Management of Data Value Chains, a Value Monitoring Capability Maturity Model

Rob Brennan[1], Judie Attard[1] and Markus Helfert[2]

[1]*ADAPT Centre, School of Computer Science and Statistics, O'Reilly Institute, Trinity College Dublin, Dublin 2, Ireland*

[2]*Lero Centre, ADAPT Centre, School of Computer Applications, Dublin City University, Dublin 9, Ireland*

{rob.brennan, attardj}@cs.tcd.ie, markus.helfert@dcu.ie

Keywords: Data Value, Business-IT Alignment, Business Value of IT, Governance of IS organisation

Abstract: This paper identifies management capabilities for data value chains as a gap in current data value research. It specifies a data value management capability framework and a first data value monitoring capability maturity model (CMM). This framework and CMM will enable organisations to identify and measure the current state of their data value monitoring processes, and show how to take steps to enhance value monitoring in order to exploit the full data value potential in their organisation. This new approach to data value management is needed since, despite the success of Big Data and the appeal of the data-driven enterprise, there is little evidence-based guidance for maximising data value creation. To date, most data value optimisation has focused on technological gains such as data platforms or analytics, without bridging the gap to organisational knowledge or human factors research. The evidence of best practice gathered here from the state of the art shows that there is a hierarchy of data value dimensions for data value monitoring, starting with cost and peaking with utility (understanding value creation). The models are validated by a case study of three organisations that are managing data value and using it to support strategic decision-making.

1 INTRODUCTION

With the advent of Big Data and the trend of maintaining ever larger datasets in the hope of supporting value creation, the concept of data value is gaining increasing traction (The Economist, 2017). Despite the uptake in data use across all sectors of society, there is a severe lack of guidance for data value management for effective data exploitation. The data value chains depicted in the literature hark back to manufacturing value chains (Crié & Micheaux, 2006; Latif, Us Saeed, Hoefler, Stocker, & Wagner, 2009; Lee & Yang, 2000; Miller & Mork, 2013; Peppard & Rylander, 2006), and focus on the acquisition, manipulation, exploitation, and distribution of data. These manufacturing-inspired data value chains cater for the use of intangible data as a product, but they do not cater for the vital aspect of managing the value creating processes within the chain (Rayport & Sviokla, 1995). It has been observed (Otto, 2015) that measures for managing data as a strategic resource have focused on technological endeavours such as data architecture or analytics, but topics such as providing evidence for the effectiveness of organisational measures and practices have remained relatively unexplored.

Effective data value chain management, and hence optimised value creation, depends on an understanding of the context of use, the value creation process, data value measures, and hence the nature of data value. However there is a lack of understanding of how data value dimensions combine into data valuations (Viscusi & Batini, 2014) or contribute to undoubtedly complex data value creation processes (Moody & Walsh, 1999). Nonetheless data value assessment metrics are available in the literature (Higson & Waltho, 2010) and have been used for specific applications. End-to-end value chain monitoring of data value is an important step towards data value management processes.

This paper explores the research question "*What capabilities are required for enterprises to best support data value management processes?*" To address this question, we bring together the diverse experiences of leading practitioners in data value management to identify the key capabilities and processes for data value maximisation, and arrange them in a new capability model. The data value monitoring and assessment capability is then examined in detail, as it acts as a basis for many subsequent data value management capabilities. Then, a monitoring capability maturity model with associated data value metrics is specified. The data value monitoring capability model is then evaluated in a case study of three data-centric organisations (a public knowledge base publisher, a data-driven retailer, and an open government data initiative) that use data valuations to drive operational performance and strategic decision-making.

The contributions of this paper are; the description of a novel data value chain management capability framework, the specification of a capability maturity model for data value monitoring (including identification of practices, outcomes, value dimensions and metrics for each maturity level), and evidence supporting this model provided by a synthesis of the state of the art and the case study analysis.

The rest of this paper is structured as follows: section 2 presents background work on data value chains and capability maturity models, section 3 provides related work and defines management capabilities for data value chains and the monitoring capability maturity model, section 4 provides a case study of data value monitoring in practice and section 5 provides a concluding discussion and future work.

2 BACKGROUND

In this section we first discuss value chains in the context of data value and then the application of capability maturity models to managing value.

Value chains have been used for decades, particularly in the manufacturing industry, to identify what manufacturer activities gave additional value to the product being created, and therefore resulting in a competitive advantage. The term "Value Chain" was introduced by Porter, where he defined a value chain to be the strategically relevant interdependent activities undertaken by a firm in order to achieve its goal (Porter, 1985). Porter lists five main activities in his value chain, namely (i) inbound logistics, (ii) operations, (iii) outbound logistics, (iv) marketing and sales, and (v) service. Therefore, the value chain was an indispensable tool that enabled the analysis of the interactions between the different activities in order to identify where value was being created. However, despite the success of the value chain concept for this aim, in recent years products and services have become increasingly digital, and therefore intangible. For this reason, a number of authors have provided newer definitions of the value chain concept whilst catering for this digital dimension. This revised data value chain provided stakeholders with a new perspective on creating value on a digital, intangible product. For example, Lee and Yang define a value chain for knowledge (Lee & Yang, 2000), Crié and Micheaux provide a more generic value chain that considers raw data (Crié & Micheaux, 2006), Miller and Mork define a data value chain with a big data perspective (Miller & Mork, 2013), Latif et al. target value creation on Linked Data (Latif, Us Saeed, Hoefler, Stocker, & Wagner, 2009), The Big Data Value Association (BDVA) provides a data value chain specifically for Big Data (see Figure 1), and Attard et al. define a data value network having a structure that caters for the fluid nature of data (Attard, Orlandi, & Auer, 2017).

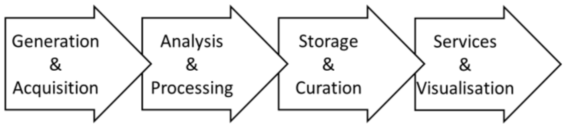

Figure 1: BDVA Big Data Value Chain adapted from (Zillner, S., Curry, E., Metzger and A., Auer, S., 2017)

As useful as all the aforementioned data value chains are, none of them consider the management of value creating processes or the capabilities required to achieve this. Whilst some publications (Crié & Micheaux, 2006; Lee & Yang, 2000) consider the management of the data product itself, the literature on data value chains described here does not specify data value management capabilities. In order to exploit the highest value of data, and obtain the highest impact, we deem the data value management capabilities of any entity to be vital in any data exploitation endeavour.

Unlike current (data) value chain models, Capability Maturity Models (CMMs) are specifically created to identify how an organisation manages its processes with the aim of becoming more successful in achieving its goals. This is usually done by assessing the current situation of an organisation, deriving and prioritising improvement measures, and controlling their implementation progress (Iversen, Nielsen, Nerbjerg, & Norbjerg, 1999). Thus CMMs guide organisations in the definition of progressive, qualitative self-assessments that identify areas of gradual improvement.

Humphrey defined one of the first CMMs that covered practices for planning, engineering, and managing software development and maintenance (Humphrey, 1989). This CMM consists of five levels of software process maturity, namely:

1. **Initial Level** - the organisation typically does not provide a stable environment for developing and maintaining software;
2. **Repeatable Level** - policies for managing a software project and procedures to implement those policies are established;
3. **Defined Level** - the standard process for developing and maintaining software across the organisation is documented;
4. **Managed Level** - the organisation sets quantitative quality goals for both software products and processes; and
5. **Optimising Level** - the entire organisation is focused on continuous process improvement.

Since Humphrey's CMM, there have been many developments on the topic. Maturity models have become more popular in Information Systems and also cater for the maturity analysis of various domains. For example, de Bruin and Rosemann discuss a maturity model that can assist organisations to improve in business process management (De Bruin & Rosemann, 2005), Gottshalk explores the digital government domain (Gottschalk, 2008), Paulk focuses on software engineering (Paulk, 2002), Ali et al cover inter-organisational systems (Ali, Kurnia, & Johnston, 2008), de Bruin et al. discuss knowledge management (de Bruin et al., 2005), Lacerda discuss usability engineering (Lacerda, Gresse, & Wangenheim, 2017), etc.

The mentioned CMM literature provides varying approaches for analysing the current maturity of an organisation, product, service or initiative, and guidance on how to improve. Despite this, these maturity models do not focus on the capabilities required to manage data value explicitly. The emerging recognition of data as an organisational asset that spans many products or services must change this. The lack of data value-oriented CMMs may be because: (i) data value is a multi-dimensional quantity that is highly context dependent (Viscusi & Batini, 2014) and (ii) data value creation processes are still not well understood. Another dependency of CMM approaches is the need for the documentation of best practice in terms of practices, outcomes and metrics in order to formulate an evidence-based maturity model and, before the work presented here, no-one has collected this evidence.

Thus, there is a need for identification of the required data value chain management capabilities to maximise value creation from data, and a capability maturity model to assess organisational readiness and guide organisational development.

3 DATA VALUE CHAIN MANAGEMENT CAPABILITIES

Defining data value chain management depends on examining how value grows from organisational capabilities, identifying a set of data value management capabilities and specifying capability maturity models for them. Due to space limits we only specify a full CMM for the fundamental data value monitoring capability as an example of our approach.

3.1 Organisational Capabilities

Organisational capabilities focus on internal processes, functions, and systems to meet customer needs. Thus, organisational capabilities foster unique service-specific routines (Felin, Foss, Heimeriks, & Madsen, 2012) which enable competencies to harness competitive advantage and are typically directed towards achieving defined goals and strategies (Winter, 2000). There are numerous definitions throughout management literature which examines the multidimensional nature of organisational capabilities. Collis (Collis, 1994) classifies many of the definitions into three broad categories:

- **Performance (capabilities):** capabilities that reflect the ability to perform the basic functional activities of the organisation. This suggests that capabilities are developed in functional areas which have defined business processes;
- **Repeated processes (capabilities)**: capabilities are repeated processes that are responsive to market trends and short lifecycles. This suggests that an organisation must be agile to adapt to customer needs by refocusing organisational capability deployment;
- **Strategise value creation (capabilities)**: capabilities that realise the value of resource allocation to execute and enable novel strategies by deploying organisational resources.

Thus, from a technological viewpoint, we can view capabilities as being socially embedded routines which may be captured through technological means, i.e. automated processes. They support the transformation of inputs into outputs. These outputs are the product of each individual process that are networked together to contribute to the entire service system.

3.1.1 Routines

An organisational capability may be described as “a high level routine (or collection of routines) that, together with its implementing input flows, confers upon an organisation’s management a set of decision options for producing significant outputs of a particular type” (Winter, 2000). It is interesting that Winter identifies ‘routine’ as the contributing factor of capabilities since it conjures a notion of learned behaviour which follows a specific execution pattern that is repetitive in nature. Routines are relatively fixed, unchanging objects which reflect an agreement of how things are done, i.e. imposing a control mechanism through repetitive patterns (Feldman & Pentland, 2003). Thus, routines encode organisational capabilities and knowledge in a learning cycle (see Figure 2).

3.1.2 Skills and Dynamic Capabilities

Chandler defines organisational capabilities as an organisation’s collective physical facilities and skills of employees, and the expertise of top management layers (Chandler, 1994). Thus, capabilities provide a building block towards organisational core competencies (Coulter, 2002), for example, research and development. In addition, considering the concept of ‘dynamic capabilities’, Collis suggests that dynamic capabilities govern the rate of change of ordinary capabilities (Collis, 1994). Dynamic capabilities are traditionally believed to involve patterned activities which originate from objectives.

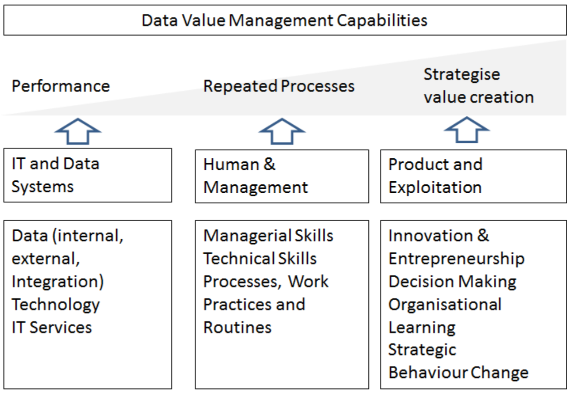

Figure 2: Data Value Management Capabilities (Adapted from (Gupta & George, 2016))

According to (Winter, 2003) (p.2) dynamic capabilities “*are those that operate to extend, modify or create ordinary capabilities…[and] involves a patterning of activity*”.

We **define capabilities** as a partial representation of the collective ability *to carry out specific business processes across a network in a cyclical, efficient, and relatively predictable manner to contribute towards organisational performance*. Figure 2 describes the three key building blocks that form the basis for performance, repeated processes and strategic value creation for data value management.

There have been some efforts to understand the role of IT in value creation. For example, transaction cost theory (Clemons & Row, 1991) is one approach to examine how technology reduces transaction costs. Thus, resources are central to the sustainability of competitive advantage. This is true for a resource-based view of IT business value. (Melville, Carroll, Kraemer, & Gurbaxani, 2004) propose that “*IT and non-IT resources and the business processes of electronically connected trading partners shape the focal firm's ability to generate and capture organisational performance*” (Melville Carroll, Kraemer, & Gurbaxani., 2004) (p. 307). IT business value research examines how organisational performance results from IT investment and to what extent the application of IT leads to improved performance. By performance, scholars have referred to productivity enhancement, profitability improvement, cost reduction, competitive advantage, inventory reduction, and other measures of performance (for example, (Hitt & Brynjolfsson, 1996). Nowadays, the emphasis often lays on optimising internal routines and capabilities though individual processes. In most cases, managers opt for CMM and IT Service Management (ITSM) as traditional approaches to evaluate the utilisation and alignment of IT resources in service optimisation.

It is well established that value can flow from data and that technology solutions can enhance the value creation process. In addition if an organisation has effective IT services to leverage data and technology then it has a multiplier effect on value creation. Most of the work to date has focused on optimising the performance of these data production capabilities (Otto, 2015).

3.1.3 Data Value Management Capabilities

We must identify additional data value management capabilities that will span the data value chain and optimise value creation. These capabilities must be capable of: (1) driving data value performance-related capabilities such as data integration by dynamically re-configuring tools and workflows to optimise value creation (an enterprise objective) rather than to optimise correctness or speed (technical objectives); (2) define and promulgate effective repeated data value processes, work practices and routines throughout the organisation, especially in ways that support automation; and (3) will support a strategic control loop for data value and thus enable new product innovation and exploitation based on data value management.

In order to support these requirements for data value chain management we propose that the following data value management capabilities are required across the length of the data value chain, split into performance, processes, and strategy groupings (Figure 3). In addition, we identify a minimal set of capabilities that will be built upon by the other management capabilities. These are termed fundamental **data value management capabilities**: data value monitoring, profiling and prediction. It must be observed that it is expected that more capabilities will be identified in future work. A brief definition of each capability is given next.

The fundamental capabilities are as follows:

- **Data value monitoring capability** – supports data value assessment and reporting at all stages in the data value chain;
- **Data value profiling capability** – the ability to specify sources of value for a specific organisation;
- **Data value prediction capability** – capacity to analyse and predict data value trends or sources from current indicators, patterns and context.

The performance capabilities for data value management are:

- **Dataset optimisation capability** – ability to acquire, curate, maintain and enhance datasets based on value maximisation;
- **Infrastructure optimisation capability** – ability to design, deploy and reconfigure the enterprise data infrastructure to maximise data value;
- **IT services optimisation capability** – the ability to evolve, adapt and orchestrate the data processes, support services and workflows towards value-based goals and outcomes.

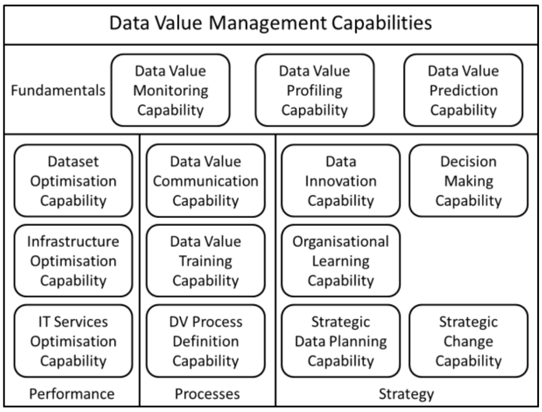

Figure 3: Data Value Management Capability Framework

The processes and human factors oriented capability for data value management are:

- **Data value communication capability** – level of integration of data value into IT leadership and governance infrastructures;
- **Data value training capability** – the ability to enhance organisational management and technical skills about data value issues;
- **Data value process definition capability** – value-centric data process design and integration of data value considerations into wider organisational processes.

The strategy capabilities are:

- **Data innovation capability** – the ability to innovate in services or products based on novel use of (or insights from) data;
- **Decision-making capability** – organisational capacity to create a data value-based control loop that feeds into organisational level decisions;
- **Organisational learning capability** – the extent to which the organisation can evolve based on data value considerations;
- **Strategic data planning capability** – identification of longer-term data acquisition, data architecture, system integration, IS priorities, business process and people required;
- **Strategic change capability** – identification of the changes in organisational goals based on data value.

In the next section we provide a capability maturity model for the data value monitoring capability as an example of a fundamental capability that must be specified in order to realise our vision.

3.2 Data Value Monitoring Capability

The data value monitoring capability focuses on assessing and reporting of data value throughout the value chain by gathering metrics on datasets, the data infrastructure, data users, costs and operational processes. Monitoring forms a fundamental function of any control capability whether embodied as the fault management or performance management functions in a telecommunications network (*3GPP*, 2009) or a MAPE (monitor, assess, plan, execute) control loop in IBM's autonomic computing initiative (Jennings et al., 2007). Monitoring data value is essential to providing feedback to any data value control system, whether automated, semi-automated or manual.

There have been a number of documented uses of data value monitoring for enhanced control of elements of the data value chain, especially in the application areas of file-storage management (Turczyk, Heckmann, & Steinmetz, 2007), information lifecycle management (Chen, 2005), information pricing (Rao & Ng, 2016), data governance (Stander, 2015), and data quality management (Even, Shankaranarayanan, & Berger, 2010). We use those and other examples from practice as the basis for our approach.

This capability aims to be holistic in assessing the key dimensions of data value for a particular organisation. Unfortunately, there are a wide range of known dimensions of data value (Viscusi & Batini, 2014) and there is not yet a consensus on their definitions, how they are related, or how practical data value metrics in information systems relate to monetary value, as measured in accounting-based measures of value. Viscusi and Batini break data value down into information capacity and information utility (Viscusi & Batini, 2014). Capacity is then subdivided into quality, structure, diffusion and infrastructure. In their scheme, utility is based on financial value, pertinence and transaction costs. In contrast, the models of (Moody & Walsh, 1999) and (Tallon, 2013) strongly emphasise usage as a key dimension of value. It is in the area of usage-based data value that the most progress has been made for practical data value monitoring systems. Hence, we must give it most prominence in the following analysis. Given a lack of a comprehensive metric framework for measuring data value, it has been found that having at least some data value measurement capability is better than none.

Ease of measurement is another important concept to consider when deciding how to monitor data value in an organisation. Some data value dimensions have well known metrics and may even have recommended data or metadata formats, for example the W3C's data quality vocabulary (Albertoni et al., 2016). Other data value dimensions, such as business utility or impact are very difficult to measure since they depend on having models and information about the business processes, outcomes and data dependencies in order to identify measurable metrics for the contribution of datasets to profit or operating efficiencies.

One important distinction to be made when creating a monitoring framework is the identification of data value metrics that can be evaluated independently of any knowledge of the context of use of the dataset, for example frequency of access, dataset volume or dataset purchase cost, and context-dependent metrics such as appropriate data quality metric thresholds or the relevance of a particular dataset for a specific task. We label this latter class of metrics as "contextual".

Table 1 defines a capability maturity model for data value monitoring based on our analysis of the existing literature for data value monitoring or metrics discussed above. As can be seen from the table, lower levels of maturity are associated with more limited collection of data value dimensions and there is a hierarchy of monitoring dimensions that emerges from the literature on data value metrics. The second major trend seen in the table is the increasing connection between technology-centric metrics such as data volume, and enterprise-centric metrics such as average financial contribution per record. Ultimately, we posit that attaining a level five capability maturity level is dependent on being able to understand or model how data value chain processes or steps convert data into value (utility). This corresponds to understanding the value creation process in a business, at least to the extent of a reliable black box predictor function.

One challenge in tackling this level of business monitoring is the need to converge the views of the technology-centric data administration functions of the business, with the financial and operations functions (and ultimately the strategic planning functions). This approach has a natural foundation in the increasing importance of data governance within overall governance structures (Tallon, 2013).

4 CASE STUDIES

4.1 Aims and Methodology

The case study comprises three organisations. The three cases represent four different industries, with headquarters located in three different countries and a scope of operations ranging from a single country, to a handful of through worldwide. For an organisation to be included in the case study, three criteria had to be met: (1) the organisation should have experience in managing data value chains and data value monitoring, (2) the organisation's data value management concept should cover multiple data value monitoring dimensions, and (3) the organisations should allow the authors of the article to get in touch with subject matter experts.

For each organisation, the data value monitoring CMM (Table 1) was applied to assess their current processes. This involved investigating their methodology and assigning their metrics to specific data value dimensions. For data collection, various methods and data sources were used, ranging from structured questioning of subject matter experts to the analysis of publicly available material (annual reports, for example), research papers, presentations, and internal documents from joint research projects that focused on data quality, data management, governance and data process improvement for increased agility and productivity. It must be observed that the diversity of the cases selected make the use of "natural controls" as often required for case study research (Lee, 1989) difficult to achieve, and this is left for future work.

Table 1: A Capability Maturity Model for Data Value Monitoring including practices, outcomes, and metrics pertaining to data value monitoring

Maturity Level	Practices	Outcomes	Value Monitoring Dimensions	Example Metrics
1	No formal practices are expected at this level. Potential for collection of pertinent metrics for non-value reasons. No coherent analysis performed.	Inconsistent approach to data value monitoring. Data value monitoring is not formalised. Ad hoc local assessments based on perceived business unit needs may exist.	No value-specific monitoring	No metrics
2	Only perform reactive data value assessment based on crisis events and execute reviews of crisis handling as needed.	Can establish links between service or product failures and data. Know the costs associated with specific datasets.	Outage events, Cost	Time since last file access (Chen, 2005), Cost of procurement (Moody & Walsh, 1999)
3	Active monitoring of data value using a mix of context-independent measures.	Know who is using what data and for what purposes. Have defined a set of metrics for value assessment. Can assess a value for each data resource.	Usage, Quality metrics, IT Operations, Intrinsic	Data volume (Short, 2014), storage cost (Tallon, 2013), completeness (Laney, 2011), Trustworthiness (Viscusi & Batini, 2014)
4	Characterise the links between data, value, business operations and revenue. Specify thresholds for business context-dependent dimensions of data value.	Can attribute specific process or product outcomes to a specific data resource. Can distinguish between data resources that are fit for business purposes and those that are not.	Contextual, Value contribution	Business user satisfaction (Higson & Waltho, 2010), Shelf life of data (Moody & Walsh, 1999), Data quality thresholds, intrinsic record value (Even & Shankaranarayanan, 2005)
5	Monitoring supports the optimised exploitation of data value. Monitoring feedback is provided to other data value management capabilities.	Understand value creation processes at each stage in value chain. Data value monitoring feeds back to whole organisation and impacts strategy.	Utility	Estimated benefit to business from using data (Higson & Waltho, 2010), Change in KPIs for business objectives (Laney, 2011), Potential lost revenue (Tallon, 2013)

4.2 Case Studies

4.2.1 DBpedia - Public Data Infrastructure for a Large, Multilingual, Semantic Knowledge Graph

DBpedia (Lehmann et al., 2015) is the centre of the current web of data. It extracts, integrates, manages and publishes authoritative open RDF-based datasets that are used as a common point of reference for interlinking and enriching most of the structured data on the web today. It relies on an automated data extraction framework to generate RDF data from Wikipedia documents, and is now mining new sources like WikiData, EU Open Data portals, DBpedia association member data and the open web. The data is published in 125 languages in the form of file dumps, Linked Data and SPARQL (SPARQL Protocol And RDF Query Language) endpoints. With the help of the H2020 ALIGNED project (Gavin et al., 2016) it has semi-automated the data publishing process through the use of the dataset metadata specification DataID (Freudenberg et al., 2016) and a supporting toolchain for generation, validation, knowledgebase staging and dockerisation. DBpedia supports higher quality knowledge extraction through tools for community generated mappings, a repository of links to other open data sets and a human-curated reference ontology. Quality management of interlinks to other datasets also remains a crucial problem in web-scale data (Meehan, Kontokostas, Freudenberg, Brennan, & O'Sullivan, 2016). Managing data value is critical in the DBpedia Association (http://wiki.dbpedia.org/dbpedia-association) which is a non-profit that is responsible for overall governance of DBpedia activities. As part of ongoing automation, innovation support and quality improvement initiatives in DBpedia, for example through the ALIGNED project a number of data value monitoring activities have been initiated.

4.2.2 myVolts – a data-driven online retailer

myVolts (http://www.myvolts.com/) is a data-driven SME that relies on high quality, agile data integration and business intelligence analytics to operate extremely efficiently. MyVolts develops and operates a highly automated internet retail and business intelligence system (Frodo) that reduces their costs and staffing. They have served over 1 million customers and are a leading source of consumer device power supplies in the markets where they operate: the USA, Ireland, UK, France and Germany. MyVolts collect, manage and analyse data on their customers, the evolving market of power supply device specifications, and the power supply needs of all consumer electronics devices. This involves 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. MyVolts use data value estimates to drive data acquisition, cleaning and enrichment, and product logistics or marketing decisions.

4.2.3 Data.gov.uk – an open government data initiative

Data.gov.uk is the United Kingdom government's effort towards opening up its data to encourage its re-use. This effort is led by the Data team in the Cabinet Office, working across government departments to ensure that data is released in a timely and accessible way. The main aims behind this effort are to enable the government to be more transparent, to support businesses and academics, to foster innovation, and to enable citizens to become more informed. Data.gov.uk focuses on publishing non-personal, non-sensitive data in a single searchable website. Available datasets are sourced from all central government departments, and a number of other public-sector bodies and local authorities. This data includes environment data, government spending data, transport data, crime data, health data, etc. There are currently close to 43 thousand datasets currently published on the portal. Data.gov.uk uses data value management to sustain the open government data portal. Decisions on maintenance of datasets or the publishing of new datasets are based mostly on usage and data quality statistics.

4.3 Case Analysis

All organisations in the case study are assessing several dimensions of the value of their data resource (see Table 2) and seeking to apply data value in order to perform strategic data planning.

In DBpedia data value monitoring started in 2015 as part of the H2020 ALIGNED project quality improvement program for DBpedia processes, data and tools. It had been found that due to the success of DBpedia as a research platform that provided an extensive formal knowledge base as a free service that industrial entities such as Wolters Kluwer Germany (Kontokostas et al., 2016) had adopted DBpedia as a datasource. This in turn highlighted data quality issues and the difficulty of maintaining innovation and a stable service offering at DBpedia[1] and led to the formation of the DBpedia Association to oversee the development of DBpedia from a research infrastructure to a public infrastructure. The Association needed to perform strategic planning and the data quality, data value, and productivities studies were used as an input to

[1] See for example DBpedia director Sebastian Hellmann's announcement on the public DBpedia mailing list on the 25th Sept 2017 https://sourceforge.net/p/dbpedia/mailman/dbpedia-discussion/?viewmonth=201709&viewday=25

that as well as automation and process improvement research. The baseline studies took place in May-June 2015 and have continued on a periodic basis since then. The data collection methods included analysis of the DBpedia data assets in the form of the DBpedia ontology, the mappings wiki and quality test reports on the DBpedia data. In addition to these direct measurements of the dataset a number of other monitoring initiatives were setup to get a wider view, these included: analysing GitHub repositories for DBpedia extraction code, monitoring Wikipedia logs (the source of DBpedia data), a staff survey, a user survey and analysis of traffic on the public DBpedia mailing list.

In contrast MyVolts Ltd. have only recently started to monitor data value sources. As a small enterprise, information naturally flows easier within their tight-knit teams and the organisation has seen many years of increasing success despite processes that have been less formalised. Nonetheless the high level of automation and the use of a single integrated IT system in their business has facilitated rapid analysis of the sources of value. As an SME they have in the past relied on informal but effective qualitative staff transfer of knowledge about the context of use and utility of data to direct their strategic data management and ultimately management of their data value chain. They are now taking the first steps to automation of data value monitoring and collecting statistics on the number of electronic device records, the growth rate of this data, the number of data quality-related trouble tickets and the costs directly associated with dataset creation. This will be combined with existing informal knowledge on data utility and context of use to enable data value to be estimated.

Data value monitoring was part and parcel of the Data.gov.uk open government initiative from the very start of the project. The initiative required the selection of key datasets for initial publishing. These key datasets comprised National Health Service, Education and Skills, Criminal Justice, Transport, and Government financial information datasets. These datasets where considered to be key datasets since they directly targeted the specified goals of the data.gov.uk initiative. Therefore, data value was here used to enable more informed decision making. Thereafter, and throughout the lifetime of the initiative, data value monitoring was implemented to ensure sustainability. Dataset usage (through number of views and downloads) and dataset publishing frequency are constantly being tracked. This enables all stakeholders, whether publishers or consumers, to identify popular or highly-requested datasets, and their update rate. In turn, dataset maintenance can be planned accordingly. Data quality is also being monitored, both from the initiative itself, and also from the dataset consumers. Focus is usually stronger on openness and usability of the datasets. The financial aspects of the initiative, whilst monitored, are mostly estimates. In addition, the utility of the published datasets is also estimated

Table 2: Data Value Monitoring Case Study Evaluations

Value Dimensions Monitored	Monitoring CMM Level	Case A: DBpedia	Case B: MyVolts	Case C: Open Government Data
Usage	2	Log analysis	Identification of data consumers	Number of views, Number of downloads, Publisher activity log
Cost	2	Creation cost, maintenance cost	Dedicated staff costs	Publishing cost, maintenance cost
Quality	3	Test driven assessment, Interlink validation, user surveys	Number of trouble tickets	Openness, Re-usability of data, Broken links, Missing resources
Intrinsic	3	Volume, number of classes, number of properties, number of Wikipedia pages mapped	Number of records, record growth rate	Number of available datasets,
IT Operations	3	User surveys, trouble ticket analysis, staff surveys	Not measured	Not measured
Contextual	4	Not measured	Not measured	Not measured
Utility	5	Not measured	Not measured	Economic Value

through the resulting contribution to the economy. For example, the National Audit Office estimated that the datasets published by April 2012 could potentially add economic value in the region of 1.6 to 6 billion Pounds a year.

5 DISCUSSIONS AND CONCLUSIONS

With this paper we discuss management capabilities for data value chains and argue that the monitoring capability is fundamental. Underpinned by three cases the paper presents a first data value monitoring capability maturity model (CMM) and data value metrics based on current best practice. The cases illustrate the wide variety of ways to operationalise the data value monitoring component, and demonstrate capabilities along five maturity levels. The analysis of our data value monitoring capability model supports the general hypothesis, i.e. that there is a hierarchy of data value dimensions when it comes to monitoring - usage and cost are the easiest to implement. It also shows the fundamental character of Data Value Monitoring, Profiling and Prediction capabilities in order to increase overall data value management capabilities.

The presented cases in this paper demonstrate that monitoring data value is a necessary prerequisite to strategic data management and ultimately data value chain management within the enterprise. Our research, as evident for example from the Data.gov.uk open government initiative, stresses the importance of the Data Value Monitoring capability even on a lower maturity level. Furthermore, the DBpedia case highlighted data quality issues and the difficulty of maintaining innovation, which again emphasises the importance of Data Value Monitoring. An example of the use of data value estimates to drive data acquisition, cleaning and enrichment is illustrated in the MyVolts case. We have shown that it is possible to assess the maturity of data value monitoring processes and use our capability maturity model to guide enterprises on the dimensions of data value to assess and what metrics to use. We also validated our maturity model by demonstrating that the hierarchy of data value monitoring dimensions we envisage is seen in nature.

The proposed Data Value Monitoring component is embedded into our overarching Capability model that provides valuable guidelines to industry. However further research in needed on: metrics to assess data value, how metrics can be combined to give data value estimates, data value monitoring infrastructure; formal models describing metrics, dimensions and how they relate (ontologies or data models), techniques for comparison of assessment methods, understanding the relationship between other CMM models to the data value CMM proposed here. However, as our research indicates, a specific data value oriented maturity framework provides value, and this new value exploitation perspective can complement the other, often output or service oriented maturity frameworks.

ACKNOWLEDGEMENTS

This research was partially supported by Science Foundation Ireland and co-funded by the European Regional Development Fund through the ADAPT Centre for Digital Content Technology [grant number 13/RC/2106] and grant 13/RC/2094 co-funded under the European Regional Development Fund through the Southern & Eastern Regional Operational Programme to Lero - the Irish Software Research Centre (www.lero.ie)

The authors wish to thank Dr Pieter De Leenheer of Collibra Inc. for feedback on this work.

REFERENCES

3GPP. (2009). (V8.1.0, 3GPP TS 32.111-2). Retrieved from http://www.qtc.jp/3GPP/Specs/32111-2-810.pdf

Albertoni, R., Isaac, A., Guéret, C., Debattista, J., Lee, D., Mihindukulasooriya, N., & Zaveri, A. (2016). *Data Quality Vocabulary ({DQV})*. Retrieved from http://www.w3.org/TR/vocab-dqv/

Ali, M., Kurnia, S., & Johnston, R. B. (2008). A Dyadic Model of Interorganizational Systems (IOS) Adoption Maturity. In *Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008)* (pp. 9–9). IEEE. https://doi.org/10.1109/HICSS.2008.18

Attard, J., Orlandi, F., & Auer, S. (2017). Exploiting the value of data through data value networks. In *Proceedings of the 10th International Conference on Theory and Practice of Electronic Governance ICEGOV'17* (pp. 475–484). https://doi.org/10.1145/3047273.3047299

de Bruin, T., Rosemann, M., Freeze, R., Kulkarni, U., & Carey, W. (2005). Understanding the Main Phases of Developing a Maturity Assessment Model. In *Australasian Conference on Information Systems (ACIS)*,.

Higson, C. & Waltho, D.. (2010). Valuing Information as an Asset. *Sas the Power To Know*, (November), 1–17. Retrieved from http://www.eurim.org.uk/activities/ig/InformationAsset.pdf

Chandler, A. D. (1994). *Scale and scope : the dynamics of industrial capitalism*. Belknap Press of Harvard University Press. Retrieved from http://www.hup.harvard.edu/catalog.php?isbn=9780674789951&content=reviews

Chen, Y. (2005). Information Valuation for Information Lifecycle Management. In *Second International Conference on Autonomic Computing (ICAC'05)* (pp. 135–146). https://doi.org/10.1109/ICAC.2005.35

Clemons, E. K., & Row, M. C. (1991). Sustaining IT Advantage: The Role of Structural Differences. *MIS Quarterly*, *15*(3), 275. https://doi.org/10.2307/249639

Collis, D. J. (1994). Research Note: How Valuable are Organizational Capabilities? *Strategic Management Journal*, *15*(S1), 143–152. https://doi.org/10.1002/smj.4250150910

Coulter, M. K. (2002). *Strategic management in action*. Prentice Hall.

Crié, D., & Micheaux, A. (2006). From customer data to value: What is lacking in the information chain? *Journal of Database Marketing & Customer Strategy Management*, *13*(4), 282–299. https://doi.org/10.1057/palgrave.dbm.3240306

De Bruin, T., & Rosemann, M. (2005). Towards a business process management maturity model, 26–28. Retrieved from https://eprints.qut.edu.au/25194/1/25194_rosemann_2006001488.pdf

Economist (2017), *"The world's most valuable resource is no longer oil, but data," Economist, May 6, 2017*. Retrieved from https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource

Even, A., & Shankaranarayanan, G. (2005). Value-driven Data Quality Assessment. *... Conference on Information Quality*. Retrieved from http://mitiq.mit.edu/ICIQ/Documents/IQ Conference 2005/Papers/ValueDrivenDQAssessment.pdf

Even, A., Shankaranarayanan, G., & Berger, P. D. (2010). Evaluating a model for cost-effective data quality management in a real-world CRM setting. https://doi.org/10.1016/j.dss.2010.07.011

Feldman, M. S., & Pentland, B. T. (2003). Reconceptualizing Organizational Routines as a Source of Flexibility and Change. *Administrative Science Quarterly*, *48*(1), 94. https://doi.org/10.2307/3556620

Felin, T., Foss, N. J., Heimeriks, K. H., & Madsen, T. L. (2012). Microfoundations of Routines and Capabilities: Individuals, Processes, and Structure. *Journal of Management Studies*, *49*(8), 1351–1374. https://doi.org/10.1111/j.1467-6486.2012.01052.x

Freudenberg, M., Brümmer, M., Rücknagel, J., Ulrich, R., Eckart, T., Kontokostas, D., & Hellmann, S. (2016). The Metadata Ecosystem of DataID. In *Special Track on Metadata & Semantics for Open Repositories at 10th International Conference on Metadata and Semantics Research, (2016)*. Retrieved from http://aksw.org/Groups/KILT

Gavin, O., Kontokostas, D., Dirschl, C., Koller, A., Davies, J., Francois, P., … Brennan, R. (2016). The ALIGNED Project - Aligned, Quality-centric Software and Data Engineering Driven by Semantics. In *EU Project Networking Session at 13th ESWC*. Retrieved from http://www.tara.tcd.ie/handle/2262/76242

Gottschalk, P. (2008). Maturity levels for interoperability in digital government. *Government Information Quarterly*, *26*, 75–81. https://doi.org/10.1016/j.giq.2008.03.003

Gupta, M., & George, J. F. (2016). Toward the development of a big data analytics capability. *Information and Management*, *53*(8), 1049–1064. https://doi.org/10.1016/j.im.2016.07.004

Hitt, L. M., & Brynjolfsson, E. (1996). Productivity, Business Profitability, and Consumer Surplus: Three Different Measures of Information Technology Value. *MIS Quarterly*, *20*(2), 121. https://doi.org/10.2307/249475

Humphrey, W. S. (1989). *Managing the software process*. Addison-Wesley. Retrieved from https://dl.acm.org/citation.cfm?id=64795

Iversen, J., Nielsen, P. A., Nerbjerg, J., & Norbjerg, J. (1999). Situated assessment of problems in software development. *ACM SIGMIS Database*, *30*(2), 66–81. https://doi.org/10.1145/383371.383376

Jennings, B., van der Meer, S., Balasubramaniam, S., Botvich, D., O Foghlu, M., Donnelly, W., & Strassner, J. (2007). Towards autonomic management of communications networks. *IEEE Communications Magazine*, *45*(10), 112–121. https://doi.org/10.1109/MCOM.2007.4342833

Kontokostas, D., Mader, C., Dirschl, C., Eck, K., Leuthold, M., Lehmann, J., & Hellmann, S. (2016). Semantically Enhanced Quality Assurance in the JURION Business Use Case. In *13th International Conference, ESWC 2016, Heraklion, Crete, Greece, May 2016* (pp. 661–676). https://doi.org/10.1007/978-3-319-34129-3_40

Lacerda, T. C., Gresse, C., & Wangenheim, V. (2017). ARTICLE IN PRESS Systematic literature review of usability capability/maturity models. *Computer Standards & Interfaces*, *13*(0), 17–1.

Laney, D. (2011). Infonomics : The Economics of Information and Principles of Information Asset Management. *The Fifth MIT Information Quality Industry Symposium, July 13-15, 2011*, 590–603.

Latif, A., Us Saeed, A., Hoefler, P., Stocker, A., & Wagner, C. (2009). The Linked Data Value Chain: A Lightweight Model for Business Engineers. In *Proceedings of International Conference on Semantic Systems* (pp. 568–576).

Lee, A. S. (1989). A Scientific Methodology for MIS Case Studies. *MIS Quarterly*, *13*(1), 33. https://doi.org/10.2307/248698

Lee, C. C., & Yang, J. (2000). Knowledge value chain. *Journal of Management Development*, *19*(9), 783–794.

Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P. N., … Bizer, C. (2015). DBpedia – A large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web*, *6*(2), 167–195. https://doi.org/10.3233/SW-140134

Meehan, A., Kontokostas, D., Freudenberg, M., Brennan, R., & O'Sullivan, D. (2016). Validating Interlinks Between Linked Data Datasets with the SUMMR Methodology (pp. 654–672). Springer, Cham. https://doi.org/10.1007/978-3-319-48472-3_39

Melville, N., Carroll, W. E., Kraemer, K., & Gurbaxani, V. (2004). Review: Information Technology And Organizational Performance: An Integrative Model Of It Business Value 1. *MIS Quarterly*, *28*(2), 283–322.

Miller, H. G., & Mork, P. (2013). From Data to Decisions: A Value Chain for Big Data. *IT Professional*, *15*(1), 57–59. https://doi.org/10.1109/MITP.2013.11

Moody, D., & Walsh, P. (1999). Measuring The Value Of Information: An Asset Valuation Approach. *Seventh European Conference on Information Systems (ECIS'99)*, 1–17. https://doi.org/citeulike:9316228

Otto, B. (2015). Quality and Value of the Data Resource in Large Enterprises. *Information Systems Management*, *32*(3), 234–251. https://doi.org/10.1080/10580530.2015.1044344

Paulk, M. (2002). Capability Maturity Model for Software. In *Encyclopedia of Software Engineering*. Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/0471028959.sof589

Peppard, J., & Rylander, A. (2006). From Value Chain to Value Network: *European Management Journal*, *24*(2–3), 128–141. https://doi.org/10.1016/j.emj.2006.03.003

Porter, M. E. (1985). *Competitive Advantage:Creating and sustaining superior performance*. *New York* (Vol. 15). https://doi.org/10.1182/blood-2005-11-4354

Rao, D., & Ng, W. K. (2016). Information Pricing: A Utility Based Pricing Mechanism. *Proceedings - 2016 IEEE 14th International Conference on Dependable, Autonomic and Secure Computing, DASC 2016, 2016 IEEE 14th International Conference on Pervasive Intelligence and Computing, PICom 2016*, 754–760.

Rayport, J. F., & Sviokla, J. J. (1995). Exploiting the Virtual Value Chain. *Harvard Business Review*, *73*, 75.

Short, J. (2014). *Valuing Enterprise Data : CLDS Research Project Brief*. Retrieved from http://www.sdsc.edu/Events/ipp_webinars/ipp_webinar_data_value_project_feb_13_2014.html

Stander, J. B. (2015). The Modern Asset : Big Data and by, (December).

Tallon, P. P. (2013). Corporate governance of big data: Perspectives on value, risk, and cost. *Computer*, *46*(6), 32–38. https://doi.org/10.1109/MC.2013.155

Turczyk, L. A., Heckmann, O., & Steinmetz, R. (2007). File Valuation in Information Lifecycle Management. *Managing Worldwide Operations & Communications with Information Technology*, 347–351.

Viscusi, G., & Batini, C. (2014). Digital Information Asset Evaluation: Characteristics and Dimensions (pp. 77–86). Springer, Cham. https://doi.org/10.1007/978-3-319-07040-7_9

Winter, S. G. (2000). The Satisficing Principle in Capability Learning. *Strategic Management Journal*, *21*(10–11), 981–996.

Winter, S. G. (2003). Understanding dynamic capabilities. *Strategic Management Journal*, *24*(10), 991–995.

Zillner, S., Curry, E., Metzger, A., Auer, S. (2017), *European Big Data Value Partnership Strategic Research and Innovation Agenda*. (2017). Retrieved from http://www.bdva.eu/sites/default/files/EuropeanBigDataValuePartnership_SRIA__v3_0.pdf