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 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 [12]. 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, 6]. 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. 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 [10], but also as a part of automated control systems for data lifecycles [7] and file retention [20]. 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 [13]. 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.

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 is therefore aimed towards 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.

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. They perform 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, and device manufacturer homepages. New consumer electronic devices must be discovered, categorised, profiled for potential sales value and have their power supply technical specifications mined from open web data.

– **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 My-Volts.

– **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 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).

In general, all four data value processes described above 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 a **value-driven data governance model**.

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" [6]. 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 [7], different consumers [12]) depending on a number of dimensions [18]. Ahituv [2], for example, suggests timeliness, content, and cost as data value dimensions. 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 [10]. 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 [7]. 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 [15].

The above dimensions, are but a few of the existing dimensions that are used in literature to characterise data value. 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 [4], this heterogeneity of dimensions in literature indicates the need for terminological unification, which would also aid in developing a common understanding of the domain.

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 [5]. 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 [7].

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. [18] recently reconfirmed Moody and Walsh's [14] earlier assertion that there is no consensus on how to measure information value.

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 two measurable and observable metrics; usage and time [7]. Wijnhoven et al. extended Chens usage-based data valuation approach with a utility-based estimation based on file metadata [20]. Turczyk et al. also calculate the value of files from usage information [17], and Jin et al. similarly also measure data usage through information sharing and number of users accessing the information [11].

Various cost metrics are used in literature to measure data value. Stander breaks 'cost' into two categories; (i) the purchase price of the data asset, and (ii) the direct costs attributed to preparing the data for use [16]. DAMA International also focus on cost as a data value characterising dimension [8]. 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 [10]. 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 [16]. Other literature such as [9, 18] also mention quality aspects as data value dimensions, yet they do not specifically mention 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 [12], 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 [3]. 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 [19]. This approach however focuses on roles and responsibilities rather than information system architectures, interfaces, processes or algorithms. The view of DAMA International [8] 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]. Brous et al. document a systematic review of data governance principles [6]. 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. 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]. We therefore identify the requirement for an interoperable, standardised, machine readable data governance model that caters for data assets, roles, and processes.

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"* [8]. 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 [7, 16, 17] 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 can include data storage, where more valuable data can be stored in more reliable, more secure

storage, whilst less valuable data can be stored using cheaper options; and data acquisition, where data can be acquired depending on whether its value for the enterprise is worth its cost.

4   Conclusion

The aim of this paper is to raise awareness about the potential impact of value-driven data governance, and guide further research. 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 1 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. 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.
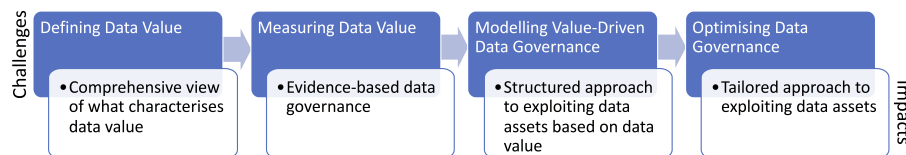

Fig. 1. Overview of challenges and impacts that can be achieved with relevant solutions

Acknowledgements

References

1. el Abed, W.: Data Governance: A Business Value-Driven Approach. Global Data Excellence pp. 1–16 (2011)
2. Ahituv, N.: A Systematic Approach toward Assessing the Value of an Information System. MIS Quarterly **4**(4), 61 (dec 1980)

3. 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)
4. 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,. SciTePress (2018)
5. Brennan, R., Attard, J., Helfert, M.: Management of data value chains, a value monitoring capability maturity model. In: ICEIS (2018)
6. Brous, P., Janssen, M., Vilminko-Heikkinen, R.: Coordinating Decision-Making in Data Management Activities: A Systematic Review of Data Governance Principles. In: International Conference on Electronic Government and the Information Systems Perspective. Springer (2016)
7. 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
8. Dama International: The DAMA Guide to The Data Management Body of Knowledge. Dama p. 430 (2009)
9. Engelsman, W.: Information Assets and their Value (2007)
10. Even, A., Shankaranarayanan, G.: Value-driven Data Quality Assessment. In: Proceedings of the 2005 International Conference on Information Quality (2005)
11. 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. IEEE (2008)
12. 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)
13. 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
14. 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
15. 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)
16. Stander, J.B.: The Modern Asset: Big Data and Information Valuation. Ph.D. thesis, Stellenbosch University (2015)
17. Turczyk, L.A., Heckmann, O., Steinmetz, R.: File Valuation in Information Lifecycle Management. In: Proceedings of the Thirteenth Americas Conference on Information Systems, Keystone, Colorado (2007)
18. Viscusi, G., Batini, C.: Digital Information Asset Evaluation: Characteristics and Dimensions. In: Smart Organizations and Smart Artifacts - Fostering Interaction Between People, Technologies and Processes, pp. 77–86. Springer (2014)
19. Weill, P., Ross, J.W.: IT Governance How Top Performers Manage IT Decision Rights for Superior Results. Harvard Business School Press Boston, Massachusetts (2004)
20. 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