute from our analysis and interpretation of the dimensions. In sum, we interpreted dimension three as the amount of "technological effort" involved in the data-driven business model. Note the negative sign of the regression coefficient which means that a high value on dimension three refers to little amount of technological effort involved and vice versa.

In order to illustrate the dimension label, we discuss two examples of particular business models with high absolute scores on dimension three. "Company 6" with a score of -.675 on this dimension offers a service for consumers that helps them to buy flight tickets at the best price. Therefore, the company uses massive amounts of pricing data gathered from the Internet to automatically predict the development of ticket prices to support consumers. Accordingly, this offering requires high technological efforts to predict accurate results using intelligent algorithms. In contrast, "Company 32" that scored .900 on dimension three offers a simple register of health clubs in a database that helps business owners to promote their health club and consumers to find an appropriate one. Thereby, rather little technological efforts are necessary.

4 Results

To sum up, the most relevant dimensions to distinguish data-driven business models according to the perspective of business model experts refer to the data source utilized, the target audience, and the technological effort required. Every dimension is shaped using two extreme points. Specifically, a data-driven business model can be based on non-user or user data, can focus on consumers or organizations, and can require a high or low technological effort. Using this differentiation, eight ideal-typical categories result. In order to foster a deep understanding, we provided several examples from our data record to illustrate how particular business models regarding these extreme points could look like. In order to present our findings, we provide a visualization using a decision tree that is shown in Figure 1.
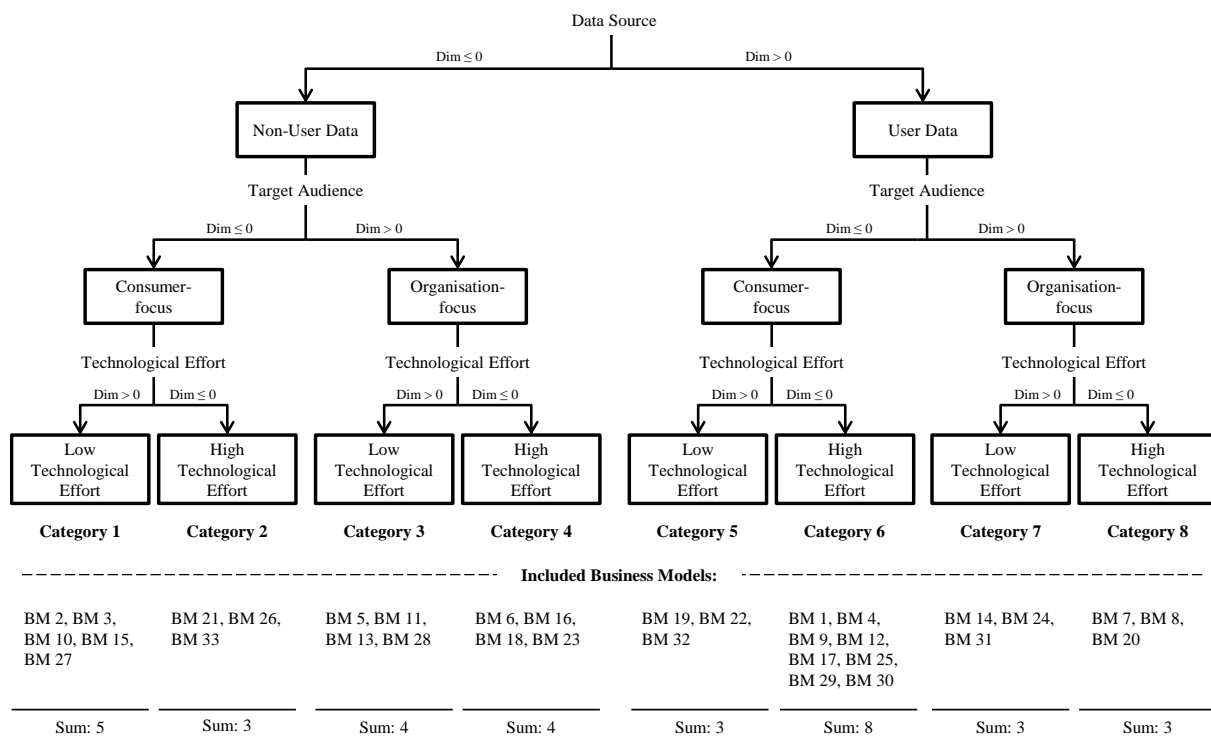

Figure 1.   Proposed taxonomy of data-driven business models illustrated using a decision tree.

We also use this visualization to provide a mapping of the particular business models to the categories of our taxonomy. This was done by analyzing the position of each business model regarding the three dimensions. We want to emphasize that, obviously, hybrid business models exist with values on the dimensions in between the extrema. For example, a business can provide a service for both consumers

and organizations. At this point, we focus on the identification of ideal-typical categories and therefore concentrate on highlighting the meaning of the different endpoints but acknowledging the possibility of hybrid scenarios. As the business models are distributed rather equally across the classes, the dimensions resulting from this research are suitable to distinguish the record of business models. The minimum number of business models in category is three, the maximum is eight. In order to analyze the most important categories regarding our taxonomy, we take a further look at category six, which contains the highest number of business models according to our taxonomy. This category can be described by the following characteristics: the business model is based on user data and focuses on offering a product or a service to consumers. Furthermore, it requires a high technological effort. Building on these characteristics, we can find offerings created with complex algorithms on the basis of big user data. An example for this category is "Company 4" that computes a data-driven credit score for consumers on the basis of social media data provided by them. As a result of an intelligent usage of this data, the credit score can be calculated in a few minutes and is independent from a lot of critical information that is required in traditional credit scores.

5 Discussion

5.1 Theoretical contributions

By revealing the perception of business model experts and the most important dimensions to distinguish data-driven business models, this research contributes by providing a foundation for this increasingly relevant research area. Therefore, in the spirit of Posey et al. (2013), we focus on the "science of diversity" that investigates a population of objects by highlighting and understanding the similarities and differences of the objects in question (McKelvey, 1978; McKelvey, 1982). In contrast, the "science of uniformity" seeks to discover the "universal laws governing the behavior, function, and processes of a population of objects" (McKelvey, 1982, p. 12). As the "science of uniformity" obviously depends on the "science of diversity", our approach is crucial to further study data-driven business models. Accordingly, the presented results may form the foundation for various theory building efforts regarding data-driven business models as we help to create a common understanding based on which important attributes and categories of data-driven business models should be considered. While this common understanding is essential for the consideration of several research questions, we will highlight two particular aspects.

We argue that our resulting dimensions emphasize the importance of an integrative consideration of data-related (e.g., big data) and business model research as we revealed relevant dimensions from both research areas. Therefore, we empirically support the literature's suggestion that it is essential to combine relevant insights from both areas (e.g., Buhl et al., 2013; Loebbecke and Picot, 2015; Veit et al., 2014). Thereby, our results can help to instantiate general business model representations that typically do not consider any technical or data-driven aspects (e.g., Al-Debei and Avison, 2010; Hedman and Kalling, 2003; Osterwalder et al., 2005). Consequently, data-specific attributes (e.g., the data source and its corresponding characteristics) may be integrated into established representations.

Moreover, our results can be a useful foundation to guide design science research efforts that develop entirely new methods, which support organization in identifying new business models (Gregor and Hevner, 2013). While literatures points to the relevance of business model innovation (e.g., Hanelt et al., 2015), existing methods (e.g., Gassmann et al., 2015; Osterwalder and Pigneur, 2010) might be too abstract to guide business model developers to work in data-driven areas. In contrast, our study is based on business model experts' perception of data-driven business models. Therefore, we argue that this perspective is particularly valuable as new methods drawing on these results can consider the experts' way of thinking about data-driven business models and hence, help to support them in the best possible way.

5.2   Practical contributions

Our study strongly builds on the perceptions and experiences of business model experts and we argue that this focus on the experts' mind-set leads to a high practical relevance of our results. Indeed, one of the initial aims of this study was to support this group of people in their daily activities. As discussed before, the development of new data-driven business models may be very beneficial for organizations to create new value. While most organizations lack a fundamental understanding of this new topic, we contribute to practice by providing an overview of eight ideal-typical categories of data-driven business models and the dimensions distinguishing them. Thereby, we help to establish a fundamental understanding that allows organizations to purposefully develop data-driven business models in a more structured manner. This is especially relevant as the business model innovation process is usually rather unstructured (e.g., Schneider and Spieth, 2013). Accordingly, our results support organizations in identifying a possible target state (i.e., a possible business model) by giving an overview of different kinds of data-driven business models and thereby showing organizations which paths they can follow. In addition to guiding the organization's own path to develop a new data-driven business model, the proposed taxonomy can also be used to segment the market by identifying comparable providers or possible competitors.

As soon as an organization has identified in which category it aims to develop a data-driven business model, the organization can benefit from looking at other companies whose business models fall into the same category. In this way, the organization can get inspired by the examples of others and analyze how the corresponding providers handle category-specific challenges. For instance, an organization might want to develop a business model that can be characterized by the utilization of user data, a consumer-focus, and high technological effort. Examples of this category are services that provide personalized content using complex recommendation algorithms, such as Netflix. Netflix offers personalized movie and TV show recommendations based on an extensive analysis of viewing preferences. These personalized offerings face the significant challenge that a user may become isolated from content that does not fit to his or her profile (i.e., filter bubble). Therefore, an organization developing its business in this category can try to learn how to handle this challenge by analyzing how these existing companies operate. Accordingly, such comparisons point to areas of expertise, which are required depending on the type of business model that is intended to be developed.

5.3   Limitations and future research

In this section we will discuss the limitations of our research and point to avenues for further research. One limitation concerns the population of business models considered in this research. On the one hand, it seems possible that the chosen data source (i.e., CrunchBase) does not cover all types of data-driven business models as it focuses on start-ups. Hence, traditional companies may have different business models requiring resources that cannot be provided by start-ups. On the other hand, our definition of data-driven business models excludes those business models that do not necessarily require digital data. Consequently, there might be businesses that use data in complementary business functions, which are not regarded in this study. Therefore, future research could examine data-driven business models in a broader context and analyze if additional types of business models emerge. Furthermore, it is also possible that new kinds of data-driven business models will be developed in the future that are not considered in our sample to date. As a consequence, it could be helpful to validate our results using different populations of data-driven business models. In addition, developing taxonomies is associated with the challenge to trade-off between generic and specific dimensions. Relying on established methods, our study has extracted rather generic dimensions. Therefore, future research may contribute by further exploring these dimensions.

6 References

Al-Debei, M. M. and D. Avison (2010). "Developing a unified framework of the business model concept." *European Journal of Information Systems,* 19 (3), 359-376.

Bharadwaj, A., O. A. El Sawy, P. A. Pavlou and N. Venkatraman (2013). "Digital business strategy: Toward a next generation of insights." *MIS Quarterly,* 37 (2), 471-482.

Bhimani, A. (2015). "Exploring big data's strategic consequences." *Journal of Information Technology,* 30 (1), 66-69.

Buhl, H. U., M. Röglinger, F. Moser and J. Heidemann (2013). "Big data." *Business & Information Systems Engineering,* 5 (2), 65-69.

Burkhart, T., J. Krumeich, D. Werth and P. Loos (2011). "Analyzing the business model concept – a comprehensive classification of literature." *Proceedings of the 32th International Conference on Information Systems.* Shanghai, China.

Chen, H., R. H. Chiang and V. C. Storey (2012). "Business intelligence and analytics: From big data to big impact." *MIS Quarterly,* 36 (4), 1165-1188.

Constantiou, I. D. and J. Kallinikos (2014). "New games, new rules: Big data and the changing context of strategy." *Journal of Information Technology,* 30 (1), 58-59.

CrunchBase (2016a). *Venture program.* URL: https://info.crunchbase.com/about/crunchbase-venture-program/ (visited on 03/15/2016).

CrunchBase (2016b). *About.* URL: https://info.crunchbase.com/about/ (visited on 3/15/2016).

CrunchBase (2016c). *The business graph.* URL: https://info.crunchbase.com/the-business-graph/ (visited on 3/15/2016).

CrunchBase (2016d). *FAQ.* URL: https://info.crunchbase.com/about/faqs/ (visited on 3/15/2016).

Davenport, T. H. (2006). "Competing on analytics." *Harvard Business Review,* 84 (1), 98-108.

Davenport, T. H., P. Barth and R. Bean (2012). "How 'big data' is different." *MIT Sloan Management Review,* 54 (1), 21-24.

Gassmann, O., K. Frankenberger and M. Csik (2015). *The business model navigator: 55 models that will revolutionise your business.* Harlow, England, Pearson.

George, G., M. R. Haas and A. Pentland (2014). "Big data and management." *Academy of Management Journal,* 57 (2), 321-326.

Goes, P. B. (2014). "Big data and IS research." *MIS Quarterly,* 38 (3), iii-viii.

Gregor, S. and A. R. Hevner (2013). "Positioning and presenting design science research for maximum impact." *MIS Quarterly,* 37 (2), 337-356.

Hanelt, A., B. Hildebrandt and J. Polier (2015). "Uncovering the role of is in business model innovation – a taxonomy-driven aproach to structure the field." In*: Twenty-Third European Conference on Information Systems (ECIS).* Münster, Germany.

Hartmann, P. M., M. Zaki, N. Feldmann and A. Neely, (2014). "*Big data for big business? A taxonomy of data-driven business models used by start-up firms.*" URL: http://cambridgeservicealliance.eng.cam.ac.uk/resources/Downloads/Monthly%20Papers/2014_March_DataDrivenBusinessModels.pdf (visited on 3/20/2016).

Hedman, J. and T. Kalling (2003). "The business model concept: Theoretical underpinnings and empirical illustrations." *European Journal of Information Systems,* 12 (1), 49-59.

Kruskal, J. B. and M. Wish (1984). *Multidimensional scaling.* Beverly Hills, California, Sage Publications.

Loebbecke, C. and A. Picot (2015). "Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda." *The Journal of Strategic Information Systems,* 24 (3), 149-157.

Lycett, M. (2013). "'Datafication': Making sense of (big) data in a complex world." *European Journal of Information Systems,* 22 (4), 381-386.

Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh and A. H. Byers. 2011. "*Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute." URL: http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation (visited on 3/15/2016).

McKelvey, B. (1978). "Organizational systematics: Taxonomic lessons from biology." *Management Science,* 24 (13), 1428-1440.

McKelvey, B. (1982). *Organizational systematics: Taxonomy, evolution, classification.* Berkeley and Los Angeles, California, University of California Press.

Nickerson, R. C., U. Varshney and J. Muntermann (2013). "A method for taxonomy development and its application in information systems." *European Journal of Information Systems,* 22 (3), 336-359.

Orlikowski, W. J. and J. J. Baroudi (1991). "Studying information technology in organizations: Research approaches and assumptions." *Information Systems Research,* 2 (1), 1-28.

Osterwalder, A. and Y. Pigneur (2010). *Business model generation: A handbook for visionaries, game changers, and challengers.* Hoboken, New Jersey, John Wiley and Sons.

Osterwalder, A., Y. Pigneur and C. L. Tucci (2005). "Clarifying business models: Origins, present, and future of the concept." *Communications of the Association for Information Systems,* 16 (1), 1-25.

Padgett, D. and M. S. Mulvey (2007). "Differentiation via technology: Strategic positioning of services following the introduction of disruptive technology." *Journal of Retailing,* 83 (4), 375-391.

Pateli, A. G. and G. M. Giaglis (2004). "A research framework for analysing ebusiness models." *European Journal of Information Systems,* 13 (4), 302-314.

Posey, C., T. Roberts, P. B. Lowry, B. Bennett and J. Courtney (2013). "Insiders' protection of organizational information assets: Development of a systematics-based taxonomy and theory of diversity for protection-motivated behaviors." *MIS Quarterly,* 37 (4), 1189-1210.

Robinson, S. L. and R. J. Bennett (1995). "A typology of deviant workplace behaviors: A multidimensional scaling study." *Academy of Management Journal,* 38 (2), 555-572.

Schief, M. and P. Buxmann (2012). "Business models in the software industry." In*: 45th Hawaii International Conference on System Science (HICSS).*

Schiffmann, S. S., M. L. Reynolds and F. W. Young (1981). *Introduction to multidimensional scaling – theory, methods, and applications.* Orlando, Florida, Academic Press, Inc.

Schneider, S. and P. Spieth (2013). "Business model innovation: Towards an integrated future research agenda." *International Journal of Innovation Management*, 17 (1), 1-34.

Sharma, R., S. Mithas and A. Kankanhalli (2014). "Transforming decision-making processes: A research agenda for understanding the impact of business analytics on organisations." *European Journal of Information Systems,* 23 (4), 433-441.

Veit, D., E. Clemons, A. Benlian, P. Buxmann, T. Hess, D. Kundisch, J. Leimeister, P. Loos and M. Spann (2014). "Business models." *Business & Information Systems Engineering,* 6 (1), 45-53.

Woerner, S. L. and B. H. Wixom (2015). "Big data: Extending the business strategy toolbox." *Journal of Information Technology,* 60 (1), 60-62.

Zott, C., R. Amit and L. Massa (2011). "The business model: Recent developments and future research." *Journal of Management,* 37 (4), 1019-1042.