Supporting Data Selection for Decision Support Systems: Towards a Decision-Making Data Value Taxonomy

1st Mathieu Lega
Namur Digital Institute (NaDI)
University of Namur
Namur, Belgium
0000-0003-1682-4920

2nd Christian Colot
Namur Digital Institute (NaDI)
University of Namur
Namur, Belgium
0000-0002-0520-7364

3rd Corentin Burnay
Namur Digital Institute (NaDI)
University of Namur
Namur, Belgium
0000-0002-0325-1732

4th Isabelle Linden
Namur Digital Institute (NaDI)
University of Namur
Namur, Belgium
0000-0001-8034-1857

***Abstract*—Data are ubiquitous and generate unprecedented opportunities to help making decisions within organizations. This has led so-called data-driven Decision Support Systems (DSS) to become critical, if not vital, systems for most companies. The design of such DSS raises important methodological challenges, since data-driven DSS should expose only useful information to decision makers, but data available in a company's database are numerous and not equally supportive. Failing to provide the right data to the right decision-maker may reduce the usefulness of a DSS, and can lead to lower quality decision outputs. This is particularly striking in the case of Self-Service Business Intelligence (SSBI) where users build DSS outputs themselves. In this paper, we elaborate on this idea of data profusion and propose a data selection criterion, namely the decision-making data value. To do this, we discuss the concept of value and its application to data and decision making, we review existing literature and propose a taxonomy of the dimensions of data value in the context of decision making. We also validate this taxonomy with semi-direct interviews and discuss the future research we plan to conduct as a way to apply this approach for the specification of high-value data-driven DSS.**

***Index Terms*—Data value, Data quality, Data utility, Decision making, Data selection, Decision Support System**

I. Introduction

The profusion of data available in organizations is clearly established and objectified. In 2020, an estimated 47 zettabytes (a thousand billions of gigabytes) of data has been produced all over the world [24]. By contrast, in 2010, that amount was only about 2 zettabytes. This exponential growth is expected to continue in the next 15 years to reach 2000 zettabytes in 2035 [24]. This inflation of accessible data has numerous – positive and negative – effects on many activities within an organisation [3]. This includes the process of supporting decision making, on which we focus our attention in this paper.

Every modern organisation likely generates significant quantity of data, needs to support decision makers who face more and more complex decision settings and dynamic environments, yet often struggles to provide them with timely and relevant decision-making support and notably data reporting. In practice, it is not uncommon for decision makers to face data sets that are too complex to be helpful, that have too many quality problems (missing values, encoding errors, etc.) to be useful, that are too isolated from other piece of data to generate real insight, etc. [11]. The purpose of so-called data-driven Decision Support Systems (DSS in the rest of this paper) is to mitigate those risks and provide smooth decision support to members of a company [22].

The proposition developed in this paper emerges in the context of Self-Service Business Intelligence (SSBI), a type of DSS where end-users have to select the pieces of data and visuals by themselves [1]. SSBI empowers end-users with the responsibility to produce dashboards by themselves, thereby reducing the time-to-delivery while improving alignment with business requirements. The question of selecting pieces of data to expose to the business users is central in DSS, and becomes even more critical in the case of SSBI [2]. Business-users indeed have little technical knowledge and little to no understanding of databases underlying a business application. As a result, the need to provide guidance and to help business users find what they really need in order to produce their own dashboards is significant.

This guidance is even more important considering three risks related to SSBI [17]; (i) dashboards may be overloading due to the presence of too much data, (ii) sub-optimal selection criteria may be applied to data by users, which may result in the omission of useful pieces of data or the inclusion of

This work was initiated in collaboration with the company Loth-Info and is partially financed by the Fonds Spécial de Recherche (FSR).
DOI: 10.18293/SEKE2022-104

irrelevant ones and (iii) bad quality data may be incorporated in the dashboard, resulting in reduced insights. All these risks involve spending resource and time on the implementation of unsupportive dashboards.

To address this problem, there is a need to find a criterion to select the most important pieces of data in the best possible way, in order to facilitate the process of designing SSBI dashboards and more globally DSS. This selection criterion should be understandable by business-users (it should not be too technical) yet incorporate important technical aspects whenever those aspects impact decision support. To the best of our knowledge, such data selection criterion has never been formalized in the literature on information management. As an answer, we advance our definition of a Decision-Making Data Value, or simply DMDV, criterion. This raises the following research questions:

1) What is *DMDV* in the context of data-driven Decision Support Systems and SSBI?
2) What are the different *dimensions* that compose our *DMDV* criterion?

In the remainder of this paper, we expand on these two questions. In section II, we position the concept of DMDV in the literature on software engineering in general. section III details our methodology to develop a taxonomy of relevant, distinct, measurable and comprehensive dimensions influencing DMDV. In section IV, we present this DMDV taxonomy. section V presents the future works. We conclude and present the limitations of this work in section VI.

II. Related Works

In this section, we review papers around three main topics: the concept of value in general, data quality and data value. We highlight some important observations to guide the reader.

A. Concept of Value

Before discussing the notion *data value*, it seems important to clarify our understanding of the notion of *value* since different definitions may be used depending on the situation [8], [20]. In [20], four main value situations are presented; (i) the exchange value that represents the amount of money needed to get a product, (ii) the esteem value that represents the price a customer is ready to pay for prestige or appearance, (iii) the use value that measures the functionalities of a product and (iv) the other value situations that group more particular situations such as the aesthetic value, the judicial value, the moral value or the religious value.

These definitions help us clarify what is meant by DMDV in the context of SSBI. Data is by definition a product, i.e. something that is produced. As a reminder, we focus in this paper on the value of data in the context of decision making within a company. We thus exclude operating or acquisition costs; for instance, we do not take into account the cost of purchasing the data from an external provider, neither do we account for the selling price of a data item. Similarly, we do not include in our DMDV the cost of collecting and recording the data, of maintaining a database or of using the data in a SSBI solution. Our goal is to support the selection of already available data for reporting purposes, not to help in determining if the use of a given piece of data is profitable for the company, i.e. if it generates more revenues than it generates costs. Our conceptualisation of value is thus not an exchange value. Neither the concept of "esteem value", nor the value situations (aesthetic, judicial, moral and religious) do relate to data aimed for internal decision making. However, the remaining concept (the use value) seems to be adequate to describe DMDV.

The concept of use value has been defined by Karl Marx in [19] when speaking about commodities. As defined by Marx, a commodity is "an object outside us, a thing that by its properties satisfies human wants of some sort or another". Based on this, he defines the use value of a commodity (citing [18]) as the "fitness to supply the necessities, or serve the conveniencies of human life". In [8], the use value is defined as "the specific qualities of the product perceived by customers in relation to their needs".

Based on these definitions, we can say that the decision-making value of data is a use value. Indeed, data may be seen as a thing that satisfies human wants, i.e. help for decision making. This brings us to define the decision-making value of a data item as: *its fitness to help decision making*. This brings us to the following observation:

> *Observation 1* - A data item is valuable if its use allows the decision maker to improve its decision-making.

B. Concept of Data Quality

The quality of a product or service likely influences its value. It is a commonsense observation in our daily lives, and this also applies in the specific context of data. Data quality has been discussed in the literature for a long time now, with many researchers trying to understand what data quality really represents and how it can be decomposed and/or measured. We can distinguish two important topics of research about data quality.

A first line of research is the *identification of the dimensions* that should be taken into account to decompose quality of data [6], [13], [15], [21], [26]. This issue has been found to be quite challenging and no standard emerged. A review of a large number of propositions has been conducted by [6]. It emphasized that some dimensions were more discussed than others in the literature, namely: accuracy, completeness, consistency, timeliness and currency. Moreover, the number of dimensions taken into account by authors varies quite a lot. While some authors derive only 5 dimensions from data quality, others decompose it into more than 15 dimensions.

> *Observation 2* - There is no strong agreement on the dimensions that influence the quality of data.

Another line of research *proposes a set of metrics* that may be used in order to assess the quality of data [6], [13], [15], i.e., to operationalize data quality. Various propositions have been made but no standard emerged. In [6], the authors also include this second aspect in their review of several papers about data quality. While most dimensions have different propositions of metrics, some of these propositions are more used than others [6]. For example, accuracy, completeness and consistency all have one specific proposition of metric that is significantly more discussed than the other propositions for the same dimension. This is not the case for timeliness and currency.

Observation 3 - Some dimensions have metrics suggested in the literature, others not.

C. Concept of Data Value

The concept of data value is more recent in the literature than data quality. The works presented below typically address data value in different ways.

In [12], the authors define the concept of intrinsic value of data to support data quality assessment. Based on the idea that quality assessment must be contextual, this value is measured starting from the records of the data. Each record receives a specific value depending on its frequency of use. More concretely, most accessed records receive a high value while less used records get a low value. These records values are then translated into weights to compute data quality metrics such as completeness or accuracy. The idea is that all records do not deserve the same attention from a data quality point of view. Most used records are more critical than less used records to assess data quality for a specific organisation. This approach offers a first step to take into account the value of the data. However, it does not discriminate this value across features of the same dataset. Hence, this approach might lead to focus on curating potential redundant and noisy features if their data quality is low for frequently used records. Another work [23] considers data value to optimize data management. For this purpose, the authors define the concept of disparity of data value and suggest ways to measure it. The value of a record in a dataset is attributed based on the context.

Observation 4 - Data value is context-dependent.

A specific line of research proposes general methodologies to derive the value of data. These methodologies can be based on human input [3], [10], [16], data processing [4], [7] or both [5]. This ability to materialize the value of data in concrete numbers is all the more important as it is a crucial ingredient for a sound data governance and more generally for any decision supported by data [3], [9]. What stands out in the literature is the fact that there is no unique and non overlapping definition of the constituents of data value [3], [16]. However, based on these methodologies, four observations may be derived for the selection of the constituents of data value:

Observation 5 - The selected dimensions should impact data value in the context of use [3].

Observation 6 - Redundancy among dimensions should be minimized [6].

Observation 7 - The selected dimensions should be measurable [3].

Observation 8 - The selected dimensions should take a maximum of data value aspects into account [7].

III. Methodology

Our DMDV taxonomy was built on a 4-steps methodology.

The first step consisted in gathering the dimensions of data value already identified in the literature. The aim here was not to conduct a systematic literature review but to gather the most discussed dimensions of data value. We searched for articles on the search engine Scopus using the following query ("data value dimensions" OR "data value assessment") with the search being performed on the title, abstract and keywords of the articles. We also included the literature review of [6] on the dimensions of data quality to the analyzed articles because data quality is highly related to data value. We then extracted the dimensions of the retrieved articles.

The aim of the second step was to create a taxonomy of all the retrieved data value dimensions that are applicable to the context of databases. In order to realize this, we first dropped the dimensions that were out of this scope. Then, we used an open card sorting approach performed by the 4 authors of the paper [28]. We shuffled all the retrieved dimensions randomly and each author classified the dimensions individually. We then discussed our results together to obtain a final data value taxonomy. The card sorting was performed by the authors because it required to understand each dimension in the context of data value and thus to have knowledge about the literature.

The third step was the selection of the dimensions to include in our DMDV taxonomy based on the global data value taxonomy. The observations 5, 6, 7 and 8 that have been identified in the previous section allowed us to guide this selection. In the remaining of this article, we respectively refer to these observations as: relevance criterion, distinctiveness criterion, measurability criterion and comprehensiveness criterion. First, the lowest-level classes of the data value taxonomy were screened for their relevance for decision making. Then, for each selected class, we only kept the dimensions that are measurable, i.e. the dimensions that already have measures proposed in the literature. Finally, we applied our comprehensiveness criterion by selecting the dimension that was assessed by the four authors as the most representative for each class based on a discussion and a vote. In other words, we kept the dimension that represents the broadest concept. This allowed us to select a global dimension that takes into account the information of the whole class. In this methodology, 3 of our criteria are explicitly used i.e. relevance, measurability and comprehensiveness, while the remaining criterion is implicit. Indeed, the distinctiveness criterion is also part of the method as we selected only one dimension by class. It is important to note that the comprehensiveness criterion is also taken into account at the beginning of the process as we started from all the dimensions retrieved from the literature.

In the final step, we validated our DMDV taxonomy with data experts, from both the scientific and the business worlds. The criterion to choose them was that they must have at least 5 years of experience in the field of data processing and notably reporting. We presented them our taxonomy during semistructured interviews and asked them to review this taxonomy according to three axes: (i) the identification of potential missing or superfluous dimensions; (ii) the identification of potential missing or superfluous classes and (iii) the identification of potential misclassified dimensions.

IV. Results

We now detail the application of our methodology to derive and validate a DMDV taxonomy from the literature.

A. Identification of the dimensions of data value

The dimensions retrieved in step 1 are presented in table I. An expected observation is that some dimensions are not applicable to the context of a database. For example, site access is a dimension directly related to the context of data available on a website and is thus not applicable to a database. We thus dropped these dimensions for step 2.

B. Data value taxonomy

The result of step 2 is presented in figures 1 and 2. The dimensions with an asterisk next to their name have a measure proposed in the reviewed literature (this will be used in the next subsection). We separate the data value taxonomy in two figures for clarity purpose. Our data value taxonomy classifies the retrieved dimensions in two main classes: data quality and data utility. Data quality encompasses the dimensions about the data itself and the way it is encoded (such as completeness or correctness) and is divided into 4 subclasses: completeness, correctness, technical aspect and time aspect. Data utility encompasses the dimensions about the use of the data (such as ease of operation and relevance) and is divided into 5 subclasses: ease of use, legal aspect, monetary aspect, uniqueness and usability. For clarification purposes, the subclass ”ease of use” encompasses the dimensions about the extent to which data may be used in an easy manner while the subclass ”usability” is about the goals that may be achieved with the data.

C. DMDV taxonomy

Our step 3 was applied in turn to data quality and data utility. The 4 subclasses of data quality dimensions displayed in figure 1 were thus assessed for their potential impact on decision making. As the subclass "technical aspects" does not directly relate to this purpose, it was consequently rejected. Then, following our step 2, only the dimensions having measures proposed in the literature were kept. Finally, the broadest dimension was selected for each subclass, leading to the selection of the 3 following dimensions, that we define in the context of decision-making: (i) Completeness: the extent to which the available data is complete, are there enough values or is the data empty? (ii) Correctness: the extent to which the available data contains errors, can we believe what is encoded? (iii) Timeliness: the extent to which the available data is up-to-date, are the values still valid?

Turning now to the utility factor, the relevance of subclasses for internal decision making was checked. Monetary aspect and legal aspect do no contribute to this goal and they were accordingly discarded. Indeed, these 2 subclasses impact the ease to get and to use data but, in this work, we focus on data already available and usable for the organisation. For the 3 remaining subclasses, we kept the following dimensions that are assessed as measurable by the literature and having the broadest scope of their subclass, and we define them in the context of decision-making: (i) Interpretability: the extent to which the available data may be interpreted, do we understand what is encoded? (ii) Uniqueness: the extent to which the information embedded in the available data is unique, as several data items may contain the same information, do we already possess this information? (iii) Usability: the extent to which the available data contains useful information for decision making (this takes into account the current usage of data and the future objectives that could impact how this data is used), is the data used for decision making?

Integrating the results of this double application of the dimension selection process on both data quality and data utility, our DMDV taxonomy is presented in figure 3.

D. Validation of the DMDV taxonomy

In order to validate our DMDV taxonomy, we realized interviews until we reached a saturation threshold in the answers. This led us to conduct 7 interviews of both researchers and practitioners (3 researchers in the field of information management, 3 IT consultants and 1 data manager).

For the first axis, "identification of potential missing or superfluous dimensions", the respondents identified 3 potential missing dimensions: granularity (cited four times), the ability to be visualized (cited three times) and quantity (cited one time). One respondent also suggested a division of interpretability into format and meaning. Granularity and the ability to be visualized are in fact encompassed respectively in our dimensions usability and interpretability. Indeed, the granularity of the data directly impacts its possible usages and the ease to visualize the data impacts its interpretability. We thus revise our definitions of usability and interpretability to better express these aspects. Usability is "the extent to which the available data contains useful information and has the right level of granularity to be used for decision making". Interpretability is "the extent to which the available data may be interpreted and notably visualized". Quantity is not identified in the literature as a data value dimension and we argue that it is more an element that determines the need to find a data selection criterion than a DMDV dimension. We also argue that interpretability does not need to be divided at this point due to our comprehensiveness criterion and that this division should be kept in mind for the eventual design of an interpretability metric. Our dimensions are thus validated.

The second axis, "identification of potential missing or superfluous classes", only generated one comment as all the respondents except one completely agreed with the classes quality and utility. One participant suggested that usability could be a third class between quality and utility based on the cognitive process she follows when designing a dashboard.

TABLE I

RETRIEVED DIMENSIONS OF DATA VALUE IN THE LITERATURE

Source	Dimensions
Batini et al. [6]	accuracy, completeness, consistency, timeliness, currency, volatility, uniqueness, appropriate amount of data, accessibility, credibility, interpretability, usability, derivation integrity, conciseness, maintainability, applicability, convenience, speed, comprehensiveness, clarity, traceability, security, correctness, objectivity, relevance, reputation, ease of operation, interactivity
Brennan et al. [9]	usage, cost, quality, intrinsic, IT operations, contextual, utility
Brennan et al. [10]	operational impact/utility, dataset replacement costs, competitive advantage, regulatory risk, timeliness
Wang et al. [27]	content, credibility, critical thinking, copyright, citation, continuity, censorship, connectivity, comparability, context, site access and availability, resource identification and documentation, author identity, author authority, information structure and design, content relevance and scope, content effectiveness, accuracy and balance of content, navigation within documents, link quality, aesthetic and emotional aspects, information source, scope, discussion, technology factors, text format, information organization, price, availability, user support system, authority, credibility, accuracy, reasonableness, support, timeliness, integrity, consistency, acquisition cost
Attard et al. [5]	usage, quality, data, infrastructure
Holst et al. [14]	usage, quality, monetization, data sourcing costs, data processing and analysis needs, importance for business model and decisions
Stein et al. [25]	usage, quality, costs, completeness, conciseness, relevance, correctness, reliability, accuracy, precision, granularity, currency, timeliness
Bendechache et al. [7]	volume, usage, utility, replacement cost, legislative risk, timeliness, competitive advantage, quality, security

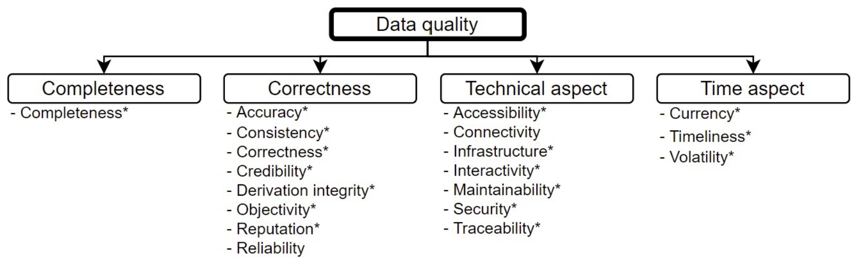

Fig. 1. Data quality part of the data value taxonomy

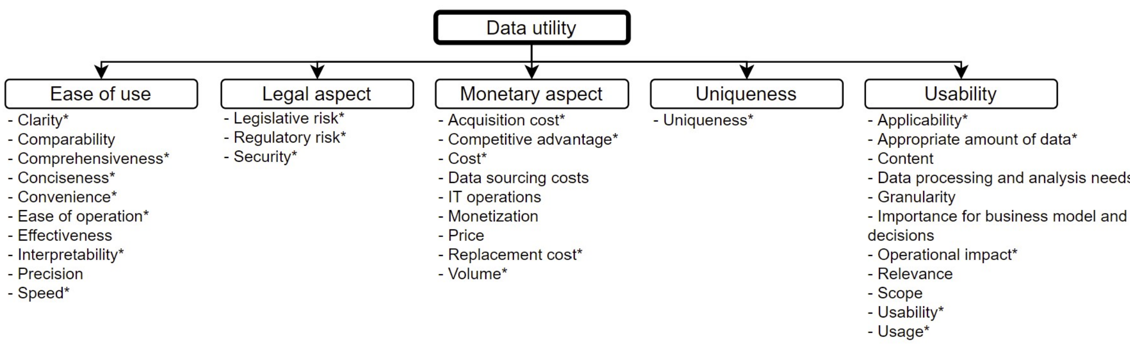

Fig. 2. Data utility part of the data value taxonomy

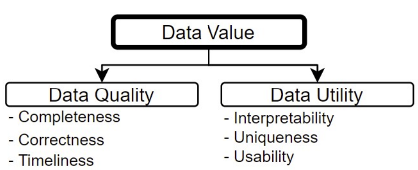

Fig. 3. DMDV taxonomy after dimensions selection

This respondent however recognized that usability may also be included in utility, so that we do not feel the need to update our taxonomy. Our two classes are thus validated.

Our final axis, "identification of potential misclassified dimensions", did not generate any comment as all the respondents agreed with the way the dimensions are classified. Our classification of our dimensions in our classes, and consequently our DMDV taxonomy, are thus validated.

V. FUTURE WORKS

This work, by proposing the concept of Decision-Making Data Value as data selection criterion and building a DMDV taxonomy, is a first necessary step towards a main objective: supporting data selection for decision making. To complete our work and achieve this objective, a lot of future works may be considered. This section discusses the main ones.

A first way to extend our work is to develop a DMDV assessment framework based on our taxonomy. For this purpose, a metric should be developed for each dimension of our taxonomy and a way to aggregate these metrics in a DMDV indicator should be proposed. This indicator of DMDV could be designed at different levels of granularity (e.g. the column level, the database level,...). This would for example allow to rank columns in terms of importance for reporting.

Then, the next step would be to test the proposed framework in real-world situations. More specifically, the choice of the metrics should be tested and adjusted if needed. In order to do this, it would be interesting to develop use cases and to apply the framework to test how it performs in comparison with the assessment of a data specialist. This would also allow to detect some particular cases that are not taken into account.

Finally, a final step would be to integrate our DMDV concept into a data-driven DSS design process. Indeed, this would allow to develop an integrated DSS design process guiding the user in the data selection. This can be quite challenging because it requires to center the data selection part of the DSS design process around the concept of DMDV. This process could then be tested with end-users to discover all the practical possibilities offered by our proposition in a data-driven DSS design process.

VI. Conclusion

In this paper, we tackle the problem of finding a data selection criterion for decision-making support. Organizations have so much data that it becomes nearly impossible for decision makers to intuitively select the most important ones. To address this problem, we suggest to use the decision-making value of data (DMDV) as data selection criterion. We thus define the concept of DMDV and develop a taxonomy of the dimensions having an impact on this concept. For this purpose, we first create a global data value taxonomy from which we derive our DMDV taxonomy, based on four criteria: relevance, distinctiveness, measurability and comprehensiveness. Our taxonomy decompose DMDV dimensions into two classes: data quality and data utility. We present data quality as the combination of completeness, correctness and timeliness dimensions, while data utility is composed of interpretability, uniqueness and usability dimensions. We conduct several interviews to validate our taxonomy and discuss the results. We also elaborate on the future works.

In terms of limitation, even if we try to objectify the selection of dimensions as much as possible, the application of our criteria may include a small part of subjectivity when selecting the final dimension to keep for each subclass. However, these criteria allow to find a set of dimensions that are relevant, distinct, measurable and comprehensive. This means that, even if an other dimension is selected for a particular subclass, its characteristics and meaning should be very similar, resulting in an equivalent set of dimensions.

References

[1] P. Alpar and M. Schulz. Self-service business intelligence. *Business & Information Systems Engineering*, 58(2):151–155, 2016.

[2] D. Arnott and G. Pervan. A critical analysis of decision support systems research revisited: the rise of design science. In *Enacting Research Methods in Information Systems*, pages 43–103. Springer, 2016.

[3] J. Attard and R. Brennan. Challenges in value-driven data governance. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*, pages 546–554. Springer, 2018.

[4] J. Attard and R. Brennan. A semantic data value vocabulary supporting data value assessment and measurement integration. 2018.

[5] J. Attard, J. Debattista, and R. Brennan. Saffron: a data value assessment tool for quantifying the value of data assets. 2019.

[6] C. Batini, C. Cappiello, C. Francalanci, and A. Maurino. Methodologies for data quality assessment and improvement. *ACM computing surveys (CSUR)*, 41(3):1–52, 2009.

[7] M. Bendechache, N. Sudhanshu Limaye, and R. Brennan. Towards an automatic data value analysis method for relational databases. 2020.

[8] C. Bowman and V. Ambrosini. Value creation versus value capture: towards a coherent definition of value in strategy. *British journal of management*, 11(1):1–15, 2000.

[9] R. Brennan, J. Attard, and M. Helfert. Management of data value chains, a value monitoring capability maturity model. 2018.

[10] R. Brennan, J. Attard, P. Petkov, T. Nagle, and M. Helfert. Exploring data value assessment: a survey method and investigation of the perceived relative importance of data value dimensions. In *ICEIS 2019-21st International Conference on Enterprise Information Systems*, pages 200–207. SciTePress, 2019.

[11] L. Cai and Y. Zhu. The challenges of data quality and data quality assessment in the big data era. *Data science journal*, 14, 2015.

[12] A. Even and G. Shankaranarayanan. Value-driven data quality assessment. In *ICIQ*, 2005.

[13] A. Even and G. Shankaranarayanan. Dual assessment of data quality in customer databases. *Journal of Data and Information Quality (JDIQ)*, 1(3):1–29, 2009.

[14] L. Holst, V. Stich, G. Schuh, and J. Frank. Towards a comparative data value assessment framework for smart product service systems. In *IFIP International Conference on Advances in Production Management Systems*, pages 330–337. Springer, 2020.

[15] M. Kaiser, M. Klier, and B. Heinrich. How to measure data quality?-a metric-based approach. *ICIS 2007 Proceedings*, page 108, 2007.

[16] K. Kannan, R. Ananthanarayanan, and S. Mehta. What is my data worth? from data properties to data value. *arXiv preprint arXiv:1811.04665*, 2018.

[17] C. Lennerholt, J. Van Laere, and E. Söderström. User-related challenges of self-service business intelligence. *Information Systems Management*, 38(4):309–323, 2021.

[18] J. Locke and W. Engels. *Some Considerations of the Consequences of the Lowering of Interest, and Raising the Value of Money*. Verlag Wirtschaft und Finanzen, 2018.

[19] K. Marx. *Le capital*. Librarie du Progres, 1875.

[20] H. S. Neap and T. Celik. Value of a product: A definition. *International Journal of Value-Based Management*, 12(2):181–191, 1999.

[21] L. L. Pipino, Y. W. Lee, and R. Y. Wang. Data quality assessment. *Communications of the ACM*, 45(4):211–218, 2002.

[22] D. J. Power. Decision support systems: a historical overview. In *Handbook on decision support systems 1*, pages 121–140. Springer, 2008.

[23] G. Shankaranarayanan, A. Even, and P. D. Berger. Optimizing data management with disparities in data value. *Journal of International Technology and Information Management*, 24(3):1, 2015.

[24] Statista. Le big bang du big data, 2019.

[25] H. Stein, L. Holst, V. Stich, and W. Maass. From qualitative to quantitative data valuation in manufacturing companies. In *IFIP International Conference on Advances in Production Management Systems*, pages 172–180. Springer, 2021.

[26] R. Y. Wang and D. M. Strong. Beyond accuracy: What data quality means to data consumers. *Journal of management information systems*, 12(4):5–33, 1996.

[27] X. Wang, C. Dong, W. Zeng, Z. Xu, and J. Zhang. Survey of data value evaluation methods based on open source scientific and technological information. In *International Conference of Pioneering Computer Scientists, Engineers and Educators*, pages 172–185. Springer, 2019.

[28] J. R. Wood and L. E. Wood. Card sorting: current practices and beyond. *Journal of Usability Studies*, 4(1):1–6, 2008.