A Conceptual Model of Benchmarking Data and its Implications for Data Mapping in the Data Economy

Markus Spiekermann[1], Sven Wenzel[1], Boris Otto[1,2]

[1]Fraunhofer Institute for Software and Systems Engineering ISST, Dortmund, Germany
{markus.spiekermann,sven.wenzel,boris.otto}@isst.fraunhofer.de
[2]TU Dortmund, Chair of Supply Net Order Management, Dortmund, Germany
{boris.otto}@tu-dortmund.de

Abstract. Digitalization of the economy requires enterprises from all industries to revisit their current business models and prepare their organizations for the digital age. One task is the (re-)design of hybrid and digital products and services. The foundation builds the improved interchangeability of data and the availability of external data sources through data markets and platforms. This leads to the requirement of a structured decision-making while mapping data sources to digital products. In order to successfully transform their business and develop valuable new products, companies require methodological help. This paper proposes a high-level conceptual model for the assessment of data sources value. It consists of an approach for comparing data sources based on a common description of data and individual metrics definition enable a benchmark process. The development of the model and its practicability has been validated in a case study with an industrial partner.

Keywords: Data Economy, Data Assets, Data Valuation, Digital Transformation, Data Mapping

1 Introduction

The digital transformation challenges companies to review their business models and consider their digitalization by developing hybrid or purely digital products. Thereby, the foundational resource is data. While data, as a corporate resource seems to grow to infinity from the perspective of volume, variety and velocity, companies face challenges on the technical as well as on the organizational side. Data becomes a product itself [1, 2], and since the improved possibilities to exchange data in a secure and sovereign way, the amount of external data sources extends the internal available pool of master data, operational transaction data and other self-created data sources. On the organizational side, challenges are reflected by the digital transformation of business models and business strategies [3–5] which are strongly related to the topics of data valuation and monetization [1, 2, 6]. On the technological side integration of data sources, improving and controlling the data quality, provide traceability and provide access to possible data citizens within the company presents the greatest

Multikonferenz Wirtschaftsinformatik 2018,
March 06-09, 2018,Lüneburg, Germany

challenges to data management [5, 7–9]. To solve these challenges, a tight integration of IT and business is required before companies can benefit from the vast amount of data. Structures that welcome data-driven business engineering and mapping the right data to their digital products have to be created. This paper deals with the research question on how to evaluate the most economical data source for a given business model in times of digital economy.

Research in the fields of digital transformation and the assessment of data or information has existed separately so far. This paper suggested a "data benchmark" approach to provide a concept for data mapping activities within the creation of digital business models, combine both topics, and improve practicability.

The goal is a theoretically sound and practically useful approach that may help companies to address the challenges of mapping data sources to their digital products and structure the decision-making by providing an applicable model.

Methodologically, the paper belongs to the domain of design science research (DSR), which strives at developing "IT artifacts intended to solve identified organizational problems" [10]. The model is developed based on a review of scientific literature and on ongoing action design research. The contribution of the proposed conceptual model for the academic community lies in getting deeper insights of challenges in the future business engineering and the derivation of methods to manage data resources. For practitioners, the model helps to implement a structured process of data assessment for mapping and selection of data sources while the company's digital transformation.

The remainder of the paper is structured as follows. Section 2 establishes the necessary background on research in the fields of digital economy and the changing role of data and its valuation. Section 3 presents the current version of the data-benchmark methodology and the corresponding conceptual approach. Section 4 then shows a real-life case to demonstrate the practicability and the evaluation of the method. Section 5 discusses the theoretical and managerial implications and limitations of the research. Section 6 concludes the paper and gives an outlook on future research.

2 Background

The theoretical background for this paper is given by two fields of research. First, the paper gives the necessary background for the topic of digital business and especially the activity of data mapping in the data economy. Second, the paper describes the changing role of data and the current situation of determining an economic value.

2.1 Digital Business

Digital Business is experiencing a renaissance at present. Initially coined in the 1990s, it is today used in broader context [3]. Traditionally the understanding of digital business was very much influenced by the debate around treating information as an enterprise asset [11–13]. This perspective on digital assets acknowledged the growing importance that data plays for enterprises business. Digitalization as the term of today's

transformation and advancement of company's business model considers the opportunities and challenges, data and digital goods bring to business models. Research and practice broadly discuss the new Digital Business and the effects on business strategies and models. Examples are the MIT Center for Digital Business [14], the Digital Business Transformation at the University of St. Gallen [15] and the Industrial Data Space as consortium of Fraunhofer Society and the Industrial Data Space Association to develop an architecture for a digital ecosystem [16]. One design principle that is fundamental of digital business models is consumer-centricity and individualization, which is closely related to the independence of time and place by providing multi-channel integration [15]. Business models in the digital economy [3, 5, 14] are characterized by an evolvement of products into hybrid or purely digital services. The close integration of IT and physical products in combination with the vast amount of available data enables smart and data-driven service offerings for traditional products as well as new innovations to gain more added value around the core product [17].

One key action within the design of data-driven business is the data mapping, "making sure that the business objects required in the end-to-end customer process are transparent and that corresponding data objects are identified and described" [3]. Since the vast amount of available internal as well as external data sources, an economical selection of data is business critical for digital services. The economical selection of the right data sources gets harder with the growing volume and indeed cannot be done intuitive without a structured process in the age of digitalization and data economy [18, 19].

2.2 Data Assets and their Valuation

Information as a resource has been researched over a longer period. First approaches were made in the concepts of Data Resource Management, where influencing factors for information resources are described [20]. Literature in the 1990s suggested that information should be viewed as intangible assets [11, 21, 22], motivated by the development of information becoming an important corporate resource. While in the 1960s and 1970s data was the result of business processes, it enables a company-wide business process management in the 1980s and 1990s. Since the turn of the millennium, data plays a key role in enabling products and (digital) services, up to the emerging of data as a stand-alone product [16].

The management of these resources is still insufficient compared to other intangible as well as tangible resources [6, 23]. A direct transfer of known and established methods for managing other tangible as well as intangible assets to the data domain is problematic, because of differences in the characteristics of data. As an example, compared to other assets, data has no abrasion, is infinitely shareable and the value increases with its use [6, 12, 13]. While available research considers information assets, we propose to distinguish between information and data, and consider data as the building blocks of information. Consequently, this paper suggests to value data as the raw material for creating digital business instead of information. We assume that data plays an important role as an input for different business processes and has an important

impact on the success of businesses [24], especially in the data economy where exchange of data resources between different companies gets more important to create products and provide better services. This makes the product data an important asset of the future company.

Valuation approaches were researched by considering the use value, the cost-of production value and the exchange value of information [25–28].

Despite of the research done in the past years, there is still a lack of practical concepts and methodologies for valuing information assets [2, 29]. Nevertheless current research shows the general possibility to value data as assets [2, 29]. Authors propose to apply existing valuation concepts (e.g. the cost-based approach, the market-based approach, and the income-based approach), without elaborate on new or modified approaches for the data valuation [29].

The cost-based approach quantifies either the cost of sourcing, exchange or reproducing data and managing it through the entire data lifecycle, by consider costs for data quality management and maintenance [2, 26, 28]. The approach of determine a use-based value quantifies the reduced uncertainty in decision-making [25, 27]. The basic idea is to determine the value of the business opportunities a company gains from using data in different business processes [29].

In the market-based approach, data assets are measured by observing the market value on competitive, active markets. The value of the data is then derived from the existing market demand. While the market-based approach revealed strong limitations for being used in the context of data valuation due to missing markets, this changes within the data economy, where platforms and marketplaces being established and can be used to compare different data products [16, 30]. . Comparable products mean that the value has to be adjusted concerning differences in for example data quality, volume, or content. A practical approach for the valuation of data sources would enable an economic data mapping for digital businesses [3]. In addition, it could provide transparency for the trade in data goods on arising data marketplaces and platforms [30, 31] and therefore implicates sourcing of external data sources to meet their business requirements.

In order to find the most economical data source, companies need to make a structured decision about which data they want to buy, based on key indicators and their metrics.

3 Research Approach

This research aims at developing a conceptual model that outlines a high-level process to structure the selection and mapping of data sources to digital business models from an economical perspective. A conceptual model therefore specifies the generally valid elements that are representative for a real system. The conceptual model can serve as a reference for designing company-specific models [32]. We designed the resulting artifact in a research project with a large pharmaceutical company since September 2016, following the Action Design Research (ADR) methodology. ADR combines design science (DS) and action research (AR) and constitutes “a research method for

generating prescriptive design knowledge through building and evaluating ensemble IT artifacts in an organizational setting" [33].

After an initial discussion in September 2016 with industry partners about the current lack of a concept for an economic data mapping in digital business, the first phase considers the corresponding requirements for such an approach (*Problem Formulation*). Experienced data managers, that identified the need of action, initiated the research activities. Although these data managers are aware of the value, data provides to their business users, they lacking guidance for the data-related challenge of determining a specific economic value for a KPI-based decision-making. As suggested by Sein et al. [33], the first BIE cycle aims at developing an early alpha version of the artifact, which will be instantiated, repeatedly tested, and continuously refined in a second cycle. In the first cycle, the emerging artifact based on an implementation of a proof of concept, comparing a list of attributes of data sources with a scoring model, presented to a focus group end of September 2016 as depicted in **Table 1**.

Table 1. Participants of initial focus group

Date	*Location*	*Focus Topics*	*Participants*
2016-09	Duisburg (D)	Business in the data economy, value of data, innovation and digital transformation	9 participants from two companies including experienced data managers, architects and scientists

The discussion results, described in more detail in section 4.1, formed the basis for further adjustments of the artifact. The concept was repeatedly discussed and refined (*Building, Intervention, and Evaluation*) in a case study between November 2016 and July 2017. The first and second session focused on the general structure and procedure, while a further session considered practicability and visualization of the conceptual model.

In addition to the focus group discussions, we conducted a software-based evaluation started in March 2017, where we implemented the alpha version of the conceptual model as a usable software-tool for practitioners. The results of the evaluation are presented in section 5. In future BIE cycles, we will instantiate the artifact in selected companies. Based on the interventions and evaluations from these cases, a beta version of the artifact will be developed (*Reflection and Learning*). The research team has initiated the activities of this cycle and we are currently applying the conceptual model in several companies to ensure that learning from our company-specific instantiations are further developed and documented (*Formalization of Learning*) as a general solution.

4 The Data Benchmark Model – Conceptual Model for Data Valuation in the Digital Economy

4.1 Purpose and Guiding Requirements

The conceptual model of data benchmarking aims at structuring the decision-making for mapping data sources to data-driven products and digital business models. It addresses the challenges of the data economy in terms of purchasing and economic selection of external data. External data extends the existing internal data pool and has to be purchased from data marketplaces or platforms to improve company's products and services. Identifying these data needs of business requires – apart from the technical capabilities – close alignment between data management, sourcing and the data citizens as the consumer of data driven insights. Therefore, a concept of determine a value for data sources, that enables a structured decision-making and mapping the most economic data sources, is required. The concept represents the real system on a high-level model guided by identified key requirements and the mentioned research question. We derived these requirements and elements from the initial focus group discussion and a subsequent literature research following Webster and Watson [34]. We searched for "data asset", "information asset", "data assessment", and "data valuation" on Scopus, as well as Google Scholar database to retrieve our information. The key elements of the pursued conceptual model are summarized in Table 2.

Within the data economy, data is business critical and has to be managed as strategic asset within the companies that has an economic value (1) [1, 2, 6, 11]. Furthermore, data assets have attributes that can describe the characteristics of data or data sources (2) [20, 35, 36]. Attributes of data can be categorized and structured in a metadata model to unify a description (3) [36]. With economic, qualitative, technologic, and competitive we identified four categories of attributes. While economic attributes e.g. refer to the price, license and ownership of the data source, qualitative attributes refer to the completeness, timeliness, accuracy, or integrity. The technologic category sums up attributes like format, type, or availability, while the category of competitive attributes refer to accessibility, novelty, and publicity level. On base of the identified attributes, metrics can be defined that enable an assessment and valuation (4). Metrics have to be individualizable by requirements of the specific use cases (5) [29]. Benchmarking methodology for data sources with a peer group leads to a structured decision-support for data mapping (6).

Based on the identified requirements and taking into account the surrounding environment (technology and organization), we designed the alpha version of our artifact.

Table 2. Requirements for data mapping using a data benchmark approach

No.	*Requirement*	*Description*
(1)	Data is a corporate asset managed as a stand-alone product.	The role of data has changed and data is traded as a stand-alone commodity on data marketplaces and data spaces.
(2)	Data is described by its attributes.	Data has attributes and characteristics that describes the data. Attributes are related to data sources, data sets or to single values.
(3)	Attributes can be used to form a metadata model for common description.	The overall attributes and characteristics of data sources can be organized with a metadata model that serves as a standard description template and group attributes in different categories (e.g. economical, technological, qualitative, competitive)
(4)	Metrics can be defined to determine an economic value.	By creating metrics and indicators on base of the data's attributes, an economic value can be determined.
(5)	Specific use cases influence the value based on their requirements	Specific business use cases influence the requirements placed on the data, which has to be considered when selecting a data source to a targeted business model.
(6)	Metrics enable a comparison of data with peers	The combination of a unified description and the defined metrics enable a benchmark of a data source within its peer group.

4.2 Design Process

Given the understanding of data as a strategic resource for the digital economy, the structure of the conceptual model builds on established procedures of benchmarking.

Benchmarking is as primarily a tool for improvement, achieved through comparison. It identifies the highest standards of excellence for products, services, or processes [37]. This paper proposes an adaption of the benchmarking wheel [38], which is well suited to the dynamic changes in the data ecosystem, to develop a product benchmarking concept for data sources. It follows the structure of the five steps proposed by the benchmarking wheel to enable a comparison between data sources:

1. *Plan*: Determine what to benchmark
2. *Find*: Identify benchmarking partners
3. *Collect*: Understand and document the benchmarking partners' information and attributes
4. *Analyze*: Analyze benchmarking information due metrics and indicators

5. *Improve*: Use results to take action and improve for example the decision-making

To design a reusable artifact regarding the ADR methodology we took the results of our focus group discussions and literature review. We adapt the identified requirements of the considered use case of data mapping in digital business to develop a conceptual model. Therefore, we based our approach on the identified steps of a benchmarking process and then assign the necessary actions regarding a data source comparison as depicted in **Figure 1**.

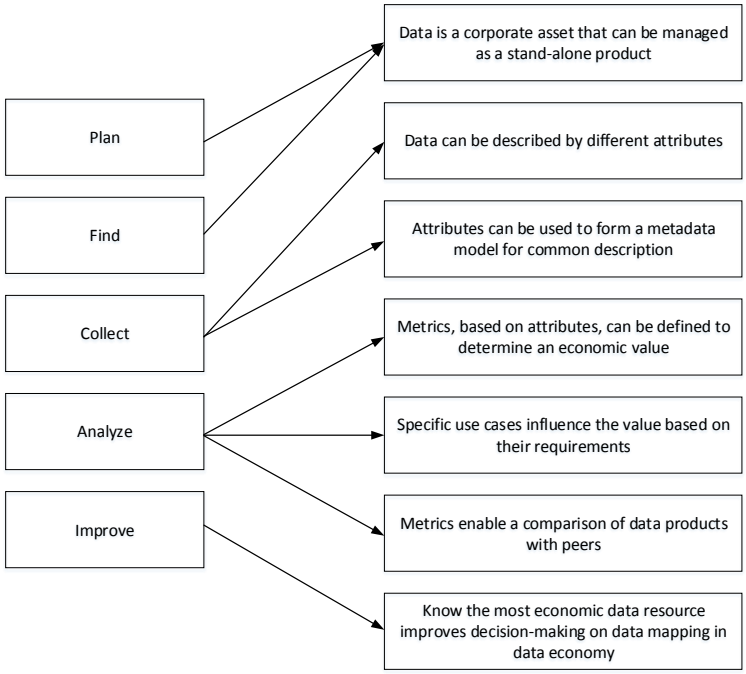

Figure 1. Assignment of data mapping requirements to the benchmarking process

Our conceptual model describes the benchmarking of data sources, resulting in the data sources to be the main object of consideration in the *"Plan"* phase of the benchmarking process. Since data becomes a stand-alone commodity in the digital ecosystem, data sources, as our peers for the benchmark, can be found on data marketplaces and (open) data platforms, which refers to the "*Find*" phase. To enable a benchmarking, available information of data sources have to be *"Collected"* and stored in a common information model. As our metadata model, we used the Metadata Model for Data Goods (M4DG)[1].

The phase *"Analyze"* in our approach is based on individualizable metrics that can be defined out of the attributes of the data sources regarding the information model. Regarding the identified requirements of the focus group discussions, we take into

[1] Metadata Model for Data Goods (M4DG) is a research project of the Fraunhofer ISST to develop a standardized model for data assets.

account that the specific business needs can influence the metrics while mapping data sources. The final *"Improve"* phase is about a comparison of data sources using the defined metrics. This finally leads to a determination of a data sources value and supports a structured decision-making, which data source to select.

As a final step in our design process, we concentrated on the graphical visualization to depict the key elements, relations and process flow of our conceptual model. The activities of the design process result in the conceptual model, depicted in **Figure 2**. The model is company-unspecific and intended to be reused in a design process of an information system.

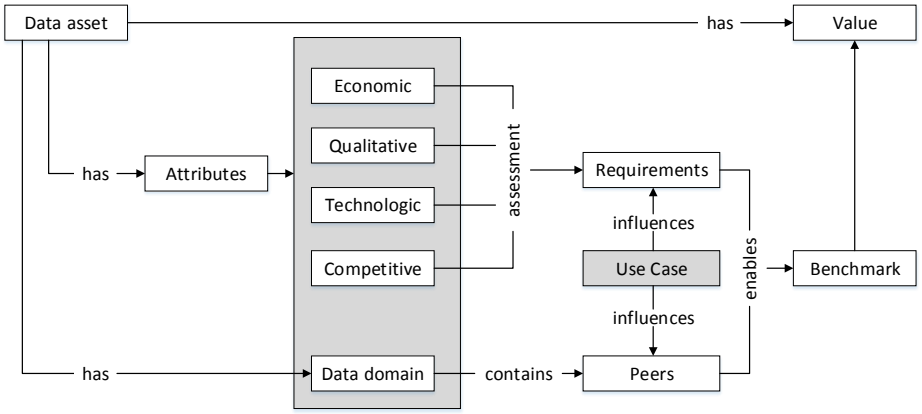

Figure 2. Conceptual model for data benchmarking

5 Evaluation

"Action research is the application and testing of ideas developed in an academic environment in real world situations under participation of the researching individuals" [39]. Therefore, we tested the validity of the designed conceptual model, the benchmarking approach, and the key elements in a case study with our industrial partner. We implemented the designed concept in a practical information system and presented the application to data management experts and data citizens, who need to map data to their digital business, for an evaluation. The information system, a web-based application, allows collecting data sources, searching for data sources, evaluating quality attributes, and defining metrics to be used for a comparison.

After a presentation of the conceptual model and the information system to different business areas, participants were asked to evaluate the structure (i.e. the completeness, simplicity, or clarity), the adaptability (i.e. robustness, practicability), and the environmental fit (i.e. utility, personal and organizational fit) of our approach.

We got a positive feedback for the chosen structure of our model. The idea behind our approach was straightforward and clearly understandable. Participants notice, that defining roles and responsibilities would make the model more complete in a next step. Regarding the environmental fit, the feedback was also very positive. The participants

think it is very good that the topic of data mapping is examined and solutions are sought. Even when the problem is not yet critical for many business users, the view of future necessity in data economy is shared for personal as well as organizational interest. The problem relates to every business unit we presented our approach and our solution was applicable in all considered cases, since all plan or actually do purchase external data sources for their business processes. This leads to the assumption of a positive adaptability of our concept having in mind, that a high effort of collect the required information and connect to platforms and marketplaces is necessary. That is why we get the feedback that the collection of data sources and its information could be time consuming and should be automatized as much as possible, in terms of the practicability of our approach.

6 Conclusion and Outlook

Within the data economy and the requirement to purchase external data this topic gets even more important to the practitioners. This paper presents a conceptual model for a data benchmark to meet the challenges in mapping the most economic data sources while designing digital business models. The emerging model was systematically developed following the four steps proposed by Action Design Research. The conceptual model reflects the key elements, assumptions and requirements related to the identified real life problems. These requirements were collected through discussions with experienced data architects, data managers, and data citizens. Because we see data as a stand-alone commodity of future business, we then designed a conceptual model based on the process of a product benchmarking reflecting data sources as products within the data economy. Using a common description for data sources and the definition of metrics related to the data sources, our concept enables a comparison between a data source and its peers to determine a value.

The evaluation of the alpha version of the artifact as well as a first prototype implementation have demonstrated its utility and organizational fit. Hereby limitations stem from the research context as only industrial partner was involved while the first evaluation phase. Despite being a conceptual model that strives for a transferable and company-unspecific degree as a scientific artifact [40], we will need more empirical testing to evaluate and complement our concept. The idea behind our concept was well received by the data managers as well as data citizens we presented our approach within the case study.

Further research activities are currently ongoing to instantiate the artifact in further companies and refine the concept in another BIE cycle. Planned results of these research activities include instantiations as well as using the conceptual model as basis to develop a framework for an economic valuation of data as corporate resource.

7 Acknowledgments

This work has been funded by the German Federal Ministry of Education and Research (BMBF) (project InDaSpace, grant no. 01IS15054).

References

1. Wang, R. Y., Lee, Y., Pipino, L., Strong, D.: Managing your information as a product. Sloan Manag. Rev. 95–106 (1998).
2. Otto, B.: Quality and Value of the Data Resource in Large Enterprises. Inf. Syst. Manag. 32, 234–251 (2015).
3. Otto, B., Bärenfänger, R., Steinbuß, S.: Digital Business Engineering: Methodological Foundations and First Experiences from the Field. Proc. 28th Bled eConference. 58–76 (2015).
4. Brown, B., Sikes, J.: Minding your digital business: McKinsey Global Survey results. McKinsey Bus. Technol. 27, 24–31 (2012).
5. Zimmermann, H.: Understanding the Digital Economy : Challenges for new Business Models. AMCIS 2000 Proc. Paper 402, 729–732 (2000).
6. Moody, D., Walsh, P.: Measuring The Value Of Information: An Asset Valuation Approach. Seventh Eur. Conf. Inf. Syst. 1–17 (1999).
7. Agrawal, D., Bernstein, P., Bertino, E.: Challenges and Opportunities with Big Data. Proc. VLDB Endow. 5, 1–15 (2012).
8. Labrinidis, A., Jagadish, H. V.: Challenges and opportunities with big data. Proc. VLDB Endow. 5, 2032–2033 (2012).
9. Ammu, N., Irfanuddin, M.: Big Data Challenges. Int. J. Adv. Trends Comput. Sci. Eng. 2, 613–615 (2013).
10. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design Science in Information Systems Research. MIS Q. 28, 75–105 (2004).
11. Horne, N.W.: Information as an asset—The board agenda. Comput. Audit Updat. (1995).
12. Oppenheim, C., Stenson, J., Wilson, R.M.S.: The attributes of information as an asset. New Libr. World. 102, 458–464 (2001).
13. Stenson, J.: The Attributes of Information As an Asset, (2006).
14. The MIT Center for Digital Business: MIT initiative on the digital economy, http://ide.mit.edu/.
15. Leimeister, J.M., Österle, H., Alter, S.: Digital services for consumers. Electron. Mark. 24, 255–258 (2014).
16. Otto, B., Auer, S., Cirullies, J., Jürjens, J., Menz, N., Schon, J., Wenzel, S.: Industrial Data Space: Digital Sovereignty over Data. (2016).
17. Yoo, Y., Henfridsson, O., Lyytinen, K.: Research Commentary—The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research. Inf. Syst. Res. 21, 724–735 (2010).
18. Newman, D.: How to Plan, Participate and Prosper in the Data Economy. Gartner, Stamford, CT. (2011).
19. Dennedy, M.F., Fox, J., Finneran, T.R.: Value and Metrics for Data Assets. In: The Privacy Engineer's Manifesto: Getting from Policy to Code to QA to Value. pp. 279–298. Apress, Berkeley, CA (2014).
20. Goodhue, D.L., Quillard, J. a, Rockart, J.F.: Managing the Data Resource: A Contingency Perspective. MIS Q. 12, 373–392 (1988).

21. Glazer, R.: Measuring the Value of Information: The Information-Intensive Organization. IBM Syst. J. 32, 99–110 (1993).
22. Fisher, T.: The Data Asset: How Smart Companies Govern Their Data for Business Success. John Wiley & Sons, Inc., Hoboken, New Jersey (2009).
23. Oppenheim, C., Stenson, J.: Studies on information as an asset I: defnitions. J. Inf. Sci. 29, 159–166 (2003).
24. Wilson, R.M.S., Stenson, J.A.: Valuation of information assets on the balance sheet. Bus. Inf. Rev. 25, 167–182 (2008).
25. Hirshleifer, J.: Where Are We in the Theory on Information? Am. Econ. Rev. 63, 31–39 (1973).
26. Hilton, R.W.: The Determinants of Information Value: Synthesizing Some General Results. Manage. Sci. 27, 57–64 (1981).
27. Stigler, G.J.: The Economics of Information, (1961).
28. Repo, A.J.: The value of information: Approaches in economics, accounting, and management science. J. Am. Soc. Inf. Sci. 40, 68–85 (1989).
29. Zechmann, A.: Assessing the Economic Value of Data Assets. Work Rep. (2016).
30. Schomm, F., Stahl, F., Vossen, G.: Marketplaces for Data: An Initial Survey. SIGMOD Rec. 42, 15–26 (2013).
31. Heckman, J., Peters, E., Kurup, N.G., Boehmer, E., Davaloo, M.: A Pricing Model for Data Markets. iConference 2015 Proc. (2015).
32. Brocke, J., Buddendick, C.: Reusable Conceptual Models – Requirements Based on the Design Science Research Paradigm. Int. Conf. Des. Sci. Res. Inf. Syst. Technol. - DESTRIST 2006. 576–604 (2006).
33. Sein, M.K., Henfridsson, O., Purao, S., Rossi, M., Lindgren, R.I.: Action design research. MIS Q. 30, 611–642 (2011).
34. Webster, J., Watson, R.T.: Analyzing the Past to Prepare for the Future : Writing a Literature Review. 26, (2002).
35. Ahituv, N.: Assessing the Value of Information: Problems and Approaches. ICIS 1989 Proc. Paper 45 (1989).
36. Garifova, L.F.: Infonomics and the Value of Information in the Digital Economy. 2nd Glob. Conf. BUSINESS, Econ. Manag. Tour. 23, 738–743 (2014).
37. Bhutta, K., Huq, F.: Benchmarking–best practices: an integrated approach. Benchmarking An Int. J. Vol. 6, 254–268 (1999).
38. Camp, R.C.: Benchmarking: The Search for Industry Best Practices that Lead to Superior Performance. (1989).
39. Recker, J.: Conceptual model evaluation. Towards more paradigmatic rigor. In: CEUR Workshop Proceedings. pp. 183–194 (2005).
40. Vom Brocke, J.: Design Principles for Reference Modeling: Reusing Information Models by Means of Aggregation, Specialisation, Instantiation, and Analogy. In: Innovations in Information Systems Modeling: Methods and Best Practices. pp. 47–76 (2009).