Information Systems and e-Business Management
https://doi.org/10.1007/s10257-025-00701-w

ORIGINAL ARTICLE

Data valuation as a business capability: from research to practice

Markus Hafner[1] · **Miguel Mira da Silva**[2] · **Henderik Alex Proper**[3]

Received: 29 July 2024 / Revised: 19 December 2024 / Accepted: 3 February 2025

Abstract

In our data-centric society, the imperative to determine the value of data has risen. Therefore, this paper presents a taxonomy for a data valuation business capability. Utilizing an initial taxonomy version, which originated from a systematic literature review, this paper validates and extends the taxonomy, culminating in four layers, twelve dimensions, and 59 characteristics. The taxonomy validation was accomplished by conducting semi-structured expert interviews with eleven subject matter experts, followed by a cluster analysis of the interviews, leading to a taxonomy heatmap including practical extensions. This paper's implications are manifold. Firstly, the taxonomy promotes a common understanding of data valuation within an enterprise. Secondly, the taxonomy aids in categorizing, assessing, and optimizing data valuation endeavors. Thirdly, it lays the groundwork for potential data valuation standards and toolkits. Lastly, it strengthens theoretical assumptions by grounding them in practical insights and offers an interdisciplinary research agenda following the taxonomy dimensions and characteristics.

Keywords Taxonomy · Data value · Data valuation · Business capability

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

Miguel Mira da Silva
mms@tecnico.ulisboa.pt

Henderik Alex Proper
henderik.proper@tuwien.ac.at

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

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

[3] Business Informatics, TU Wien, Favoritenstraße 9-11, 1040 Vienna, Austria

Published online: 05 March 2025

1 Introduction

More and more enterprises are augmenting their conventional value chain by integrating data, along with related data products and use cases. This extension from a traditional to a data-driven enterprise follows a purpose: generating value with data to foster fact-based decision-making and long-term competitive advantages (Faroukhi et al. 2020; Pei 2022; Hafner and Mira da Silva 2023; Coyle and Manley 2023).

Consequently, enterprises are increasingly adopting a value-oriented perspective on data, focusing on how to prepare and contextualize data, assess the value of data-driven use cases and products, forecast their future value contributions, allocate this value to specific organizational units and products, as well as implement these data-driven use cases and products while actively monitoring value forecasts and outcomes. These activities collectively establish what is known as *data valuation* (Brennan et al. 2018; Debattista et al. 2018; Holst et al. 2020; Stein et al. 2021; Hafner et al. 2024a).

Despite the increasing demand for data valuation, the field remains in its infancy, with several gaps between academic approaches and practical requirements (Li et al. 2019; Cong et al. 2022; Meierhofer et al. 2022). In the year 2020, it was observed that only about one in five enterprises engaged in data valuation, while nearly twice as many were involved in data monetization through products and services (Thieullent et al. 2020), which underlines the challenge of capturing the data value of data-driven monetization endeavors. Research, particularly in the areas of business and information systems, has been exploring the topic of data valuation to address this issue. However, it is noteworthy that various concepts developed for data valuation exhibit markedly different objectives, scopes, and structures, challenging real-world enterprises to select, adapt, and implement the most suitable data valuation approach per their requirements (Thieullent et al. 2020; Hafner and Mira da Silva 2023).

These challenges manifest in several forms. Firstly, the definitional aspect of comprehensively delineating what constitutes data valuation and what does not is pivotal to properly implementing data valuation as a business capability in real-world settings. It is emphasized by Wu et al. (2022, p.24150) that "developing trustworthy data valuation methods that are explainable, fair, and robust is extensively required to measure the value of data and also decide how to use them in real-world applications." Secondly, another challenge arises from data valuation's interdisciplinary nature and complexity. Sidgman and Crompton (2016, p.176) assert that "one of the most important areas for researchers is the one that will ensure data are valued at an appropriate level […]." Consequently, striking a balance between the requisite level of granularity necessary to adequately capture data value and the embedment of data valuation across technology-driven, business-driven, and organization-driven domains is essential for effective and efficient data valuation. This is underlined by Brennan et al. (2018, p.582) arguing that “data value monitoring infrastructure; formal models describing metrics, dimensions and how they relate (ontologies or data models)” are required.

In response to these challenges, Hafner and Mira da Silva (2023) have devised a taxonomy that classifies theoretical data valuation approaches based on four layers, nine dimensions, and 36 characteristics. This taxonomy aims to render theoretical data valuation approaches practical-oriented as a business capability, facilitating sustainable integration into an enterprise architecture. In this context, a business capability is defined as a proficiency comprising four layers: information, resources, roles, and processes (Offerman et al. 2017; Gonzalez et al. 2018; Hafner and Mira da Silva 2023). However, this data valuation business capability (DVBC) taxonomy also has a core limitation: validating and extending theoretical layers, dimensions, and characteristics based on a systematic literature review (SLR) with inputs from real-world enterprises.

This synergy between theory and practice is essential when an artifact such as a DVBC taxonomy is intended to address real-world problems following design science research (Hevner et al. 2004; vom Brocke et al. 2020). More precisely, the existing DVBC taxonomy provides a basis for adequately comprehending data valuation. Nevertheless, its validation and extension with real-world enterprise professionals have not yet been conducted. Consequently, real-world enterprises may not adopt the existing taxonomy, perceiving it as another encapsulated academic artifact with limited practical applicability. Therefore, this study revisits an iteration within the process of empirical-to-conceptual taxonomy development based on Nickerson et al. (2017), while focusing on the subsequent research questions (Table 1):

A frequently employed approach in information systems, namely the interpretative qualitative approach (Kaplan and Duchon 1988; Iyamu 2018) through expert interviews (Hove and Anda 2005; Myers and Newman 2007; Bearman 2019), will be applied to perform a validating and extending empirical-to-conceptual iteration of the taxonomy development method of Nickerson et al. (2017) targeting the DVBC taxonomy by Hafner and Mira da Silva (2023). After the subsequent Sect. 2, which offers an elucidation of the research background, Sect. 3 delineates the methodology employed. The study results will be presented in Sect. 4, discussed in Sect. 5, and concluded in Sect. 6.

Table 1 Research questions

ID	Research question
1	What key dimensions and characteristics do real-world enterprises emphasize when determining data value?
2	How do scientific approaches in the context of data valuation business capabilities align with practices observed in real-world enterprises?

2 Research background

This section establishes theoretical groundwork for the current research on data valuation and its integration as a business capability within enterprise architectures according to *The Open Group Architecture Framework*, also known as TOGAF (The Open Group 2022).

2.1 Data valuation business capability

An SLR according to Okoli (2015) and Webster and Watson (2002) was conducted in a preceding study to cultivate a comprehensive and in-depth understanding of the most recent scholarly work, presently undergoing final stages in a peer-review process. The primary objective of this SLR was to examine the constituent elements of data value, explicitly focusing on the drivers and theories associated with assessing data value and how these elements are embedded within an enterprise architecture (Anonymized for review 2024b).

The analysis of the 102 identified papers highlights the growing momentum of data valuation in scientific research and emphasizes the increasing recognition of data as a core asset within enterprises. However, it is evident that the concept of data value is not uniformly defined (Anonymized for review 2024b). Instead, scholars differentiate between the more apparent economic value of data and other forms of value, such as socio-ecological value (e.g., the intangible benefits of data-driven initiatives that contribute to sustainability, such as waste reduction due to circular economy use cases), functional value (e.g., improvements in data-driven decision-making), and perceived value (e.g., the significance individuals attribute to their personal data). Other definitions of data value originate from the big data area, where the value of data is defined at the junction of the expanded 3 V (Liang et al. 2018) to 7 V model (Khan et al. 2014), encompassing dimensions such as velocity, variety, volume (Liang et al. 2018), veracity, validity, and volatility (Khan et al. 2014). Further, niche definitions of data value may utilize proxy metrics, such as the reduction of uncertainty in a particular use case (Li et al. 2017; Wang et al. 2021) or the performance of AI models in AI-driven applications to define data value (Schneider et al. 2022; Xu et al. 2022).

The various definitions of data value suggest that the criteria affecting data value and underlying theories differ significantly. The most prominent categories of criteria affecting data value, which encompass a range of metrics, include business utility and use cases (Brennan et al. 2019; Holst et al. 2020; Meierhofer et al. 2022), information entropy (Li et al. 2017; Wang et al. 2021), costs (Brennan et al. 2019; Stein et al. 2021; Cheong et al. 2023), data quality (Batini et al. 2018; Brennan et al. 2019; Mendizabal-Arrieta et al. 2023), data security (Gkatzelis et al. 2015; Huang et al. 2020; Shen et al. 2022), data lifetime (Robinson 2017; Pei 2022; Kang and Guo 2023), as well as various sentiment dimensions (Brennan et al. 2019; Busch-Casler and Radic 2022; Meierhofer et al. 2022). These criteria are processed through different theoretical perspectives, such as game theory

(Tian et al. 2022; Wang et al. 2024), decision theory (Stahl and Vossen 2016a; Lim et al. 2024), and others.

This fragmented landscape of data value presents several challenges for researchers and enterprises, beginning with the difficulty of establishing a clear definitional foundation. Therefore, in this study, data value is defined as the monetary and non-monetary (Elia et al. 2020; Hafner et al. 2024a), as well as qualitative and quantitative (Stein et al. 2021; Hafner et al. 2024a), benefits derived from the application of data within a specific use case (Brennan et al. 2019; Holst et al. 2020; Meierhofer et al. 2022), contributing to enterprises of various types. The process of realizing value with data is referred to as data valuation and involves core steps from preparing and contextualizing the data, determining and allocating its value (Brennan et al. 2018; Stein et al. 2021;Hafner et al. 2024a), realizing the data-driven use cases (Hafner et al. 2024a) and monitoring their results (Brennan et al. 2018; Debattista et al. 2018), as well as accompanying the entire process with user-oriented change management (Hafner et al. 2024a).

In addition to the definitional challenges surrounding data value and the associated data valuation, enterprises face significant hurdles such as identifying the most suitable data valuation approach tailored to their specific requirements, effectively aligning and integrating data valuation into their daily operations, architectural frameworks, and standards, as well as connecting the dots between data valuation theories and approaches originating from diverse research domains (Sidgman and Crompton 2016; Enders 2018; Noshad et al. 2021; Pei 2022; Anonymized for review 2023). One way to tackle these challenges is to set up data valuation not as an ad hoc endeavor but as an actively managed business capability properly integrated into an enterprise architecture. This business capability, referred to as *DVBC*, encompasses multiple dimensions and layers, as illustrated in the DVBC taxonomy in Fig. 1.

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 and Lifetime	Data Quality	Data Security and Privacy	Sentiment and Perception	Others	N
Resources	Data Valuation Theory	Economic	Cooperative Game Theory	Non-Cooperative 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
Processes	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

* E = Exclusive   *N = Non-Exclusive

Fig. 1 Data valuation business capability taxonomy (Anonymized for review 2023) to be validated and extended

2.2 Related taxonomies

In developing the theory-based DVBC taxonomy, Anonymized for review (2023) analyzed related taxonomies. Hence, only a brief mention of the associated taxonomies is provided. In previous work, Seufert et al. (2021) developed a taxonomy for classifying value catalogs, aiming to link the performance of an enterprise with its investments in information technology, which is related to data. Additionally, Engel et al. (2022) address business value specifically in the context of data-driven use cases, particularly in the AI domain. Thus, the resulting taxonomy addresses the data value induced by AI use cases for enterprises. Moreover, Lega et al. (2022) developed a taxonomy that contextualizes data value within decision-making, particularly considering data quality and utility. The content of these three related taxonomies and the results of the conducted SLR served as the foundations for the DVBC taxonomy (Anonymized for review 2023) as illustrated in Fig. 1.

The DVBC taxonomy undergoing the present validation and extension comprises four layers, nine dimensions, and 36 characteristics (Anonymized for review 2023). To organize and cluster the data valuation approaches from academia, the taxonomy employs the four business capability layers information, resource, role, and process (Gonzalez et al. 2018) as a bracket beneath which the dimensions and characteristics are allocated. The layers are aligned with TOGAF (Gonzalez et al. 2018; The Open Group 2022) to ensure the taxonomy conforms to established industry best practices in enterprise architecture. The dimensions and characteristics within the four layers incorporate various perspectives focused on business, as well as those centered on data and technology (Anonymized for review 2023).

The dimensions and characteristics of the DVBC taxonomy meet quality criteria based on Nickerson et al. (2017), providing a theoretically and scientifically grounded solution space for enterprises aiming to address data valuation as a comprehensive business capability. Enterprises can integrate and further develop modular elements from the DVBC taxonomy according to their requirements in data valuation endeavors. The DVBC taxonomy distinguishes between exclusive and non-exclusive taxonomy dimensions. In exclusive dimensions, only one associated characteristic can be selected. For instance, within the exclusive dimension *result*, an enterprise may calculate either the *specific absolute data value*, *approximate absolute data value*, or *relative data value* for a given use case. Non-exclusive dimensions, such as *data value driver*, operate differently. In these cases, one or more data value drivers can be included within a single data valuation use case, as factors like *costs*, *data quality*, as well as *sentiment and perception* may all be relevant for determining the data value.

In their concluding remarks, Anonymized for review (2023) assert that the DVBC taxonomy necessitates additional validation, particularly in practical, real-world environments, in addition to the already conducted theoretical validation.

3 Methodology

This section describes the symbiosis of the applied methodologies, especially semi-structured expert interviews as well as textual cluster analysis.

3.1 Methodological overview

In information systems, soliciting insights derived from practical field experience either quantitatively or qualitatively is paramount to a) enable a comprehensive interpretation of the generated artifact and b) facilitate the incorporation of diverse perspectives (Kaplan and Duchon 1988; Szopinski et al. 2019). This approach fosters the ability of the developed artifact to solve practical challenges (Hevner et al. 2004; vom Brocke et al. 2020) while complying with taxonomy quality standards such as conciseness, robustness, and extendibility (Nickerson et al. 2017; Szopinski et al. 2020).

To gather the required practical field experiences, a four-phase interpretative qualitative approach based on Mingers (2001) is employed to validate and extend the DVBC taxonomy. This approach (see Fig. 2) is complemented by activities for the creation, execution, and documentation of expert interviews following Hove and Anda (2005) and Myers and Newman (2007).

The initial phase, *appreciation*, is the central phase for extracting information (Mingers 2001). This will be achieved by applying the dramaturgical interview model outlined by Myers and Newman (2007). In the second phase, referred to as *analysis*, the data gathered from the interviews must be comprehended and organized (Mingers 2001). This is achieved through the transcription of the interviews, as advised by Hove and Anda (2005) and Myers and Newman (2007). Subsequently, in phase 3, *assessment*, the insights gained from the interviews are evaluated, interpreted, and discussed. Phase 4, *action*, involves preparing and disseminating the results, thereby facilitating the spreading of new knowledge to the readers of this research paper (Mingers 2001; Myers and Newman 2007).

Utilizing qualitative expert interviews, originating in sociology (Hove and Anda 2005), has emerged as a common practice in the research area of information systems to observe phenomena such as data valuation from various angles (Carruthers 1990). A specific application area of qualitative expert interviews, as seen in manuscripts published in top-tier journals and conferences, is the validation and extension of artifacts (Schultze and Avital 2011) such taxonomies (Szopinski et al. 2019; Omair and Alturki 2020), with examples including the decision-making data value taxonomy (Lega et al. 2022), the taxonomy of information systems for corporate carbon risk management (Körner et al. 2023), the taxonomy for non-fungible tokens

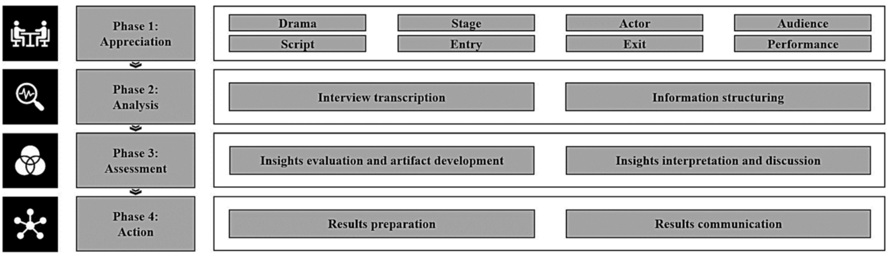

Fig. 2 Applied research approach based on Mingers (2001), Hove and Anda (2005), Myers and Newman (2007) for validating the data valuation business capability taxonomy

(Hartwich et al. 2024), the cyber-physical taxonomy (Jiang et al. 2023), the agile IT setup taxonomy (Jöhnk et al. 2017), the user-generated content taxonomy (Weingart et al. 2023), or the electronic records management system adoption taxonomy (Mukred et al. 2018).

Following these scientific best practices, semi-structured expert interviews are utilized, characterized by a customized interaction with the interviewee shaped by their specific area of expertise and corresponding responses. These semi-structured interviews, which are specifically suitable for complex and interdisciplinary subjects (Carruthers 1990; Abraham et al. 2013; Bearman 2019) like data valuation, are framed by a general interview guide (see Appendix). This guide offers a broad structure for the discussion, but deviations are permitted to accommodate the natural flow of the expert's responses (Myers and Newman 2007). Moreover, the interview guide facilitates the integration of the interviewee's relevant perspectives, experiences, and opinions, ensuring the accurate derivation of appropriate conclusions (Carruthers 1990). Despite the advantages of semi-structured interviews, a key drawback is the substantial time and resource investment required for preparation, execution, and post-processing (Szopinski et al. 2019). To address this challenge, proactive interview planning is essential, coupled with using assistive technologies such as the transcription function of *Microsoft Teams*, which was applied in this context. The following subsections will describe the applied methodology (see Fig. 2).

3.2 Phase 1: appreciation

According to Myers and Newman (2007) qualitative semi-structured interviews can be seen as dramas, which consist of various building blocks such as the *stage, actors, audience, script, entry, exit,* and the *performance* itself. To facilitate flexibility in scheduling and organization, the authors, respectively interviewers, opted to conduct expert interviews virtually using *Microsoft Teams* meetings as the stage. In order to mitigate the disadvantages of remote meetings, such as the potential limitation in interpreting non-verbal communication (Iyamu 2018), expert interviews are conducted with a camera-on policy.

Each interview was comprised of two actors: one interviewer and one interviewee, who performed the interview and produced the resulting outcomes for the audience of data and enterprise architecture professionals in practice. To validate the taxonomy, one expert interviewer in data valuation used transcription software to reduce documentation and focus on moderation. The chosen interviewees had to meet specific criteria, including a minimum number of years of experience, expertise within a particular domain, and geographic location (Iyamu 2018).

For this study, the authors concentrated on specialized professionals from Germany (DE), the Netherlands (NL), and Switzerland (CH), specifically those in consulting or corporate positions adjacent to data value. Moreover, the experts are currently not employed in the same organization and have at least three years of expertise (to be considered senior experts in an area) in data value, enterprise architectures, or both. Furthermore, the interviewees were not asked to pre-analyze the DVBC taxonomy for validation in order to avoid biased responses. However, it cannot be ruled out that the interviewees may have encountered the taxonomy under validation. An overview of the interviewees is presented in Table 2.

In addition to establishing the stage and actors, it is imperative to construct a script for semi-structured interviews, referred to as the interview guide (Carruthers 1990; Hove and Anda 2005; Myers and Newman 2007). The interview guide (see Appendix) was structured into three main phases.

The first phase begins with an introduction, the so-called entry, to the interview, placing particular emphasis on familiarizing the participants and obtaining consent for activities such as recording, transcription, and information processing, along with some initial questions to establish common ground. Once all formalities are addressed, the second phase delves deeper into the core interview theme. The exit is marked by a reflection phase, during which hypothetical questions can be posed, further comments solicited, and the subsequent steps clarified (Myers and Newman 2007; Bearman 2019).

The questions employed in the interview guides primarily adopt open formulations, encouraging the interviewees to provide more extensive and detailed insights about their experiences (Bearman 2019). Nonetheless, specific questions incorporate closed formulations to elicit precise statements (Bearman 2019).

The interaction among the aforementioned drama building blocks is regarded as the performance itself (Myers and Newman 2007). Three success factors were considered to ensure high-quality performance. Firstly, all participants needed to comprehend the purpose and content of the interview (Carruthers 1990). To achieve this, the preliminary interview guide was attached to the invitation email, a gesture for which the experts expressed gratitude. Secondly, despite recording the interview for transcription purposes, the interviewer established a secure environment of anonymity and compliance for the interviewees (Carruthers 1990). This was achieved through transparent and frequent communication regarding the recording's purpose before, during, and after the interview, further reinforced by the explicit opt-in agreement from the interviewees for the recording. Thirdly, the questions were framed clearly, non-judgmentally, and non-offensively (Carruthers 1990; Leech 2002; Hove and Anda 2005), using terminology widely understood or explained in advance.

3.3 Phase 2: analysis

After executing the expert interviews, it is imperative to capture the data systematically gathered for subsequent analysis. This process involves creating transcriptions, a customary procedure in both qualitative research and the field of information systems (Mingers 2001; Hove and Anda 2005; Myers and Newman 2007). Transcribing expert interviews aims to enhance the transparency and comprehensibility of the obtained results. An essential prerequisite for transcribing is obtaining explicit consent to record the interviews while ensuring the confidentiality of sensitive information (Myers and Newman 2007). The transcriptions of the interviews are carried out in a manner that conceals the true identities of the participants.

Table 2 Interviewed participants

Role	Expertise years	Perspective	Description	Country
Data value expert	4	Consulting	Data value consulting for financial departments	DE
Co-founder & CEO	6	Consulting	Data strategy and management for automotive, financial services, manufacturing, and med-tech	DE
Head of data & intelligence	20	Consulting	Business intelligence, AI, strategy, and data science (new tech) consulting	DE
Head of data & advanced analytics	10	Corporate & consulting	Data science, data strategy, and management consulting, mainly automotive	CH
Director solutions	10	Corporate & consulting	Data / AI strategy, data architectures, governance, and monetization models	DE
CIO & COO	34	Corporate	IT, data, process, and efficiency steering mainly in automotive and manufacturing	DE
Director digitalization	20	Corporate	Data analytics, data-driven transformation, AI, and data strategy, mainly in manufacturing	DE
Executive & professor	21	Consulting	Digital transformation and data management, mainly in finance, government, and utilities	NL
Directing enterprise architect	30	Consulting	Data and data management in an enterprise architecture context, mainly in governance	NL
Managing director	27	Consulting	IT transformation, IT strategy, IT governance, and enterprise architectures in many sectors	NL
Manager data strategy	6	Corporate & consulting	Data strategy, AI, data analytics, data products, and services in telecommunication	DE

After transcription of all interviews, the gathered data is subjected to clustering for subsequent analysis (Mingers 2001). To accomplish this, the first step involved the use of textual cluster analysis, an empirical method closely linked to the development of taxonomies (Hayashi et al. 2019). Cluster analysis serves multiple purposes, including analyzing and processing large and complex datasets, such as textual information (Hayashi et al. 2019), organizing text to enhance evidence retrieval (Aggarwal and Zhai 2012), and visualizing textual data to highlight the significance and interdependencies of terminologies (Bukar et al. 2023). Therefore, cluster analysis organizes the textual interview data into clusters of logically related words. Employing cluster analysis as an exploratory empirical method helps minimize a priori biases during data analysis and description (Hayashi et al. 2019), such as confirmation bias, which can arise when there is a tendency to believe statements from personally favored interviewees or those holding specific positions (Montibeller and von Winterfeldt 2015). These biases may emerge from pre-existing theories, assumptions, or personal tendencies that have influenced the DVBC taxonomy under validation or grown from the conducted expert interviews. Therefore, performing an initial cluster analysis to group the interview data into preliminary clusters facilitates a more comprehensive description of the interview findings in subsequent stages of analysis (Hayashi et al. 2019), while minimizing biases and avoid constraining the findings to fit within the predefined dimensions and characteristics of the taxonomy to be validated. This macro-level perspective enables the subsequent validation of the theory-based taxonomy for a DVBC by Anonymized for review (2023) and facilitates its augmentation with practical dimensions and characteristics.

One approach to conducting cluster analysis is by using *VOSviewer* (Visualization of Similarities), initially developed for bibliometric datasets but now proven to apply to complex text-based data, such as interview transcripts (Zhang et al. 2021; Lin et al. 2022; Bukar et al. 2023). Following recommendations for future research to "explore VOSviewer's application for analyzing text networks in [researchers'] respective domains," (Bukar et al. 2023, p.7) this study employs VOSviewer as an initial method for analyzing the expert interview data.

To ensure uniformity in language, the interviews conducted in German were translated to English using *ChatGPT*. The interviewer cross-checked the translated interview guides to prevent substantive errors or erroneous interpretations during translating interviews from German to English using ChatGPT.

3.4 Phase 3 and 4: assessment and action

During the assessment phase, the raw data and the grouped, pre-analyzed data obtained from the cluster analysis are employed to validate the DVBC taxonomy. The authors inspect whether and, if so, to what degree the taxonomy's dimensions and characteristics are addressed within the expert interviews.

The latter process entails the examination of transcribed interviews combined with the context within which interviewees articulated their statements. Therefore, a *Microsoft Excel* spreadsheet was generated, wherein taxonomy layers, dimensions, and characteristics are represented horizontally and corresponding interviewees vertically in an assessment matrix. The assessment was executed at the characteristic level of the taxonomy, wherein once an interviewee mentioned a characteristic, whether explicitly or implicitly, it was marked as *relevant/validated*. Explicit mentions refer to the direct reference by the interviewee to the dimensions and characteristics of the DVBC taxonomy. Implicit mentions, on the other hand, involve indirect references to these taxonomy dimensions and characteristics, which are subsequently categorized based on synonyms (e.g., the mention "good data" (Anonymized Interviewees 2023, p.16) is assigned to the characteristics *data quality*) or through contextual abstraction and allocation through the research team (e.g., the mentions "CEO" and "CDO" (Anonymized Interviewees 2023, p.13) are grouped under the characteristic *top-management*, or the mention "increase in revenue" (Anonymized Interviewees 2023, p.31) is assigned to the characteristic *topline growth*). Care was taken to ensure that repeated mentions of a characteristic within one interview were not counted multiple times but recorded uniquely as *relevant/validated*, aiming for harmonization across interviews.

Given the total of eleven expert interviews conducted, each taxonomy characteristic could receive a maximum of eleven instances of the tag *relevant/validated*, as depicted in the heatmap representation of the taxonomy in Fig. 4. It is assumed that the validation of taxonomy characteristics indicates an implicit validation of the taxonomy dimensions and layers above them. Further, to guarantee the quality- and content-related validation of the taxonomy, a final check regarding objective and subjective ending conditions based on (Nickerson et al. 2017) was executed (see chapter 5.2).

Depending on the findings of this assessment, potential expansions to the taxonomy are created. This is done to enhance the DVBC taxonomy with real-world perspectives and to highlight potential constraints or avenues for future research (Mingers 2001).

In the final stage *action*, as recommended by Mingers (2001), the findings were bundled into a scholarly paper.

4 Findings

This section presents the findings of the taxonomy validation and extension process, commencing with a cluster analysis aimed at elucidating the prominent interrelationships among word clusters within the domain of data valuation. Subsequently, the validation of taxonomy dimensions and characteristics is conducted, culminating in expanding the taxonomy through insights gleaned from expert interviews.

4.1 Cluster analysis

The cluster analysis of the expert interview transcripts is designed to illustrate the interrelationships among the key terminologies and topics highlighted in these interviews. This is fundamental for comprehending the overarching context of data valuation and its associated taxonomy from a real-world standpoint. It is important to note that the cluster analysis broadly categorizes relevant themes within the context of data value, to extend and validate the DVBC taxonomy. The cluster analysis is explicitly not the focus of a standalone bibliometric text analysis.

The authors tested various cluster sizes in VOSviewer (indicating the number of mentions of a terminology). However, a cluster size of ten yielded the most precise results based on the context of the interviews and the transcriptions. This choice was made with consideration for achieving a meaningful level of granularity that accurately reflects the spoken word in the interviews while avoiding excessive detail that might obscure the overall context.

Figure 3 depicts the outcomes of the cluster analysis. The size of each bubble corresponds to the occurrences and, consequently, the significance of a term across all interviews. Additionally, the distance between the bubbles indicates the degree of association between the terms. Specifically, closer proximity in combination with the line thickness indicates a higher degree of association among the related terms (van Eck and Waltman 2011, 2018).

As anticipated, based on the research question and the accompanying interview questions, the core of the cluster analysis revolves around the terms *data*, *value*, and *company*, determined by their frequency of mention and their proximity to one

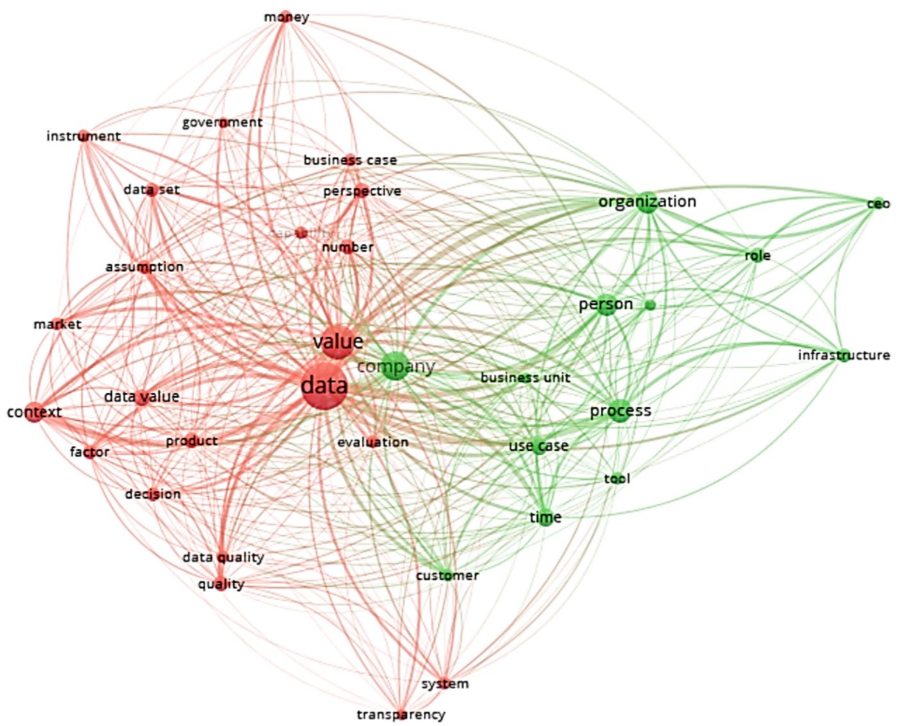

Fig. 3 Cluster analysis of core terminologies and topics, including their interrelationships

another. The authors categorize the terminologies into two broad clusters based on the analysis.

The red cluster, centered around *data* and *value*, particularly highlights the interrelations between *data* and its associated *data value* with factors such as the *context* or the *product* in which the data is applied. Another primary driver influencing data value includes, on the one hand, *data quality* and, on the other, the *perspective* from which one assesses the value of data, which aligns with the data value driver of sentiment and perception. Furthermore, a strong connection between *data* and *assumption* is notable, underscoring one of the core enterprise challenges of relying on gut-based data valuation rather than a more robust determination of data value. The relatively significant distance of the term *money* from the core of the cluster analysis suggests that for most practitioners, data and data value are not necessarily strongly associated with value in terms of money.

Furthermore, the red cluster exhibits similarities to the green cluster by making it evident that the value derived from data is heavily contingent on the underlying *use case* or *context*. Additionally, the green cluster highlights pertinent terminologies in the organizational context, such as *business unit*, *person*, and *process*. It emphasizes the relevance of interpersonal collaboration in data valuation and its setup across an enterprise, e.g., in the form of a business capability.

In addition to frequently mentioned and related terms, it is beneficial to shed light on less frequently occurring terms. The cluster analysis reveals that, from the practitioner's perspective, terms related to enterprise architecture management are rarely explicitly mentioned, suggesting a lack of direct association between enterprise architectures and data value. Additionally, interviewees tend to refer to proprietary data valuation approaches, where present, without directly referencing scientific methods and approaches. Both less frequently mentioned areas suggest that data valuation in companies is often still in its early stages and is conducted in a rather non-structured or unsystematic manner.

4.2 Taxonomy validation and extension

After the cluster analysis, the interviews were individually examined, and the responses were assigned to the corresponding dimensions and characteristics of the theory-based taxonomy for a DVBC. The number of implicit and explicit mentions of the dimensions and characteristics results in a field-tested taxonomy backed with a heat map, as illustrated in Fig. 4.

It is important to note that the absence or low frequency of mentions of dimensions and characteristics should not be confused with a falsification of the taxonomy. Instead, practitioners' mentions underscore the relevance of specific topics for data valuation based on their current knowledge and historical practical experiences. At the same time, the theory-based taxonomy for a DVBC also considers a forward-looking perspective on emerging patterns, which may not be prominent in enterprises yet. Consequently, the decision has been made not to eliminate any unmentioned dimensions and characteristics. Instead, the depicted symbiosis provides a

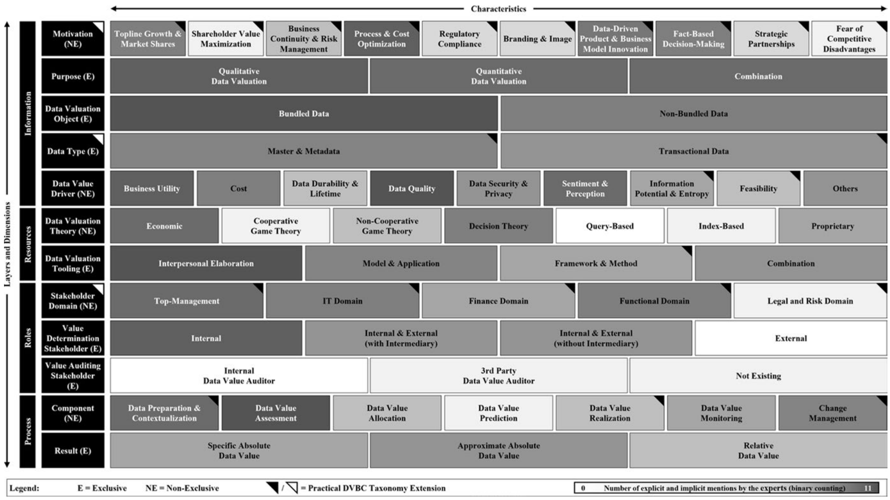

Fig. 4 Heatmap data valuation business capability taxonomy based on Anonymized for review (2023) including extension

broad and deep solution space combining scientific and practical insights, which can be tested in future case studies.

Upon examining the heatmap taxonomy as depicted in Fig. 4, it is apparent that a substantial proportion of experts referenced many taxonomy characteristics. This leads to the overarching inference that the theory-driven DVBC taxonomy is substantiated by empirical validation from real-world practitioners. Nevertheless, although acknowledged by practitioners, certain dimensions were not encompassed in the initial DVBC taxonomy. Consequently, an extension to the taxonomy was devised, denoted by triangular markers located at the upper right corners of the respective dimensions and characteristics. A detailed description of the validation and extension results is presented in the following sections. Furthermore, to enhance the understandability of the DVBC taxonomy, all dimensions and characteristics are briefly explained in Table 5 in the Appendix.

4.2.1 DVBC layer: information.

The information layer of the DVBC taxonomy originally consisted of the *purpose*, the *data valuation object*, and the *data value driver*. Two additional dimensions were introduced based on the interviews: *motivation* and *data type*. The following paragraphs will further explain these extensions and the initial dimensions.

One extension of the DVBC taxonomy is the dimension of *motivation*. When asked about the purpose practitioners aim to achieve with the determination of data value, few focused on the characteristics of qualitative or quantitative data valuation, as originally anticipated. Instead, their focus was primarily on strategic motivations for the enterprise, such as "the main purpose, of course […] for most organizations either is to increase their market share or at least defend their market share"

(Anonymized Interviewees 2023, p.12). Other motivations mentioned in the interviews are “cost-saving potential or new sales” (Anonymized Interviewees 2023, p.2) as well as “faster decisions, more efficient and predictable processes, compliance, and innovation.“ (Anonymized Interviewees 2023, p.6). In summary, the non-exclusive characteristics of *process and cost optimization, data-driven product and business model innovation, fact-based decision-making,* as well as *topline growth and market shares* are particularly relevant in this context.

Additionally, when examining the DVBC taxonomy dimensions of *motivation* and *purpose*, the analysis of interview transcripts showed that most practitioners observe data value as more than just a strictly quantifiable or monetary measure. Specifically, the interviewees state that “[data value] could be monetary, but not necessarily” (Anonymized Interviewees 2023, p.5) and even more explicitly that “data value is much more than the pure monetary value that might be embedded in the data” (Anonymized Interviewees 2023, p.23). Instead, incorporating *qualitative data valuation* or combining both *qualitative* and *quantitative data valuation* can yield additional benefits for an enterprise. For instance, one interviewee noted that in their enterprise, the determination of data value is conducted in a highly qualitative manner. However, output factors such as increased productivity or reduced delivery time resulting from data are quantitatively measurable metrics, which can, at least partially, be attributed to the data being utilized. These indirect quantitative measures, or so-called proxies, offer a way to make the purpose of data valuation more tangible. Nevertheless, other interviewees emphasize that “in practice, maybe we can put a number to [data value], but then we have to be really aware of the fact that it is a proxy” (Anonymized Interviewees 2023, p.23) “and you always have to consider the person, the data set, and the context “ (Anonymized Interviewees 2023, p.19). Moreover, the necessary integration of *quantitative* and *qualitative data valuation* was emphasized multiple times, as data value often contains a significant degree of subjectivity and is highly context-dependent, making it difficult to determine a purely quantitative figure. As one interviewee noted, “the actual value is not something that we can pin down, I think that is subjective and personal “ (Anonymized Interviewees 2023, p.18).

To determine the data value, either qualitatively or quantitatively, influencing parameters or so-called *data value drivers* may be employed. Many practitioners assert that the *business utility*, deeply linked to the specific use case and context of the data being valued, constitutes one of the pivotal *data value drivers*. As an example, one interviewee mentioned, “data does not have value without taking into account the context” (Anonymized Interviewees 2023, p.12) which was further expanded upon by other interviewees, who emphasized that data value should be approached in a "use case or product-oriented way" (Anonymized Interviewees 2023, p.18), as many enterprises today, if they engage with data valuation at all, typically "assessing data based on a use case” (Anonymized Interviewees 2023, p.18).

Additionally, the majority of practitioners identify *data quality* as a fundamental data value driver, encompassing subcomponents like transparency and completeness, aligning closely with existing academic research (Otto 2015; Yu and Zhang 2017; Anonymized Interviewees 2023). Moreover, interviewees establish a correlation between *data quality* and its value, suggesting that if "the value is high, then the data quality requirements should also be high" (Anonymized Interviewees 2023, p.7). This leads to the conclusion that high data quality is a prerequisite for data valuation and “the quality decides: how valuable is [the data] actually?” (Anonymized Interviewees 2023, p.31). If the data quality were not sufficiently high, enterprises "would not need to evaluate and would not need to use it" (Anonymized Interviewees 2023, p.26).

Nevertheless, most practitioners agree that data value determination cannot rely solely on objective *data value drivers*, such as *data quality*. Instead, interviewees “claim the value of data is subjective and situational” (Anonymized Interviewees 2023, p.18). This underscores the notion that data value may vary according to the *sentiments and perceptions* of stakeholders, including their experiences, perceived risks, and expectations related to the data under consideration.

In addition to the aforementioned *data value drivers*, identified as the most critical based on expert interviews, practitioners also brought up other data value drivers such as *cost (e.g., “*We definitely consider costs. This includes both monetary costs and the time investment for our employees.” (Anonymized Interviewees 2023, p.10)*), data durability and lifetime (e.g., “*I would rather have fairly good data now then really good data tomorrow.*”* (Anonymized Interviewees 2023, p.10)), as well as *data security and privacy (e.g., “H*ygiene factors […] are also things like data privacy.” (Anonymized Interviewees 2023, p.13)). However, these were not necessarily emphasized as top priorities.

Furthermore, the expert interviews revealed the necessity to incorporate two additional *data value drivers*, namely *information potential and entropy*, as well as *feasibility*, into the DVBC taxonomy. As a practical *data value driver, feasibility* pertains to the likelihood and extent to which a data product or data-driven use case can be successfully implemented and realized. Some practitioners highlight that "the biggest impact on data value […] is ultimately the feasibility" (Anonymized Interviewees 2023, p.26), emphasizing that assessing data value is irrelevant without implementation and realization from their perspective. Specifically, finding a “balance between feasibility and impact” (Anonymized Interviewees 2023, p.28) of data-driven use cases and data products is crucial for practitioners. Conversely, *information potential and entropy* represent a theoretical *data value driver*. *Information potential and entropy* pertain to the ability of data to offer value across diverse domains, use cases, and contexts. The entropy aspect of the data value driver focuses explicitly on how data can mitigate uncertainty in specific events, consequently providing an indirect enhancement of value to the organization (Shen et al. 2019; Mendizabal-Arrieta et al. 2023). However, *information potential and entropy* is not only a central *data value driver*, but also one specific aspect, future uncertainty, is highlighted as "the most prominent challenge." (Anonymized Interviewees 2023, p.12).

Regardless of the *data value drivers* applied, it is essential to identify specific entities for data valuation, known as *data valuation objects*. According to the interviewees, *data valuation objects* primarily manifest as *bundled data*, which is why value allocation frequently occurs “along the use cases or along a data product” (Anonymized Interviewees 2023, p.2). Furthermore, practitioners indicate that “even a single data point already has meaning and it could potentially has value”

(Anonymized Interviewees 2023, p.16), demonstrating that *non-bundled data* can also serve as a basis for data valuation. However, the vast majority emphasize that the value of data must always be considered in the context of its application. Therefore, *bundled data* serve as the central foundation for determining data value in the respective use case, data product, or application. Furthermore, during the interviews, the taxonomy dimension *data type* was added, which consists of the exclusive characteristics of *master and metadata*, as well as *transactional data*. Looking at the frequency of mentions, it is notable that both *master and metadata*, as well as *transactional data*, are suitable for data valuation, and this suitability is also contingent on context and use cases.

4.2.2 DVBC layer: resources

Unlike the aforementioned dimensions within the DVBC taxonomy, the dimension of *data valuation theory* has received comparatively less attention from practitioners. Nevertheless, when discussing data valuation, practitioners primarily consider *economic data valuation* in the context of cost–benefit analysis ("if we have to make a decision, if and where we are going to invest […] then of course you have to do cost–benefit analysis" (Anonymized Interviewees 2023, p.17), as well as *decision theory* data valuation in terms of prioritization ("we probably won't do all of them [AI use cases] anyway, but prioritization becomes important" (Anonymized Interviewees 2023, p.28)), as the prevailing theories in use. These are supplemented by more *proprietary approaches* that do not adhere to a standardized process or business capability.

The predominantly *proprietary approaches* to data valuation in enterprises are also evident because some enterprises employ specially "developed […] tools and frameworks" (Anonymized Interviewees 2023, p.10) to conduct rudimentary data valuation. Consequently, the characteristic *framework and method* have been incorporated into the dimension of *data valuation tooling*. However, specific tools for data valuation are hardly developed or utilized within enterprises. Instead, practitioners assert that data valuation is primarily conducted through *interpersonal elaboration* among stakeholders, coupled with basic *applications and models* like business case templates in associated tools such as *Microsoft Excel*.

4.2.3 DVBC layer: roles

As interpersonal elaboration has been identified as a central cornerstone in determining the value of data, a diverse range of stakeholders plays a pivotal role in this endeavor. Most practitioners highlight the significance of *internal stakeholders*, potentially incorporating *external partners* such as "clients and suppliers" (Anonymized Interviewees 2023, p.13) or data brokers as intermediaries. On the other hand, the practitioners did not mention purely external data valuation. A taxonomy dimension labeled *stakeholder domain* has been introduced to delineate the specific internal stakeholders to be engaged. This dimension encompasses non-exclusive characteristics, including *top-management* (e.g., CEO, CFO, CDO), *IT domain* (e.g., enterprise architect)*, finance domain* (e.g., controller), *legal and risk domain* (e.g., legal and risk experts), as well as *functional domain*, such as production, sales, or product management. However, the dimension of *value auditing stakeholders* was scarcely brought up, and in some cases, it was either deemed irrelevant or not yet implemented by the practitioners.

4.2.4 DVBC layer: process

In examining the *process components* of data valuation, it can be observed that "theoretically, there is monitoring and evaluation" (Anonymized Interviewees 2023, p.30) as crucial parts of data valuation. While these core components may be theoretically sound, they are "not always so coherent" (Anonymized Interviewees 2023, p.30) in practice. However, *data value assessment*, which determines data value as the foundation of data valuation, is often combined with *data value monitoring*, defined as comparing initial assumptions with actual outcomes (Hafner et al. 2024a).

In contrast, *data value allocation* and *data value prediction* were mentioned minimally, if at all. Notably, practitioners identified three additional characteristics within the non-exclusive taxonomy dimension of *component* as particularly relevant, thus warranting their inclusion in the taxonomy. As a preliminary step, many practitioners consider *data preparation and contextualization* crucial. This process is essential for making the data discoverable, accessible, as well as available, and for organizing it into logically coherent clusters, such as *bundled data* within the dimension of *data valuation object*. Moreover, *data value realization*, entailing the implementation of corresponding data products or data-driven use cases, is classified as a key process component. These relatively linear process components are complemented by an iterative *change management* approach, as data- and data-value-oriented organizational transformations "will bring a big change for many people" (Anonymized Interviewees 2023, p.8). Thus, it is crucial to enable enterprises to design formalized data valuation processes and business capabilities and engage employees in embodying and implementing them actively.

The analysis of the last DVBC taxonomy dimension *result* effectively consolidates the insights anticipated by the preceding dimensions. Given that data value is neither a fixed concept nor a clearly measurable figure, most interviewees do not expect a precisely measurable *absolute data value* as an outcome of data valuation. Notably, the uncertainty and subjectivity, as well as context dependency in determining data value, introduce a degree of fuzziness to the concept of data value. In this regard, one interviewee highlighted this by stating that while one "has to be really aware of the fact that it is a proxy" (Anonymized Interviewees 2023, p.19). This highlights that practitioners, particularly with regard to the practical applicability of data valuation, often anticipate an *approximated data value* or even a *relative data value*. This *relative data value* is exemplarily characterized by the ability to "compare the value of data in the context for a person with the value of something else in the same context for the same person" (Anonymized Interviewees 2023, p.16).

In addition to addressing taxonomy-related questions with practitioners, the broader question of the usefulness of a DVBC for enterprises was posed (see Fig. 5).

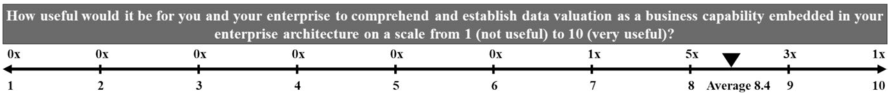

Fig. 5 Indication of the usefulness of a data valuation business capability for enterprises

As depicted in Fig. 5, ten out of eleven practitioners expressed confidence in assessing the usefulness of a DVBC. Notably, most practitioners provided two ratings, with the more conservative number serving as the basis for verification. Even when employing conservative values, it is evident that practitioners, on average, rate the usefulness of a DVBC at 8.4 out of 10. This underscores the practical significance of the research topic and thus contributes to verifying the academic assertions regarding the importance of data and its valuation (Brennan et al. 2019; Stein et al. 2021; Anonymized for review 2023). The practitioners argue their ratings with five significant interpretation clusters regarding the usefulness of a DVBC.

Firstly, the interviewees state that a properly set-up DVBC, embedded within an enterprise architecture (Anonymized Interviewees 2023), is a competitive differentiator. It supports the enterprise in reliably prioritizing an arbitrary range of versatile, data-driven use cases and data products, ultimately enabling fact-based investment decisions (Anonymized Interviewees 2023). However, it is imperative to interpret the data value correctly, primarily when data value is expressed as a number, to account for uncertainties and to derive the right decisions (Anonymized Interviewees 2023).

Secondly, the usefulness of a DVBC is highly valued because, according to the interviewees, it can help create transparency regarding the reasons behind data-driven activities and their added value. This ultimately contributes to actively shaping the change management process and engaging affected individuals, from decision-makers to operational employees, on the journey toward becoming a data-driven enterprise (Anonymized Interviewees 2023).

Thirdly, the DVBC is considered useful for enterprises as it enables the development of data-driven business models for both data consumers and providers. Additionally, the DVBC offers opportunities for consulting firms to strengthen their data-related consulting portfolios with tangible tools and frameworks that support their clients in data valuation (Anonymized Interviewees 2023).

Fourthly, the DVBC facilitates comprehensive data analysis and accurate valuation, which is linked to examining the processes from which data originate or for which processes data are utilized. This examination enables the identification and realization of process optimization opportunities, thereby enhancing overall operational efficiency (Anonymized Interviewees 2023).

Fifthly, the DVBC is classified as useful because it can form a component of an international data valuation standard that minimizes valuation uncertainties (Anonymized Interviewees 2023).

5 Discussion

This section focuses on discussing the research that was conducted. Accordingly, the discussion addresses the content, method, and quality of the validation and extension of the DVBC taxonomy.

5.1 Content-related discussion

The interviews, each lasting approximately 30–35 min, expanded the academic perspective on data valuation by adding insights from real-world practitioners within consulting and corporate enterprises. While data valuation is increasingly recognized as strategically important – particularly in data-intensive sectors like finance and insurance – many enterprises are still in their infancy of developing effective data valuation practices (Mavrogiorgou et al. 2023; Veldkamp 2023). This supports the argument of various scholars that, while data value is a highly relevant topic, the approaches to data valuation offered in academic literature often do not meet the practical requirements of real-world enterprises and their ecosystems (Li et al. 2019; Cong et al. 2022; Meierhofer et al. 2022). Therefore, the present study provides a valuable contribution by outlining a solution space for data valuation as a business capability that integrates both theoretical concepts and practical insights.

The frequently noted gap between scientific data valuation approaches and practical applicability (Li et al. 2019; Meierhofer et al. 2022) and its associated oversimplification (Cong et al. 2022) stems from various challenges currently faced by enterprises, as highlighted in the interviews conducted. These challenges reflect the complexity and interdisciplinary nature of data valuation, making it intricate to establish a straightforward monetary value for data, which instead “must be calculated more creatively” (Fleckenstein et al. 2023a). For example, interviewees emphasized that diverse stakeholders across business units, both internally and externally, and at both strategic and operational levels, including their experiences and requirements should participate in the data valuation process. This requirement alone introduces an arbitrarily high level of complexity. As anticipated based on existing literature, the roles of data providers, users, owners, and brokers are particularly important for data valuation (Pei 2022). In contrast, value auditing stakeholders (Holst et al. 2020) represent a more academic concept for which most enterprises currently lack sufficient maturity in the context of data valuation (Zeleti and Ojo 2017). Based on the conducted interviews, enterprises are currently primarily focused on establishing fundamental aspects of reusability and traceability in data valuation before conceptualizing and implementing data value governance processes such as auditing.

Although many enterprises are still in the early stages of data valuation (Li et al. 2019; Thieullent et al. 2020), it can still be highly beneficial to progress toward a full-stack DVBC incrementally. As highlighted by the DVBC taxonomy and emphasized by the interviewees, data valuation is “not really a sequential process, [but] an iterative process” (Anonymized Interviewees 2023, p.14). Particularly, when data value is determined using approaches that engage multiple stakeholders with diverse perspectives and incorporate various criteria, it is recommended to iteratively compare the estimated data value with the actual value generated (Anonymized Interviewees 2023). This iterative process, encompassing data value determination, realization, and monitoring (Hafner et al. 2024a), facilitates the continuous refinement of models, frameworks, and algorithms, enhancing the enterprise's overall data maturity.

A further notable outcome of the interviews was the strong consensus that data value is highly context-dependent, heavily influenced by the underlying use case, data product, organizational objectives, as well as the individual requirements of the involved stakeholders (Brown and Escobar 2019; Schneider et al. 2022; Mendizabal-Arrieta et al. 2023). While this is highlighted by the connections between the words *data*, *value*, *use case*, *product*, *context*, and *perspective* within the cluster analysis, the subsequent detailed analysis of the interviews reveals that the addition of a dimension, namely motivation, is essential in the DVBC taxonomy. As motivation in this context describes the specific drive behind an enterprise's engagement with data valuation, these motivations play a critical role in determining which other dimensions of the DVBC taxonomy are classified as particularly relevant. For instance, if an enterprise places high importance on generating topline growth, expanding market share, and/or optimizing processes, a quantitative assessment of data value and its business utility becomes especially relevant to perform a cost–benefit analysis. Conversely, a different motivation, such as building strategic partnerships or enhancing positive branding and corporate image, may emphasize other dimensions within the DVBC taxonomy, such as the internal and external stakeholders to be engaged or focused attention on reputation-sensitive factors like data security and privacy. In summary, the context of data valuation and the accompanying motivation of companies and their employees significantly influence the further development of a DVBC.

Regardless of the data valuation context, the data value driver data quality is regarded not merely as a central influencing factor for data value, but as a fundamental prerequisite for data valuation. This perspective aligns with current research, where numerous studies identify data quality as a primary driver of data value (Stahl and Vossen 2016b; Yu and Zhang 2017; Schneider et al. 2022; Mendizabal-Arrieta et al. 2023). The cluster analysis conducted further indicates that data quality is closely and consistently associated with the *data* itself and its *value*. Nonetheless, the interviews also highlight other crucial value drivers, including costs (Schneider et al. 2022), the potential of data to reduce uncertainty (Mendizabal-Arrieta et al. 2023), and the feasibility of implementing use cases and data products. These factors are deemed essential for bridging the gap between scientific data valuation frameworks and practical, real-world requirements.

Further, in light of the analyzed dimensions of the DVBC taxonomy, based on the cluster analysis and in-depth interview examination, it becomes evident that data valuation goes beyond simply attaching a price tag to a dataset. Rather, data valuation spans multiple layers of enterprise architecture (Hafner et al. 2024a). Following the TOGAF standard and its layers business, data, application, and technology (The Open Group 2022), a review of the cluster analysis reveals that all these layers are relevant to data valuation. The business layer is pertinent through the inclusion of *people*, *organizations*, *roles*, and *processes*; the data layer involves aspects such as *data quality*, *datasets*, and *data value*; the application layer incorporates *tools* and *systems* for data valuation; and the technology layer encompasses the *infrastructure* required for data valuation. Consequently, defining data valuation as a business capability and embedding it within an enterprise architecture as "consistent technical and organizational practices associated with the management of data" (Queiroz et al. 2024) represents a valid and promising approach to making data valuation accessible for enterprises. It is essential that this is not achieved solely through technological means; rather, a cultural shift towards becoming a data-driven enterprise is necessary. This shift requires comprehensive change management, as highlighted in scientific literature (Windt et al. 2019; Mendizabal-Arrieta et al. 2023) and confirmed by the conducted interviews.

Furthermore, a meta-analysis of all conducted interviews, especially concerning the evaluation of a DVBC's usefulness within the interviewees' enterprises, uncovers several insights that align with existing scientific literature, including the observation that determining and realizing value through data extends beyond quantitative benefits (Elia et al. 2020; Hafner et al. 2024a). One of the primary advantages of implementing data valuation as a business capability within an enterprise architecture lies in significantly enhancing transparency not only in operational processes but also in investment and resource allocation decisions (Hafner and Mira da Silva 2023). This transparency goes beyond just making things visible; it uncovers hidden or unknown information and patterns and helps to reduce uncertainties related to the value of data. In essence, implementing a DVBC promotes more profound engagement with data (Hafner and Mira da Silva 2023), leading to greater transparency, decreased uncertainty, and risks (Wang et al. 2021; Veldkamp 2023), and ultimately, more reliable data through iterative monitoring (Hafner and Mira da Silva 2023). In addition to the internal benefits for enterprises, a DVBC-enabled increase in transparency also brings external advantages. "Confusion about data value is one of the main reasons why market participants are not willing to share their data," (Wang et al. 2021, p.2) so improving transparency about data value, its scale, and the methods to assess it could drive entire industries and ecosystems into the data-driven era.

Concerning industries and ecosystems, it is also evident that "today, there is no standard to measure the value of data" (Fleckenstein et al. 2023b, p.1). This results in the determination of data value being heavily dependent on the underlying methodology, the individuals involved, the auditors or facilitators engaged in the process, and other contributing factors. A DVBC can serve as a critical trigger point and foundation to intensify efforts in developing an "internationally acclaimed open standard" (Hafner and Mira da Silva 2023, p.19) such as an ISO standard for data valuation. An essential component of a data valuation standard could involve making the abstract concept of data value more tangible through clear definitions, guidelines, and best practices. Such a standardization framework could facilitate the development of tools designed to support and guide individuals in implementing and operating the DVBC. This, in turn, could prove beneficial for enterprises, as "most companies still lack theoretical and practical tools for quantifying the value of the data in their ecosystem" (Meierhofer et al. 2022, p.10).

Ultimately, the synergy between highly data-mature enterprises, industries, and ecosystems that embrace the concept of data value as a business capability can potentially drive the development of new business models and associated *business cases* (see cluster analysis) both within and beyond enterprise boundaries. Specifically, "novel business models such as data-driven businesses […] transformed entire industries" (Recker et al. 2021, p.270) and are contributing to an "entire data economy […], which offers far greater opportunities on the market than the usual business model of the various companies themselves" (Hafner and Mira da Silva 2023, p.23).

5.2 Methodology-related discussion

The methodology employed in this study involves collecting and analyzing empirical data. Specifically, the authors utilized a theory-based DVBC taxonomy (Hafner and Mira da Silva 2023) as a foundation, validated through semi-structured interviews with practitioners in real-world settings, primarily employing open-ended questions. Upon examination of the transcribed interviews, it became apparent that using semi-structured interviews with open-ended questions was highly influential in understanding the complex and multifaceted subject of data valuation within the context of enterprise architecture.

However, challenges emerged, including a certain level of fuzziness in the responses due to the open-ended and semi-structured nature of the questions and interview guide (Carruthers 1990). This resulted in explicit and implicit references to specific dimensions and characteristics within the taxonomy. Nevertheless, given that one objective of this study was to obtain unbiased perspectives on data valuation from actual enterprises, the use of semi-structured open-ended interviews can be considered the most suitable methodology, recognizing that specific theory-based dimensions, such as data valuation theory or value auditing stakeholder, may have received less attention.

Moreover, the taxonomy, as well as its validation and extension, align with the TOGAF definition of business capabilities across the four layers: information, resources, roles, and processes (Gonzalez et al. 2018). It is essential to acknowledge, as emphasized in the initial theory-based taxonomy proposed by Hafner and Mira da Silva (2023), that although TOGAF is a widely adopted and extensively tested standard in both theory and practice (Bui 2017; Al-Turkistani et al. 2021; Anonymized for review 2023), alternative definitions and structural elements of a business capability certainly have their raison d'être (Offerman et al. 2017).

Furthermore, it should be noted that based on eleven expert interviews, preliminary foundations were established for the cluster analysis. The number of eleven conducted interviews raises the question of whether the authors have achieved information saturation. While the cluster analysis would be even more robust with additional interviews or iterations, it is important at this juncture to consider the objective of the cluster analysis. The objective of the cluster analysis is to demonstrate connections between relevant terminologies of data valuation with practical insights from in-depth subject matter experts, serving as one foundation for the validation and extension of the DVBC taxonomy. It is acknowledged at this point that the cluster analysis has served its purpose despite the limited number of expert interviews because the interviewees have solid backgrounds in both consulting and corporate functions, leading to a broad and deep perspective on data valuation across sectors and enterprises. However, future research involving the further analysis of spoken words in the interview using other techniques, such as latent semantic analysis (Deerwester et al. 1990) or latent dirichlet allocation (Blei et al. 2003), could be advantageous in strengthening the findings.

5.3 Quality-related discussion

In the context of the quality-related discussion, the validation and expansion of the DVBC taxonomy not only incorporates empirical observations from real-world enterprises but also establishes connections with other taxonomies, such as Attard and Brennan (2018), along with its recommendations for practical evaluation of data value-related taxonomies. However, it is essential to assess the extent to which the validated and extended DVBC taxonomy aligns with the objective and subjective ending conditions according to the taxonomy development process (Nickerson et al. 2017). Table 3 provides an overview of the degree to which the ending conditions are fulfilled, followed by a detailed rationale.

The first objective ending condition, namely *objectivity*, pertains to determining whether the validated and extended DVBC taxonomy has examined a representative sample of elements (Nickerson et al. 2017; Hafner and Mira da Silva 2023). During the process of validating and extending the DVBC taxonomy, all dimensions of the taxonomy were explored in expert interviews, indicating that interviewees could comprehensively address the taxonomy's dimensions and characteristics based on their experiences. By combining this practical validation and extension with the original theory-based taxonomy, it can be verified that objectivity is achieved.

The second objective ending condition, namely *granularity*, pertains to the requirement that at least one object can be categorized under a characteristic of a taxonomy dimension (Nickerson et al. 2017; Hafner and Mira da Silva 2023). Each object within the DVBC taxonomy is assigned to a broader cluster, namely a dimension or layer. Moreover, considering the examples provided by interviewees for the extended taxonomy dimensions and characteristics, it is evident that at least one object can also be classified under the taxonomy characteristics. Thus, it can be inferred that the objective ending condition of granularity is met.

The third objective ending condition, *uniqueness*, ensures that all dimensions and characteristics are free of duplications (Nickerson et al. 2017; Hafner and Mira da Silva 2023). Given that all characteristics and dimensions are free of duplicates and are allocated to a single higher-order cluster, it is evident that the objective ending condition of uniqueness is satisfied.

The fourth objective ending condition, *stability*, pertains to ensuring that the final iteration of the taxonomy development process is free from additions, mergers, and splits of taxonomy elements (Nickerson et al. 2017; Hafner and Mira da Silva 2023). This condition is partially met. Expert interviews were conducted to both extend and validate the initial DVBC taxonomy. While the initial taxonomy demonstrated stability, the validation process further solidified its stability. However, the extension of the taxonomy during this iteration does not fully adhere to the definition of

Table 3 Quality assessment based on ending conditions following (Nickerson et al. 2017)

Ending condition type	Ending condition	Status	Impact / Explanation
Objective	Objectivity	Fully met	The DVBC taxonomy covers a representative sample of elements (layers, dimensions, characteristics)
Objective	Granularity	Fully met	The DVBC taxonomy ensures that each taxonomy dimension has at least one characteristic under which an object can be categorized
Objective	Uniqueness	Fully met	The DVBC taxonomy ensures that it contains no duplicates
Objective	Stability	Partially met	The DVBC taxonomy is considered stable enough for its intended use, as expert validation and extensions confirmed the initial stability, even though it does not fully meet strict stability definitions
Subjective	Robustness	Fully met	The DVBC taxonomy allows for clear differentiation among elements, which was further enhanced by incorporating practical insights from the interviewees
Subjective	Explainability	Fully met	The DVBC taxonomy ensures explainability, as solidified through the interview process and the subsequent adjustments to element definitions and naming
Subjective	Extendibility	Fully met	This study demonstrates that the DVBC taxonomy possesses the capability to add, remove, or modify its elements
Subjective	Conciseness	Not met	The DVBC taxonomy does not meet the conciseness condition due to the addition of essential dimensions for practitioners, exceeding the recommended limit, but this is justified to provide a comprehensive view of the DVBC taxonomy from a practical perspective
Subjective	Comprehensiveness	Fully met	The DVBC taxonomy achieves a high level of comprehensiveness by integrating insights from both scientific literature and real-world practitioners

stability according to Nickerson et al. (2017) and Hafner and Mira da Silva (2023). Nevertheless, given the stability of the initial DVBC taxonomy and the validation and extension of its elements by experts, the taxonomy developed in this study is deemed stable enough for its intended purpose.

In addition to objective ending conditions, subjective ending conditions (Nickerson et al. 2017; Szopinski et al. 2020; Hafner and Mira da Silva 2023) must also be considered when evaluating the quality of the contents of the validated and extended DVBC taxonomy. The first subjective ending condition, *robustness*, focuses on determining whether the elements of the taxonomy enable clear differentiation from one another. Following the executed validation and extension iteration, the robustness of the initial taxonomy is further strengthened. Expert interviews revealed opportunities for proper differentiation among the taxonomy elements, which were subsequently incorporated, thus enhancing the taxonomy's robustness. Moreover, the interview-based approach to validating and extending the taxonomy revealed and addressed potential misunderstandings, thereby satisfying another subjective ending condition, namely *explainability* (Nickerson et al. 2017; Szopinski et al. 2020; Hafner and Mira da Silva 2023).

The enhanced robustness of the taxonomy connects with the third subjective ending condition, *extendibility*, which inquires whether the taxonomy can be expanded to accommodate additional insights and perspectives (Nickerson et al. 2017; Szopinski et al. 2020;Hafner and Mira da Silva 2023). Since interviewees indicated the need to further include both dimensions (e.g., *stakeholder domain* or *motivation*) and characteristics (e.g., *top-management* or *IT domain*), the condition of extendibility is confirmed to be met.

The fourth subjective ending condition, *conciseness*, which concerns achieving an appropriate balance in the quantity of taxonomy dimensions to ensure a sufficient yet manageable number (Nickerson et al. 2017; Szopinski et al. 2020; Hafner and Mira da Silva 2023), is no longer satisfied. This is attributed to the addition of crucial dimensions for practitioners, exceeding the maximum recommended limit of nine taxonomy dimensions (Miller 1994; Nickerson et al. 2017). However, this deviation from the ending condition is justified in light of the study's objective to offer a comprehensive perspective on DVBC from the viewpoint of real-world practitioners. Consequently, imposing a constraint on the depth and breadth of taxonomy detailing is deemed non-value-adding and is therefore disregarded for this study.

The fifth subjective ending condition, *comprehensiveness*, assesses the taxonomy's ability to differentiate various concepts related to data valuation (Nickerson et al. 2017; Szopinski et al. 2020; Hafner and Mira da Silva 2023). Since the initial DVBC taxonomy, which underwent validation and extension, already met the subjective ending condition of comprehensiveness based on a SLR incorporating forward and backward research, the validated and extended DVBC taxonomy is deemed even more comprehensive. This enhanced comprehensiveness stems from not only considering other scientifically related taxonomies but also incorporating the perspective of real-world enterprises, thereby facilitating the symbiosis of academia and practice.

In summary, it can be concluded that the objective and subjective ending conditions are predominantly fulfilled exhaustively or partially, which is considered acceptable in the context of the study objectives.

6 Conclusion

In light of the extensive academic discourse surrounding data valuation and its constrained practical implementation, this study endeavors to validate a theory-driven artifact, namely the DVBC taxonomy by Hafner and Mira da Silva (2023). To this end, the paper empirically verifies the DVBC taxonomy through interviews with eleven experts in the fields of data and enterprise architecture management following the methodological approaches according to Mingers (2001), Hove and Anda (2005), Myers and Newman (2007). These semi-structured interviews are transcribed and form the basis for a cluster analysis conducted using VOSviewer. The result of the interview and cluster analysis process is the validation of the DVBC taxonomy in the form of a heatmap, along with a taxonomy extension encompassing additional dimensions and characteristics required by real-world practitioners. The resulting field-tested DVBC taxonomy comprises four layers, twelve dimensions, and 59 characteristics. Notably, three dimensions and 23 characteristics pertain to the practice-oriented extension of the taxonomy. Concerning the usefulness of a DVBC for the interviewed practitioners, they indicate an average score of 8.4 out of 10 when interpreted conservatively, underscoring the significance of data valuation in academic discourse and within enterprises.

More specifically, the practical impacts of this study are threefold. Firstly, both a previously conducted SLR and the executed interviews revealed a divergence in the language used by different individuals when discussing data value. Consequently, the validated DVBC taxonomy aids in establishing a shared comprehension and knowledge transfer, utilizing commonly accepted terminology about data value across an enterprise and its associated domains and stakeholders. Secondly, the validated DVBC taxonomy serves as a strategic tool for enterprises to categorize their data valuation initiatives, assess their maturity level, and systematically enhance these efforts. By allowing organizations to modularly combine taxonomy elements based on specific business requirements, the DVBC taxonomy enables data-related endeavors to evolve beyond digital transformation initiatives based on gut feeling into truly value-oriented data-driven transformations. This approach supports enterprises in both measuring and realizing the value of their data through actionable steps. Engaging with key decision-makers, such as the CDO or CIO, can further refine this process by identifying which taxonomy elements align best with organizational goals and determining the necessary resources for effective implementation in particular use cases. Thirdly, the DVBC taxonomy serves as a foundation for potential future and present data valuation standards, frameworks, and toolkits, which the interviewed practitioners highly requested.

From a theoretical standpoint, this study carries two significant implications. Firstly, the validated DVBC taxonomy and its extension provide empirical support for theoretical assumptions by grounding them in practical insights. This serves to fortify the theoretical assertions put forth by various researchers in the field of data value. Secondly, the taxonomy, including its dimensions and layers, affords a broad scope for interdisciplinary research. This is because the taxonomy encompasses both technological and business-oriented viewpoints, facilitating the convergence of research domains such as information systems, business, management, and accounting, as well as decision sciences. This interdisciplinary collaboration is essential for pursuing research agendas at their intersections, which can be derived from the validated DVBC taxonomy, its dimensions, and its characteristics.

Acknowledging study limitations, the study involved eleven subject matter experts with consulting or corporate experience in German, Dutch, or Swiss enterprises across sectors. Thus, providing a universally applicable statement regarding the taxonomy's broad relevance and geographical generalizability is challenging. Additionally, due to the semi-structured open-question format of the interviews, the study explores practitioners' experiences without scrutinizing each dimension and characteristic in granular detail. Additionally, considering the sample size of eleven interviews, there is potential to expand the sample size to achieve greater information saturation.

Considering the interview contents and limitations, the research directions for future scientific work consider four areas. Firstly, to effectively integrate data valuation as a business capability within enterprise architectures, it is advisable to construct a conceptual model incorporating the taxonomy elements and their interrelations, ideally rooted in foundational ontologies such as UFO (Guizzardi et al. 2022). Secondly, it is recommended to elaborate on the taxonomy characteristics, transforming them into a concrete and measurable metrics catalog, potentially through implementing a wide-ranging survey across enterprises. This will serve as the basis for developing data valuation standards, frameworks, and tools that assist enterprises in systematically assessing the value of their data while also considering potential uncertainties through approaches like multi-criteria decision analysis models. Thirdly, there is a suggestion to design an end-to-end data valuation process, encompassing various process components and involving different stakeholders. This approach aims to enhance data valuation from a procedural standpoint, which is a crucial factor enabling enterprises to transition towards a data-driven business. Finally, it is recommended to further strengthen the proposed DVBC taxonomy and its related models through large-scale case studies and/or controlled experiments in real-world enterprises.

Appendix

See Tables 4 and 5.

Table 4 Interview guide

Phase	Question
Intro	Could you kindly introduce yourself (position, years of expertise, sectors you have worked in, which serve as foundation for your insights) and give permission to record the session for transcription purposes?
Intro	If you think about data value, what is data value for you and what is it not?
Intro	*Corporate perspective:* Does your enterprise determine the value of its data? If so, is the endeavor of determining data value established as a systemized process or business capability? *Consulting perspective:* Do the enterprises you have been working with determine the value of their data? If so, is the endeavor of determining data value established as a systemized process or business capability?
Intro	*Corporate perspective:* What are the main pitfalls and challenges you see in your enterprise regarding data valuation? *Consulting perspective:* What are the main pitfalls and challenges you see in the enterprises you have been working with regarding data valuation?
Taxonomy evaluation	*Corporate perspective:* When talking about data value, what kinds of data does your enterprise deal with? *Consulting perspective:* When discussing data value, what kinds of data do the enterprises you have been working with deal with?
Taxonomy evaluation	*Corporate perspective:* What are the main purposes for your enterprise when determining the value of your data? *Consulting perspective:* What are the main purposes of the enterprises you have been working with when determining the value of their data?
Taxonomy evaluation	*Corporate perspective:* What are parameters affecting the value of data in your enterprise and which of them would you consider the most relevant? *Consulting perspective:* What are parameters affecting the value of data in the enterprises you have been working with and which of them would you consider the most relevant?
Taxonomy evaluation	*Corporate perspective:* How would you describe a typical process or approach your enterprise uses to determine the value of data? *Consulting perspective:* How would you describe a typical process or approach the enterprises you have been working with use to determine the value of data?
Taxonomy evaluation	*Corporate perspective:* Which activities/processes occur before and/or after the determination of data value in your enterprise? *Consulting perspective:* Which activities/processes occur before and/or after the determination of data value in the enterprises you have been working with?
Taxonomy evaluation	*Corporate perspective:* Which tools, frameworks, and standards does your enterprise use to determine the value of data? *Consulting perspective:* Which tools, frameworks, and standards do the enterprises you have been working with use to determine the value of data?
Taxonomy evaluation	*Corporate perspective:* Which stakeholders (internal and external) are involved in the data valuation endeavor of your enterprise and what roles do they play in the data valuation endeavors? *Consulting perspective:* Which stakeholders (internal and external) are involved in the data valuation endeavor of the enterprises you have been working with what roles do they play in the data valuation endeavors?

Table 4 (continued)

Phase	Question
Taxonomy evaluation	*Corporate perspective:* Imagine having a sound data valuation business capability in place in your enterprise. What would you expect to be its outcomes?
	Consulting perspective: Imagine having a sound data valuation business capability in place in the enterprises you have been working with. What would you expect to be the outcomes of it?
Outlook and closing	*Corporate perspective:* How useful would it be for you and your enterprise to comprehend and establish data valuation as a business capability embedded in your enterprise architecture on a scale from 1 (not useful) to 10 (very useful)? Please explain your rating
	Consulting perspective: How useful would it be for you and the enterprises you have been working with to comprehend and establish data valuation as a business capability embedded in your enterprise architecture on a scale from 1 (not useful) to 10 (very useful)? Please explain your rating
Outlook and closing	*Make a wish:* What requirements do you have regarding comprehensive data valuation how would you like to be equipped?

Table 5 Explanation of DVBC taxonomy dimensions and characteristics based on (Hafner and Mira da Silva 2023, 2024b; Anonymized Interviewees 2023)

Hierarchy	Taxonomy element	Description
Dimension	Motivation	Groups various reasons why enterprises engage with data valuation
Characteristic	Topline growth and market shares	The motivation behind data valuation and its associated activities is to increase revenue and profits while expanding market share
Characteristic	Shareholder value maximization	The motivation behind data valuation and its associated activities is to satisfy investors and shareholders with promising financial results
Characteristic	Business continuity and risk management	The motivation behind data valuation and its associated activities is to utilize valued data for risk identification and mitigation, ensuring the seamless operation of business processes
Characteristic	Process and cost optimization	The motivation behind data valuation and its associated activities is to use data to increase process efficiency and effectiveness while minimizing waste and improving overall productivity
Characteristic	Regulatory compliance	The motivation behind data valuation and its associated activities is to ensure compliance with external regulations, such as the EU-GDPR or the EU Data Act, as well as internal compliance policies
Characteristic	Branding and image	The motivation behind data valuation and its associated activities is to enhance the enterprises' external image to attract skilled professionals and gain positive media and investor attention
Characteristic	Data-driven product and business model innovation	The motivation for data valuation and its associated activities is to enhance an enterprise's capability to expand its traditional product portfolio with data-driven offerings and integrate data-driven elements into its business model
Characteristic	Fact-based decision-making	The motivation behind data valuation and its associated activities lies in increasing the proportion of decisions grounded on data and evidence rather than relying on intuition or subjective judgment
Characteristic	Strategic partnerships	The motivation behind data valuation and its associated activities is to promote collaboration and partnerships, both internally across business unit boundaries and externally within data ecosystems
Characteristic	Fear of competitive disadvantages	The motivation behind data valuation and its associated activities is to remain competitive and aligned with society and industries that are rapidly evolving in the data-driven era
Dimension	Purpose	Groups the objectives of enterprises concerning data valuation, with a particular focus on the determination of data value
Characteristic	Qualitative data valuation	The purpose is to gather contextual insights about the value of data
Characteristic	Quantitative data valuation	The purpose is to gather numerical insights about the value of data
Characteristic	Combination	The purpose is to integrate contextual and numerical insights to derive a mathematical data value embedded within its application context or use case
Dimension	Data valuation object	Groups potential entities of data that may be subject to data valuation
Characteristic	Bundled data	Bundled data represents data consolidated into clusters, such as data products or assets, aligned with specific use cases

Table 5 (continued)

Hierarchy	Taxonomy element	Description
Characteristic	Non-bundled data	Non-bundled data consists of standalone raw data points that remain unorganized
Dimension	Data type	Groups different types of data that constitute the data valuation object
Characteristic	Master- and metadata	Data that remains relatively stable over time and may include information about other data
Characteristic	Transactional data	Dynamic data undergoes frequent changes over time, reflecting the current state of processes, events, or activities, among others
Dimension	Data value drivers	Groups various clusters of criteria influencing data value, each containing various metrics to assess their impact on data value
Characteristic	Business utility	A data value driver that evaluates how the use case or application context influences the value of a data valuation object
Characteristic	Cost	A data value driver that evaluates how various costs, such as data management, storage, processing, and opportunity costs, among others, impact the value of a data valuation object
Characteristic	Data durability and lifetime	A data value driver that evaluates how the lifespan and recency of data impact the value of a data valuation object
Characteristic	Data quality	A data value driver that evaluates how various data quality dimensions, such as accuracy, completeness, and others, impact the value of a data valuation object
Characteristic	Data security and privacy	A data value driver that evaluates how security requirements and data privacy level influence the value of a data valuation object
Characteristic	Sentiment and perception	A data value driver that evaluates how subjective, sentiment-based experiences and judgments impact the value of a data valuation object
Characteristic	Information potential and entropy	A data value driver that evaluates how the reduction of uncertainty impacts the value of a data valuation object
Characteristic	Feasibility	A data value driver that evaluates how the efforts and likelihood of successfully implementing and utilizing a data valuation object influence its value
Characteristic	Others	Proprietary or proxy data value drivers that may be introduced by enterprises based on specific requirements and cannot be allocated to the above-mentioned clusters
Dimension	Data valuation theory	Groups various scientific and non-scientific approaches and theories to determine the value of data
Characteristic	Economic	Economic data valuation subsumes conventional cost-price data valuation approaches tailored to data
Characteristic	Cooperative game theory	Cooperative game theory subsumes interpersonal game theory concepts such as the Shapley value, among others, to determine the data value
Characteristic	Non-cooperative game theory	Non-cooperative game theory subsumes interpersonal game theory concepts such as Nash bargaining or the Stackelberg game, among others, to determine the data value

Table 5 (continued)

Hierarchy	Taxonomy element	Description
Characteristic	Decision theory	Decision theory subsumes complex and multi-dimensional concepts such as the Analytic Hierarchy Process or Multi-Criteria Decision Modelling to determine the data value
Characteristic	Query-based	Query-based approaches determine the data value by pricing technical queries to a database or data marketplace
Characteristic	Index-based	Index-based data valuation approaches determine the value of data by employing proxy metrics or indices, such as Likert scores, to evaluate criteria influencing data value
Characteristic	Proprietary	Proprietary data valuation approaches are predominantly used in enterprises with low data maturity and encompass enterprise-specific models, approaches, and frameworks
Dimension	Data valuation tooling	Groups equipment used to effectively apply data valuation theories and data value drivers for assessing and realizing data value
Characteristic	Interpersonal elaboration	Data valuation is achieved by bringing together the appropriate individuals in relevant roles to determine and realize the value of data collaboratively
Characteristic	Model and application	Data valuation is achieved by applying algorithms and logic provided through technical applications to determine and realize the value of data
Characteristic	Framework and method	Data valuation is achieved by applying conceptual frameworks and methods such as workshop concepts and ontologies to determine and realize the value of data
Characteristic	Combination	Data valuation is achieved by combining interpersonal collaboration, technical models and applications, as well as conceptual frameworks and methods to determine and realize data value
Dimension	Stakeholder domain	Groups organizational clusters of an enterprise that are essential for data valuation
Characteristic	Top-management	The enterprise's executive level comprises roles such as CEO, CDO, CIO, CFO, CISO, among others
Characteristic	IT domain	An enterprise's IT cluster comprises areas such as data and advanced analytics, platforms, infrastructure, and others
Characteristic	Finance domain	An enterprise's finance cluster comprises areas such as mergers and acquisitions, controlling, financial planning, among others
Characteristic	Functional domain	The functional clusters of an enterprise consist of departments directly impacting the enterprise's value chain, such as logistics, production, and maintenance, among others
Characteristic	Legal and risk domain	The legal and risk cluster of enterprises comprises areas such as data security and protection, data ethics, among others
Dimension	Value determination stakeholder	Groups individuals involved in determining the data value and distinguishes if these individuals are of an internal or external nature
Characteristic	Internal	Only enterprise-internal stakeholders determine the value of data

Table 5 (continued)

Hierarchy	Taxonomy element	Description
Characteristic	Internal and external (with intermediary)	Internal and external stakeholders determine the value of data while collaborating and communicating via an intermediary such as a data broker
Characteristic	Internal and external (without intermediary)	Internal and external stakeholders determine the value of data while collaborating and communicating directly with each other without an intermediary such as a data broker
Characteristic	External	Only enterprise-external stakeholders determine the data value and dictate this value to the enterprise
Dimension	Value auditing stakeholder	Groups individuals that observe and moderate the process of data valuation and distinguishes if these individuals are of an internal or external nature
Characteristic	Internal data value auditor	An enterprise-internal auditor performs the data valuation auditing
Characteristic	3rd party data value auditor	An enterprise-external auditor performs the data valuation auditing
Characteristic	Not existing	A data valuation auditor does not exist
Dimension	Component	Groups main processes that are part of a data valuation business capability and can be executed
Characteristic	Data preparation and contextualization	Gathering, preparing, clustering, and contextualizing data are prerequisites for determining its value
Characteristic	Data value assessment	Determining the value of a data valuation object
Characteristic	Data value allocation	Allocating the data value of a data valuation object to a logical entity, such as a business unit, a project, or an employee, among others
Characteristic	Data value prediction	Forecasting the future value of a data valuation object
Characteristic	Data value realization	Developing and implementing a data valuation object, such as a data product or a data-driven use case
Characteristic	Data value monitoring	Analyzing the determined or forecasted data value and the real value after realization, including gap analysis
Characteristic	Change management	Supporting data valuation processes and the individuals involved through training, requirements management, and knowledge transfer, among others
Dimension	Result	Groups various outcomes of a data valuation business capability, with a particular focus on determining data value
Characteristic	Specific absolute data value	A precise and explicit indication of data value, which is associated with particular data valuation objects and is not comparable to other data valuation objects in relation
Characteristic	Approximate absolute data value	A roughly estimated and explicit indication of data value, which is associated to particular data valuation objects and is not comparable to other data valuation objects in relation
Characteristic	Relative data value	An indication of data value that compares the value of a data valuation object to other data valuation objects

Acknowledgements It is acknowledged that *Grammarly* and *ChatGPT* have been utilized as translation tools and orthography optimization.

Funding Open access funding provided by FCT|FCCN (b-on). This study is partially funded by an Erasmus + grant (211 / SMP / 2022).

Declarations

Conflict of interest There are no competing interests to declare.

References

Abraham R, Faller H, Kinderen S, Aier S (2013) Can boundary objects mitigate communication defects in enterprise transformation? findings from expert interviews. In: lecture notes in informatics (LNI), Proceedings–Series of the Gesellschaft fur Informatik (GI). St. Gallen

Aggarwal CC, Zhai C (2012) A survey of text clustering algorithms. In: Aggarwal CC, Zhai ChengXiang (eds) Mining text data. Springer, Boston, MA, pp 77–128. https://doi.org/10.1007/978-1-4614-3223-4_4

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

Hafner M, Mira da Silva M (2023) Towards a taxonomy for business capabilities determining data value. Knowl Inf Syst. https://doi.org/10.1007/s10115-023-01994-4

Hafner M, Mira da Silva M, Proper HA (2024) Towards a reference ontology for a data valuation business capability. Enterp Inf Syst. https://doi.org/10.1080/17517575.2024.2358920

Anonymized for review (2024b) Anonymized title. Currently under peer-review in J Manag Analy

Anonymized interviewees (2023) Interview transcripts

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. SciTePress, Funchal, 2:133–144. https://doi.org/10.5220/0006777701330144

Batini C, Castelli M, Viscusi G et al (2018) Digital information asset evaluation: a case study in manufacturing. Adv Inf Sys 49:19–33. https://doi.org/10.1145/3242734.3242737

Bearman M (2019) Focus on methodology: eliciting rich data: a practical approach to writing semi-structured interview schedules. Focus Health Prof Educ Multi-Prof J 20:1–11. https://doi.org/10.11157/fohpe.v20i3.387

Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. J Mach Learn Res 3:993–1022. https://doi.org/10.1162/jmlr.2003.3.4-5.993

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

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, Heraklion, pp 188–195. https://doi.org/10.5220/0007723402000207

Brown ME, Escobar VM (2019) NASA's early adopter program links satellite data to decision making. Remote Sens. https://doi.org/10.3390/rs11040406
Bui QN (2017) Evaluating enterprise architecture frameworks using essential elements. Commun Assoc Inf Syst 41:121–149. https://doi.org/10.17705/1CAIS.04106
Bukar UA, Sayeed MS, Razak SFA et al (2023) A method for analyzing text using VOSviewer. MethodsX 11:102339. https://doi.org/10.1016/j.mex.2023.102339
Busch-Casler J, Radic M (2022) Personal data markets: a narrative review on influence factors of the price of personal data. In: Guizzardi R, Ralyté J, Franch X (eds) Research challenges in information science. Springer International Publishing, Cham, pp 3–19. https://doi.org/10.1007/978-3-031-05760-1_1
Carruthers J (1990) A rationale for the use of semi-structured interviews. J Educ Adm. https://doi.org/10.1108/09578239010006046
Cheong H, Kim B, Vaquero IU (2023) Data valuation model for estimating collateral loans in corporate financial transaction. J Risk Financ Manag. https://doi.org/10.3390/jrfm16030206
Cong Z, Luo X, Pei J et al (2022) Data pricing in machine learning pipelines. Knowl Inf Syst 64:1417–1455. https://doi.org/10.1007/s10115-022-01679-4
Coyle D, Manley A (2023) What is the value of data? A review of empirical methods. J Econ Surv. https://doi.org/10.1111/joes.12585
Debattista J, Attard J, Brennan R (2018) Semantic data ingestion for intelligent, value-driven big data analytics. In: 4th international conference on big data innovations and applications. IEEE, Barcelona, pp 1–8. https://doi.org/10.1109/Innovate-Data.2018.00008
Deerwester S, Dumais ST, Furnas GW et al (1990) Indexing by latent semantic analysis. J Am Soc Inf Sci 41:391–407
Elia G, Polimeno G, Solazzo G, Passiante G (2020) A multi-dimension framework for value creation through big data. Ind Mark Manag. https://doi.org/10.1016/j.indmarman.2019.08.004
Enders T (2018) Exploring the value of data–a research agenda. In: Satzger G, Patrício L, Zaki M, Kühl N, Hottum P (eds) Exploring service science. IESS 2018 Lecture notes in business information processing. Springer International Publishing, Cham, pp 274–286. https://doi.org/10.1007/978-3-030-00713-3_21
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: 55th Hawaii international conference on system sciences. AIS. https://doi.org/10.24251/HICSS.2022.723
Faroukhi AZ, El AI, Gahi Y, Amine A (2020) An adaptable big data value chain framework for end-to-end big data monetization. Big Data Cogn Comput 4:1–27. https://doi.org/10.3390/bdcc4040034
Fleckenstein M, Obaidi A, Tryfona N (2023) A review of data valuation approaches and building and scoring a data valuation model. Harv Data Sci Rev. https://doi.org/10.1162/99608f92.c18db966
Gkatzelis V, Aperjis C, Huberman BA (2015) Pricing private data. Electron Mark 25:109–123. https://doi.org/10.1007/S12525-015-0188-8
Gonzalez S, Hansen K, Hendricky H, et al (2018) TOGAF ® Series Guide Business Capabilities
Guizzardi G, Benevides A, Fonseca C et al (2022) UFO: unified foundational ontology. Appl Ontol. https://doi.org/10.3233/AO-210256
Hartwich E, Ollig P, Fridgen G, Rieger A (2024) Probably something: a multi-layer taxonomy of non-fungible tokens. Internet Res 34:216–238. https://doi.org/10.1108/INTR-08-2022-0666
Hayashi Y, Friedel JE, Foreman AM, Wirth O (2019) A cluster analysis of text message users based on their demand for text messaging: a behavioral economic approach. J Exp Anal Behav 112(3):273–289. https://doi.org/10.1002/jeab.554
Hevner AR, March ST et al (2004) Design science in information systems research. MIS Q 28:75–105. https://doi.org/10.5555/2017212.2017217
Holst L, Stich V, Schuh G, Frank J (2020) Towards a comparative data value assessment framework for smart product service systems. In: Lalic B, Majstorovic V, Marjanovic U, von Cieminski G, Romero D (eds) Advances in production management systems. Towards smart and digital manufacturing. APMS 2020. IFIP advances in information and communication technology. Springer International Publishing, Cham, pp 330–337. https://doi.org/10.1007/978-3-030-57997-5_39

Hove SE, Anda B (2005) Experiences from conducting semi-structured interviews in empirical software engineering research. In: 11th IEEE international software metrics symposium (METRICS'05). IEEE, Como, pp 10 pp.–23. https://doi.org/10.1109/METRICS.2005.24
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
Iyamu T (2018) Collecting qualitative data for information systems studies: the reality in practice. Educ Inf Technol (Dordr) 23:2249–2264. https://doi.org/10.1007/s10639-018-9718-2
Jiang Y, Jeusfeld MA, Ding J, Sandahl E (2023) Model-based cybersecurity analysis. Bus Inf Syst Eng 65:643–676. https://doi.org/10.1007/s12599-023-00811-0
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). AIS, Guimarães. https://aisel.aisnet.org/ecis2017_rp/98/
Kang H, Guo D (2023) Value evaluation of data resources based on artificial neural network in digital economy. Soft Comput. https://doi.org/10.1007/s00500-023-08361-z
Kaplan B, Duchon D (1988) Combining qualitative and quantitative methods in information systems research: a case study. MIS Q 12:571–586. https://doi.org/10.2307/249133
Khan MA, Uddin MF, Gupta N (2014) Seven V's of big data understanding big data to extract value. In: Zone 1 conference of the American society for engineering education. IEEE, Bridgeport, pp 1–5. https://doi.org/10.1109/ASEEZone1.2014.6820689
Körner M-F, Michaelis A, Spazierer S, Strüker J (2023) Accelerating sustainability in companies: a taxonomy of information systems for corporate carbon risk management. In: 31st European conference on information systems. AIS, Kristiansand. https://aisel.aisnet.org/ecis2023_rp/248
Leech BL (2002) Asking questions: techniques for semistructured interviews. Polit Sci Polit 35:665–668. https://doi.org/10.1017/S1049096502001129
Lega M, Colot C, Burnay C, Linden I (2022) Supporting data selection for decision support systems: towards a decision-making data value taxonomy. In: 34th international conference on software engineering and knowledge engineering, SEKE. Knowledge systems institute graduate school, Pittsburgh, pp 487–492. https://doi.org/10.18293/SEKE2022-104
Li X, Yao J, Liu X, Guan H (2017) A first look at information entropy-based data pricing. In: 37th international conference on distributed computing systems (ICDCS). IEEE, Atlanta, pp 2053–2060. https://doi.org/10.1109/ICDCS.2017.45
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). IEEE, Suzhou, pp 88–93. https://doi.org/10.1109/ICBDA.2019.8713240
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
Lim S, Lee C-H, Bae J-H, Jeon Y-H (2024) Identifying the optimal valuation model for maritime data assets with the analytic hierarchy process (AHP). Sustainability. https://doi.org/10.3390/su16083284
Lin C-L, Chen Z, Jiang X et al (2022) Roles and research trends of neuroscience on major information systems journal: a bibliometric and content analysis. Front Neurosci. https://doi.org/10.3389/fnins.2022.872532
Mavrogiorgou A, Kiourtis A, Makridis G, et al (2023) FAME: Federated decentralized trusted data marketplace for embedded finance. In: international conference on smart applications, communications and networking (SmartNets). IEEE, Istanbul, pp 1–6. https://doi.org/10.1109/SmartNets58706.2023.10215814
Meierhofer J, Benedech R, Schweiger L et al (2022) Quantitative modelling of the value of data for manufacturing SMEs in smart service provision. ITM Web Conf. https://doi.org/10.1051/itmconf/20224104001
Mendizabal-Arrieta G, Castellano-Fernández E, Rapaccini M (2023) A pricing model to monetize your industrial data. Front Manuf Technol. https://doi.org/10.3389/fmtec.2023.1057537
Miller GA (1994) The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol Rev. https://doi.org/10.1037/0033-295X.101.2.343
Mingers J (2001) Combining is research methods: towards a pluralist methodology. Inf Syst Res 12:240–259. https://doi.org/10.1287/isre.12.3.240.9709
Montibeller G, von Winterfeldt D (2015) Cognitive and motivational biases in decision and risk analysis. Risk Anal 35:1230–1251. https://doi.org/10.1111/risa.12360

Mukred M, Yusof ZM, Mokhtar UA, Fauzi F (2018) Taxonomic framework for factors influencing ERMS adoption in organisations of higher professional education. J Inf Sci 45:139–155. https://doi.org/10.1177/0165551518783133
Myers MD, Newman M (2007) The qualitative interview in IS research: examining the craft. Inf Organ 17:2–26. https://doi.org/10.1016/J.INFOANDORG.2006.11.001
Nickerson RC, Varshney U, Muntermann J (2017) A method for taxonomy development and its application in information systems. Eur J Inf Syst. https://doi.org/10.1057/ejis.2012.26
Noshad M, Choi J, Sun Y et al (2021) A data value metric for quantifying information content and utility. J Big Data 8:82. https://doi.org/10.1186/s40537-021-00446-6
Offerman T, Stettina CJ, Plaat A (2017) Business capabilities: a systematic literature review and a research agenda. In: 2017 international conference on engineering, technology and innovation (ICE/ITMC). IEEE, Madeira, pp 383–393. https://doi.org/10.1109/ICE.2017.8279911
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
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
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
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
Queiroz M, Tallon P, Coltman T (2024) Data value and the search for a single source of truth: What is it and why does it matter? In: 57th Hawaii international conference on system sciences. AIS, Waikiki, pp 6636–6645. https://aisel.aisnet.org/hicss-57/os/practice-based_research/2
Recker J, Lukyanenko R, Jabbari M, Samuel BM, Castellanos A (2020) From representation to mediation: a new agenda for conceptual modeling research in a digital world. MIS Q 45(1):269–300. https://doi.org/10.25300/MISQ/2021/16027
Robinson SC (2017) What's your anonymity worth? Establishing a marketplace for the valuation and control of individuals' anonymity and personal data. Digital Policy, Regul Gov 19:353–366. https://doi.org/10.1108/DPRG-05-2017-0018
Schneider J, Abraham R, Meske C, Vom Brocke J (2022) Artificial intelligence governance for businesses. Inf Syst Manag 40:229–249. https://doi.org/10.1080/10580530.2022.2085825
Schultze U, Avital M (2011) Designing interviews to generate rich data for information systems research. Inf Organ 21:1–16. https://doi.org/10.1016/j.infoandorg.2010.11.001
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. AIS, Marrakech. https://aisel.aisnet.org/ecis2021_rip/34
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
Shen Y, Guo B, Shen Y, et al (2019) Pricing personal data based on information entropy. In: 2nd international conference on software engineering and information management. ACM, New York, USA, pp 143–146. https://doi.org/10.1145/3305160.3305204
Sidgman J, Crompton M (2016) Valuing personal data to foster privacy: a thought experiment and opportunities for research. J Inf Syst 30:169–181. https://doi.org/10.2308/isys-51429
Stahl F, Vossen G (2016a) Fair knapsack pricing for data marketplaces. In: Pokorný J, Ivanović M, Thalheim B, Šaloun P (eds) Advances in databases and information systems. Springer International Publishing, Cham, pp 46–59. https://doi.org/10.1007/978-3-319-44039-2_4
Stahl F, Vossen G (2016b) Data quality scores for pricing on data marketplaces. In: Nguyen NT, Trawiński B, Fujita H, Hong T-P (eds) Intelligent information and database systems: 8th Asian Conference, ACIIDS 2016. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp 215–224. https://doi.org/10.1007/978-3-662-49381-6_21
Stein H, Holst L, Stich V, Maass W (2021) From qualitative to quantitative data valuation in manufacturing companies. In: Dolgui A, Bernard A, Lemoine D, von Cieminski G, Romero D (eds) Advances in production management systems. Artificial intelligence for sustainable and resilient production systems: IFIP WG 5.7 international conference. Springer International Publishing, Cham, pp 172–180. https://doi.org/10.1007/978-3-030-85902-2_19

Szopinski D, Schoormann T, Kundisch D (2019) Because your taxonomy is worth it: towards a framework for taxonomy evaluation. In: 27th European conference on information systems. AIS, Stockholm-Uppsala. https://aisel.aisnet.org/ecis2019_rp/104

Szopinski D, Schoormann T, Kundisch D (2020) Criteria as a prelude for guiding taxonomy evaluation. In: 53rd Hawaii international conference on system sciences. Maui. https://doi.org/10.24251/HICSS.2020.622

The open group (2022) The TOGAF® Standard–10th Edition

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

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

van Eck NJ, Waltman L (2011) Text mining and visualization using VOSviewer. ISSI Newsletter 50–54. https://doi.org/10.48550/arXiv.1109.2058

van Eck NJ, Waltman L (2018) VOSviewer Manual

Veldkamp L (2023) Valuing data as an asset. Rev Financ 27:1545–1562. https://doi.org/10.1093/rof/rfac073

vom Brocke J, Hevner A, Maedche A (2020) Introduction to design science research. In: vom Brocke J, Hevner A, Maedche A (eds) Design science research. cases. Springer International Publishing, Cham, pp 1–13. https://doi.org/10.1007/978-3-030-46781-4_1

Wang B, Guo Q, Xia T et al (2024) Cooperative IoT data sharing with heterogeneity of participants based on electricity retail. IEEE Internet Things J 11:4956–4970. https://doi.org/10.1109/JIOT.2023.3301731

Wang B, Guo Q, Yang T et al (2021) Data valuation for decision-making with uncertainty in energy transactions: a case of the two-settlement market system. Appl Energy 288:116643. https://doi.org/10.1016/j.apenergy.2021.116643

Webster J, Watson RT (2002) Analyzing the past to prepare for the future: writing a literature review. MIS Q 26:xiii–xxiii. http://www.jstor.org/stable/4132319

Weingart P, Wambsganns T, Soellner M (2023) A taxonomy for deriving business insights from user-generated content. In: 31st European conference on information systems. AIS, Kristiansand. https://aisel.aisnet.org/ecis2023_rp/401

Windt B, Borgman H, Amrit C (2019) Understanding leadership challenges and responses in data-driven transformations. In: 52nd Hawaii international conference on system sciences. Maui. https://doi.org/10.24251/HICSS.2019.599

Wu Z, Shu Y, Low BKH (2022) DAVINZ: Data valuation using deep neural networks at initialization. In: Chaudhuri K, Jegelka S, Song L, et al. (eds). In: Proceedings of the 39th international conference on machine learning. PMLR, pp 24150–24176

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. IEEE, pp 850–859. https://doi.org/10.1109/INFOCOM48880.2022.9796669

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

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

Zhang Y, Ge L, Xiao L et al (2021) A bibliometric review of information systems research from 1975–2018: setting an agenda for is research. J Glob Inf Manag 29:1–24. https://doi.org/10.4018/JGIM.287631