REGULAR PAPER

Towards a taxonomy for business capabilities determining data value

Markus Hafner[1] · Miguel Mira da Silva[1]

© The Author(s) 2023

Abstract
Data and its valuation have gained vital significance in academia and enterprises, coinciding with diverse data valuation approaches encompassing various layers, dimensions, and characteristics. This paper assesses data value determination through a business capability lens based on the TOGAF standard. The paper encompasses (a) constructing a Data Valuation Business Capability (DVBC) taxonomy and (b) validating the taxonomy using two existing data valuation concepts from academia. The methodology involves information systems taxonomy development techniques backed by a previously conducted systematic literature review of 64 articles. The resultant taxonomy comprises four business capability layers, nine dimensions, and 36 characteristics. These layers and dimensions offer business, technology, and organizational perspectives, reflecting the interdisciplinary nature of data valuation alongside an enterprise architecture. Characteristics within these layers and dimensions are either exclusive or non-exclusive based on their contents. The compiled findings meet both objective and subjective quality criteria. The implications of the DVBC are multifaceted, influencing scholars and professionals alike. Scholars gain a cohesive tool enhancing transparency in the extensively debated data value domain, fostering linkages among information systems, enterprise architecture management, and data management. This empowers the progress in developing comprehensive data valuation concepts. Additionally, professionals may employ the DVBC taxonomy as a lighthouse and guiding tool, fostering internal dialog on data valuation. This entails elevating data valuation to a pivotal business capability, necessitating collaborative, regular assessment, and enhancement involving business and technological stakeholders. By adopting this taxonomy, the challenge of consistently determining data value can be effectively addressed in both academia and enterprises.

Keywords Data value · Data valuation · Business capability · Taxonomy

✉ Markus Hafner
markus.hafner@tecnico.ulisboa.pt

Miguel Mira da Silva
mms@tecnico.ulisboa.pt

[1] Instituto Superior Técnico, Av. Rovisco Pais 1, 1049-001 Lisbon, Portugal

🙂 Springer

Abbreviations

AI	Artificial intelligence
CIO	Chief information officer
C2E	Conceptual-to-empirical
DVBC	Data valuation business capability
E2C	Empirical-to-conceptual
IT	Information technology
Iter	Iteration
SLR	Systematic literature review
TOGAF	The open group architecture framework

1 Introduction

Generating value from and with data is not just a short-term hype but a paradigm firmly established in various industries and enterprises [1–7]. For this emerging paradigm, it is particularly relevant to determine the value of data, the associated data initiatives, and data-driven use cases systematically and standardized [2, 8, 9].

The current solutions for determining the data value in science and practice are scattered over a broad range with a wide variety of content depth. While several enterprises do not determine their data value at all [2], others use estimation techniques based on knowledge and gut feeling of subject matter experts [10]. Furthermore, a previous and currently submitted literature review unveiled that there are data valuation concepts from many research areas, especially computer science, decision science, as well as business, management, and accounting. It is noticeable that these data valuation approaches pursue different, if not opposing, targets and often occur encapsulated from each other. Consequently, it is challenging for enterprises to identify, classify, as well as to implement data valuation approaches in their ecosphere [11–15].

This paper defines data valuation as a holistic business capability to streamline the ideas, concepts, and approaches regarding data value determination and make them applicable to enterprises. This not only includes the pure determination of the data value but also expands the focus to include processes, people, resources, and information [16–18] in data valuation endeavors. Furthermore, the authors hypothesize that it is beneficial for professionals and scholars to obtain a classification and differentiation key for data valuation business capabilities (DVBC) [19, 20]. Consequently, within the scope of this paper, a DVBC taxonomy is developed, which supports practitioners in analyzing and understanding the complex topic of data valuation and deriving targeted conclusions for their individual data monetization paths. For developing the DVBC taxonomy, the following research question serves as a compass:

RQ What are the main dimensions and characteristics of a data valuation business capability?

To contribute closing to the creation of value from and with data and close the existing research gap, we answer the research question by describing the scientific background in Sect. 2. In particular, Sect. 2 discusses the concepts of business capabilities and data value, including their relationship to one another, and describes related taxonomies. Section 3 elaborates on the systematic literature review (SLR) as the foundation for the subsequent taxonomy development process tailored to information systems, according to [19]. The final

DVBC taxonomy is presented in Sect. 4 before being tested in Sect. 0 using two data valuation approaches provided by academia. In Sect. 6, the findings are discussed before Sect. 7 completes with a conclusion.

2 Research background

This section elaborates on the concepts of business capabilities and data value. Furthermore, adjacent taxonomies are described.

2.1 Business capabilities

The definition of a business capability is not consistently expressed in academia. A literature analysis carried out by [18] underlines this conclusion and therefore bundles the various definitions of a business capability as follows: "A particular ability that a business may possess or exchange to achieve a specific corporate goal" [18].

Since the above definition is rather generic, we use the more detailed and practical definition of a business capability according to the *TOGAF* standard. *TOGAF* is considered scientifically sound [18, 21, 22] and is practically in use [23], which is crucial for the applicability of a business capability taxonomy in practice. According to *TOGAF*, a business capability comprises *information, processes, roles*, and *resources*. *Information*, meaning the knowledge associated with a business capability [16, 17], is crucial to perform various *processes* [16, 24] within a business capability. These processes may be executed by people [17] associated with *roles* [16] using tangible and intangible *resources* [16, 24].

2.2 Data valuation business capability

The delimitation of the term data value serves as the foundation to understand what a data valuation business capability means. Data and the associated information imply a value in diverse shades that can be generated by implementing data-driven use cases or through the direct or indirect sale of data [2–5]. This data value can be of social-ecological, economical [25, 26], functional, and/or symbolic nature [10, 26] to add a measurable business value [13]. Thereby, data value is determined by a multitude of value drivers [27, 28], underlying theories [29–33], as well as frameworks [34, 35].

The contents and purpose of the above-mentioned data value definitions suggest applying the business capability concept to achieve the purpose of determining the data value. Thus, the idea of bundling isolated approaches from the field of data value emerges, including the data value drivers, theories, as well as frameworks and expanding them in the context of a business capability. Therefore, the resulting data valuation business capabilities comprise information, processes, roles, and resources (see Sect. 2.1).

2.3 Related taxonomies

In academic literature, there have already been attempts to explain the traits of data value and its determination approaches, for example, in the form of taxonomies. These existing concepts have been elaborated especially in the two domains (IT) business value and data value.

One taxonomy for classifying value catalogs was developed by Seufert et al. [36] to intensify the linkage between enterprise performance and IT investments. Value catalogs are therefore defined as reference lists that determine the economic impact of the utilization of information technology. In the value catalog taxonomy, dimensions are introduced as a foundation for the DVBC taxonomy. Specifically, the dimension relating to *methods for quantifying the value* via value catalogs is particularly close to the DVBC taxonomy. Thereby, methods under certainty and uncertainty encompassing static, dynamic, and qualitative characteristics are distinguished. However, the development of additional IT business value assessment approaches is noted as an open research gap toward the DVBC taxonomy can contribute.

Engel et al. [37] also adapted the concept of business value and concretized it to artificial intelligence (AI) use cases. The resulting taxonomy aims to identify dimensions and characteristics that enable the value contribution of AI use cases at an organizational scope. The dimensions of *source of business value improvement* and *benefit to business value* are particularly relevant for the DVBC taxonomy. The *source of business value improvement* classifies the effect implied by an AI use case. This effect, either of automation, transformation, or information nature, can serve as a basis for formulating value drivers and theories for determining the value. Furthermore, the *benefit to business value* dimension distinguishes which type of improvement, as a result of an AI use case contributes to the increase in business value. The characteristics of *cost, quality, revenue*, as well as *risk and compliance performance* can all impact business value. Thus, these characteristics are declared non-exclusive since an AI use case can cover several performance improvements simultaneously. The DVBC taxonomy uses a similar logic.

In addition to taxonomies in the area of business value and information technology, Lega et al. [38] have developed a taxonomy regarding data value in the context of decision-making. Two main dimensions of data value are defined, namely *data quality* and *data utility*, which are also included as non-exclusive characteristics in the DVBC taxonomy under the dimension *data value driver*. In the decision-making data value taxonomy, the data value dimensions and their characteristics are subdivided into more fine-grained metrics that promote the applicability of the taxonomy. The decision-making data value taxonomy thus promotes an understanding of the scope and limits of data value but does not address the classification of data valuation approaches in the context of business capabilities in an enterprise architecture. As a further research step, it is proposed to concretize data value assessment frameworks, to which the DVBC taxonomy also contributes.

3 Research design

The following subsections describe the methods used to develop the DVBC taxonomy. In addition, relevant literature for the taxonomy design is analyzed.

3.1 Literature review

The first step of our research design is adhering to [39] recommendations for SLRs combined with the forward- and backward search, according to [40].

As a basis for our SLR, the authors developed a search query concerning our research question. The search query ("Compan*" OR "Enterprise*") AND ("Data Pric*" OR "Data Valu*") AND ("Approach*" OR "Architecture*" OR "Capabilit*" OR "Method*" OR "Model*") NOT ("Data Values") combines data valuation approaches with enterprise architectures and

Definition of inclusion / exclusion criteria and development of search query	Querying in abstracts **n = 304**	Deleting duplicates **(-133)** **n = 171**	Screening abstracts in detail **(-140)** **n = 31**	Executing Backward Search **(+23)** **n = 54**	Executing Forward Search **(+10)** **n = 64**

Fig. 1 Systematic literature review process based on [39, 40]

capabilities. Furthermore, related synonyms are covered, while data values, which refer to technical values of storages, are explicitly excluded.

To consider a complete set of relevant published full-text access literature (journal articles, conference proceedings) in the English language from 2012 onwards, the search query was applied to the databases *ACM Digital Library, AIS eLibrary Ebsco, Emerald Insight, IEEE Xplore, ScienceDirect, Scopus, Web of Science,* and *Wiley Online Library*. Figure 1 illustrates the SLR and related findings.

In total, 304 scientific contributions were recognized as raw sample. Out of the raw sample, 133 duplicates were removed. An abstract reading was used for the ensuing 171 scientific contributions, which led to the deletion of another 140 articles and a sample size of 31 articles. According to [40], both backward (+ 23) and forward search (+ 10) were applied to fill in potential gaps in the search query. Thus, the authors expanded the size of the final sample from 33 to 64 scientific contributions.
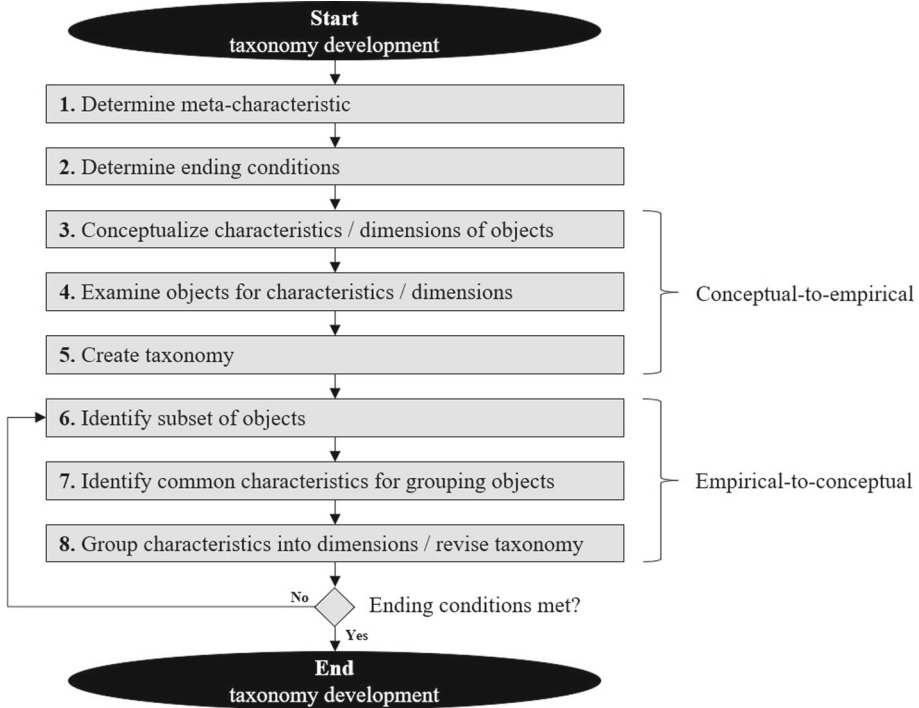
3.2 Taxonomy development process

After the first foundational step, the systematic literature review, the taxonomy development method according to [19] was applied, which is tailored to the research area information systems. As a wide range of domains addresses the research topic data value, the iterative approach of the method is particularly well suited to encompass the complexity and interdisciplinarity of the topic.

Nickerson et al. [19] propose two taxonomy development approaches. While the conceptual-to-empirical approach is more suitable when little data is available on the objects to be classified (in this case, DVBC), the empirical-to-conceptual (E2C) approach is more ideal for a taxonomy with a broad data foundation. In this paper, a combination of both approaches was used. In the first iteration (iter.), the conceptual-to-empirical (C2E) approach was applied to establish guiderails for the proper definition of business capabilities. The depth of the content of a DVBC in relation to dimensions and characteristics was created using the empirical-to-conceptual approach in further iterations, as many data valuation approaches are available due to the SLR carried out. Figure 2 below shows the applied method.

The first step in taxonomy development, according to [19] is to determine the meta-characteristics of the taxonomy. It is of particular relevance that the meta-characteristic is aligned with the purpose of the taxonomy (see Sect. 1). Therefore, the meta-characteristic may be described as the identification of the especially pertinent layers, dimensions, and characteristics of a DVBC. In addition, the applied methodology in step 2 defines ending conditions of subjective and objective nature, as illustrated in Table 1.

In the context of this paper, *conciseness* (consider a maximum of five to nine dimensions [19, 41]), *robustness*, *comprehensiveness*, *extendibility*, and *explainability* are subjective ending conditions. To meet scientific requirements, [19] further specify eight objective ending conditions. These ending conditions have been condensed into the four applicable ending

Fig. 2 Taxonomy development method for information systems by [19] applied to this paper

conditions *completeness*, *granularity*, *uniqueness*, and *stability,* to ensure a simplified under-standing of this taxonomy development without significantly changing their content. lists and explains the ending conditions and shows the development of the degree of fulfillment of the ending conditions after each iteration.

3.2.1 Conceptual-to-empirical iteration

In the first iteration of the taxonomy development process, the framework for detailing a DVBC was created using the conceptual-to-empirical approach (see steps three to five in Fig. 2). Specifically, the framework serves as a logical bracket of the underlying dimensions as well as characteristics. The technique of determining a framework using layers has already been applied in various taxonomies [42–44] and is therefore defined as state-of-the-art.

The defined layers of *information*, *resources*, *roles*, and *processes* are based on the com-ponents of a business capability according to *TOGAF* (see Sect. 2.1) [16, 45]. From this, the hypothesis is derived: every DVBC includes the four dimensions of *information*, *resources*, *roles*, and *processes*.

3.2.2 Empirical-to-conceptual iterations

The second iteration is based on the SLR and therefore follows the empirical-to-conceptual approach in the taxonomy development process. Based on the SLR, we found that the litera-ture on data valuation is examined from various angles, especially from the research areas of

Table 1 Degree of fulfillment of taxonomy development ending conditions per iteration

Ending conditions	Explanation	Iter. 1	Iter. 2	Iter. 3	Iter. 4	Iter. 5	Iter. 6
		C2E	E2C	E2C	E2C	E2C	E2C
Subjective							
Conciseness	Does the taxonomy's quantity of dimensions allow it to be relevant without becoming r overwhelming?			X	X	X	X
Robustness	Do the dimensions and characteristics allow for enough distinction between objects to be classified?		X	X	X	X	X
Comprehensiveness	Are all dimensions of the objects of interest recognized?					X	X
Extendibility	Is it simple to add a new dimension or a new characteristic to an existing dimension?		X	X	X	X	X
Explainability	Do the dimensions and characteristics enlighten about an object to be classified?					X	X
Objective							
Completeness	Has a representative sample of objects been investigated?		X	X	X	X	X
Granularity	Is at least one object categorized under each characteristic of each dimension?					X	X
Uniqueness	Are dimensions, characteristics, and their combinations free of duplications?		X	X	X	X	X
Stability	Is the final iteration free of additions, mergers, and/or splits of dimensions, characteristics, and/or objects?						X

business, management, and accounting as well as *computer science* according to [46]. The high granularity level in the literature regarding approaches and concepts to determine the data value was lifted to a higher level of abstraction and therefore formed the three taxonomy dimensions *data value driver*, *data valuation theory*, as well as *data valuation tooling*.

In the third iteration, the *information* layer was completed. In this context, the *purpose* dimension is particularly relevant to adequately frame a business capability [16, 18, 24, 47]

as well as to serve as a starting point for data valuation [27]. In addition, a DVBC requires information on objects, which will be classified in the taxonomy. Consequently, the *data valuation object* dimension was included.

The fourth iteration sought to complete the *roles* layer. The analysis of the SLR findings illustrates that two main groups particularly stood out about their frequency of mentions concerning past and future research. The first, most frequently, and intensely discussed group includes all stakeholders involved in determining the data value and is therefore summarized under the *value determination stakeholder* dimension in this taxonomy [29, 48]. The second relevant group is derived particularly from the relevance for future research and can be subsumed in the dimension *value auditing stakeholder* [48–51].

The fifth iteration deepened the functional content and outcomes of the *process* layer of a DVBC. From this, the dimension *component* is formed, which details the content of a DVBC [52]. Further, the outcomes are considered in the *result* dimension [35, 49, 53].

In the sixth iteration, all of the ending conditions of a subjective and objective nature (see Table 1) were met, and thus the taxonomy development process was terminated. The outcome of the last sixth iteration is declared as a final DVBC taxonomy and is further described below.

4 Taxonomy: data valuation business capability

The final taxonomy for DVBC is described in the following sections. In total, the developed taxonomy in Table 2 has four layers, nine dimensions, and a subset of 36 characteristics.

During the taxonomy development process, the authors learned that it is impossible to limit the choice of characteristics to be mutually exclusive (indicated with letter E) for some dimensions since important information would be lost. The dimensions of *data value driver*, *data valuation theory,* and *component*, in particular, imply the most diverse forms of expression per scientific study, per enterprise, and per data valuation approach. Consequently, non-exclusive characteristics (indicated with letter N) in these dimensions are supported in accordance with other recently published taxonomies in the research areas of digital transformation and information systems [37, 44, 78, 79]. Further, the visualization style of a morphological box is chosen to increase the intuitiveness and usability of the taxonomy [80, 81].

4.1 Layer 1—information

The first layer of a business capability, and therefore also of this taxonomy, comprises information or so-called knowledge that a business capability requires and consumes to determine a data value [16]. Specifically, the information layer encompasses three dimensions *purpose*, *data valuation object,* and *data value driver*.

4.1.1 Purpose

The *purpose* dimension characterizes the goal of the data valuation and thus sets the guide rails for the detailing of a DVBC. Two characteristics and their combination can therefore be exclusively defined.

While *qualitative data valuation* focuses on generating contextual knowledge about the data value [34, 38, 50], *quantitative data valuation* concentrates on numerical information [34, 35, 54]. The existing literature shows that a *combination* [34] of both characteristics to different extents is also possible.

Table 2 Taxonomy for data valuation business capabilities (*E, Exclusive; N, Non-Exclusive)

Layer	Dimension	Characteristics							E/N*
Information	Purpose	Qualitative data valuation [34, 38, 50]	Quantitative data valuation [34, 35, 54]				Combination [34]		E
	Data valuation Object	Bundled Data [15, 55–57]				Non-bundled data [57]			E
	Data value Driver	Business Utility [12, 26, 58]	Cost [29, 34, 59, 60]	Data durability & Lifetime [61, 62]		Data Security & Privacy [60, 65–67]	Sentiment & Perception [28, 32, 57, 68–70]	Others	N
Resources	Data Valuation Theory	Economic [29, 32, 54]	Cooperative Game Theory [30, 31]	Non-Coop. Game Theory [57]	Decision theory [51, 62, 71]	Query-based [29, 60, 72–74]	Index-Based [59, 62]	Proprietary [10]	N
	Data valuation Tooling	Interpersonal Elaboration [12, 35, 50]		Model and application [48, 61, 75]			Combination [34, 59]		E
Roles	Value determination stakeholder	Internal [34, 51]	Internal and External (with Intermediary) [34, 48, 50, 57]	Internal and External (without Intermediary) [50, 57]			External [34, 50, 51]		E

Table 2 (continued)

Layer	Dimension	Characteristics				E/N*
	Value auditing stakeholder	Internal data value auditor [50, 51]	3rd Party data value auditor [50, 51]	Not existing		E
Process	Component	Data value Assessment [12, 50, 76]	Data value Allocation [52]	Data value Prediction [34, 50, 52]	Data value Monitoring [52, 76]	N
	Result	Specific absolute data value [34, 53]	Approximate absolute data value [73, 77]		Relative data value [35]	E

4.1.2 Data valuation object

After the question of why data valuation should take place (purpose), the dimension *data valuation object* aims at the question whose value should be determined.

From this, the two exclusive characteristics of *bundled information* and *non-bundled information* can be defined. Bundled information refers to data clusters that can be logically grouped according to their content in each use case. Examples of this are data products, data assets, and datasets in various forms. In contrast, non-bundled information are isolated raw data points that have not yet undergone any logical clustering [15, 55, 56].

4.1.3 Data value driver

The third dimension of *data value driver* poses the question of which parameters affecting the data value are considered for the data value determination. Since data valuation is an interdisciplinary topic with an arbitrarily high degree of complexity, it is evident that the *data value driver* dimension cannot have exclusive characteristics. Instead, DVBC can consider a variety of data value drivers. At this point, it is hypothesized that the more data value drivers are considered, the more accurate the determined data value will be. This hypothesis underlines the non-exclusive nature of the dimension, although the hypothesis needs to be validated in another study.

We distinguish six data value drivers and, to underline the non-exclusive nature of this dimension, add the characteristic *other*. The *other* characteristic includes both proprietary data value drivers and potential additional data value drivers outside the analyzed literature.

The characteristic *business utility* includes the impact of a data-driven use case on a process or an enterprise [12, 26, 58]. In addition, the characteristic *cost* may consist of expenses associated with the data being evaluated, such as data collection, processing, analysis, and management costs [29, 34, 59, 60] as well as opportunity costs [35, 58, 82]. Furthermore, we complement data management related data value drivers and tailor them into the characteristics of *data durability and lifetime* [61, 62], *data quality* [51, 58, 60, 63, 64], and *data security and privacy* [60, 65–67]. As a final characteristic, data value drivers are considered, which imply a particular subjectivity. The *sentiment and perception* characteristic includes, for example, the perceived data value and, thus the willingness-to-pay [28, 32, 57, 68, 69] as well as risks associated with the valuation and monetization of data [70].

4.2 Layer 2—resources

The second layer in the DVBC taxonomy considers the associated resources of tangible and intangible types, which are required for determining the value of data [16]. As an intangible resource, the dimension *data valuation theory* is introduced. Furthermore, the dimension *data valuation tooling* represents the tangible dimension of the resource layer.

4.2.1 Data valuation theory

Similar to the data value drivers in Sect. 4.1.3, the dimension *data valuation theory* is defined as a cluster of non-exclusive characteristics since DVBC can be based on numerous theories to determine the data value. Here, we distinguish seven data valuation theories.

The first characteristic of *economic* comprises all theories that determine the data value based on price-quantity diagrams, cost curves, and conventional hardware-oriented pricing (cost, competition, customer) [29, 32, 54]. In addition, *game theory* can be used as a data valuation theory, which can be divided into two characteristics, *cooperative* [30, 31] and *non-cooperative game theory* [57]. The fourth characteristic *decision theory* summarizes approaches, e.g., analytic hierarchy process [51, 62] or fair knapsack [71], that assess the data value while considering uncertainty and vagueness. A more technical data valuation theory deals with the valuation of database queries of different types and can therefore be subsumed under the term *query-based* [29, 60, 72–74]. Furthermore, the sixth characteristic, *index-based* deals with the indexation of data value drivers for determining an indexed data value [59, 62]. In addition to the aforementioned data valuation theories, which represent certain paradigms in data value determination, the seventh characteristic clusters all *proprietary* theories that have not been considered in the taxonomy dimension so far or are based on an expert's gut feeling only [10].

4.2.2 Data valuation tooling

To combine and apply data valuation theories and data value drivers, scholars propose different, rather personal as well as rather application-based approaches. Consequently, we define three exclusive characteristics in our taxonomy's dimension data valuation tooling.

The first characteristic of the data valuation tooling dimension is *interpersonal elaboration*. Interpersonal elaboration is a vehicle to support data value determination used by multiple scholars. The assessment of the data value and their use cases by domain experts is particularly relevant for this characteristic [12, 35, 50].

In contrast, some models and applications support and facilitate data value determination. Models and applications can occur in various forms. More theoretical constructs such as economic cost or price curves are considered, as well as ecosystem-oriented intermediary solutions, such as data marketplaces [48, 61, 75].

As a third characteristic in the data valuation tooling dimension, a *combination* of interpersonal elaboration as well as models and applications is introduced, since particularly practical research approaches suggest these combined data valuation approaches [34, 59].

4.3 Layer 3—roles

The third layer of the DVBC taxonomy focuses on the roles and responsibilities that stakeholders, individuals, and organizational units play to determine data value [16]. In concrete terms, the roles layer consists of two role dimensions that deal with the determination of data value (*value determination stakeholder*) on the one hand and its auditing (*value auditing stakeholder*) on the other.

4.3.1 Value determination stakeholder

The stakeholders included in the determination of data value play a central role in a DVBC since they affect an enterprise's ecosystem. Four exclusive characteristics in the form of a continuum are established in this taxonomy, ranging from the inclusion of purely *internal* stakeholders to the inclusion of strictly *external* stakeholders. Mixed forms of internal and external stakeholders are defined as intermediate characteristics. These diverse forms include direct collaboration *without an intermediary* as well as *with an intermediary* such as a data broker or data marketplace [29, 48, 70]. Regardless of the internal or external nature of data valuation stakeholders, such as data providers [29, 48, 83] or data buyers [29, 48, 83], all are relevant to some extent in a variety of data valuation approaches, which underscores their necessity in this taxonomy.

4.3.2 Value auditing stakeholder

The second dimension of the roles layer containing three exclusive characteristics is *value auditing stakeholder*. Auditing the data value is relevant for validating its determination process and result. At this point, *internal data value auditors* or *third-party data value auditors* can occur in science and practice [48, 50, 51]. Moreover, no audit can be performed, leading to the third characteristic *not existing*.

4.4 Layer 4—processes

The fourth layer of the DVBC taxonomy focuses on related processes and patterns to accomplish a certain output of a business capability [16]. To be more precise, in this case, *components* and *results* are defined as dimensions under the process layer.

4.4.1 Component

The components dimension describes the leading practices of a business capability, which in turn include individual sub-processes, activities, and functional modules. The results of the executed SLR suggest that four main characteristics of a DVBC can be formed. The components and, thus characteristics in this dimension can also occur in a combined manner. Consequently, the component dimension, including its characteristics, is also defined as non-exclusive.

A predominant number of data valuation approaches focus on the data value assessment, i.e., the pure determination of the data value, which is why the *data value assessment* is recorded as a characteristic in this taxonomy [12, 50, 76]. In addition, some approaches assign the data value to dedicated entities. The resulting characteristic is described as *data value allocation* [52]. Two further components include the temporal dimension in their scope of functions. While the *data value prediction* characteristic forecasts a future data value [50, 52], e.g., via customer lifetime value [34], the *data value monitoring* characteristic compares the planned and actual data value [52, 76].

4.4.2 Result

The results dimension contrasts three exclusive characteristics that describe the outcome of the DVBC. On the one hand, a *relative data value* can be determined [35], which compares

individual data values, data initiatives, and use cases with each other in relative terms and outputs them, for example, in order of the corresponding value. On the other hand, an absolute data value can be defined as the result of a data valuation. A distinction can be made between a *specific absolute data value* and an *approximate absolute data value*. While the specific absolute data value aims at an exact determination of the data value [34, 53], the approximate absolute data value focuses on a corresponding estimation of the data value, for example, considering uncertainty or the necessary computing power [73, 77].

5 Application of the DVBC taxonomy

This section assesses the DVBC taxonomy's utility and exemplifies its intended usage, as suggested by [19]. To test our DVBC taxonomy, we have selected two contemporary data value determination approaches that follow a more technology-supported approach by [35] on the one hand and a more ecosystem-supported multilayered approach by [34] on the other hand.

Table 3 illustrates the findings of the application, which are further described in Sects. 5.1 and 5.2.

5.1 Approach 1—automatic data value analysis method for relational databases

Approach 1 by [35] is an automated, metric-based data value assessment technique that provides a scoring mechanism for automatically determining the business value of bundled data, aka datasets data in relational databases.

To calculate the data value, a multitude of metrics are considered that are represented as data value drivers in the DVBC taxonomy, which legitimates the non-exclusive definition of its data value driver dimension. In concrete terms, the metrics *utility*, *volume*, and *usage* of approach 1 can be assigned to the taxonomy characteristic *business utility*. The approach of [35] also reflects the characteristic *data quality* by the metrics *timeliness* and *quality*. The *cost* characteristic is embodied by Bendechache et al. via the metric *replacement costs*. In addition, [35] also reflect the characteristic *sentiment and perception* by including legal risk and competitive advantage metrics.

The resource layer of the DVBC taxonomy is characterized by a combination of *interpersonal elaboration* as well as *models and applications* for determining the data value for the example of approach 1. On the one hand, Bendechache et al. apply a survey-based questionnaire, which is validated using a query-based approach and assign corresponding data value indices in the form of a scoring system to the datasets. Therefore, *internal stakeholders* are required to determine the data value, even though no *audit* takes place.

As a result of the approach 1, a *relative data value* is issued by the scoring system, which serves solely for the *data value assessment*.

5.2 Approach 2—from qualitative to quantitative data valuation

In contrast to approach 1, the data valuation approach of [34] combines both qualitative and quantitative elements. The object of valuation in approach 2 is bundled data, described as datasets, which are based on use cases. In addition to considering *business utility* as a data value driver, data attributes in general are considered, which are represented by the characteristics of *data durability and lifetime*, *data quality*, as well as *data security and privacy*

Table 3 Application of the DVBC taxonomy to data valuation approaches (● = [34]; ○ = [35])

Layer	Dimension	Characteristics						E/N
Information	Purpose	Qualitative Data Valuation	Quantitative Data Valuation ○			Combination ●		E
	Data valuation Object	Bundled Data ●○	Non-Bundled Data					E
	Data value Driver	Business Utility ●○	Cost ●○	Data Durability & Lifetime ●	Data Quality ●○	Data Security & Privacy ●○	Sentiment & Perception ○	Others	N
Resources	Data valuation Theory	Economic ●	Cooperative game theory	Non-coop. game theory	Decision theory	Query-based ○	Index-Based ○	Proprietary ●	N
	Data valuation Tooling	Interpersonal elaboration	Model and application			Combination ●○		E
Roles	Value determination stakeholder	Internal ○	Internal and External (with Intermediary) ●	Internal and External (without Intermediary)	External			E
	Value auditing stakeholder	Internal Data Value Auditor	3rd Party Data Value Auditor		Not Existing ○			E
Process	Component	Data value assessment ●○	Data value allocation	Data Value Prediction ●	Data Value Monitoring ●			N
	Result	Specific absolute data value ●	Approximate Absolute Data Value	Relative Data Value ○				E

in the DVBC taxonomy. Furthermore, *costs* relating to the data value are considered, which are processed, among other things, in a data valuation theory with economic characteristics. This is the so-called *cost-based* and *transaction-based data valuation* approach. To complete the data valuation framework, *proprietary* theories are included, such as defining threshold metrics, which increase or decrease the data value to a dedicated percentage.

To enable data value determination, [34] apply both *interpersonal elaboration,*, e.g., in the form of an initial segmentation of relevant use cases, as well as models and applications, e.g., data marketplaces as vehicles for transaction-based data valuation.

The value determination stakeholders can be *internal and external*, considering intermediaries, e.g., data marketplaces. Similar to approach 1, there is no data value audit in approach 2.

As a result of the data valuation framework, according to Stein et al., a *specific absolute data value* is to be ascertained, which is used for *data value assessment* on the one hand and for *data value prediction* on the other, for example in the form of the customer lifetime value.

6 Discussion

In order to better contextualize the DVBC taxonomy within existing literature, the contents of this paper are discussed and categorized both in terms of content and methodology at this juncture. This serves to render the quality and potential limitations of the paper transparent.

6.1 Methodology-related discussion

The methodological approach adheres to established academic standards for the construction of the systematic literature reviews [39] and taxonomies within the domain of information systems, specifically emphasizing digitalization, data, its value, and the associated ecosystems [42–44, 84].

However, it must be noted that although data valuation approaches and concepts are described in the present literature sample, they are not yet defined as business capabilities. Instead, this DVBC taxonomy serves to raise the multitude of data valuation approaches to the level of abstraction of a business capability to categorize them in the DVBC taxonomy. The latter becomes evident in the overview of capabilities in handling data value by Zeleti and Ojo [85]. Notably absent from this list is the inclusion of data value determination as a capability. This omission can be attributed to two factors. Firstly, the paper by Zeleti and Ojo [85] was published in 2017, and since then new insights into data valuation have emerged. Furthermore, the capabilities identified by Zeleti and Ojo [85] intersect significantly with the DVBC taxonomy, indicating that the present DVBC taxonomy implies many elements of Zeleti and Ojo [85], thus serving as a meaningful complement.

To comprehend and compare existing data valuation concepts within business capabilities, it is crucial to generalize fine-grained concepts and detail generic concepts. For this DVBC taxonomy, the *TOGAF* standard served as a frame construct providing the layers of *information, processes, roles*, and *resources*. At this point, however, it is essential to note that other notions for describing a business capability [17, 18] also have their reason for existence and could result in a different structuring of the taxonomy.

6.2 Content-related discussion

The content-related discussion of the DVBC taxonomy assesses the extent to which it complements existing literature, addresses research gaps, and justifies its raison d'être.

Initially, the DVBC taxonomy utilizes and complements existing taxonomies in literature. Firstly, the DVBC taxonomy leverages the "taxonomy of incentive mechanisms for data sharing in data ecosystems" of [84] as an impulse generator and builds upon this taxonomy by initiating a step prior to data sharing, specifically focusing on data value determination. However, the layers, dimensions, and characteristics identified by Gürpinar et al. [84] are partially utilized and further developed in the DVBC taxonomy.

Secondly, the DVBC taxonomy complements the reference value catalog taxonomy of [36] by introducing a data-centric perspective. This includes, for instance, addressing methods for determining value in general (qualitatively and quantitatively) and augmenting them within the data context in the DVBC taxonomy. Additionally, [36] emphasize uncertainty within value determination, which, within the DVBC taxonomy, is only partially represented in the dimension of *result*. While the goal of complexity reduction has been achieved, it may be necessary to provide further detail on uncertainty for future expansion and application of the DVBC taxonomy.

Thirdly, the "decision-making data value taxonomy" by [38] defines data value based on data utility and quality. Concerning the analyzed literature, it can be noted that while these two dimensions are indeed components of data value and a business capability for its determination, they are not entirely comprehensive. Hence, this DVBC taxonomy integrates data utility and data quality as characteristics under the dimension of *data value driver*. Additionally, the DVBC taxonomy incorporates further characteristics, dimensions, and layers to ensure a more applicable taxonomy in practical contexts.

Due to the DVBC taxonomy's emphasis on the practical applicability and implementation of a DVBC capability within an enterprise, two perspectives are not explicitly included in the taxonomy but serve as essential supplements: data value definition and external influences. The scientific literature provides a multitude of definitional frameworks and delineations of the term data value [11, 25, 26], which are recommended to be considered when discussing DVBC. However, their differentiation does not appear as a dimension or characteristic in the DVBC taxonomy to reduce complexity at this point and render the DVBC taxonomy applicable to professionals. Furthermore, the external perspective on data value determination, particularly the market and competitive structure of an enterprise [75], can be particularly relevant for implementing and applying a DVBC. This is why the work of [75] is viewed as a complementary pillar to the present DVBC taxonomy.

To conclude the content-related discussion, a fundamental aspect can be noted, which is often emphasized in the scientific literature and underscored by the DVBC taxonomy. It is the necessity of a symbiosis between business and technological perspectives in the purposeful management of data, and consequently, in the determination of data value [3, 14]. For instance, while DVBC taxonomy dimensions such as *data valuation object, data value driver*, or *data valuation tooling* predominantly address technological aspects of data valuation, the dimensions of *value determination, value auditing*, as well as *stakeholder* and *component* focus more on the business or procedural perspectives.

6.3 Quality-related discussion

To ensure high quality of the taxonomy and identify potential weaknesses, as well as areas for future research, it is imperative to subject the DVBC taxonomy to quality standards. Therefore, scholars have shown that the taxonomy evaluation criteria *(a) usefulness, (b) applicability, (c) comprehensiveness, (d) robustness, (e) conciseness, (f) extensibility*, and *(g) explanatory* are of particular relevance [86] and met by the DVBC taxonomy (see Table 1).

Usefulness refers to how the DVBC taxonomy, serving as the examined artifact, aids practitioners in accomplishing a specific objective [86]. Given that one objective of this study is categorizing diverse data valuation concepts as business capabilities, it is plausible to characterize the resultant DVBC taxonomy as useful.

Applicability, denoting the extent to which the artifact is put to practical use [86]. In Sect. 5, the DVBC taxonomy is assessed by applying two data valuation concepts, indicating that the developed taxonomy possesses scientific applicability. Nevertheless, the applicability of the DVBC taxonomy in real-world settings remains an area warranting further investigation in subsequent research endeavors.

Comprehensiveness, in the sense of the DVBC taxonomy's ability to adequately differentiate various data valuation concepts [86], is indeed achieved. The DVBC taxonomy encompasses technological and business viewpoints, mirroring the depth and breadth found in other scientifically robust taxonomies within information systems [42, 44].

Robustness, defined as the durability of the DVBC taxonomy over time [86], can also be affirmed within the scope of this paper. This is because the DVBC taxonomy exhibits a consistent nature from iteration two of the taxonomy development process onwards. However, it is worth noting that future research should focus on scrutinizing the robustness of the DVBC taxonomy through real-world case studies.

Conciseness, which pertains to the simplicity of the DVBC taxonomy [86], can also be verified at this juncture. This is substantiated by the strictly defined upper limit of nine dimensions [19, 41] and by the deliberate exclusion of factors such as external elements that could divert and complicate the focus of the DVBC taxonomy. To prevent any potential gaps in understanding the DVBC taxonomy, additional complementary taxonomies and frameworks from academia were described and related to the DVBC taxonomy.

Extensibility, defined as the capacity to incorporate additional dimensions and characteristics [86], is met, especially concerning potentially new characteristics. The DVBC taxonomy does not impose restrictions on the inclusion of new characteristics within existing dimensions. The same holds true for the extension of further dimensions. However, it is worth noting that in doing so, it is ideal not to compromise the quality criterion of conciseness or if necessary, to do so only with a valid justification.

Explanatory applies when the DVBC taxonomy can describe the objects under evaluation [86], such as data valuation concepts. At this juncture, the dimension of being explanatory can also be verified. This is because the scope of nine dimensions and a total of 36 characteristics provides the requisite breadth and depth for explaining and categorizing data valuation concepts.

In addition, [86] define six guidelines to ensure the compliance of excellence standards regarding the development of taxonomies. *Guideline 1* (scoping of taxonomy evaluation), *guideline 2* (justification of objective ending conditions), *guideline 3* (justification of subjective ending conditions), and *guideline 4* (demonstration of DVBC taxonomy applicability) have been successfully carried out based on this paper. Further, *guideline 5* (evaluation of

DVBC taxonomy usefulness) and *guideline 6* (long-term re-evaluation) are planned for future work.

Consequently, the excellence standards for methodology, content, and quality of the developed DVBC taxonomy are considered as given with the note of future field testing and long-term validation.

7 Conclusions

Our research is motivated by several contributions highlighting the significance of data valuation and demand for more in-depth study in this area [11–15, 38]. Therefore, this paper focuses on developing a data valuation business capability to provide a concept to classify and structure existing and emerging data valuation approaches from the perspective of a business capability. To achieve this, the taxonomy development method for information systems by [19] is applied and tested against two contemporary data valuation approaches [34, 35], and the taxonomy is reviewed for its raison d'être in terms of methodology, content [87] and quality based on the evaluation criteria by [86]. The result is an excellence-assured data valuation business capability taxonomy consisting of four layers, nine dimensions, and 36 characteristics.

The implications of the developed DVBC taxonomy are considerable versatility. For academic scholars, these implications manifest in two primary ways: Firstly, the taxonomy establishes a vital connection between the realms of information systems, enterprise architecture management, and data management. This connection is instrumental in catalyzing the growth of these domains for fostering the advancement of comprehensive data valuation concepts. Secondly, the taxonomy enhances clarity and insight into extensively debated subjects concerning the nature of data value and the strategies for its operationalization. By incorporating enterprise architecture management as a central bracket, this taxonomy contributes to transparency and enables the transfer of data valuation theories from academia to practical deployment in real-world enterprises.

For professionals, the implications of this taxonomy are threefold. Firstly, considering that a substantial number of enterprises continue to encounter challenges in effectively gauging data value, this taxonomy serves as an illuminating lighthouse. It shifts the perspective on data valuation away from a sporadic and non-standardized notion toward an integral business capability firmly embedded within an enterprise's architecture. Secondly, the layers, dimensions, and characteristics outlined in the DVBC taxonomy may serve as a bedrock for professionals, facilitating the development of enterprise-specific assessments of data valuation maturity. This, in turn, assists enterprises in systematically constructing their data valuation capabilities while pinpointing pivotal action areas in this domain. Lastly, professionals operating within the CIO domain can leverage the taxonomy to comprehend and effectively communicate to stakeholders that data valuation transcends being solely a technological or business matter. Instead, it represents a symbiosis of these facets, underscoring the need for cohesive collaboration. This realization empowers professionals to engage the necessary stakeholders in a manner that fosters value generation with and through.

Our research has limitations. One limitation relates to the structure of the taxonomy, which is based on the practice-approved and scientifically sound *TOGAF* standard, knowing that other structuring concepts exist for describing a business capability. Another content limitation relates to the dimensions and characteristics based on the results of a systematic literature review. It is possible that particular concepts in the literature may not be covered by

this taxonomy. Due to the breadth and scope of existing data valuation approaches and their meaningful integration in this taxonomy, some taxonomy dimensions have been defined as non-exclusive, which may dilute the strict delineation of data valuation business capabilities to some degree. Another limitation lies in the explicit exclusion of external elements related to data valuation, which is aimed at maintaining the focus and simplicity of the taxonomy. Nevertheless, it is important to note that this may necessitate an expansion of the DVBC taxonomy in future practical settings, a possibility facilitated by the design of the taxonomy itself.

With a view to future scientific work, three thematic areas can be formulated. On the one hand, it is recommended to validate the developed taxonomy in a field study of long-term nature. In addition, a data value ontology, which follows scientifically sound standards such as *OntoClean* [88] could be developed based on the taxonomy. Furthermore, the concept for defining data valuation as a business capability should be tested and refined in real-world scenarios following the design science research paradigm [87].

8 Data availability statement and declarations

The information utilized for performing both the upstream SLR and the existing classification scheme, aimed at establishing a data valuation business capability, has been sourced from the institutional licenses of Instituto Superior Técnico, Lisbon. The search engines employed within the scholarly databases *ACM Digital Library, AIS eLibrary Ebsco, Emerald Insight, IEEE Xplore, ScienceDirect, Scopus, Web of Science,* and *Wiley Online Library* were implemented following the protocol detailed in Sect. 3.1 and illustrated in Table 4 below to foster a more comprehensive outline promoting scientific transparency and replicability [89].

Table 4 Details regarding the systematic literature transparency

Search engine	Link	Search query
ACM digital library	https://dl.acm.org/search/advanced	("Compan*" OR "Enterprise*") AND ("Data Pric*" OR "Data Valu*") AND ("Approach*" OR "Architecture*" OR "Capabilit*" OR "Method*" OR "Model*") NOT ("Data Values")
AIS eLibrary Ebsco	https://aisel.aisnet.org/do/search/advanced	
Emerald insight	https://www.emerald.com/insight/advanced-search	
IEEE Xplore	https://ieeexplore.ieee.org/search/advanced	
ScienceDirect	https://www.sciencedirect.com/search/entry	
Scopus	https://www.scopus.com/search/form.uri?display=advanced	
Web of science	https://www.webofscience.com/wos/woscc/advanced-search	
Wiley online library	https://onlinelibrary.wiley.com/search/advanced	

Furthermore, the data obtained from the aforementioned databases were analyzed and visualized using commonly accessible *Microsoft Office* software, such as *Excel* and *Power-Point*. The relevant scholarly references were automatically curated, standardized, as well as managed, employing the *Mendeley Reference Manager* version 2.95.0. Finally, the premium version of the software *Grammarly* (version: v1.0.40.855) from the corresponding manufacturer in combination with *ChatGPT* was employed to ensure orthographic accuracy only.

Lastly, no research funds and cooperations influencing the research are to be declared. Further, no competing interests are directly or indirectly related to the paper.

Author contributions Markus Hafner, a doctoral candidate at the Instituto Superior Técnico Lisbon, is recognized as the corresponding author for both the data valuation business capability taxonomy and the related paper. Additionally, Markus Hafner was responsible for developing the graphics based on the findings from this research, as well as a previous systematic literature review that had been conducted and submitted. Miguel Mira da Silva, a full professor at the Instituto Superior Técnico Lisbon, acted as a contributing author and provided valuable expertise as a subject matter expert and trusted advisor in regards to the methodological and content-related aspects of the research.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. KPMG International Cooperative (2020) KPMG's Global Automotive Executive Survey 2020. https://assets.kpmg.com/content/dam/kpmg/be/pdf/2020/07/global_automotive_executive_survey_2020.pdf. Accessed 8 Sep 2023
2. Thieullent A-L, Jiang Z, Perhirin V, et al (2020) The data-powered enterprise. https://www.capgemini.com/wp-content/uploads/2020/11/Data-powered-enterprise-report.pdf. Accessed 8 Sep 2023
3. Kaufmann M (2019) Big data management canvas: a reference model for value creation from data. Big Data Cogn Comput 3:1–18. https://doi.org/10.3390/bdcc3010019
4. Faroukhi AZ, El Alaoui I, Gahi Y, Amine A (2020) Big data monetization throughout big data value chain: a comprehensive review. J Big Data 7:1–22. https://doi.org/10.1186/s40537-019-0281-5
5. Gartner Inc (2022) Definition of Data Monetization—IT Glossary I Gartner. In: Gartner Glossary. https://www.gartner.com/en/information-technology/glossary/data-monetization. Accessed 16 Feb 2023
6. Eggers J, Hein A (2020) Turning big data into value: a literature review on business value from process mining. In: 28th European conference on information systems (ECIS). pp 1–21
7. Hafner M, Hirsch M, Willemsen M et al (2022) Transformation towards a data-driven business. https://www2.deloitte.com/content/dam/Deloitte/de/Documents/consulting/Deloitte_Transformation%20towards%20a%20Data-Driven%20Business.pdf. Accessed 16 Feb 2023
8. Leitner-Hanetseder S, Lehner OM (2022) AI-powered information and big data: current regulations and ways forward in IFRS reporting. J Appl Acc Res. https://doi.org/10.1108/JAAR-01-2022-0022
9. Pei J (2022) A survey on data pricing: from economics to data science. IEEE Trans Knowl Data Eng 34:4586–4608. https://doi.org/10.1109/TKDE.2020.3045927

10. Grover V, Chiang RHL, Liang TP, Zhang D (2018) Creating strategic business value from big data analytics: a research framework. J Manag Inf Syst 35:388–423. https://doi.org/10.1080/07421222.2018.1451951

11. Elia G, Polimeno G, Solazzo G, Passiante G (2020) A multi-dimension framework for value creation through big data. Ind Mark Manag 90:508–522. https://doi.org/10.1016/j.indmarman.2019.08.004

12. Brennan R, Attard J, Petkov P et al (2019) Exploring data value assessment: A survey method and investigation of the perceived relative importance of data value dimensions. In: 21st International conference on enterprise information systems. SciTePress, pp 188–195. https://doi.org/10.5220/0007723402000207

13. Enders T (2018) Exploring the value of data—a research agenda. In: International conference on exploring services science. https://doi.org/10.1007/978-3-030-00713-3_21

14. Brinch M, Gunasekaran A, Fosso Wamba S (2021) Firm-level capabilities towards big data value creation. J Bus Res 131:539–548. https://doi.org/10.1016/J.JBUSRES.2020.07.036

15. Fricker S, Maksimov Y (2017) Pricing of data products in data marketplaces. In: International conference on software business. https://doi.org/10.1007/978-3-319-69191-6_4

16. Gonzalez S, Hansen K, Hendricky H et al (2018) TOGAF ® Series Guide Business Capabilities

17. Brits J-P, Botha G, Herselman M (2007) Conceptual Framework for Modeling Business Capabilities. In: Informing Science and IT Education Joint Conference. https://doi.org/10.28945/3148

18. Offerman T, Stettina CJ, Plaat A (2017) Business capabilities: a systematic literature review and a research agenda. In: International conference on engineering, technology and innovation (ICE/ITMC). IEEE, Madeira. https://doi.org/10.1109/ICE.2017.8279911

19. Nickerson RC, Varshney U, Muntermann J (2017) A method for taxonomy development and its application in information systems. Eur J Inf Syst 22:336–359. https://doi.org/10.1057/ejis.2012.26

20. Omair B, Alturki A (2020) An improved method for taxonomy development in information systems. Int J Adv Comput Sci Appl. https://doi.org/10.14569/IJACSA.2020.0110470

21. Bui QN (2017) Evaluating enterprise architecture frameworks using essential elements. Commun Assoc Inf Syst 41:121–149. https://doi.org/10.17705/1CAIS.04106

22. Al-Turkistani HF, Aldobaian S, Latif R (2021) Enterprise architecture frameworks assessment: capabilities, cyber security and resiliency review. In: 2021 1st International conference on artificial intelligence and data analytics (CAIDA). pp 79–84. https://doi.org/10.1109/CAIDA51941.2021.9425343

23. Cameron BH, McMillan EN (2013) Analyzing the current trends in enterprise architecture frameworks. J Enterp Archit 9:60–71. https://doi.org/10.5220/0006321002820293

24. Wissotzki M (2015) An exploration of capability research. In: IEEE International enterprise distributed object computing workshop, EDOCW. Institute of Electrical and Electronics Engineers Inc., pp 179–184. https://doi.org/10.1109/EDOC.2015.33

25. Attard J, Brennan R (2018) A semantic data value vocabulary supporting data value assessment and measurement integration. In: 20th International conference on enterprise information systems vol. 2, pp. 133–144. https://doi.org/10.5220/0006777701330144

26. Meierhofer J, Benedech R, Schweiger L, et al (2022) Quantitative modelling of the value of data for manufacturing SMEs in smart service provision. In: ITM Web of conferences. EDP Sciences, p 4001. https://doi.org/10.1051/itmconf/20224104001

27. Green A (2012) Understanding the value of customer data. J Direct Data Digit Mark Pract 13:221–233. https://doi.org/10.1057/dddmp.2011.47

28. Spiekermann S, Korunovska J (2017) Towards a value theory for personal data. J Inf Technol 32:62–84. https://doi.org/10.1057/jit.2016.4

29. Tian Y, Ding Y, Fu S, Liu D (2022) Data boundary and data pricing based on the shapley value. IEEE Access 10:14288–14300. https://doi.org/10.1109/ACCESS.2022.3147799

30. Ghorbani A, Kim MP, Zou J (2020) A distributional framework for data valuation. In: 37th International conference on machine learning. PMLR, pp 3535–3544. https://doi.org/10.48550/arXiv.2002.12334

31. Xu A, Zheng Z, Wu F, Chen G (2022) Online data valuation and pricing for machine learning tasks in mobile health. In: Proceedings—IEEE INFOCOM. Institute of electrical and electronics engineers Inc., pp 850–859. https://doi.org/10.1109/INFOCOM48880.2022.9796669

32. Koutris P, Upadhyaya P, Balazinska M et al (2015) Query-based data pricing. J ACM 62:1–44. https://doi.org/10.1145/2770870

33. Verma N, Singh V (2022) Query-based data valuation strategy: an exploratory view. In: Saraswat M, Roy S, Chowdhury C, Gandomi AH (eds) Lecture notes in networks and systems. Springer Science and Business Media Deutschland GmbH, pp 687–702. https://doi.org/10.1007/978-981-16-5120-5_52

34. Stein H, Holst L, Stich V, Maass W (2021) From qualitative to quantitative data valuation in manufacturing companies. In: Bernard A, Lemoine D, Dolgui A et al (eds) Advances in production management systems. Artificial intelligence for sustainable and resilient production systems. Springer International Publishing, Cham, pp 172–180. https://doi.org/10.1007/978-3-030-85902-2_19

35. Bendechache M, Limaye N, Brennan R (2020) Towards an automatic data value analysis method for relational databases. In: 22nd International conference on enterprise information systems (ICEIS-2020). https://doi.org/10.5220/0009575508330840

36. Seufert S, Wulfert T, Wernsdörfer JE (2021) Towards a reference value catalogue for a company-specific assessment of the IT business value—proposing a taxonomy to select IT impacts from existing catalogues. In: 29th European conference on information systems. Duisburg-Essen

37. Engel C, Buschhoff J, Ebel P (2022) Structuring the quest for strategic alignment of artificial intelligence (AI): a taxonomy of the organizational business value of AI use cases. In: Hawaii international conference on system sciences. https://doi.org/10.24251/HICSS.2022.723

38. Lega M, Colot C, Burnay C, Linden I (2022) Supporting data selection for decision support systems: towards a decision-making data value taxonomy. In: International conference on software engineering and knowledge engineering, SEKE. Knowledge Systems Institute Graduate School, pp 487–492. https://doi.org/10.18293/SEKE2022-104

39. Okoli C (2015) A guide to conducting a standalone systematic literature review. Commun Assoc Inf Syst 37:879–910. https://doi.org/10.17705/1cais.03743

40. Webster J, Watson RT (2002) Analyzing the past to prepare for the future: writing a literature review. MIS Quarterly 26:xiii–xxiii

41. Miller GA (1994) The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychol Rev 63:343–352

42. Gelhaar J, Groß T, Otto B (2021) A taxonomy for data ecosystems. In: 54th Hawaii international conference on system sciences. University of Hawai'i at Manoa. http://hdl.handle.net/10125/71359

43. Gomes SB, Santoro FM, Mira Da Silva M (2021) A taxonomy for digital technology. In: 27th Americas conference on information systems

44. Lis D, Otto B (2021) Towards a taxonomy of ecosystem data governance. In: 54th Hawaii international conference on system sciences. University of Hawai'i at Manoa. https://doi.org/10.24251/HICSS.2021.733

45. The Open Group (2022) The TOGAF® Standard—10th Edition

46. Scimago Lab, Scopus (2022) SJR: Scientific journal rankings. In: Scimago Journal & Country Rank. https://www.scimagojr.com/journalrank.php. Accessed 16 Feb 2023

47. Amiri AK, Cavusoglu H, Benbasat I (2015) Enhancing strategic IT alignment through common language: using the terminology of the resource-based view or the capability-based view? In: The 2015 International conference on information systems. Fort Worth

48. Bataineh AS, Mizouni R, El Barachi M, Bentahar J (2016) Monetizing personal data: a two-sided market approach. In: International conference on ambient systems, networks and technologies. Elsevier, pp 472–479. https://doi.org/10.1016/J.PROCS.2016.04.211

49. Wang Q, Zhao H, Wang Q et al (2020) Enabling secure wireless multimedia resource pricing using consortium blockchains. Futur Gener Comput Syst 110:696–707. https://doi.org/10.1016/J.FUTURE.2019.09.026

50. Holst L, Stich V, Schuh G, Frank J (2020) Towards a comparative data value assessment framework for smart product service systems. In: IFIP Advances in information and communication technology. Springer, pp 330–337. https://doi.org/10.1007/978-3-030-57997-5_39

51. Li Z, Ni Y, Gao X, Cai G (2019) Value evaluation of data assets: progress and enlightenment. In: 4th IEEE International conference on big data analytics, ICBDA 2019. Institute of Electrical and Electronics Engineers Inc., pp 88–93. https://doi.org/10.1109/ICBDA.2019.8713240

52. Brennan R, Attard J, Helfert M (2018) Management of data value chains, a value monitoring capability maturity model. 20th International conference on enterprise information systems. SciTePress, New York, pp 573–584. https://doi.org/10.5220/0006684805730584

53. Barthel S, Schallehn E (2013) The monetary value of information: a leakage-resistant data valuation. In: Lecture Notes in informatics (LNI), proceedings—series of the Gesellschaft fur Informatik (GI). pp 131–138

54. Liang J, Yuan C (2021) Data price determinants based on a hedonic pricing model. Big Data Res 25:100249. https://doi.org/10.1016/J.BDR.2021.100249

55. Yu H, Zhang M (2017) Data pricing strategy based on data quality. Comput Ind Eng 112:1–10. https://doi.org/10.1016/J.CIE.2017.08.008

56. Rix C, Frank J, Stich V, Urban D (2021) Pricing models for data products in the industrial food production. In: Dolgui A, Bernard A, Lemoine D et al (eds) Advances in production management systems. Artificial intelligence for sustainable and resilient production systems. Springer International Publishing, Cham, pp 553–563. https://doi.org/10.1007/978-3-030-85914-5_59

57. Liang F, Yu W, An D et al (2018) A survey on big data market: pricing, trading and protection. IEEE Access 6:15132–15154. https://doi.org/10.1109/ACCESS.2018.2806881

58. Schneider J, Abraham R, Meske C, Vom Brocke J (2022) Artificial intelligence governance for businesses. Inf Syst Manag. https://doi.org/10.1080/10580530.2022.2085825
59. Li J, Wang D, Qi G et al (2022) Alliance chain-based simulation on a new clinical research data pricing model. Ann Transl Med 10:836–836. https://doi.org/10.21037/atm-22-3671
60. Shen Y, Guo B, Shen Y et al (2022) Personal big data pricing method based on differential privacy. Comput Secur 113:102529. https://doi.org/10.1016/J.COSE.2021.102529
61. Robinson SC (2017) What's your anonymity worth? Establishing a marketplace for the valuation and control of individuals' anonymity and personal data. Digit Policy Regul Gov 19:353–366. https://doi.org/10.1108/DPRG-05-2017-0018
62. Lin Z, Wu Y (2016) Research on the method of evaluating the value of data assets. In: International conference on education, E-learning and management technology. https://doi.org/10.2991/iceemt-16.2016.95
63. Otto B (2015) Quality and value of the data resource in large enterprises. Inf Syst Manag 32:234–251. https://doi.org/10.1080/10580530.2015.1044344
64. Lu J, Li S, Zhang X (2021) A study on the business data evaluation method of the power grid value-added service. 5th International conference on power and energy engineering, ICPEE 2021. Institute of Electrical and Electronics Engineers Inc., pp 288–292. https://doi.org/10.1109/ICPEE54380.2021.9662594
65. Huang Y, Milani M, Chiang F (2020) Privacy-aware data cleaning-as-a-service. Inf Syst 94:101608. https://doi.org/10.1016/J.IS.2020.101608
66. Wagner A, Wessels N, Buxmann P, Krasnova H (2018) Putting a price tag on personal information—a literature review. Hawaii international conference on system sciences. IEEE Computer Society, pp 3760–3769. https://doi.org/10.24251/HICSS.2018.474
67. Duan Y, Liu P, Feng Y (2022) Pricing strategies of two-sided platforms considering privacy concerns. J Retail Consum Serv. https://doi.org/10.1016/j.jretconser.2021.102781
68. Heckman J, Boehmer E, Peters E, et al (2015) A pricing model for data markets. In: iConference 2015 Proceedings
69. Fehrenbach D, Herrando C (2021) The effect of customer-perceived value when paying for a product with personal data: a real-life experimental study. J Bus Res 137:222–232. https://doi.org/10.1016/J.JBUSRES.2021.08.029
70. Yang J, Xing C (2019) Personal data market optimization pricing model based on privacy level. Information (Switzerland) 10:123. https://doi.org/10.3390/info10040123
71. Stahl F, Vossen G (2016) Fair knapsack pricing for data marketplaces. In: East European conference on advances in databases and information systems. Springer Verlag, New York. https://doi.org/10.1007/978-3-319-44039-2_4
72. Deep S, Koutris P (2017) QIRANA: a framework for scalable query pricing. ACM SIGMOD international conference on management of data. Association for Computing Machinery, pp 699–713. https://doi.org/10.1145/3035918.3064017
73. Wang X, Wei X, Liu Y, Gao S (2018) On pricing approximate queries. Inf Sci (N Y) 453:198–215. https://doi.org/10.1016/J.INS.2018.04.036
74. Miao G, Yao Y, Chen L et al (2022) Towards query pricing on incomplete data. IEEE Trans Knowl Data Eng 34:4024–4036. https://doi.org/10.1109/TKDE.2020.3026031
75. Zhang M, Beltrán F, Liu J (2022) A survey of data pricing for data marketplaces. SSRN Electron J. https://doi.org/10.2139/ssrn.3609120
76. Debattista J, Attard J, Brennan R (2018) Semantic data ingestion for intelligent, value-driven big data analytics. In: International conference on big data innovations and applications. Institute of electrical and electronics engineers Inc., pp 1–8. https://doi.org/10.1109/Innovate-Data.2018.00008
77. Deep S, Koutris P (2017) The design of arbitrage-free data pricing schemes. In: 20th International conference on database theory (ICDT 2017). Schloss Dagstuhl- Leibniz-Zentrum fur Informatik GmbH, Dagstuhl Publishing. https://doi.org/10.4230/LIPIcs.ICDT.2017.12
78. Berger S, Denner M-S, Roeglinger M (2018) The nature of digital technologies—development of a multi-layer taxonomy. In: 26th European conference on information systems (ECIS)
79. Jöhnk J, Röglinger M, Thimmel M, Urbach N (2017) How to implement agile IT setups: A taxonomy of design options. In: 25th European conference on information systems (ECIS)
80. Szopinski D, Kundisch D, Schoormann T (2020) Visualize Different: Towards Researching the Fit Between Taxonomy Visualizations and Taxonomy Tasks. In: 15th International Conference on Wirtschaftsinformatik. GITO Verlag, pp 1255–1261. https://doi.org/10.30844/wi_2020_k9-szopinski
81. Ritchey T (2006) Problem structuring using computer-aided morphological analysis. J Oper Res Soc. https://doi.org/10.1057/palgrave.jors.2602177
82. Ghorbani A, Zou J (2019) Data shapley: equitable valuation of data for machine learning. In: 36th International conference on machine learning. https://doi.org/10.48550/arXiv.1904.02868

83. Zhang Z, Song W, Shen Y (2021) A reasonable data pricing mechanism for personal data transactions with privacy concern. Asia-Pacific Web (APWeb) and web-age information management (WAIM) joint international conference on web and big data. Springer Science and Business Media Deutschland GmbH, pp 64–71. https://doi.org/10.1007/978-3-030-85899-5_5
84. Gürpinar T, Henke M, Otto B (2021) Towards a taxonomy of incentive mechanisms for data sharing in data ecosystems. In: 25th Pacific Asia conference on information systems
85. Zeleti FA, Ojo A (2017) Open data value capability architecture. Inf Syst Front 19:337–360. https://doi.org/10.1007/s10796-016-9711-5
86. Szopinski D, Schoormann T, Kundisch D (2020) Criteria as a prelude for guiding taxonomy evaluation. In: 53rd Hawaii international conference on system sciences
87. vom Brocke J, Hevner A, Maedche A (2020) Introduction to design science research. Design science research cases 1–13. https://doi.org/10.1007/978-3-030-46781-4_1
88. Guarino N, Welty CA (2009) An overview of OntoClean. In: Staab S, Studer R (eds) Handbook on ontologies. Springer-Verlag, Berlin Heidelberg, pp 201–220. https://doi.org/10.1007/978-3-540-92673-3
89. Matarese V (2022) Kinds of replicability: different terms and different functions. Axiomathes 32:647–670. https://doi.org/10.1007/s10516-021-09610-2

Markus Hafner (M. Eng.) is a Ph.D. student in the Engineering and Management department at Instituto Superior Técnico in Lisbon. His research focuses on data value, data monetization, and enterprise architecture management. Markus is an accomplished researcher who has contributed to the field through both the submission of scientific papers in peer-reviewed international journals as well as practical papers in cooperation with leading enterprises from the technology and consulting sector. His research aims to bridge the gap between academia and practical application by providing scientific findings in the field of enterprise information management that can be implemented in enterprises.

Miguel Mira da Silva is a full professor of information systems at the Instituto Superior Técnico (University of Lisbon, Portugal) and the coordinator of the Digital Transformation group at the INOV research institute. Miguel also coordinates an online master's degree in enterprise information systems, as well as several training programs and courses. He graduated and received an MSc degree in computer engineering from the Instituto Superior Técnico, a Ph.D. in computing science from University of Glasgow, a Sloan Fellowship master's degree in management from the London Business School, and more recently an habilitation in information systems from Instituto Superior Técnico.