

Received 1 September 2023, accepted 11 September 2023, date of publication 14 September 2023,
date of current version 28 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3315588

 SURVEY

A Systematic Survey of Data Value: Models, Metrics, Applications and Research Challenges

MALIKA BENDECHACHE¹, JUDIE ATTARD², MALICK EBIELE³,
AND ROB BRENNAN³, (Senior Member, IEEE)

¹ADAPT and Lero Research Centres, School of Computer Science, University of Galway, Galway, H91 AEX4 Ireland

²Malta Information Technology Agency, Blata l-Bajda HMR9010, Malta

³ADAPT Research Centre, School of Computer Science, University College Dublin, Belfield, Dublin 4, D04 V1W8 Ireland

Corresponding author: Malika Bendeche (malika.bendeche@universityofgalway.ie)

This work was supported in part by the Science Foundation Ireland (SFI) under Grant 13/RC/2094_P2 (Lero) and Grant 13/RC/2106_P2 (ADAPT), and in part by the European Regional Development Fund (ERDF).

ABSTRACT Data is central to modern decision making and value creation. Society creates, consumes and collects data at an increasing pace. Despite advances in processing power, data is expensive to maintain and curate. So, it is imperative to have methods and tools to distinguish between data based on its value. Yet, there is no consensus on what characterises the value of data or how this data value should be assessed. This results in heterogeneous data value models and inconsistent measurement techniques that are siloed in specific application domains. This limits the formalisation and exploitation of these concepts. We present in this paper a methodical literature analysis that discusses data value models, assessment metrics and current applications. We also highlight challenges hindering the development and exploitation of data value as concept. This leads to the identification of a set of research questions to help researchers contribute to this emerging field. The aim of this article is to stimulate further research and deployment of quantitative data value models and value-driven applications.

INDEX TERMS Data value, data value characterization, data value quantification, data governance, data value metrics.

I. INTRODUCTION

Data has become an indispensable commodity for society [1]. For example, the European Union has given high importance to data from strategic, legal and regulatory perspectives.¹ Big data, and especially big data analytics, is increasingly prevalent as a driver of business value [2]. This is part of the trend where intangible assets have become an important source of value for businesses [3]. These data assets can range from market intelligence reports to sensor readings. Such data assets can be acquired or generated internally, either directly or as a by-product of offering goods or services [4].

Data lags behind other intangible assets by not being added to the balance sheet [5] and part of the reason

The associate editor coordinating the review of this manuscript and approving it for publication was Alba Amato¹.

¹https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en

for this is the challenge in assigning a value to the data [6]. The importance of valuing intangible assets, however, is frequently pointed out in relevant research [5], [7], [8], [9], [10]. The EU Open Data Directive² stresses the importance of *high-value datasets*: identified as datasets covering a number of domains that provide important benefits to the society, the environment, and the economy through their use and reuse. But what is exactly meant by “data value”? Many publications explore this term [11], [12], [13], but there is currently *no consensus on the definition of data value, its component dimensions or on how this data value can be assessed or quantified*.

A major challenge in valuing data assets is the heterogeneous nature of value, often broken down into dimensions of value [14]. This is exacerbated by the observation that data value, like data quality, is both subjective and highly

²<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32019L1024>

context dependent [13], [15], [16]. In the literature numerous approaches are considered to characterise data value. For example, Moody and Walsh [17] describe 7 “laws” of information which are applicable to data assets:

- 1) Information [data] is infinitely sharable;
- 2) The value of information [data] increases with use;
- 3) Information [data] is perishable;
- 4) The value of information [data] increases with accuracy;
- 5) The value of information [data] increases when combined with other information [data];
- 6) More is not necessarily better; and
- 7) Information [data] is not depletable.

These properties contrast strongly with other tangible assets that have a more easily defined value in use or exchange. Other early authors consider additional dimensions of the value of data, including the resulting business impact of new data [18], usage [6], and monetary costs or benefits [19].

The conceptual heterogeneity of data value leads to numerous approaches to quantify the value of data. For example, cost-based, market-based, and income-based approaches are frequently used to monetarily value intangible assets and data assets [9], [20], [21]. Data value measurement has also been applied in numerous domains as part of decision support or automated systems. For instance, consider data valuation for information life cycle management [22], [23]. In this context, data valuation becomes an indispensable tool for effectively governing the entire trajectory of data – from its inception to archival or disposal. By assigning a tangible value to data at each stage, organisations can make well-informed choices about data retention, archival frequency, and disposal strategies. This not only optimises data resource allocation but also ensures that valuable information is appropriately managed throughout its lifecycle. Another example is data markets and data pricing [24]. Much like traditional marketplaces where goods and services are traded, data markets have emerged as platforms where data can be exchanged, bought, and sold. The value attached to various datasets forms the basis for these transactions. Organisations seeking specific datasets can engage in a market-driven approach to acquire the data they require, thereby fostering a data-driven economy. This innovative approach relies heavily on accurate data valuation to facilitate fair and mutually beneficial transactions. business decision making in agriculture [25] is also another example. With advancements in precision agriculture and data-driven decision making, farmers now leverage data to optimise crop yield, manage resources efficiently, and make informed business choices. Data value assessment allows farmers to prioritise which data streams are most critical for their operations, ensuring that investments in data collection and analysis align with tangible benefits. This is pivotal in a sector where data-driven insights can directly impact productivity and profitability. Spanning this work requires an interdisciplinary approach as it draws on information systems deployments in many fields as well as computer science research.

There are many reasons beyond balance sheet concerns to quantify the value of the data itself, and to treat data independently from specific data-driven applications or businesses. These include: Data can be shared between many applications or even organisations; It is more efficient to develop reusable assessment methods, tools and metrics that treat data in an application-independent fashion, as seen with data quality assessment methods [26]. Common models for data value will enable heterogeneous tools and data value-consuming applications, such as value-driven data governance systems, to operate seamlessly in terms of data value quantification [27], [28]. The literature identifies many purposes for measuring the value of data. For example to provide knowledge of the value of data/information as an asset for merger negotiations [12]; to improve organisational accountability for data by raising awareness [5]; to justify the costs of creating, maintaining or purchasing data [29], [30]; to identify relevant data for an application [5], [31]; and to enable data-driven decision-making about data like file retention [32]. Most of the time, data’s true value is not recognised or exploited as an asset by organisations [5], [10], [17]. This is particularly evident in the latest trend of hoarding data [33] where organisations blindly capture all possible data with little means to discriminate between the data being accumulated, despite the fact that data storage still comes with costs [34]. Robust, easily interpretable data value assessment techniques will give us the tools to address this problem.

It is notable that despite the width of literature available on data value, only two publications provide a comprehensive overview through systematic surveys: (1) Viet et al. [35] surveyed works published between 2006 and 2017 and focused on the related concept of value of information (VoI) in supply chain decisions; (2) Alawad and Kraemer [36] authored a systematic review focusing on VoI in wireless sensor networks and Internet of Things. Both of these papers are limited to their respective application domains. In addition, these works did not examine other related terms for VoI, such as data value, information value, data valuation, or information asset. Other publications provide a more unstructured approach to reviewing literature. For example, Faroukhi et al. [37] presented an unstructured survey of the literature related to data monetisation in big data value chains. Again this work was limited in scope and we eliminate data monetisation from our focus to minimise the overlap. Yanlin and Haijun [38] provide a timely but non-systematic survey of data value concepts focusing only on the Chinese literature. Fleckenstein et al. [39] provide a useful framework of approaches to data valuation models that we reuse but their work is not supported by a systemic search as presented here, lacks details on assessment metrics useful for automation and instead focuses on qualitative approaches to assessment.

This shows that there is both a need and a gap in the literature for a wider structured survey aimed at unifying the field and identifying a wide-ranging research agenda. In this paper we provide a systematic survey that comprehensively

analyses the existing literature covering the domain of data value in terms of data value models, assessment metrics, and applications.

The rest of this paper is structured as follows: Section II defines some terminology and concepts used throughout the paper. In Section III, we describe our research method for the systematic survey. Section IV provides an analysis of the primary studies resulting out of the systematic survey with a focus on data value models, metrics and applications. We discuss our research agenda and the highlighted relevant research questions in Section V and we provide our conclusions and future work in Section VI.

II. TERMINOLOGY AND CONCEPTS

In this section we provide a number of definitions of terms related to the topic of data value, and used throughout this article. This is to provide consistency to the discussion and more broadly to the domain of data value research.

A. DATA

The Oxford dictionary defines data to be “[f]acts and statistics collected together for reference or analysis”.³ Data is therefore facts about the world. Data can be collected (e.g. using sensors or surveys) or calculated (e.g. age from date of birth). Data can be represented in different formats, including text, images, spreadsheets, JSON, etc. Here we do not discuss the distinction between data, information, and knowledge, and for the sake of incorporating all relevant literature, we include all three terms in the scope of our search.

B. VALUE

In the generic sense of the term, the Oxford dictionary defines value to be “[t]he regard that something is held to deserve; the importance, worth, or usefulness of something”.⁴ While this applies in most, if not all, contexts, different disciplines have more specific definitions. For example, in economics, the definition and measurement of value would be in terms of currency. Stern [40] identified two kinds of value in the context of natural resource scarcity indicators: use value and exchange value. Prices and rents are common measures of exchange value and unit costs as a measure of use value [40]. Other definitions, such as sentimental value, would be in terms of personal or emotional associations rather than material worth. It is evident that these varying definitions are tied to the subjective and contextual nature of value. With the aim of characterising the latter concept, we define value to be *a number of different data value dimensions (attributes) that in an aggregate manner represent the worth return of the thing in question.*

³<https://en.oxforddictionaries.com/definition/data> (Accessed 1st October 2019)

⁴<https://en.oxforddictionaries.com/definition/value> (Accessed 1st October 2019)

C. DATA VALUE

There is a wide range of definitions of data value across different domains and for specific use cases in the literature. For instance in ML, data value is the weight or contribution of each training sample or feature in improving a model performance [41], [42], [43]. In applied energy, data value is defined as the quantitative relationship among the data, uncertainty reduction, and profit enhancement [44]. In business, data value is typically estimated in terms of cost (e.g. collection cost, storage cost, or cost related to the loss of the data) and revenue by selling or exploiting the data [39]. This is exactly what Laney tried to do with his financial valuation models of information [data] [5]. There are also some definitions of data value which are more general and can be applied to multiple domains. For example, Khokhlov and Reznik [45] defined data value as data usefulness. Another such definition of data value is: data value is the future importance of data, it expresses a probability of further use [46], [47].

D. DATA VALUE MODEL

Representations of the value of data, either as explanatory, descriptive or predictive models. These representations define the relevant data value dimensions and relationships between the dimensions that characterise the value of value to an individual, application or organisation in a specific context.

E. DATA VALUE DIMENSIONS

Attributes of the value of data assets that are relevant to data consumers, maintainers or owners. Sometimes called data value aspects. Due to the subjective and contextual nature of value, some dimensions may be considered to be more characteristic of value than others, depending on the use case, data asset, or consumer.

F. METRICS

Metrics are specified quantitative measures of data or its context that can be used to measure the data value dimensions of a specific data item. For instance, if we consider “Usage” as a dimension in a relational database, a metric that measure this dimension could be “Number of writes in a day”. Metrics can be subjective or objective and qualitative or quantitative. All metrics can be mapped to one or more data value dimensions in a descriptive or predictive data value model. A set of observations of data value metrics for a specific data item quantify or measure the mapped data value dimensions for those metrics.

G. DATA VALUE QUANTIFICATION

The quantification, assessment or measurement of data value is the explicit calculation of the value of a specific data item based on a descriptive or predictive data value model. It is usually based on a set of observations of a set of data value metrics. Data value metrics could be the specific measurable elements that are used within the process of data value

qualification. Data value qualification involves using these metrics to qualify, assess, and make judgements about the value of data. Expert judgement can also be used in subjective and less formal quantification methods. The focus of this survey is on more formal methods based on observations which may themselves be objective or subjective.

III. REVIEW TECHNIQUE

In this survey, we follow a methodical literature survey technique with three phases of activities - (i) actively-planning, (ii) conducting and reporting the review results, and (iii) exploration of research challenges as per of the widely accepted guidelines and process outlined in Pai et al. [48] and Kitchenham et al. [49], [50]. The remainder of this section details the research question, the process for the identification of research, and the data extraction process.

A. RESEARCH QUESTIONS

The goal of this survey is to provide a comprehensive overview of the literature that provides discussions on the value of data, and its models, dimensions, assessment techniques, metrics and applications. We therefore define the following as research questions:

- Q1: What are the existing models and dimensions used to characterise the value of data?
- Q2: What metrics and measurement approaches have been created for the quantification of the value of data?
- Q3: In what application domains have data value models or assessment techniques been applied?
- Q4: What are the issues and challenges facing the data value research?

The research questions are designed to cover a wide range of aspects related to data value, from theoretical models and measurement techniques to practical applications and challenges, aligning with our goal of providing a comprehensive overview of the existing literature in this area.

B. IDENTIFICATION OF RESEARCH

Research was identified from the following four electronic databases: (i) Springerlink, (ii) ScienceDirect, (iii) IEEEExplore, and (iv) ACM Digital Library.

Figure 1 shows the search string used to query these databases based on metadata (title, abstract, keywords). In the case of IEEEExplore, the query did not produce sufficiently accurate results. As such, the query for IEEEExplore was rerun to both metadata and full texts. In addition to the papers returned by the above databases, we also included some handpicked papers [6], [29], [32], [46], [51], [52], [53], [54], [55], [56] that were recommended by the domain experts which were not returned by the search string.

The initial search returned a total of 434 research papers related to the research topic. A set of inclusion and exclusion criteria, shown in Table 1, were defined to enable the selection of papers to include in this study to be carried out in a systematic and replicable manner. In line with Kitchenham et al. [49], [50], three researchers independently

```
( 'Data value' OR 'data valuation' OR 'infonomics' OR 'data asset'
OR 'value of data' OR 'data governance' OR 'information value'
OR 'value of information' OR 'information valuation' OR 'knowledge
asset' )
AND
( 'assess' OR 'measure' OR 'evaluation' OR 'estimate' )
AND
( 'model' OR 'framework' OR 'system' OR 'theory' OR 'dimension'
OR 'metric' ).
```

FIGURE 1. Search string.

screened titles. The title-based exclusion reduced the number of papers to 121. Then, the papers' abstracts were read. The abstract-based exclusion reduced the number of papers to 73. In the next phase, the full text of the papers was read. In all three phases, in case of disagreements on whether a paper was to be included or discarded, discussions were held until an agreement was reached. Following the application of these criteria, 63 research articles were included in the final review. These are listed in Table 2.

C. DATA EXTRACTION

The papers were manually reviewed by the three authors independently. For each one of the 63 papers, the following data was extracted: bibliographic data, the contribution towards the domain of data value (e.g., data value model, data value assessment, or data value use case/domain), implemented details, the data value dimensions and metrics (if applicable) under study, and the type of validation used (if applicable). The data was then compared and aligned, with discussions taking place if any inconsistencies were found.

IV. REVIEW RESULTS

In this section, we analyse the papers obtained from our systematic review. We start by reviewing their bibliographic data such as number of publications per year and also publication venues. Then we delve into their content: data value models, data value assessment metrics, and applications.

A. PRELIMINARY RESULTS

We see in Figure 2 (extracted from Table 2) that most of the publications (34 papers) are conference papers, 21 are journal papers, 3 are symposium papers, 3 are workshop papers and 2 are books. We can also see that the search in the domain of data value goes all the way back to the year of 1980. However, most of the selected articles were published in the period starting from 2003.⁵ We notice an increased number of works from a year to another especially in the last decade. This reflects on the increase of interest in researching the data value field. The same pattern can be seen in terms of type of publications where more diversity in type of publication can be seen in recent years including journals, conferences, workshops and symposiums. An analysis of publication venue (in Table 2) suggests that there is no common venue for publication in this field. In fact,

⁵Data was only collected until second quarter of 2022.

TABLE 1. Inclusion and exclusion criteria.

Inclusion Criteria	Exclusion Criteria
Full-text	Uncompleted studies
Published at any time	Non English
Published in the above selected databases	Duplicated studies
Published in workshops (W), symposiums (W), conferences (C), books (B), and journals (J) across all disciplines	Studies that discuss value outside the context of data, information or knowledge
Papers written in English on the topic of defining, describing, characterising, modelling, measuring, or quantifying data value	Studies on the value of Information Technology

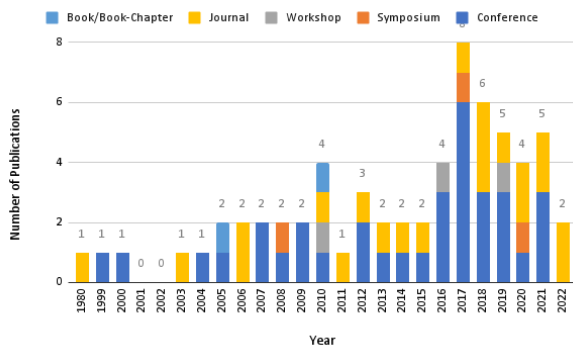


FIGURE 2. Number of publications per year.

almost all the selected papers are published in different venues. This indicates that there is no common consensus in term of venue for publication in this field. We notice also that the values are diverse in terms of discipline. Some of them are in computer science, information systems and engineering (e.g., International Conference on System Sciences, Iberian Conference on Information Systems and Technologies, Conference on Management of Engineering and Technology, Conference on management Science and Engineering, etc.), other are in science and environment such as Journal of Livestock Science and Journal of Science of The Total Environment. One paper was published in the interdisciplinary journal of space policy.

B. DATA VALUE MODELS

Fleckenstein et al. [39] identified three categories of approaches to data valuation: market-based valuation, economic models, and dimension-based models. They define market based as "estimates the value of data in terms of cost and revenue when buying and selling data or data-intensive businesses"; Economic models as "estimating economic benefit as a result of making data available [measuring impact]". The dimension-based approach "examines valuation points of a specific data set both inherent to data, like data quality ... and contextual to value [of] data [e.g. data usage]".

This work fits within the dimension-based approach of Fleckenstein et al. It focuses on new metrics and dimensions that are specific to data value and not already known from the data quality metrics literature. For example, 26 data quality dimensions and over 80 associated metrics are described by Zaveri et al. [26] and 21 data quality dimensions are

identified in the recent review of data quality dimensions by Wang et al. [99]. The most commonly referenced data quality dimensions in our data value survey sample papers were accuracy [17], [35], timeliness [35], [51], completeness [35], [45], latency [70], volume [32], [47], [96] and provenance [45], [70]. Note that volume in many models has an inverse or convex value curve in relation to value [17].

The research approach in this paper yielded 19 primary studies that focus on models including one or more dimensions which characterise data value, as shown in Table 3. All previously known data quality dimensions [26], [99] are grouped in the table into "data quality". Four new dimensions are identified: Content/Uniqueness, Usage, Utility and Financial. The most popular dimension cited is related to examining the content or uniqueness of the data and its relevance for a task. The Content dimension of data value was first identified by Even and Shankaranarayanan [32] in 2005. Usage is a well established dimension of value since [17] defined it as a distinguishing feature of data value and classed as part of the context of value by Fleckenstein et al. Utility is perhaps the original dimension of data value [31]. We extend the dimensional approach to include one market-based model aspect of Fleckenstein et al., a financial dimension that represents measurable aspects of data within an organisation like cost or price.

As can be seen from the table, most current models of data value are limited in their perspective since they only focus on a subset of the data value dimensions, such as the financial dimension, and they do not provide a comprehensive view of how dimensions are related. One feature of Table 3 is that the majority of models address three or less dimensions of data value.

It may be asked are any of these dimensions antecedents of (factors that influence) value rather than dimensions of value itself. Examining Wang's original antecedents (management responsibility, operation and assurance costs, research and development, production, distribution, personal management, and legal) it seems there is some overlap, e.g. in production costs. However this neglects the contention of Fleckenstein et al. and Laney [5] that data value, even in a dimensional approach, is a wider concept than data quality. Additionally, metrics have been identified in this search for all of the dimensions, and this suggests that they are directly useful for data value calculations.

In the following subsections we provide an overview of the models covered in the mentioned primary studies, based

TABLE 2. Final selection of publications. J:Journal, C:Conference, S:Symposium, W:Workshop, B:Book, BC:Book Chapter.

No	Title	Authors	Type	Year	Venue
1	A Systematic Approach toward Assessing the Value of an Information System	Ahituv	J	1980	MIS quarterly
2	Measuring the Value Of Information-An Asset Valuation Approach	Moody and Walsh	C	1999	European Conf. on Information Systems
3	Information value mapping for fusion architectures	Peterson and Sundareshan	C	2000	Int. Conf. on Information Fusion
4	Improving Data Quality in Practice: A Case Study in the Italian Public Administration	Missier et al.		2003	Distributed and Parallel Databases
5	Situation-based message rating in information logistics and its applicability in collaboration scenarios	Meissen et al.	C	2004	Euromicro Conf.
6	Information Valuation for Information Lifecycle Management	Chen	C	2005	2nd Int. Conf. on Autonomic Computing
7	Value-Driven Data Quality Assessment	Even and Shankaranarayanan	B	2005	ICIQ
8	The value of information: Measuring the contribution of space-derived earth science data to resource management	Macaulay		2006	Space Policy
9	How to calculate information value for effective security risk assessment	Sajko et al.		2006	J. of Information and Organizational Sciences
10	Information Assets and their Value	Engelsman	C	2007	Not a Peer Reviewed Paper
11	File valuation in information lifecycle management	Turczyk et al.	C	2007	Americas Conf. on Information Systems
12	The role and value of information for supply loop management: Framework and application for the end-of-life cell phone industry	Doctori-Blass and Geyer	S	2008	IEEE Int. Symp. on Electronics and the Environment
13	A Hierarchical Storage Strategy Based on Block-Level Data Valuation	Zhao et al.	C	2008	Int. Conf. on Networked Computing and Advanced Information Management
14	An approach for quantifying enterprise value of information (EVI)	Lu and Zhu	C	2009	Int. Conf. on Management Science and Engineering
15	The value research on information resources management of supply chains	Dang et al.	C	2009	Int. Conf. on Management Science and Engineering
16	Block-Level Data Migration in Tiered Storage System	Zhao et al.	C	2010	Int. Conf. on Computer and Network Technology
17	Chapter 28 - Measuring EIM and the Value of Information	Ladley	BC	2010	Making EIM Enterprise Information Management Work for Business
18	Understanding the dynamics of information management costs	Tallon		2010	Communications of the ACM
19	A Framework to Assess Value of Information in Future Vehicular Networks	Giordani et al.	W	2010	Workshop on Technologies, models, and Protocols for Cooperative Connected Cars
20	Asset Analysis of Risk Assessment for IEC 61850-Based Power Control Systems—Part I: Methodology	Liu et al.		2011	IEEE Trans. on Power Delivery
21	Disentangling the Value of Information and Analytics through Componentised Business Architecture	Glissmann et al.	C	2012	Hawaii Int. Conf. on System Sciences
22	Information-driven self-deployment and dynamic sensor coverage for mobile sensor networks	Jalalkamali and Olfati-Saber	C	2012	American Control Conf.
23	Knowledge-empowered agent information system for privacy payoff in eCommerce	Yassine et al.		2012	Knowledge and Information Systems
24	Enhancing Battlefield Situational Awareness through Fuzzy-Based Value of Information	Hanratty et al.	C	2013	Hawaii Int. Conf. on System Sciences
25	Use of information from monitoring and decision support systems in pig production: Collection, applications and expected benefits	Cornou and Kristensen		2013	Livestock Science
26	Value-Based File Retention	Wijnhoven et al.		2014	J. of Data and Information Quality
27	Digital information asset evaluation: characteristics and dimensions	Viscusi and Batini	C	2014	Smart Organisations and Smart Artifacts
28	Dynamic risk evaluation model as a security field	Shimazu et al.	C	2015	Glob. Conf. on Consumer Electronics
29	Quality and value of the data resource in large enterprises	Otto		2015	Information Systems Management
30	Mining online forums for valuable contributions	Scaffidi	C	2016	Iberian Conf. on Information Systems and Technologies
31	Optimal self-driven sampling for estimation based on value of information	Soleymani et al.	W	2016	Int. Workshop on Discrete Event Systems
32	Your Data in the Eyes of the Beholders: Design of a Unified Data Valuation Portal to Estimate Value of Personal Information from Market Perspective	Kassa et al.	C	2016	Int. Conf. on Availability, Reliability and Security
33	Information pricing: A utility based pricing mechanism	Rao and Ng	C	2016	Int. Conf. on Dependable, Autonomic and Secure Computing
34	Investigating the Value of Privacy within the Internet of Things	Mayle et al.	C	2017	IEEE Glob. Communications Conference
35	Assessing the Value of Data: An Approach to Evaluate the Technology Driven Benefits of Smart Product Data	Schuh et al.	C	2017	Int. Conf. on Management of Engineering and Technology
36	A fast graphic-based information valuation algorithm for cooperative information sharing	Hu et al.	C	2017	IEEE Int. Conf. on Systems, Man, and Cybernetics
37	An Evaluation Method of Data Valuation Based on Analytic Hierarchy Process	Qiu et al.	S	2017	Int. Symp on Pervasive Systems, Algorithms and Networks
38	A research on information value in energy internet based on risk theory	Tan et al.	C	2017	IEEE Conf. on Energy Internet and Energy System Integration
39	Design of information value determination method for information-sharing systems during large-scale disasters	Sonobe et al.	C	2017	IEEE Int. Conf. on Cognitive Informatics & Cognitive Computing
40	A First Look at Information Entropy-Based Data Pricing	Li et al.	C	2017	Int. Conf. on Distributed Computing Systems
41	Value of information in reservoir development projects: Technical indicators to prioritize uncertainties and information sources	Santos et al.		2017	J. of Petroleum Science and Engineering
42	Data valuation considering knowledge transformation, process models and data models	Sathananthan	C	2018	Int. Conf. on Research Challenges in Information Science
43	An Importance Measure to Assess the Value of a Component Inspection Policy	Fauriat and Zio	C	2018	Int. Conf. on System Reliability and Safety
44	The Smart Black Box: A Value-Driven Automotive Event Data Recorder	Yao and Atkins	C	2018	Int. Conf. on Intelligent Transportation Systems
45	The value of information in supply chain decisions: A review of the literature and research agenda	Viet et al.		2018	Computers & Industrial Engineering
46	The economic value of information provided by milk biomarkers under different scenarios: Case-study of an ex-ante analysis of fat-to-protein ratio and fatty acid profile to detect subacute ruminal acidosis in dairy cows	Rojo-Gimeno et al.		2018	Livestock Science
47	Value-of-information in spatio-temporal systems: Sensor placement and scheduling	Malings and Pozzi		2018	Reliability Engineering & System Safety
48	Profit Allocation for Federated Learning	Song et al.	W	2019	IEEE Int. Conf. on Big Data
49	Information Value Based on the Scenario of Wind Power Trading in Electricity Markets	Chen et al.	C	2019	Conf. on Energy Internet and Energy System Integration
50	The Importance of Data Assets and Its Accounting Confirmation and Measurement Methods	Zhang et al.	C	2019	Int. Conf. on Behavioral, Economic and Socio-Cultural Computing
51	Value Evaluation of Data Assets: Progress and Enlightenment	Li et al.	C	2019	Int. Conf. on Big Data Analytics
52	Efficient task-specific data valuation for nearest neighbor algorithms	Jia et al.		2019	arXiv preprint
53	The value of perfect and imperfect information in lake monitoring and management	Koski et al.		2020	Science of The Total Environment
54	Competitive Data Trading Model With Privacy Valuation for Multiple Stakeholders in IoT Data Markets	Oh et al.		2020	IEEE Internet of Things J.
55	Value of Information Analysis via Active Learning and Knowledge Sharing in Error-Controlled Adaptive Kriging	Zhang et al.	C	2020	IEEE Access
56	What is the Value of Data Value in Practical Security Applications	Khokhlov and Reznik	S	2020	IEEE Systems Security Symp (SSS)
57	From Qualitative to Quantitative Data Valuation in Manufacturing Companies	Stein et al.	C	2021	IFIP Int. Conf. on Advances in Production Management Systems
58	Data valuation for decision-making with uncertainty in energy transactions: A case of the two-settlement market system	Wang et al.		2021	Applied Energy
59	Equitable Valuation of Crowdsensing for Machine Learning via Game Theory	He et al.	C	2021	Int. Conf. on Wireless Algorithms, Systems, and Applications
60	A data value metric for quantifying information content and utility	Noshad et al.		2021	J. of big Data
61	Using Business Data in Customs Risk Management: Data Quality and Data Value Perspective	Hofman et al.	C	2021	IFIP EGOV Conf.
62	Optimal inspection of binary systems via Value of Information analysis	Lin et al.		2022	Reliability Engineering & System Safety
63	Value of Information in Wireless Sensor Network Applications and the IoT: A Review	Alawad and Kraemer		2022	IEEE Sensors J.

TABLE 3. Data value dimensions included in data value models studied.

Publication	Economic	Content	Quality	Usage	Utility
Shimazu et al. [71]		✓			
Zhang et al. [89]	✓				
Hanratty et al. [69]		✓	✓		
Schuh et al. [76]	✓			✓	
Yao and Atkins [84]		✓	✓		
Hofman et al. [95]		✓	✓		
Mayle et al. [75]	✓				✓
Zhao et al. [53]		✓		✓	
Sonobe et al. [79]		✓			✓
Yanlin and Haijun [38]	✓			✓	
Jia et al. [51]		✓			✓
Tan et al. [78]		✓			✓
Sajko et al. [55]	✓		✓		✓
Li et al. [80]	✓	✓	✓		
Qiu et al. [46]		✓	✓	✓	
Viscusi and Batini [14]		✓	✓	✓	✓
Lu and Zhu [61]	✓	✓	✓	✓	
Laney [5]	✓	✓	✓	✓	✓
Ahituv [31]	✓	✓	✓	✓	✓

on the main dimension that, according to the authors, the data value would contribute towards. For each dimension we provide a table of data value assessment metrics identified in the search to facilitate reuse in new applications or further research. As has been long established for data quality, to assess specific data assets, it is necessary to define metrics to quantify or measure data value dimensions [100]. Only about half of the selected papers (31 out of 63) provide specific data value metrics and this shows the relative immaturity of the field. In the tables below the metrics are classified as being suitable for subjective or objective measurement in the “Type” column. In total 44 metrics for data value assessment beyond those typically used for data quality assessment have been identified and allocated to data value dimensions. The Utility, Content-Uniqueness, Usage and Financial dimensions have 14, 10, 10 and 9 metrics identified respectively. These are candidates for the biggest departure from traditional data quality dimensions for data value assessment. Three metrics are used in more than one dimension (rival access loss, camera resolution and market price). Camera resolution is an example of a common sensor-based metric that can quantify the likely information content of data and therefore its value.

Moreover, we also provide a distinct section for market-based VoI models (12 papers), which we follow Fleckenstein et al. to be a separate type of data model (economic) since they focus on cost/benefit difference in outcome as opposed to specific dimensions of data value.

1) ECONOMIC OR FINANCIAL-BASED MODELS

The financial or economic dimension is one of the most popular dimensions used to determine the value of a data asset. This is probably due to the tangible nature of this aspect. This dimension is based on the metrics included in the “Accountancy Valuation Models” for data (information) value of Moody and Walsh [17] or the financial models of

Laney [5] which includes the realised or potential cost of data, the market value of data, and the present financial value of data. Zhang et al. [90] propose a theoretical capitalisation of data assets, based on the historical cost method, the fair value, and the current value. Li et al. [81] focus on data pricing, where an entropy-based method is proposed to measure the value of a dataset based on size and information content. A pricing function is then provided to convert from entropy-based value to price. Mayle et al. [76] propose a game-centric model of a private data exchange in return for a service. The model takes into account the priority given to data items by users and the monetary value given to users in return for their data. Schuh et al. [77] model an extended, economic value definition which is proportional to the benefits and costs associated with the product or service. Based on “technological value contributions”, the model supports manufacturing companies in evaluating if a generic set of field data generated by a smart product provides value to the user.

2) CONTENT UNIQUENESS-BASED MODELS

This dimension was identified by Even and Shankaranarayanan [32] and Viscusi and Batini [14]. Measurement can be derived by assessing the content and its applicability for business use cases. The research identified here proposes measurement methods that link impartial characteristics and contextual perception to measure the potential business value associated with the data. This is related to but distinct from utility metrics which measure the value of the data in use. Uniqueness is an important aspect of content for value and is a characterisation of the value of data based on rarity [13] or scarceness [14] of the information contained in it. Yao and Atkins [85] propose Smart Black Box (SBB) data compression decision making based on data value, calculated on data novelty and events. Shimazu et al. [72] on the other hand take a value-based approach towards setting data confidentiality. Their paper defines a method for setting data confidentiality based on risk, taking into account data value, protection level, and threat level.

3) QUALITY-BASED MODELS

Data quality dimensions are tied with Data Content Uniqueness as the most common way identified to quantify the value of data assets, and can include aspects in the Data Quality Model defined in the ISO/IEC 25012 Standard.⁶ To a certain extent this is because the quality concept of “fitness for use” is closely aligned with the concept of value, for example possessing the same strong dependence on context of use. Measurement can be derived by assessing the dataset directly and its conformance to standards or applicability for business use cases. Fleckenstein et al. [39] describe the dimension-based approach to data value as an extension of data quality methods and tools. Given the wide range of data quality metrics well known to the community

⁶<http://iso25000.com/index.php/en/iso-25000-standards/iso-25012>

TABLE 4. Data value metrics for financial dimension. S: Subjective. O: Objective.

Metric	Description	Type
Compliance cost	Estimated financial cost to business of not keeping data for compliance/regulation purposes [55].	S, O
Fixed record value	A count of records (see data volume metrics) with a fixed value per record can be used [32].	S, O
IP value	For intellectual property data assets whose value is directly proportional to its financial value (e.g. licences or patents) [55].	S, O
Maintenance cost	Estimated financial cost to business of maintaining, adapting or curating data [92].	O
Market price	Price that a user will accept for data, sometimes in return for a specific service [75, 60]	O
Replacement cost	Estimated financial cost to create, replace or reproduce this data [55, 22, 17].	S, O
Revenue loss	Estimated financial cost to business of not having access to data [55, 60] or a service based on observations in the data [96].	S, O
Rival access loss	Estimated financial cost to business of competitors having access to data [55].	S, O
Storage cost	Estimated financial cost to business of storing data [92, 30].	O

TABLE 5. Data value metrics for content-uniqueness dimension. S: Subjective. O: Objective.

Metric	Description	Type
Camera resolution	For visual sensors a measure of the potential accuracy of sensor readings from a specific source [50].	O
Customer lifetime value	Data records that describe or refer to specific customers, for example business transactions with that customer, may be weighted. Typically used with count of records [32].	S, O
Field value	The financial amount of a transaction recorded in a field of a data record may be used as a weight for the value of that record. Typically used with count of records [32].	O
Image quality	An estimate of the quality of an image based on camera type (smartphone type), brightness, contrast, resolution, noise, and exposition [44].	S, O
Data information entropy	A measure of the uniqueness of data in the dataset [80].	O
Event information measure	By classifying the contents of an image or sensor reading to a set of events with known probabilities, can calculate the information measure of the image containing a specific event [84]. Alternatively [67] specify a Fisher Information Matrix (FIM) based metric.	O
KNN Shapley value	For K-nearest neighbours classifier algorithms Jia et al. provide an efficient Monte Carlo method to calculate the Shapley value of the KNN utility of each data point [51].	O
Lexical semantic uniqueness	A measure of the fraction of most unique words in a file compared to the contents of the whole collection [46].	O
Mutual information	A qualification of the inferential gain by increasing the size or features of data when comparing two datasets [94].	O
Scarcity	The probability that other organisations (particularly competitors and partners) have the same data [5].	S

we do not provide a table of them here but instead refer readers to other surveys of data quality metrics such as [26]. In Hanratty et al. [70] base their definition of data value on data quality related aspects, namely content reliability, trust of source, and timeliness. Intended for the military decision-making domain, the authors propose a fuzzy-logic-based approach that values data. Hofman et al. [97] propose an evaluation framework for data quality and value assessments. The authors explore data quality categories and dimensions for assessing the potential value of linking different customs data sets and linking a business dataset to a customs data set. The categories of data quality used were contextual (including relevancy, value-added, timeliness, completeness, and amount of data as dimensions) and representational (including interpretability, ease of understanding, concise representation and consistent representation).

4) USAGE-BASED MODELS

Authors such as Moody and Walsh [17] argue that data has no intrinsic value, yet it only becomes valuable when people use it. Following this argument, the more the data is used, the more valuable it becomes. This is based on the economic concept of “value in use”. This dimension is a characterisation of how often or by whom [71] a dataset is used. It was developed by Chen [22] into a set

of concrete metrics. Qiu et al. [47] propose a data value measuring algorithm for data migration applications based on the usage of the data, such as the access time, and the data read and write frequency and access, and other content related aspects such as data size and file content. Zhao et al. [54] propose a model that values data blocks based on the timeliness of the data, data distribution and usage, and the association between blocks, as well as other usage related characteristics such as read and write frequency and granularity. Yanlin and Haijun [38] derive a data asset valuation framework based on data asset production cost and data asset spillover value, where the data is considered to be more valuable the more it is used, particularly if it is used in a multidimensional manner.

5) UTILITY-BASED MODELS

Utility characterises data value in terms of value in use and the benefits, usually business oriented, that can be derived from it [14], [17], [101]. This dimension is often used to classify data value metrics which are very specifically tied to a particular application, service or business process and in many cases a specific dataset. Sonobe et al. [80] define a data value model based on a utility function derived from seller/producer estimates, with the aim of enabling rapid flow of the most valuable data in a disaster recovery situation. Jia et al. [52]

TABLE 6. Data value metrics for usage dimension. *S: Subjective. O: Objective.*

Metric	Description	Type
Access interval	A time series measure weighted towards data most recently created. [46].	O
Frequency	An access time series measure based on how often data is read and written to [46].	O
Number of accesses	A measure of the degree of usage [22]. Moody et al. suggest this metric [17].	O
Potential user count	An estimate of the number of similar users in the system compared to the ones that have accessed the data, that are hence likely to also access the data [46].	S, O
Purpose	A measure of the significance of the type of usage [22].	S
Read write profile	Data valuation based on read and write granularity, data distribution, and data association [53].	O
Recency	A specialised time-based metric to capture when data was last used, defined by [22].	O
User count	Called the number of visits by Qiu et al., the number of unique users that access the data [46].	O
User role	[70] indicates user role is a measure of the significance of the usage.	S, O
Usage over time	A quantification of value based on the usage of the data over time [22].	O

also use a utility function in an attempt to answer how much each data point is worth for machine learning models based on a model of the relative value of data derived from its Shapley Value from game theory. Tan et al. [79] derive an analytic formula for information value assessment based on utility functions, data type classification, information uncertainty, and the willingness of different actors to pay for more information based on their risk preferences.

6) MULTI-DIMENSIONAL MODELS

Laney [5] discusses a number of “Information Asset Valuation Models” i.e. a set of approaches for calculating the value of data assets. Compared to the previous models, Laney’s models are more comprehensive as they comprise a larger number of dimensions. Laney [5] divides the models into fundamental models and financial models. Fundamental models are mostly based on the content of the data, including the utility, business usage, impact of information, and quality aspects, such as validity, completeness, integrity, and consistency, and are claimed to be better for information asset management applications. The financial models are based on measures of money and accountancy and claimed to be better for assessing an information asset’s business benefits. Based on the model definitions provided, Laney [5] uses more of an expert-based approach, as opposed to automated data valuation. Whilst a great basis for further research, these models in many cases lack the definition of concrete metrics.

Similar to Laney, Viscusi and Batini [14] propose a more comprehensive model for digital information asset evaluation. The authors consider information value to be based on: information capacity (comprising information quality, information structure, and information infrastructure) and utility based on information diffusion. The authors base their model on the assumption that information value can be quantified either on the basis of information utility of the IT capabilities enabled by a data asset, or otherwise on the basis of the overall capabilities the data asset may provide in the initiative in question.

Lu and Zhu [62] also consider a number of dimensions in the evaluation of Enterprise Value of Information (EVI), and construct an evaluation model of EVI based on a combination of a qualitative and quantitative approach. The authors

propose an EVI evaluation index system that takes into consideration information authenticity, timeliness, degree of coverage, degree of relevancy, degree of superposition, manager’s subjective consciousness, information flux, and information cost. Whilst only the latter two are quantitative indexes, all the previous are categorised as qualitative. The authors therefore use the cloud model evaluation to translate the qualitative indicators into the quantitative target.

Ahituv [31] defines a joint utility function that includes a set of dimensions for information value as follows: timeliness (including response time and frequency); contents (including similarity and aggregation level); format (including medium, ordering, and graphic design); and cost. Albeit the author provides this list of attributes to demonstrate an approach, the author points out that he does not intend this list to be exhaustive.

In summary, a wide range of data value dimensions are defined by existing models but only the models in [5], [14], [31] and [62] define a broad multi-dimensional approach to data value and in each case the dimensions selected differ.

7) VALUE OF INFORMATION (VoI) MODELS

Keisler et al. [102] define the value of information (VoI) as “a decision analytic method for quantifying the potential benefit of additional information in the face of uncertainty”. It is often used in decision support systems and automated sensor data fusion applications. Thus it can be seen that VoI is a term used for a specific type of data value in the literature. Since VoI calculation is always specialised to a specific decision and application domain, there are a wide range of VoI models and assessment methods in the literature. Its application depends on having an objective function to be maximised and a choice between courses of action leading to uncertain payoffs [103].

A popular approach is to define VoI in terms of monetary values related to the costs and benefits of the use case in question. The monetary definition of VoI corresponds to the difference between the cost for acquiring/collecting new data and the benefit that this would create in terms of reducing uncertainty in decision making and improved business gains. For instance, Koski et al. [53], Macauley [60], and Rojo-Gimeno et al. [86] define and estimate VoI as a

TABLE 7. Data value metrics for utility dimension. S: Subjective. O: Objective.

Metric	Description	Type
Fidelity	Usefulness of sample data for a given inferential task [94].	O
Goods condition	For end of life processing (recycle, parts reuse, product reuse) need estimate of goods condition [60].	S, O
Market price	Price that a user will accept for data, sometimes in return for a specific service [75, 60]	O
nKudos	Number of approvals (“kudos”) given to a message by the community [72].	O
Personal data differential price	A measure that calculates the relative utility of a dataset by comparing the differences between original data of a personal nature, and an obfuscated version of the same data [54].	O
Provider assessment	An assessment of the value of a data item by the provider of that data [79].	S
Relevancy	The number of business processes that use, or could benefit from this data [5].	S, O
Revenue	Revenue increases or decreases with or without access to the data [5, 43].	S, O
Rival access loss	Estimated financial cost to business of competitors having access to data [55].	S, O
Service outcome	Reduced or improved service outcomes with or without access to the data [60].	S, O
Safety distance	For vehicular sensors a measure of the distance where sensor readings become increasingly important. [50].	O
User assessment	An assessment of the value of a data item by the user of that data after consuming it [79].	S
User priority	Priority given to a data item by a specific user [75].	S, O

monetary value in three different application domains. Doctori-Blass and Geyer [61], and Dang et al. [63] also define VoI as monetary/economic value for the specific domain of supply chain. While the aforementioned papers aim to provide a single/global assessment of VoI, Giordani et al. [51] and Fauriat and Zio [84] aim to assess the VoI of different data sources separately, in an attempt to identify the most relevant information and prioritising their acquisition. However, in the former, the definition of VoI is not based on monetary value. Instead, it is based on an aggregation of three domain specific attributes (i.e., a weighted-sum of source proximity, timeliness, and quality). Santos et al. [82] and Hanratty et al. [70] aim at improving the accuracy and applicability of the defined VoI for their respective use cases (i.e., reservoir development and military operations management). References [74], [87], and [92] discuss the large computational complexity that is associated with computing and optimising VoI in various domains and propose techniques to improve the efficiency and the time complexity of these processes.

In summary, the literature on VoI is relevant for discussing more general concept of data value but VoI is limited to creating application-specific data value models where it is possible to efficiently calculate or measure the expected value of perfect information (EVPI), the expected value of sample information (EVSI) and the expected net gain of sampling (ENGS) and this is not possible in many real world deployments.

C. APPLICATION DOMAINS AND USE CASES

Approximately two thirds of the papers (45) had a specific, detailed application domain or use case for data value. A thematic analysis of these papers is provided in Table 8. This helps to better understand the current application areas for data value and it identifies gaps for further work or opportunities for cross domain exploitation of existing results. The most common application areas identified were, in order of frequency of occurrence: information management, sensors and monitoring, security and privacy, information (data) pricing for data markets, and business

decision making support. Each of these areas is discussed in detail below.

1) INFORMATION MANAGEMENT

This is the largest application area identified with 15 papers on this topic. Within this theme the most common sub-topic is a grouping of 10 papers in the discipline of information lifecycle management (ILM) looking at data migration, data storage, and file management. This theme also includes the related fields of enterprise information management and data quality management (3 papers).

ILM is an industry term for managed dynamic and efficient storage resource management for the increased digital data being managed given the availability of multi-tier storage that trades access time for cost and the increasing prevalence of data legislation forcing compliance [30]. Central to the ILM approach is the idea that “not all corporate information [data] has the same value and values change over time” [6]. Chen [6] identifies the three key ILM tasks as information valuation, information characterisation & classification, and task prioritisation & optimisation. Given the relatively early date of Chen’s work (2005), the ongoing expansion of corporate data and the key role played by data valuation in ILM, it is not surprising that there is a significant body of work in this field.

Automating ILM is dependent on defining data value metrics that can be executed either in realtime or periodically to enable files [71], blocks [64] or other data [30] to be shifted between storage types or deleted. The relatively high availability of file metadata [6] and access information [46] characterising the numbers of data accesses means that both of these have been exploited by authors. Some authors caution that that metadata-driven approaches can require great effort to collect and maintain appropriate metadata and instead propose probabilistic approaches to metrics [46]. Despite the progress on data value assessment for ILC, Wijnhoven et al. [71] conducted a large scale comparison of subjective data value assessments conducted by experts using Sajko et al. [56] questionnaire and automated assessments and found a poor correlation. However, when identifying

TABLE 8. Thematic areas for data value application domains.

Information Management	Sensors and Monitoring	Security and Privacy
Information lifecycle management (4)	Environmental monitoring (1)	Security risk assessment (3)
Data migration/storage (4)	Sensing for earth sciences(1)	Security, crowd sensing, privacy (1)
File management (3)	Livestock production (2)	Privacy, data markets (2)
Enterprise Information Management (2)	Sensor fusion (1)	Privacy, IoT markets (2)
Data Quality Management (1)	Smart Vehicles (2)	Privacy, data brokers (1)
Data Markets and Information Pricing	Decision Making	Other Applications
Privacy, data markets (2)	Petroleum Reservoir development	Web Resource identification
Information pricing (2)	Military decision-making (1)	Information overload (1)
Data markets, disaster relief (2)	Supply chain management (3)	Multi-agent communication (2)
Machine learning (1)		Social media (1)

“wastage”, the least valuable files, they found that automated methods had an 80% accuracy, far better than when identifying the most valuable files.

Given the commercial importance of data retention, the works often include extensive evaluations in real-world settings. For example, Turczyk et al. [46] describe a case study that generates file migration rules for 150,000 files, whilst Wijnhoven et al. [71] conducted a case study with 77 employees of Capgemini Netherlands. Many of the studies propose domain specific file or block-based data value metrics that would be difficult to re-purpose for other use cases or domains [54]. However the work has matured over the last decade and fully automated ILM systems are now a reality in data centres and so the focus has switched from data value metrics to decision algorithms based on the metrics enabling fully automated ILM [47].

Enterprise information management [65] has largely grown out of the discipline of records management which focuses on managing information for legal compliance and supporting efficient operations of the organisation (governance). There is a growing number of records or data sources in organisations and a prioritisation mechanism is needed to deal with them most effectively, given limited resources. However this work is less mature than ILM use cases. Ladley [65] presents a list of loosely defined metrics and ideas on how to measure the value of information assets for Enterprise information management. Tallon [30] proposes a tiered information framework that, by considering the value of information allows CIOs to comprehend the interplay of market forces that shape information costs. Laney [5] devotes a chapter to EIM driven by thinking of data as an asset. He states that most EIM metrics are for justifying, funding, prioritising and gauging the success of initiatives for either managing data and business initiatives that use the data. This provides extensive use cases for data valuation in EIM. Unfortunately the metrics presented by Laney are less formally specified and thus hard to operationalise. Finally there research on value-driven data quality management (a sub-topic of EIM) which, has strong metrics since it aims

to enable automated processing of data [32]. Given the importance of data quality to modern machine learning and analytics pipelines, this is a promising topic for further research.

2) SENSORS AND MONITORING

Eight papers were identified as applying data value to sensor deployment or communication and monitoring systems. Topics discussed include a sensing system for earth sciences like satellites [60], environmental monitoring such as lake water quality [53], livestock production [86], sensor fusion architectures [57] and vehicular networks [51].

Many of the papers make use of the Value of Information (VoI) measure discussed above. In most cases probabilistic VoI approximation techniques are used, e.g. Bayesian [60] or Monte Carlo [53], due to the computational complexity of determining VoI exactly in realistic situations. Macauley [60] summarises that the value of data is based on “(1) How uncertain decision makers are; (2) What is at stake as an outcome of their decisions; (3) How much it will cost to use the information [data] to make decisions; and (4) The price of the next-best substitute for the information [data]”. It can be seen that answering these questions for specific domains can inform decision-making about data acquisition or use. These domain-specific answers limit the re-usability of VoI methods in other domains. For example Cornou and Kristensen [25] assign a value to knowledge of pig drinking behaviour that is very closely tied to their use case.

The key applications of data value estimates for these papers are to assess the worth of paying for additional monitoring capacity or information sources (justifying ICT system expense), optimising deployments of mobile sensors, minimising inter-sensor communications network load, and optimising data storage in low resource devices. In most of these applications the data value estimate can be used to prioritise data collection or communication activities. Thus, utility-based VoI models dominate. However, for monitoring system deployment scenarios [25], [60], [86] economic benefits are also important as the decision is a longer term one

about investment rather than short-term automated decision about mobile sensor movement [68], data stream selection [57], message transmission [51], or data storage [85].

3) SECURITY AND PRIVACY

This section discusses the eight papers that were classified as dealing with data value applications for security and privacy. There is a strong distinction between the two application areas: for security data value was used as a part of the risk assessment process [45], [56], [66], [72]; for privacy the emphasis was on placing an economic value on the private data that users exchange in return for “free” services [24], [69], [75], [76]. Within the privacy papers the topic of IoT (Internet of things) was seen as an important sub-application area for two of the papers. One paper was a cross-over that considered both security and privacy aspects in security system design influenced by data value [45].

In security risk assessment, for example as defined by the ISO 27705 standard, a key question is prioritisation of data assets to be protected by security systems. Identifying the most valuable assets has long been understood as part of this process [56]. This encourages consideration of a multi-dimensional view of data value. For example the early work of Sajko et al. [56] examined assessing timeliness, utility, replacement cost, legislative risk, and competitive advantage through a qualitative, questionnaire-based approach. Not all assessments are qualitative though, as Shimazu et al. [72] propose a quantitative assessment of risk using the value of information [data value], protection level, and threat level. Khokhlov and Reznik [45] provide mappings between data quality metrics and data value as a further extension of the quantitative approach. They also show how the computed data value can be used to influence security system design for a crowd-sensing application and other one in TOR (The Onion Router) network including privacy protection. Liu et al. [66] extend ISO27005-style risk assessment of data assets to assets in power control systems (smart grids).

The privacy related studies use economic (monetary) data value assessments (sometimes called “privacy valuation”) to drive competitive data trading [24], [69], or to raise users’ awareness of the value of personal data by providing transparent access to the value realised by service providers and data brokers [75], [76]. Game theoretic models play an important role in these markets whether trying to drive the market [24] or provide more transparency for users [76]. The data value model of Yassine et al. [69] is notable because instead of more usual data value dimensions it uses “risk of sharing private data” as a dimension of data value. The privacy data value papers are collectively characterised by a lack of strong implementation or evaluation. This may be due to the relative cost of integrating with many data broker platforms [75] or the immaturity of deployments of the IoT technology being discussed [24], [76].

4) DATA MARKETS AND INFORMATION PRICING

This topic was addressed by only seven of the papers analysed. This is interesting as data markets and pricing, i.e. a direct measure of the economic value of data, have received a lot of commercial attention in recent years but comparatively little seems to have been published on this topic, at least in the technical venues that we examined. Pricing for sale was seen as an end in itself by only two of the papers [55], [81] whereas three papers deal with data pricing in the context of privacy as a way of raising the awareness of users or increasing transparency of the transaction taking place between users and platform or service providers [24], [69], [75]. We also group in this theme methods that place an emphasis on data trading or data markets, even when they are not evaluated financially. This could be considered to include many of the “Value of Information” papers that are based on a games theoretic approach to the concept of a market for information. However there are also more concrete market mechanisms specified for encouraging data transfer in disaster situations [80] and markets for personal data [24]. Finally there is one paper discussing training data pricing for machine learning models [52].

Li et al. [81] describe a data pricing strategy based on information entropy and give a useful overview of different existing pricing strategies: subscription, query-based pricing, and bundling/discrimination-based pricing. It claims none of these are actually based on assessing the information value or contents of a dataset and presents a new information entropy-based way to measure the value of a dataset based on two value dimensions: size and information content. It provides an interesting evaluation of information content based on the ability of a dataset (or subset) to train a classifier and this is tested on six research datasets. A set of three example pricing functions to convert from entropy-based value to price are also provided and validated in use cases. Rao and Ng [55] instead define a utility-based pricing mechanism, however unlike most utility measures they define a method to estimate the utility of data before and after obfuscation for privacy purposes using Kolmogorov statistics. This is applicable to personal information but it is unclear how other applications of the approach would work.

Jia et al. [52] seek to answer how much each (additional) input training data point is worth for machine learning models based on a model of the relative value of data derived from its Shapley value from game theory. The Shapley value defines an optimal distribution of value to resources in a multi-actor system such as a market. Since this is very expensive to calculate ($O(2^N)$) the paper develops an approximation for k-nearest neighbours (KNN) machine learning models and a more practical Monte Carlo approximation algorithm. This analysis is based on having a utility function for the use of the data, which is often hard to estimate. They provide a classification of types of data valuation problems. The paper is significant as they explore the issues of data valuation at scale as they seek methods to assign value to each individual data point rather than an entire data-set or information source.

Given the importance of data for ML methods, it is to be expected that there will be more work in this direction in the future.

5) DECISION MAKING AND OTHER APPLICATIONS

This is a catch-all application domain category with 11 papers. In practice data value techniques are usually applied as an input to some form of decision making: should I use this data? do I need more data to make this decision? and so forth. The only sub-topic that contains multiple papers are supply chain management with 3 papers [35], [61], [63] but other areas include military decision making [70], insurance industry case studies [67] electricity markets [89], information overload [59], petroleum reservoir development [82], smart grids [79] and identifying relevant content in social media [73]. Several of these papers are based on VoI measures discussed above and do not contribute anything unique their domain application.

Supply chain management is significant in having a number of papers and the fact that two of them are relatively early (2000s). Dang et al. [63] say “Information resources management is the basis of supply chain” and that this is a source of competitive advantage. They propose the concept of a value chain of information resources that parallels their own focus on physical supply chain management. Viet et al. [35] provide a structured review of the value of information in different supply chain decisions focusing on articles between 2006 and 2017. They find that the focus has been on data availability rather than data characteristics (quality dimensions). They lay out a research agenda. The two most relevant questions for our study are: more research is needed on information characteristics (data quality dimensions) and a need for new methods to assess the value of data. They highlight the importance of enabling multi-method modelling approaches and so semantic models of data value such as DaVe [27] could have an important contribution there.

Other significant papers include Glissmann et al. [67] who create a model to dissect data value from the value generated by analytics in business decision-making in the context of an insurance industry example. The approach is based on models of business operations and the way information (raw and analytical) is connected to the organisation and its business priorities (business architecture/enterprise architecture). This is a tempting model for data governance operations but it is very unlikely that businesses will have an up to date catalog of their data assets or comprehensive business process models in a machine-readable format due the relative immaturity of data governance in most industries [5]. Scaffidi [73] discussed the use of data value measures to support the automatic identification of valuable Web resources in social media. This use of data value as a proxy for “importance” or “interest” is surely an area of overlap with other fields such as information retrieval and recommender systems that seek to satisfy a user’s desires by supplying their data needs. More work is required to examine this overlap and identify paths to cross-fertilise the research in both areas. Scaffidi’s

approach is interesting in that it depends on collecting both positive and negative examples of value and highlights the importance of authorship of content for value (which is linked to Wijndhoven et al.’s idea of incorporating the role or seniority of the people accessing the data in their usage-based value model for files [71]).

Laney [5] discusses techniques for applying data value measurements in a business context in chapter 11. First he introduces data value improvement through the idea that there are three degrees of value: i) realised value (current economic benefits), ii) probable value (based on intended uses) and iii) potential value (if optimally applied). The “information performance gap” is then defined as the difference between realised and probable value. It is suggested that his Market Value of Information (MVI) and Economic Value of Information (EVI) models can be used to calculate realised and probable value. The “information vision gap” is then defined as the difference between probable and potential values. Laney indicates that is harder to estimate. He suggests using Business Value of Information (BVI)-based actual versus potential valuations as a prioritisation technique. Unfortunately, the estimation of BVI is based on a crude estimate for relevance, which limits its applicability. A set of use cases for business decision-making are identified and a decision method based on the application and comparison of multiple of Laney’s data value models is defined. The use cases and methods are as follows: prioritising information asset management initiative investments, proving benefits of information governance, innovation and digitisation, monetisation and analytics, to help building a business case for monetising information assets, reducing information lifecycle expenses.

All of these methods are currently limited by the subjective models available, yet they represent significant insight into the types of decision-making possible and the relevant aspects of data value that we need be able to understand and quantify in order to maximise its utility.

V. RESEARCH AGENDA

The review results showed that despite [104] identifying “Data value measurement” as a key challenge for the future scope of data governance in their well received review [104], there are still many challenges that are hindering the efficient and effective exploitation of the concept of data value. We have identified the following three main challenges for data value research: a common conceptual framework, reusable assessment methods and tools, and future applications. Table 9 summarises the open research questions and topics of interest.

A. COMMON CONCEPTUAL FRAMEWORK

The first major challenge observed in the literature is the absence of a common conceptual framework for data value that sets a common basis for communication, research, and tool interoperability in terms of data value applications, context, stakeholders, models, terminology, and data value

dimensions [35]. Currently, most data value research is use-case or application focused (see Table 8) and identifies its own terminologies, models, and even the type of value they seek to derive from data. There are opportunities to develop data value ontologies [27], catalogues and other online resources to make terminologies, metrics, use cases and data value test data easier to exchange and reuse – similar to the efforts that have been done in the data quality engineering field [26], [99].

The few unified, multidimensional data value models like [14] are incomplete when compared with scope of data value identified in this review. At present the infonomics models of Laney [5] come closest to this ideal but many of their underlying metrics rely on subjective judgement and are not suitable for scalable, automated decision making. Fleckenstein et al. [39] explains that despite the utility of their framework of three types of data value models (market-based, economic and dimensional) that the “three models overlap with each other” and issues like regulation and privacy affect all models. All of the models are immature and lack a consistent way to represent the data value context which is external to the data and yet part of the data value assessment. This hinders the development and deployment of tool chains or frameworks for data value assessment and management or novel applications of data value to driving decisions in new domains like data governance [28]. A key problem for any unified models is how to reconcile, combine, and convert between financial models of data value measured in currency and scalar metrics typically measured in the range 0-1 [105]. Lastly, since there are several dimensions that could be included as part of a data value model, it is crucial that models provide machine-readable methods to score importance and provide means to aggregating the relevant dimensions and explaining their impact on individual use cases. Fleckenstein et al. [39] describes one model but it is human-oriented rather than addressing automation directly.

B. REUSABLE ASSESSMENT METHODS AND TOOLS

It is critical that the data value community has access to known, reusable techniques and tools for data value assessment.

Although this paper has presented the most comprehensive collection of data value metrics to date (see Tables 5, 4, 6, 7.), the coverage of individual data value dimensions has high variance. Many metrics are extremely specific to particular applications, especially in VoI models (see Sec. IV-B7). The relative frequency of subjective as opposed to objective metrics in the literature [5], [32], [45], [47], [56], [61], [70], [80] and the significant number of data value metrics based on abstract and idealised utility functions [52], [74], [79], [80], [92] that are hard to connect to real applications, shows that data value assessment is still immature. Objective metrics are also more suitable for automation and the creation of reusable assessment tools, as has long been common in the data quality

domain [26]. As Viet et al. [35] says there is “a need for new methods to assess the value of data”.

As well as lowering costs, common tools and metrics foster reproducibility of research findings and will facilitate the creation of open competitions and technical challenges by the research community. At present, the proliferation of individual assessment tools and models ensures divergence of results and duplication of effort.

Furthermore, although several metrics already exist for the assessment of data value metrics on static data, we live in the big data era with constantly updated data streams, yet there is little research to date on data value metrics for this domain. It is key to accelerate stream data value research to both analyse the applicability of existing works and propose novel assessment techniques that will dynamically capture the evolution of data value over time.

C. APPLICATIONS

The potential scope of applications for robust data value assessment techniques is currently unknown. We have seen that a very wide range of application types and domains are already represented (see section IV-C) with automated data value-driven decisions is most mature in the field of Enterprise Information Management Fleckenstein et al. [39] explains. However with modern, increasingly data-centric businesses the potential for data value must be significant. Machine learning applications are heavily data dependent and although data quality has started to be addressed in this field there are very few data value papers published on machine learning to date.

It is notable from this survey that many data value models and metrics are still only tested in simulations [51], [53], [60] or case studies, and there is a lack of longitudinal case studies on the experiences and benefits of data value-driven decision making. Enterprise Information Management is an exception with several significant deployments reported but the privacy and security application papers [24], [45], [75], [76] do have not a single deployment of automated, objective data value assessment reported.

However, the diversity of current approaches demands some synthesis for more widespread applications. For example, the lack of an established pipeline, lifecycle, or workflow for the application of data value impedes its adoption. This could span data value assessment, reporting, improvement, value-driven decisions, and other stages. Early work on data value monitoring capability maturity models exists [106] but this must be expanded to all lifecycle stages.

Common data value representations, perhaps based on semantic models [27], are key to both application and domain interoperability. At the business level, they will facilitate the comparison and integration of reported data value between multiple units in the same organisation, eliminating silos and enabling management of organisation-wide data value chains [1]. Beyond the organisation, this will provide better insights and collaboration opportunities to other stakeholders, e.g.

TABLE 9. Research areas for data value and potential research questions for future research.

Topics	Challenges and Research questions
Common Conceptual Framework Challenge	
Terminology	What would a unified terminology for data value consist of and how should it best be expressed and shared?
Data value models	Is there a set of data value dimensions that can be used across many use cases? Can application and stakeholder-independent predictive models of data value be developed? How should financial and scalar data value metrics be combined in data value models?
Multi-models	What is the most appropriate basis for combining multiple model results? How can common concerns like privacy be consistently included in multiple models?
Reusable Assessment Methods Challenge	
New metrics	What new methods or metrics are needed to comprehensively and effectively assess the value of data? To what extent can utility function-based metrics be linked to observable properties of data sets or services?
Open source tools	What is the potential for open source or widely deployed data value assessment tools?
New types of data	What interfaces, formats, standards, and protocols are needed to enable data value management tool-chains? How can stream processing techniques assess dynamic data value?
Applications Challenge	
Lifecycles	Would a standard data value lifecycle foster tool development and adoption?
Deployment	What are the differences between lab-based data value applications and real world deployments?
Decision making	What are the effects of data value mechanisms on decision-making performance? How can organisations use data value to enhance decision-making and break down siloes? How does data value evolve over time for different applications?
Machine learning	Can data value approaches be applied to data collection and curation for machine learning applications? How can machine learning approaches be best applied to data value assessment?

governments, policy-makers, economists, non-governmental agencies, press.

While there are already attempts to link data value with the proliferation of machine learning [88] techniques, it is surprising how little is being done. We are witnessing several breakthroughs in machine learning, but only a tiny fraction of them are related to data value (e.g., assessment of data quality metrics [107]). On the other hand, a large proportion of machine learning outcomes relies on data (collected, synthetic, or augmented) with many data quality issues [108]. Therefore, data value could play an important role in providing guidance and decision support mechanisms for machine learning. This collaboration would no doubt also produce more machine-learning based techniques for assessing data value.

VI. CONCLUSION

In this paper, we surveyed 63 existing works defining, characterising, modelling, and applying data value as a concept and for driving decision making (Table 2). We have identified that despite data value having conceptual origins back to at least 1980 in [31], the field is still immature with a lack of commonly agreed terminologies, models, or approaches to data value.

Our analysis found that there is a lack of generalised data value models and commonly understood dimensions to properly quantify the value of data, see Section IV-B.

This contrasts strongly with the more mature but related domain of data quality. This leads to an absence of common validation platforms and tools, which limits comparison of work, reproducibility, rate of progress, and industrial deployment. Nonetheless this survey has collected the most comprehensive list of data value metrics to date (Table 8) where it can be seen that the Usage dimension is the most often measured dimension of data value that is not already a known dimension of data quality.

Despite the increase in number of works in the area in recent years, it is clear that there is still many important research questions be resolved and a set have been collected in Table 9. Addressing these challenges will help organisations better understand and exploit their data more effectively. More mature data value techniques will enable them to quantify the value of their data assets more accurately and efficiently. This will not only help mitigate any data-related risks and enhance any data governance efforts, but also enable data-based decision making, including data acquisition and investment decisions, data maintenance decisions, innovation decisions, and also business decisions (e.g. in merger and acquisition scenarios).

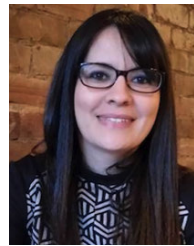
REFERENCES

- [1] E. Curry, "The big data value chain: Definitions, concepts, and theoretical approaches," in *New Horizons for a Data-Driven Economy*. Cham, Switzerland: Springer, 2016, pp. 29–37.

- [2] N. Côte-Real, T. Oliveira, and P. Ruivo, "Assessing business value of big data analytics in European firms," *J. Bus. Res.*, vol. 70, pp. 379–390, Jan. 2017.
- [3] B. P. Foster, R. Fletcher, and W. D. Stout, "Valuing intangible assets," *CPA J.*, vol. 73, no. 10, p. 50, 2003.
- [4] T. Fisher, *The Data Asset: How Smart Companies Govern Their Data for Business Success*. Hoboken, NJ, USA: Wiley, 2009.
- [5] D. B. Laney, *Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage*, 1st ed. Evanston, IL, USA: Routledge, Sep. 2017.
- [6] Y. Chen, "Information valuation for information lifecycle management," in *Proc. 2nd Int. Conf. Autonomic Comput. (ICAC)*, Jun. 2005, pp. 135–146.
- [7] S. Vidalis, "Calculating the value of information assets," Newport Bus. School Working Paper Series, Tech. Rep., Summer 2007, vol. 1.
- [8] J. Short and S. Todd, "What's your data worth?" *MIT Sloan Manage. Rev.*, vol. 58, no. 3, p. 17, 2017.
- [9] S. Powell, "Accounting for intangible assets: Current requirements, key players and future directions," *Eur. Accounting Rev.*, vol. 12, no. 4, pp. 797–811, Dec. 2003.
- [10] J. B. Stander, "The modern asset: Big data and information valuation," Ph.D. dissertation, Stellenbosch Univ., 2015.
- [11] J. Attard and R. Brennan, "DaVe: A semantic data value vocabulary to enable data value characterisation," in *Proc. 20th International Conf. Enterprise Inf. Syst. (ICEIS)*, S. Hammoudi, M. Śmiałek, O. Camp, and J. Filipe, Eds. Funchal, Portugal: Springer, 2019, pp. 239–261.
- [12] C. Higson and D. Waltho, "Valuing information as an asset," SAS, London, U.K., White Paper, 2009.
- [13] W. Engelsman, "Information assets and their value," in *Proc. 6th Twente Student Conf. IT*, The Netherlands, 2007.
- [14] G. Viscusi and C. Batini, "Digital information asset evaluation: Characteristics and dimensions," in *Smart Organizations and Smart Artifacts—Fostering Interaction Between People, Technologies and Processes*. Cham, Switzerland: Springer, 2014, pp. 77–86.
- [15] J. P. Caylor, R. J. Hammell, T. P. Hanratty, E. G. Heilman, and J. T. Richardson, "The effects of situational context on information valuation," in *Intelligent Human Systems Integration*, W. Karwowski and T. Ahram, Eds. Cham, Switzerland: Springer, 2019, pp. 465–470.
- [16] J. Mancini, "Infonomics: How do you measure the value of information?" AIIM Eur., Tech. Rep., 2016.
- [17] D. Moody and P. Walsh, "Measuring the value of information: An asset valuation approach," in *Proc. 7th Eur. Conf. Inf. Syst. (ECIS)*, 1999, pp. 1–17.
- [18] S. Al-Saffar and G. L. Heileman, "Semantic impact graphs for information valuation," in *Proc. 8th ACM Symp. Document Eng. (DocEng)*, New York, NY, USA, Sep. 2008, pp. 209–212.
- [19] B. Dalessandro, C. Perlich, and T. Raeder, "Bigger is better, but at what cost? Estimating the economic value of incremental data assets," *Big Data*, vol. 2, no. 2, pp. 87–96, Jun. 2014.
- [20] J. Houseal, "Intangible asset valuation approaches and methods," *Eur. J. Appl. Physiol.*, pp. 85–90, 2012.
- [21] K. King, "The value of intellectual property, intangible assets, and goodwill," *J. Intellectual Property Rights*, vol. 7, pp. 1–4, May 2002.
- [22] Y. Chen, "Information valuation for information lifecycle management," in *Proc. 2nd Int. Conf. Autonomic Comput. (ICAC)*, Jun. 2005, pp. 135–146. [Online]. Available: <http://ieeexplore.ieee.org/document/1498059>
- [23] H. Jin, M. Xiong, and S. Wu, "Information value evaluation model for ILM," in *Proc. 9th ACIS Int. Conf. Softw. Eng., Artif. Intell., Netw., Parallel/Distrib. Comput.*, Aug. 2008, pp. 543–548.
- [24] H. Oh, S. Park, G. M. Lee, J. K. Choi, and S. Noh, "Competitive data trading model with privacy valuation for multiple stakeholders in IoT data markets," *IEEE Internet Things J.*, vol. 7, no. 4, pp. 3623–3639, Apr. 2020.
- [25] C. Cornou and A. R. Kristensen, "Use of information from monitoring and decision support systems in pig production: Collection, applications and expected benefits," *Livestock Sci.*, vol. 157, nos. 2–3, pp. 552–567, Nov. 2013.
- [26] A. Zaveri, A. Rula, A. Maurino, R. Pietrobon, J. Lehmann, and S. Auer, "Quality assessment for linked data: A survey," *Semantic Web*, vol. 7, no. 1, pp. 63–93, Mar. 2015.
- [27] J. Attard and R. Brennan, "A semantic data value vocabulary supporting data value assessment and measurement integration," in *Proc. 20th Int. Conf. Enterprise Inf. Syst. (ICEIS)*, vol. 2, 2018, pp. 1–12.
- [28] J. Attard and R. Brennan, "Challenges in value-driven data governance," in *Proc. OTM Confederated Int. Conf. Move Meaningful Internet Syst.* Valletta, Malta: Springer, Oct. 2018, pp. 546–554.
- [29] B. Otto, "Quality and value of the data resource in large enterprises," *Inf. Syst. Manage.*, vol. 32, no. 3, pp. 234–251, Jul. 2015. [Online]. Available: <http://www.tandfonline.com/doi/full/10.1080/10580530.2015.1044344>
- [30] P. P. Tallon, "Understanding the dynamics of information management costs," *Commun. ACM*, vol. 53, no. 5, pp. 121–125, May 2010.
- [31] N. Ahituv, "A systematic approach toward assessing the value of an information system," *MIS Quart.*, pp. 61–75, Dec. 1980.
- [32] A. Even and G. Shankaranarayanan, "Value-driven data quality assessment," in *Proc. Int. Conf. Inf. Quality (MIT IQ Conf.)*, 2005, pp. 1–15.
- [33] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," *Int. J. Inf. Manage.*, vol. 35, no. 2, pp. 137–144, Apr. 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0268401214001066>
- [34] P. P. Tallon, "Corporate governance of big data: Perspectives on value, risk, and cost," *Computer*, vol. 46, no. 6, pp. 32–38, Jun. 2013.
- [35] N. Q. Viet, B. Behdani, and J. Bloemhof, "The value of information in supply chain decisions: A review of the literature and research agenda," *Comput. Ind. Eng.*, vol. 120, pp. 68–82, Jun. 2018.
- [36] F. Alawad and F. A. Kraemer, "Value of information in wireless sensor network applications and the IoT: A review," *IEEE Sensors J.*, vol. 22, no. 10, pp. 9228–9245, May 2022.
- [37] A. Z. Faroukhi, I. El Alaoui, Y. Gahi, and A. Amine, "Big data monetization throughout big data value chain: A comprehensive review," *J. Big Data*, vol. 7, no. 1, pp. 1–22, Dec. 2020.
- [38] W. Yanlin and Z. Haijun, "Data asset value assessment literature review and prospect," *J. Phys., Conf. Ser.*, vol. 1550, no. 3, May 2020, Art. no. 032133.
- [39] M. Fleckenstein, A. Obaidi, and N. Tryfona, "A review of data valuation approaches and building and scoring a data valuation model," *Harvard Data Sci. Rev.*, vol. 5, no. 1, pp. 4–36, Jan. 2023. [Online]. Available: <https://hdsr.mitpress.mit.edu/pub/1qxrkrng>
- [40] D. I. Stern, "Use value, exchange value, and resource scarcity," *Energy Policy*, vol. 27, no. 8, pp. 469–476, Aug. 1999.
- [41] A. Ghorbani and J. Zou, "Data shapley: Equitable valuation of data for machine learning," Jun. 2019, *arXiv:1904.02868*.
- [42] Y.-W. Kim and S. Lee, "Data valuation algorithm for inertial measurement unit-based human activity recognition," *Sensors*, vol. 23, no. 1, p. 184, Dec. 2022. [Online]. Available: <https://www.mdpi.com/1424-8220/23/1/184>
- [43] J. Yoon, S. O. Arik, and T. Pfister, "Data valuation using reinforcement learning," Sep. 2019, *arXiv:1909.11671*.
- [44] B. Wang, Q. Guo, T. Yang, L. Xu, and H. Sun, "Data valuation for decision-making with uncertainty in energy transactions: A case of the two-settlement market system," *Appl. Energy*, vol. 288, Apr. 2021, Art. no. 116643. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S030626192100177X>
- [45] I. Khokhlov and L. Reznik, "What is the value of data value in practical security applications," in *Proc. IEEE Syst. Secur. Symp. (SSS)*, Jul. 2020, pp. 1–8.
- [46] L. A. Turczyk, O. Heckmann, and R. Steinmetz, "Simulation of information lifecycle management," in *Proc. 18th Annu. Inf. Resour. Manag. Assoc. Conf. (IRMA)*, Vancouver, BC, Canada, 2007, pp. 1063–1066.
- [47] S. Qiu, D. Zhang, and X. Du, "An evaluation method of data valuation based on analytic hierarchy process," in *Proc. 14th Int. Symp. Pervasive Syst., Algorithms Netw., 11th Int. Conf. Frontier Comput. Sci. Technol., 3rd Int. Symp. Creative Comput. (ISPAN-FCST-ISCC)*, Jun. 2017, pp. 524–528.
- [48] M. Pai, M. McCulloch, and J. Colford, "Systematic review: A road map version 2.2. Systematic reviews group, UC Berkeley, 2002," Tech. Rep., 2004.
- [49] B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering—A systematic literature review," *Inf. Softw. Technol.*, vol. 51, no. 1, pp. 7–15, 2009.

- [50] B. Kitchenham, R. Pretorius, D. Budgen, O. P. Brereton, M. Turner, M. Niazi, and S. Linkman, "Systematic literature reviews in software engineering—A tertiary study," *Inf. Softw. Technol.*, vol. 52, no. 8, pp. 792–805, 2010.
- [51] M. Giordani, T. Higuchi, A. Zanella, O. Altintas, and M. Zorzi, "A framework to assess value of information in future vehicular networks," in *Proc. 1st ACM MobiHoc Workshop Technol., Models, Protocols Cooperat. Connected Cars*, Jul. 2019, pp. 31–36.
- [52] R. Jia, D. Dao, B. Wang, F. A. Hubis, N. M. Gurel, B. Li, C. Zhang, C. J. Spanos, and D. Song, "Efficient task-specific data valuation for nearest neighbor algorithms," 2019, *arXiv:1908.08619*.
- [53] V. Koski, N. Kotamäki, H. Hämäläinen, K. Meissner, J. Karvanen, and S. Kärkkäinen, "The value of perfect and imperfect information in lake monitoring and management," *Sci. Total Environ.*, vol. 726, Jul. 2020, Art. no. 138396.
- [54] X. Zhao, Z. Li, and L. Zeng, "A hierarchical storage strategy based on block-level data valuation," in *Proc. 4th Int. Conf. Netw. Comput. Adv. Inf. Manage.*, vol. 1, Sep. 2008, pp. 36–41.
- [55] D. Rao and W. K. Ng, "Information pricing: A utility based pricing mechanism," in *Proc. IEEE 14th Int. Conf. Dependable, Autonomic Secure Comput., 14th Int. Conf. Pervasive Intell. Comput., 2nd Int. Conf. Big Data Intell. Comput. Cyber Sci. Technol. Congr. (DASC/PiCom/DataCom/CyberSciTech)*, Aug. 2016, pp. 754–760.
- [56] M. Sajko, K. Rabuzin, and M. Bača, "How to calculate information value for effective security risk assessment," *J. Inf. Org. Sci.*, vol. 30, no. 2, pp. 263–278, 2006.
- [57] T. J. Peterson and M. K. Sundareshan, "Information value mapping for fusion architectures," in *Proc. 3rd Int. Conf. Inf. Fusion*, vol. 1, 2000, p. MOC5/18.
- [58] P. Missier, G. Lalk, V. Verykios, F. Grillo, T. Lorusso, and P. Angeletti, "Improving data quality in practice: A case study in the Italian public administration," *Distrib. Parallel Databases*, vol. 13, no. 2, pp. 135–160, 2003.
- [59] U. Meissen, S. Pfennigschmidt, T. Wahnfried, and K. Sandkuhl, "Situation-based message rating in information logistics and its applicability in collaboration scenarios," in *Proc. 30th Euromicro Conf.*, 2004, pp. 484–491.
- [60] M. K. Macauley, "The value of information: Measuring the contribution of space-derived earth science data to resource management," *Space Policy*, vol. 22, no. 4, pp. 274–282, Nov. 2006.
- [61] V. Doctori-Blass and R. Geyer, "The role and value of information for supply loop management: Framework and application for the end-of-life cell phone industry," in *Proc. IEEE Int. Symp. Electron. Environ.*, May 2008, pp. 1–6.
- [62] Y. Lu and H.-W. Zhu, "An approach for quantifying enterprise value of information (EVI)," in *Proc. Int. Conf. Manage. Sci. Eng.*, Sep. 2009, pp. 594–600.
- [63] W. Dang, Y.-M. Sun, and T.-T. Wang, "The value research on information resources management of supply chains," in *Proc. Int. Conf. Manage. Sci. Eng.*, Sep. 2009, pp. 625–629.
- [64] X. Zhao, Z. Li, X. Zhang, and L. Zeng, "Block-level data migration in tiered storage system," in *Proc. 2nd Int. Conf. Comput. Netw. Technol.*, 2010, pp. 181–185.
- [65] J. Ladley, "Measuring EIM and the value of information," in *Making EIM Enterprise Information Management Work for Business*, J. Ladley, Ed. Boston, MA, USA: Morgan Kaufmann, 2010, ch. 28, pp. 377–388. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B978012375695400028X>
- [66] N. Liu, J. Zhang, and X. Wu, "Asset analysis of risk assessment for IEC 61850-based power control systems—Part I: Methodology," *IEEE Trans. Power Del.*, vol. 26, no. 2, pp. 869–875, Apr. 2011.
- [67] S. Glissmann, J. Sanz, R. Liu, and V. Becker, "Disentangling the value of information and analytics through componentized business architecture," in *Proc. 45th Hawaii Int. Conf. Syst. Sci.*, Jan. 2012, pp. 4250–4259.
- [68] P. Jalalkamali and R. Olfati-Saber, "Information-driven self-deployment and dynamic sensor coverage for mobile sensor networks," in *Proc. Amer. Control Conf. (ACC)*, Jun. 2012, pp. 4933–4938.
- [69] A. Yassine, A. A. N. Shirehijini, S. Shirmohammadi, and T. T. Tran, "Knowledge-empowered agent information system for privacy payoff in eCommerce," *Knowl. Inf. Syst.*, vol. 32, no. 2, pp. 445–473, Aug. 2012.
- [70] T. P. Hanratty, R. J. Hammell II, B. A. Bodt, E. G. Heilman, and J. C. Dumer, "Enhancing battlefield situational awareness through fuzzy-based value of information," in *Proc. 46th Hawaii Int. Conf. Syst. Sci.*, Jan. 2013, pp. 1402–1411.
- [71] F. Wijnhoven, C. Amrit, and P. Dietz, "Value-based file retention," *J. Data Inf. Qual.*, vol. 4, no. 4, pp. 1–17, May 2014.
- [72] M. Shimazu, A. Kanai, S. Tanimoto, and H. Sato, "Dynamic risk evaluation model as a security field," in *Proc. IEEE 4th Global Conf. Consum. Electron. (GCCE)*, Oct. 2015, pp. 628–629.
- [73] C. Scaffidi, "Mining online forums for valuable contributions," in *Proc. 11th Iberian Conf. Inf. Syst. Technol. (CISTI)*, Jun. 2016, pp. 1–6.
- [74] T. Soleymani, S. Hirche, and J. S. Baras, "Optimal self-driven sampling for estimation based on value of information," in *Proc. 13th Int. Workshop Discrete Event Syst. (WODES)*, May 2016, pp. 183–188.
- [75] Y. M. Kassa, J. Gonzalez, Á. Cuevas, R. Cuevas, M. Marciel, and R. González, "Your data in the eyes of the beholders: Design of a unified data valuation portal to estimate value of personal information from market perspective," in *Proc. 11th Int. Conf. Availability, Rel. Secur. (ARES)*, Aug. 2016, pp. 701–705.
- [76] A. Mayle, N. H. Bidoki, S. Masnadi, L. Boeloeni, and D. Turgut, "Investigating the value of privacy within the Internet of Things," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2017, pp. 1–6.
- [77] G. Schuh, R. Kreutzer, and M. Patzwald, "Assessing the value of data: An approach to evaluate the technology driven benefits of smart product data," in *Proc. Portland Int. Conf. Manage. Eng. Technol. (PICMET)*, Jul. 2017, pp. 1–10.
- [78] H. Hu, Y. Xu, Y. Zhang, and M. Liu, "A fast graphic-based information valuation algorithm for cooperative information sharing," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2017, pp. 1892–1897.
- [79] S. Tan, C. Jiang, and Y. He, "A research on information value in energy Internet based on risk theory," in *Proc. IEEE Conf. Energy Internet Energy Syst. Integr. (EI)*, Nov. 2017, pp. 1–6.
- [80] T. Sonobe, A. Takahashi, and T. Sugauma, "Design of information value determination method for information-sharing systems during large-scale disasters," in *Proc. IEEE 16th Int. Conf. Cognit. Informat. Cognit. Comput. (ICCI*CC)*, Jul. 2017, pp. 482–489.
- [81] X. Li, J. Yao, X. Liu, and H. Guan, "A first look at information entropy-based data pricing," in *Proc. IEEE 37th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Jun. 2017, pp. 2053–2060.
- [82] S. M. G. Santos, A. T. F. S. Gaspar, and D. J. Schiozer, "Value of information in reservoir development projects: Technical indicators to prioritize uncertainties and information sources," *J. Petroleum Sci. Eng.*, vol. 157, pp. 1179–1191, Aug. 2017.
- [83] S. Sathanathan, "Data valuation considering knowledge transformation, process models and data models," in *Proc. 12th Int. Conf. Res. Challenges Inf. Sci. (RCIS)*, May 2018, pp. 1–5.
- [84] W. Fauriat and E. Zio, "An importance measure to assess the value of a component inspection policy," in *Proc. 3rd Int. Conf. Syst. Rel. Saf. (ICRSRS)*, Nov. 2018, pp. 368–375.
- [85] Y. Yao and E. Atkins, "The smart black box: A value-driven automotive event data recorder," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 973–978.
- [86] C. Rojo-Gimeno, V. Fievez, and E. Wauters, "The economic value of information provided by milk biomarkers under different scenarios: Case-study of an ex-ante analysis of fat-to-protein ratio and fatty acid profile to detect subacute ruminal acidosis in dairy cows," *Livestock Sci.*, vol. 211, pp. 30–41, May 2018.
- [87] C. Malings and M. Pozzi, "Value-of-information in spatio-temporal systems: Sensor placement and scheduling," *Rel. Eng. Syst. Saf.*, vol. 172, pp. 45–57, Apr. 2018.
- [88] T. Song, Y. Tong, and S. Wei, "Profit allocation for federated learning," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2019, pp. 2577–2586.
- [89] Y. Chen, B. Wang, Q. Guo, X. Feng, E. Lu, D. Xiang, and B. Wen, "Information value based on the scenario of wind power trading in electricity markets," in *Proc. IEEE 3rd Conf. Energy Internet Energy Syst. Integr. (EI)*, Nov. 2019, pp. 429–434.
- [90] Y. Zhang, Y. Huang, D. Zhang, and Y. Qian, "The importance of data assets and its accounting confirmation and measurement methods," in *Proc. 6th Int. Conf. Behav., Econ. Socio-Cultural Comput. (BESC)*, Oct. 2019, pp. 1–8.
- [91] Z. Li, Y. Ni, X. Gao, and G. Cai, "Value evaluation of data assets: Progress and enlightenment," in *Proc. IEEE 4th Int. Conf. Big Data Anal. (ICBDA)*, Mar. 2019, pp. 88–93.
- [92] C. Zhang, Z. Wang, and A. Shafieezadeh, "Value of information analysis via active learning and knowledge sharing in error-controlled adaptive kriging," *IEEE Access*, vol. 8, pp. 51021–51034, 2020.

- [93] H. Stein, L. Holst, V. Stich, and W. Maass, "From qualitative to quantitative data valuation in manufacturing companies," in *Proc. IFIP Int. Conf. Adv. Prod. Manage. Syst.* Cham, Switzerland: Springer, 2021, pp. 172–180.
- [94] B. Wang, Q. Guo, T. Yang, L. Xu, and H. Sun, "Data valuation for decision-making with uncertainty in energy transactions: A case of the two-settlement market system," *Appl. Energy*, vol. 288, Apr. 2021, Art. no. 116643.
- [95] Q. He, Y. Qiao, S. Yang, and C. Wang, "Equitable valuation of crowdsensing for machine learning via game theory," in *Proc. Int. Conf. Wireless Algorithms, Syst., Appl.* Nanjing, China: Springer, Jun. 2021, pp. 133–141.
- [96] M. Noshad, J. Choi, Y. Sun, A. Hero, and I. D. Dinov, "A data value metric for quantifying information content and utility," *J. Big Data*, vol. 8, no. 1, pp. 1–23, Dec. 2021.
- [97] W. Hofman, J. Migeotte, M. Labare, B. Rukanova, and Y.-H. Tan, "Using business data in customs risk management: Data quality and data value perspective," in *Proc. 20th IFIP WG 8.5 Int. Conf. Electron. Government (EGOV)*, Granada, Spain. Berlin, Germany: Springer-Verlag, Sep. 2021, pp. 271–287, doi: [10.1007/978-3-030-84789-0_20](https://doi.org/10.1007/978-3-030-84789-0_20).
- [98] C. Lin, J. Song, and M. Pozzi, "Optimal inspection of binary systems via value of information analysis," *Rel. Eng. Syst. Saf.*, vol. 217, Jan. 2022, Art. no. 107944.
- [99] J. Wang, Y. Liu, P. Li, Z. Lin, S. Sindakis, and S. Aggarwal, "Overview of data quality: Examining the dimensions, antecedents, and impacts of data quality," *J. Knowl. Economy*, pp. 1–20, Feb. 2023.
- [100] M. Kaiser, M. Klier, and B. Heinrich, "How to measure data quality—A metric-based approach," in *Proc. ICIS*, 2007, p. 108.
- [101] C. Maina, "Just-in-case or just-in-time library services: Option and usage valuation of libraries and information services," *Can. J. Inf. Library Sci.-Reveu Canadienne des Sciences de l'Informaton et de Bibliotheconomie*, vol. 28, no. 3, pp. 96–97, Sep. 2004.
- [102] J. M. Keisler, Z. A. Collier, E. Chu, N. Sinatra, and I. Linkov, "Value of information analysis: The state of application," *Environ. Syst. Decisions*, vol. 34, no. 1, pp. 3–23, Mar. 2014.
- [103] E. C. F. Wilson, "A practical guide to value of information analysis," *Pharmacoeconomics*, vol. 33, no. 2, pp. 105–121, Feb. 2015.
- [104] R. Abraham, J. Schneider, and J. vom Brocke, "Data governance: A conceptual framework, structured review, and research agenda," *Int. J. Inf. Manage.*, vol. 49, pp. 424–438, Dec. 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0268401219300787>
- [105] B. Heinrich, D. Hristova, M. Klier, A. Schiller, and M. Szubartowicz, "Requirements for data quality metrics," *J. Data Inf. Qual.*, vol. 9, no. 2, pp. 1–32, Jun. 2017.
- [106] R. Brennan, J. Attard, and M. Helfert, "Management of data value chains, a value monitoring capability maturity model," in *Proc. 20th Int. Conf. Enterprise Inf. Syst. (ICEIS)*, 2018, pp. 1–12.
- [107] V. Walter, A. Gyoery, and C. Legner, "Deploying machine learning based data quality controls—design principles and insights from the field," in *Proc. 17th Int. Conf. Wirtschaftsinformatik*, 2022, pp. 1–15.
- [108] V. Gudivada, A. Apon, and J. Ding, "Data quality considerations for big data and machine learning: Going beyond data cleaning and transformations," *Int. J. Adv. Softw.*, vol. 10, no. 1, pp. 1–20, 2017.



MALIKA BENDECHACHE received the Ph.D. degree in computer science from University College Dublin, Ireland. She is currently a Lecturer/an Assistant Professor with the School of Computer Science, University of Galway, Ireland, and a Funded Investigator with the ADAPT Centre for AI-Driven Digital Content Technology. Her current research interests include big data analytics, machine learning, AI governance and data governance, and security and privacy. She designs novel big data analytics and machine learning techniques to enhance the capability and efficiency of complex systems. She also leverages complex systems to improve the effectiveness, privacy, and trustworthiness of analytics/machine learning techniques.



JUDIE ATTARD received the Ph.D. degree in computer science from the University of Bonn, Germany. She was a Research Fellow with the ADAPT Research Centre, Trinity College, Dublin. She is currently a Data Governance Manager with a leading online bank in Malta. Her current interests include the extraction of value from data assets, including data governance, open data, data quality, and semantics. She was awarded an EDGE Marie Skłodowska-Curie COFUND Fellowship, in 2018.



MALICK EBIELE received the B.S. degree in computer science and computer engineering from l'Universite Nangui Abrogoua (ex Universite d'Abobo-Adjame), Côte d'Ivoire (Ivory Coast), in 2018, and the M.Sc. degree in mathematical sciences from the African Institute for Mathematical Sciences (AIMS), Rwanda, in 2020. He is currently pursuing the Ph.D. degree in data value analytics and value-driven governance with the School of Computer Science, University College Dublin, Ireland.



ROB BRENNAN (Senior Member, IEEE) received the degree from Dublin City University and the degree from Queens University Belfast. He is currently a Lecturer/an Assistant Professor with the School of Computer Science, University College Dublin, Ireland, and a Funded Investigator with the Science Foundation Ireland ADAPT Centre for AI-Driven Digital Content Technology. In ADAPT, he leads the value and risk challenge with the Transparent Digital Governance Research Strand. His current research interests include data protection, data value, data quality, semantics, data governance, and AI governance.

...