

# Challenges in Value-Driven Data Governance

Judie Attard and Rob Brennan

KDEG, ADAPT Centre, School of Computer Science and Statistics, O'Reilly  
Institute, Trinity College Dublin, Dublin 2, Ireland  
{attardj,rob.brennan}@cs.tcd.ie

**Abstract.** Data is quite popularly considered to be the new oil since it has become a valuable commodity. This has resulted in many entities and businesses that hoard data with the aim of exploiting it. Yet, the ‘simple’ exploitation of data results in entities who are not obtaining the highest benefits from the data, which as yet is not considered to be a fully-fledged enterprise asset. Such data can exist in a duplicated, fragmented, and isolated form, and the sheer volume of available data further complicates the situation. Issues such as the latter highlight the need for value-based data governance, where the management of data assets is based on the quantification of the data value. This quantification will provide an opportunity for evidence-based approaches to data governance. This paper has the purpose of creating awareness and further understanding of challenges that result in untapped data value. We identify niches in related work, and through our experience with businesses who use data assets, we here analyse four main context-independent challenges that hinder entities from achieving the full benefits of using their data. This will aid in the advancement of the field of value-driven data governance and therefore directly affect data asset exploitation.

**Keywords:** Data governance · data value · data asset · data exploitation.

## 1 Introduction

The exponential growth in the availability of data has led to an evident increase of companies that use data as an enterprise asset. While money and people have been considered to be enterprise assets for a long time, data is as yet hardly considered to be so [14]. Yet, organisations and companies are increasingly relying on their data to become more competitive, for example, by having greater knowledge of their customers, by taking more informed decisions, by finding new innovative uses for the data, by controlling risks and cutting costs, and also by innovating upon this data. Such use of data assets enables companies to not only better achieve their goals, but also to improve their financial performance.

As the volumes of data continue to rise, and enterprises and organisations are increasingly relying on data, data is being duplicated, fragmented, and isolated into various silos [1]. The sheer volume of data also affects data quality, since data standards cannot be enforced so easily. Such data may lead to additional

risks, increased costs, non efficient processes, and therefore potential business losses [1, 7]. Moreover, although there is the general understanding that data of a high quality leads to more benefits when the data is exploited, there might be a lack of understanding on the process of managing this data, as well as the resulting business impact of using it.

These issues already show a clear need to manage data assets. DAMA International define data governance to be “the exercise of authority and control (planning, monitoring, and enforcement) over the management of data assets” [9]. Data governance is therefore the management of data architecture, data quality, data security, data operations, etc. Tasks include the setting, monitoring and enforcing of policies, standards, and procedures; the coordination, maintenance, and implementation of data architecture; the acquisition of data assets and the monitoring of their costs, quality and security, and the creation of data ownership rights. Data governance therefore enables the effective use of data assets.

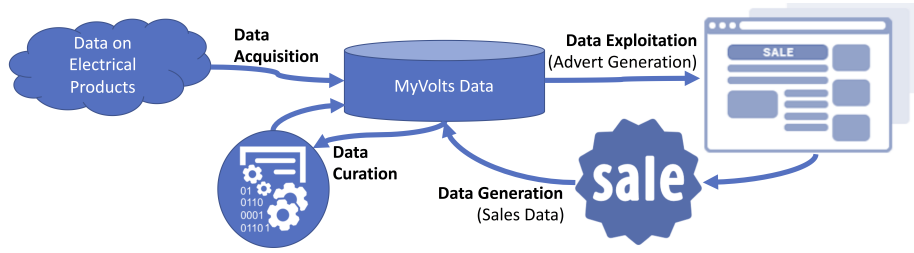
Since data governance and management efforts and investments are on the rise, it is becoming increasingly relevant to identify the economic value of data and the return on investment. Data value has been used as a basis for organisational decision making on quality [11], but also as a part of automated control systems for data lifecycles [8] and file retention [23]. Failing to value data will result in a number of consequences such as retaining information that has little to no value, reduction in data usage, and leaving data investments vulnerable to budget cuts [16]. Hence, data value is an aspect that plays a very important role in data governance. The issue is that although gaining recognition as a valuable asset, data has as yet resisted quantitative measurement, and data value is as yet mostly limited to be a notional value [17].

The aim of this paper is to identify and analyse the challenges that hinder entities from enjoying the full benefits of exploiting data as assets, in context of value-driven data governance. Existing literature cover various aspects of data value, however there is as yet no consensus on how to measure or quantify the value of data. There are also additional challenges that hinder the valuation of data, and this in turn makes data governance efforts more demanding. The contribution of this paper will therefore be invaluable for any entity that exploits data as an asset, in an effort to optimise data governance efforts.

The rest of the paper is structured as follows: in Section 2 we provide an overview of identified challenges within a real use case, where a business exploits data assets with the aim of gaining competitive advantage, in Section 3 we analyse and discuss value-driven data governance challenges (as identified in niches in related work, and through our experience with businesses who use data assets), and finally in Section 4 we deliver our concluding remarks.

## 2 Use Case

In this section we present MyVolts; a company that uses data assets in order to obtain a competitive edge, as a use case with the aim of providing a first overview of challenges in value-driven data governance.



**Fig. 1.** MyVolts Data Value Chain Example

MyVolts is a successful SME with a 15 year track record that develops and operates a highly automated internet retail and business intelligence system. They currently operate in 4 countries in Europe, namely Ireland, the UK, France, and Germany, and also in the USA. In these countries MyVolts is a leading source for consumer device power supplies. Figure 1 shows an example of a data value chain within MyVolts. As indicated in the figure, MyVolts performs the following data value processes (amongst others) on their data assets:

- **Data Acquisition:** This company gathers data which includes data on their customers, the evolving market of power supply device specifications, and the power supply needs of all consumer electronics. They collect this data by monitoring social media, web sales data such as Amazon top seller lists, customer queries and complaints, and device manufacturer homepages. New consumer electronic devices must be discovered, categorised, profiled for potential sales value and have their power supply technical specifications (voltage, polarity, tip type and dimensions) mined from open web data. There are an estimated 5.5 million consumer electronics devices on sale today and the number of powered devices is growing rapidly.
- **Data Curation:** The lack of standardised machine-readable repositories means that PDF is the dominant data publication format. Integrating this data while maintaining strict quality control is a major challenge for MyVolts. They also need to prioritise the data curation tasks to ensure optimal fitness for use for the ensuing data exploitation processes.
- **Data Exploitation:** In this process MyVolts use their data assets to create adverts for their products. This process is a decision-rich process that requires to identify which products need advertising (e.g. products with low sales, new products, products on offer) and which consumers to target. Data exploitation is therefore a process that also requires tapping into various data assets and potentially also integrating them.
- **Data Generation:** This process is an ongoing process resulting from product sales through the MyVolts website. This information, as specified above, will be used to create targeted adverts to optimise sales, and therefore also increase profits.

## 2.1 Challenges in MyVolts Use Case

The sample data value chain detailed above already provides us with a number of challenges that stem from using data assets as a basis for competitive advantage.

The first challenge is the quantification of the value of data as it is being acquired. MyVolts need to be able to **measure the value** of this data in order to identify whether this data is worth their effort and/or money. This quantification will not only enable MyVolts to reduce the risk of investing poorly in the data acquisition process, but also help target company efforts and aid decision making in the data exploitation and data curation processes.

The first challenge is directly related to the second challenge; **what makes data valuable?** In this use case, in order to be valuable, data needs to be reliable, timely, relevant, accurate, with good potential for impact once its used, and preferably even unique (not available to other competitors). A further aspect of this challenge is the context of use; what might be valuable data for one use might be irrelevant for another. Therefore, different contexts of use will have different requirements as to what makes data valuable.

In general, all four data value processes shown in Figure 1 highlight the requirement for data governance. A successful effort to exploit data assets and achieve competitive advantage requires various data governance tasks, including the definition of roles; data policies, standards, and procedures; the definition of an interoperable data architecture; and data storage and organisation. Therefore, we here identify the need of implementing **value-driven data governance model**. Such a evidence-based model would need to allow data assets to flow within the business or organisation, provide insight into what are the inputs and outputs of the existing processes, and also identify how these processes provide value to the business with regard to achieving the company goals.

Finally, we also identify the challenge of **optimising data governance** for the specific use case. An ideal data governance approach does not only encompass data governance tasks, but is also tailored to the use case in question. In this use case, decision making is a recurring process that exists throughout the data value chain. For instance, MyVolts employees need to decide which data to acquire, how much is it worth paying for, what data to maintain and what data to discard, what products to advertise, etc. Building upon the previously-mentioned challenges, the optimisation of data governance will result in efficient and effective use of data assets whilst minimising costs and achieving company goals (such as increasing profits, innovative use of data, etc).

## 3 Challenges in Value-Driven Data Governance

In the following section we discuss challenges that hinder entities from effectively exploiting data assets. We base our analysis on existing related work and on our experience with businesses exploiting data assets.



### 3.1 Defining Data Value

Data value is recognised as a “key issue in information systems management” [7]. Yet, while most research on information or data value seeks to identify dimensions that characterise it, there is still **no consensus on the definition of data value**. In fact, the multi-dimensional nature of value, as well as the role context plays in data value quantifying efforts, make the definition of data value quite challenging. The interdisciplinary nature of this field also adds to the complexity of this task.

Different data has varied value in different contexts (e.g. different points in time [8], different consumers [14]) depending on a number of dimensions [21]. Ahituv [2], for example, suggests timeliness, content, and cost as data value dimensions, which clearly parallel modern research on data quality dimensions [3]. Even and Shankaranarayanan follow a similar reasoning where they focus on the intrinsic value of data and consider data quality dimensions that are both context independent and context dependent [11]. Infonomics [14] is an economically-based discipline devoted to information valuation but, in common with many economic theories, it focuses on utility functions and other explanatory models. Laney also explores the applicability to the business and the availability to competitors as dimensions of data value [15]. Chen, on the other hand, presents an information valuation approach that quantifies the value of a given piece of information based on its usage over time [8]. Along with usage, Sajko et al. also define data value dimensions to consist of meaning to business (through profits evaluation and utility), cost, and timeliness of data [18].

The above dimensions, shown in Table 1, are but a few of the existing dimensions that are used in literature to characterise data value. Yet, most authors in literature seem to agree that context is a data value dimension that is quite relevant in any data value definition or quantification effort [10]. Whilst initial efforts have already been made with regards to providing a semantic data value vocabulary that can be used to comprehensively define data value [5], this heterogeneity of dimensions in literature indicates the need for terminological unification, which would also aid in developing a common understanding of the domain.

	Data Value Dimensions						
	Usage	Cost	Timeliness	Quality	Applicability to Business	Utility	Uniqueness
Ahituv [2]		✓	✓		✓		
Chen [8]	✓	✓	✓				
Even and Shankaranarayanan [11]		✓		✓	✓		
Laney (2011) [14]			✓	✓		✓	✓
Laney (2012) [15]					✓		✓
Sajko et al. [18]	✓	✓	✓		✓		
Viscusi and Batini [21]		✓		✓		✓	✓

**Table 1.** Overview of dimensions used to characterise data value

### 3.2 Measuring Data Value

To build adaptive, value-driven systems, it is necessary to have concrete value assessment techniques that report over time. Without assessment, the effective management of value, and hence efficient exploitation of data is highly unlikely [6]. The measurement of data value in an unbiased manner allow for better data characterisation and classification, which then enable data exploitation optimisation [8]. This requires monitoring data value dimensions within data value chains. Despite the growing literature on data as a valuable asset and on data exploitation, there is little to no work on how to **directly assess or quantify the value** of specific datasets held or used by an organisation within an information system. Moreover, existing methods for measuring the value of data often require intensive human effort and are also case-specific [8].

The lack of consensus on the definition of data value as described in Section 3.1 is hindering progress on data value assessment since entities and organisations are still fundamentally challenged to understand what characterises data value. Viscusi et al. [21] recently reconfirmed Moody and Walsh’s [17] earlier assertion that there is no consensus on how to measure information value. This issue also stems from the the diverse motivations for measuring the value of data. Examples of these purposes include; ranking of results for question answering systems [4], information life cycle management [8, 12], security risk assessment [18], and problem-list maintenance [13].

Usage, cost, and quality are three recurring data value dimensions that are measured in existing literature. Chen, for example, devises an approach to measure data value based on usage-over-time [8]. This valuation method is derived from two measurable and observable metrics; usage and time. The author here infers the value of information based on a number of usage statistics that include usage count, usage time, the source of usage, and the purpose of the usage. Wijnhoven et al. extended Chens usage-based data valuation approach with a utility-based estimation based on file metadata [23]. Through case studies in a consulting practice, they found that the frequency of use and the grade of user accessing the file were the most important predictors of value. Turczyk et al. also calculate the value of files from usage information [20], and Jin et al. similarly also measure data usage through information sharing and number of users accessing the information [12].

Various cost metrics are used in literature to measure data value. Stander breaks ‘cost’ into two categories, namely; (i) the purchase price of the data asset, and (ii) the direct costs attributed to preparing the data for use [19]. Stander also mentions some approaches for measuring data value, including; the cost approach, where data value is measured as the expenditure required to reproduce or obtain a data asset; the market approach, where data value is the price that organisations in the market are willing to pay for a data asset; and the income approach, which relies on the estimation of future income based on the exploitation of a data asset. DAMA International also focus on cost as a data value characterising dimension. They use the cost of losing a data asset and the

resulting impacts of not having the current amount and quality level of data as metrics of the value of a data asset.

Data quality metrics are the focus of Even and Shankaranarayanan’s paper, where the authors describe a quantitative approach for assessing the business value attributed to data assets [11]. The data value aspects assessed include completeness, validity, accuracy, and currency. On the other hand, Stander considers the frequency and accuracy aspects of data assets [19]. Other literature such as [10, 21] also mention quality aspects as data value dimensions, however they do not specifically mention any metrics that can be used to quantify this value.

Whilst usage, cost, and quality are three of the most popular data value dimensions that are quantified in literature, other data value dimensions are also discussed. For example, Laney focuses on utility functions [14], however this provides for abstract measures that are very challenging to realise as concrete metrics. Al-Saffar and Heileman provide an information valuation model with the aim of measuring the impact that a data asset can have [4]. The authors however acknowledge that this metric is subjective.

The existing literature therefore not only highlights the lack of existing metrics to quantify value, but also points out the need for more efforts in defining data value. Moreover, the literature also makes evident the complexity of quantifying data value, also due to its dependence on the context of use and its subjectivity. Yet, the subjective nature of some dimensions that characterise data value certainly does not rule out their quantification. Similar to some data quality aspects such as timeliness, such dimensions can still be accurately quantified in an objective manner, if only relevant for a specific context of use.

### 3.3 Modelling Value-Driven Data Governance

Whether using the ‘data governance’ term explicitly or not, many organisations and entities are exploring new strategies and approaches towards governing and managing their data assets. Such strategies may include both direct data manipulation (e.g. data quality, security, access) and also business plans on how the data asset is exploited. The issue here is that to date most data governance models are either proprietary, or otherwise human-process oriented and thus **do not support interoperable systems specification**.

Weill and Ross define an organisational approach to data governance where they establish a set of processes and delegation of authorities for making decisions and providing input [22]. This approach however focuses on roles and responsibilities rather than information system architectures, interfaces, processes or algorithms. The view of DAMA International [9] is more concrete and defines processes, roles and formal goals; for better decision-making, assuring compliance, increasing efficiency and business integration. Abed defines a framework based on four value pillars; agility, trust, intelligence, and transparency, focusing on enabling business sustainability and supporting economical growth [1]. Al-Ruithe et al. emphasise the importance of monitoring and measuring tools to support data governance [3]. Brous et al. document a systematic review of data

governance principles [7]. The authors identify four main principles based on the review; the organisation of data management, ensuring alignment with business needs, ensuring compliance, and ensuring a common understanding of data.

Realistically, it is quite doubtful to have a one size fits all data governance solution, as any data governance efforts must fit the specific organisation in question and cater for the needs of the business. That being said, current approaches lack the link between data assets and organisational value. Such a strategy is essential in exploiting data assets to achieve competitive advantage that provides both short and long term value, therefore ensuring business success and sustainability [1]. Moreover, Brous et al. state that “evidence is scant as to which data governance processes should be implemented, what data governance should be coordinating, or how data governance could be coordinated” [7]. It is therefore easy to understand the requirement for an interoperable, standardised, machine readable data governance model that caters for data assets, roles, and processes. Such a model should:

- Enable all entities within a data governance model to communicate;
- Enable outputs from different processes to be reused without requiring human intervention; and
- Integrate existing models and processes to provide an overall data governance effort.

### 3.4 Optimising Data Governance

Effective data value chain governance, and hence optimised exploitation of data assets, depends on an understanding and representation of the context of use, the exploitation processes, data value measures, and hence also the nature of data value. In fact, one of the goals of data governance is “to understand and promote the value of data assets” [9]. Many data processing systems include “black-box” processes that do not provide any insight or reasoning behind their outputs, results or motivations. This lack of data understanding undermines the specification and enforcement of data governance policies, and provision of robust auditing. Moreover, although there is some literature where data value monitoring/measurement has been used with the aim of enhancing control of processes within a data value chain [8, 19, 20] such literature focuses on the management of individual processes within the system, such as file storage, and not on overarching data governance optimisation.

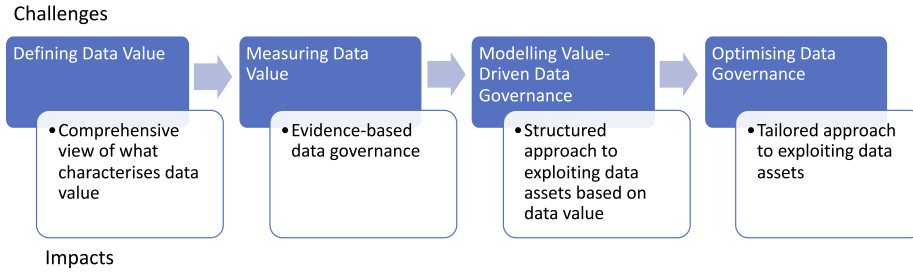
This challenge is also directly related to the challenge of measuring data value, as discussed in Section 3.2. The quantification of data value (even if just estimates) would enable the optimised governance of data assets in an enterprise. Examples of data governance processes that can be optimised include:

- Data storage: for example more valuable data can be stored in more reliable, more secure storage, whilst less valuable data can be stored using cheaper options;
- Data access: more valuable data, for example sensitive data, can be restricted to be used by employees with a higher user grade;

- Data acquisition: Data can be acquired depending on whether its value for the enterprise is worth its cost;
- Data standards: Data standards can be defined based on the existing data specifications, with the aim of achieving the data specifications that are required by the context of use; and
- Data maintenance: Data curation or maintenance processes can be prioritised according to data value. For example either prioritising data assets that are very valuable to the business, and therefore will result in the highest impact, or otherwise prioritising data assets that will benefit most from being maintained.

## 4 Conclusion

The aim of this paper is to raise awareness about the potential impact of value-driven data governance, and guide further research on the topic. Therefore, based on existing literature and experience of existing businesses that exploit data assets, we explored four main challenges that hinder value-driven data governance. As data has been established as a requirement for most businesses to remain competitive, it has become vital to implement data governance to enable successful data asset exploitation. Figure 2 provides an overview of the challenges covered in this paper, as well as the resulting impacts that can be achieved with relevant solutions. As the figure indicates, the challenges build on top of each other. Therefore, in an ideal world, the solution of the first challenge would contribute towards the solution of the next challenge.



**Fig. 2.** Overview of challenges and impacts that can be achieved with relevant solutions

Based on the contributions in this paper, our next efforts will be targeted towards defining data value. This will lay the foundations to obtaining a comprehensive description of what characterises data value, and therefore the basis to measuring each data value dimension. Whilst we envisage a comprehensive data value model, the actual data value quantification will most likely require to be context dependent, and therefore a subset of the data value dimensions would need to be used.

Any relevant solutions for the challenges explored in this paper will contribute towards an overall more efficient and effective data governance and therefore data asset exploitation. In turn, this will enable the more successful achievement of data-driven business or company goals.

## References

1. el Abed, W.: Data Governance: A Business Value-Driven Approach. Global Data Excellence pp. 1–16 (2011), [http://www.irmuk.co.uk/articles/Data Governance A Business Value-Driven Approach.pdf](http://www.irmuk.co.uk/articles/Data%20Governance%20A%20Business%20Value-Driven%20Approach.pdf)
2. Ahituv, N.: A Systematic Approach toward Assessing the Value of an Information System. *MIS Quarterly* **4**(4), 61 (dec 1980). <https://doi.org/10.2307/248961>, <http://www.jstor.org/stable/248961?origin=crossref>
3. Al-Ruithe, M., Benkhelifa, E., Hameed, K.: Key dimensions for cloud data governance. 2016 IEEE 4th International Conference on Future Internet of Things and Cloud (FiCloud) pp. 379–386 (2016)
4. Al-Saffar, S., Heileman, G.L.: Semantic Impact Graphs for Information Valuation. In: Proceedings of the Eighth ACM Symposium on Document Engineering. pp. 209–212. DocEng '08, ACM, New York, NY, USA (2008). <https://doi.org/10.1145/1410140.1410181>
5. Attard, J., Brennan, R.: A semantic data value vocabulary supporting data value assessment and measurement integration. In: Proceedings of the 20th International Conference on Enterprise Information Systems - Volume 2: ICEIS,. pp. 133–144. INSTICC, SciTePress (2018). <https://doi.org/10.5220/0006777701330144>
6. Brennan, R., Attard, J., Helfert, M.: Management of data value chains, a value monitoring capability maturity model. In: ICEIS (2018)
7. Brous, P., Janssen, M., Vilminko-Heikkinen, R.: Coordinating Decision-Making in Data Management Activities: A Systematic Review of Data Governance Principles. pp. 115–125. Springer, Cham (sep 2016)
8. Chen, Y.: Information Valuation for Information Lifecycle Management. In: Second International Conference on Autonomic Computing (ICAC'05). pp. 135–146. IEEE (jun 2005). <https://doi.org/10.1109/ICAC.2005.35>, <http://ieeexplore.ieee.org/document/1498059/>
9. Dama International: The DAMA Guide to The Data Management Body of Knowledge. Dama p. 430 (2009)
10. Engelsman, W.: Information Assets and their Value pp. 1–10 (2007)
11. Even, A., Shankaranarayanan, G.: Value-driven Data Quality Assessment. . . . Conference on Information Quality (2005), [http://mitiq.mit.edu/ICIQ/Documents/IQ Conference 2005/Papers/ValueDrivenDQAssessment.pdf](http://mitiq.mit.edu/ICIQ/Documents/IQ%20Conference%202005/Papers/ValueDrivenDQAssessment.pdf)
12. Jin, H., Xiong, M., Wu, S.: Information Value Evaluation Model for ILM. In: 2008 Ninth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing. pp. 543–548. IEEE (2008). <https://doi.org/10.1109/SNPD.2008.112>, <http://ieeexplore.ieee.org/document/4617429/>
13. Klann, J.G., Shadow, G.: Modeling the Information-value Decay of Medical Problems for Problem List Maintenance. In: Proceedings of the 1st ACM International Health Informatics Symposium. pp. 371–375. IHI '10, ACM, New York, NY, USA (2010). <https://doi.org/10.1145/1882992.1883045>

14. Laney, D.: Infonomics : The Economics of Information and Principles of Information Asset Management. The Fifth MIT Information Quality Industry Symposium, July 13-15, 2011 pp. 590–603 (2011)
15. Laney, D.: Gartner Analytic Ascendancy Model. Information Economics, Big Data and the Art of the Possible with Analytics p. 33 (2012), [https://www-950.ibm.com/events/wws/grp/grp037.nsf/vLookupPDFs/Gartner\\_Doug-Analytics/\\$file/Gartner\\_Doug-Analytics.pdf](https://www-950.ibm.com/events/wws/grp/grp037.nsf/vLookupPDFs/Gartner_Doug-Analytics/$file/Gartner_Doug-Analytics.pdf)
16. Maina, C.K.: Valuing information in an information age: The price model and the emerging information divide among individuals, societies, and nations (2003), <http://www.cais-acsi.ca/search.asp?year=3002>
17. Moody, D., Walsh, P.: Measuring The Value Of Information: An Asset Valuation Approach. Seventh European Conference on Information Systems (ECIS'99) pp. 1–17 (1999). <https://doi.org/citeulike:9316228>
18. Sajko, M., Rabuzin, K., Bača, M.: How to calculate information value for effective security risk assessment. *Journal of Information and Organizational Sciences* **30**(2), 263–278 (2006)
19. Stander, J.B.: The Modern Asset : Big Data and by (December) (2015)
20. Turczyk, L.A., Heckmann, O., Steinmetz, R.: File Valuation in Information Lifecycle Management. *Managing Worldwide Operations & Communications with Information Technology* pp. 347–351 (2007), <http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1548&context=amcis2007>
21. Viscusi, G., Batini, C.: Digital Information Asset Evaluation: Characteristics and Dimensions. pp. 77–86. Springer, Cham (2014), [http://link.springer.com/10.1007/978-3-319-07040-7\\_9](http://link.springer.com/10.1007/978-3-319-07040-7_9)
22. Weill, P., Ross, J.W.: IT Governance How Top Performers Manage IT Decision Rights for Superior Results. Harvard Business School Press Boston, Massachusetts (2004), [http://www.msu.ac.zw/elearning/material/1300172657060910\\_it\\_governance\\_matrix2535p4.pdf](http://www.msu.ac.zw/elearning/material/1300172657060910_it_governance_matrix2535p4.pdf)
23. Wijnhoven, F., Amrit, C., Dietz, P.: Value-Based File Retention. *Journal of Data and Information Quality* **4**(4), 1–17 (may 2014). <https://doi.org/10.1145/2567656>, <http://dl.acm.org/citation.cfm?doid=2628135.2567656>