Towards Process Patterns for Processing Data Having Various Qualities

Agung Wahyudi, Marijn Janssen

Towards process patterns for processing data having various qualities

Agung Wahyudi[1] & Marijn Janssen[2]

[1,2]Faculty of Technology, Policy and Management, Delft University of Technology
Jaffalaan 5, 2628 BX Delft, The Netherlands

[1]A.Wahyudi,[2]M.F.W.H.A.Janssen@tudelft.nl

Abstract. Organizations become more data-intensive and companies try to reap the benefits from this. Although there is a large amount of data available, this data has often different qualities which hinders use. Creating value from big data requires dealing with the variations in quality. Depending on their quality, data need to be processed in various ways to prepare this data for use. Although the processes vary, dealing with certain levels of data quality is a recurring challenge for many organizations. By developing generic process patterns organizations can reuse each other solutions. In this paper, process patterns for dealing with various levels of data quality are derived based on a case study of a large telecom company that employs all kinds of data to create operational value. The process patterns can possibly be used by other organizations.

Keywords: big data, data quality, data processing, process patterns, telecom

1 Introduction

Today's organizations collect more and more data due to datafication. Datafication refers to activities that digitalize all objects which are related to the organizations' processing chain [1, 2]. Data can originate from internal and external sources and might have different qualities. Data quality refers to data that are fit for use by data users or data consumers [3, 4]. The definition of data quality captures a broad perspectives by including by the quality conveyed by the data and the use of the data.

Many studies suggest that organizations can gain benefits from the data if they succeed in unlocking value from the data. This can result in greater efficiency and profits [5] as well as competitive advantages [6-8]. Therefore, organizations are seeking ways to realize the value from their big data [8].

Value creation requires the processing of data. Data can be processed in various ways (e.g. [9], [10], [11]). Although the idea of drawing value from the data seems to be straightforward, many organizations failed to do so. According to a recent study by Reid, Petley [12], two third of businesses across Europe and North America failed to unlock value from big data. In this paper, we identify generic process patterns that can be used by any organization to deal with data which have various data qualities. The

adfa, p. 1, 2011.

various data qualities of internal and external sources require organizations to deal with them in various manners. Which process should be followed depends on the data qualities. These variations have some similarities that create process patterns of how organizations deal with them. Which process should be followed depends on the data qualities. The data quality provides the initial set of conditions to select the process steps that are necessary to prepare the data for use. Such patterns can be viewed as a practice which can be reused or from which others can learn. We view a process pattern as a recurring sequence of steps that results in the accomplishment of a certain value. Given certain starting conditions, the patterns can be followed to create value form the data. Process patterns should be independent on the technology implementation. They should enable organizations to more easily create value from the data.

The objective of this research is to derive process patterns for creating value from data. For this purpose, a case study in the Telecom industry is investigated and typical process patterns are derived. The research approach is presented in section 2. On the basis of state-of-the-art literature in Section 3 and a case study at a telecom in Indonesia presented in Section 4, the patterns will be presented in Section 5. Finally, conclusions will be drawn in Section 6.

2 Research Approach

The purpose of this study is to identify process patterns. First, the concept of big data quality, data processing, and process patterns were identified from a rigorous review of literature. Since this study aims at enhancing our understanding of how an organization overcame big data quality challenge to create value from the data, a qualitative case study-based approach was used to inductively arrive at process patterns [13]. Qualitative case study research is widely used in information systems research, and is well suited to understanding the interactions between information technology-related innovations and organizational contexts [14]. According to Yin [13], the case study includes a variety of data collection instruments to ensure construct validity.

The following criteria were used for the selection of the case: 1) the case is in an information-intensive organization context; 2) the case employs and combines many datasets to create operational value; and 3) case study information should be available and accessible. We conducted a case study within the context of PT Telekomunikasi Indonesia Tbk., the biggest telecom in Indonesia, which provided the researchers with unlimited access to subject matter experts and internal documentation for all the cases. This helped to perform triangulation to ensure the construct validity of the case study [13].

3 Literature Review

3.1 Data Quality (DQ)

Data quality (DQ) is a prominent challenge mentioned in the big data literature [8, 15-21]. As described by Redman [22], low DQ impacts on operational level, tactical level,

and strategic level of organizations, e.g. cost increase (to 8-12% of revenue), poorer decision making, and difficulties in setting strategies.

Wand and Wang [23] define DQ as "data that are fit for use by data users or data consumers" (p. 6). This definition draws the attention to the view that DQ is not only related to the data it conveys, but also to the use of the data. Wang and Strong [4] classify DQ into four types based on data of consumers' point of view, namely 1) intrinsic DQ that denotes that data have quality in their own right (e.g. accuracy); 2) contextual DQ that highlights the requirement that DQ must be considered within the context of the task at hand (e.g. value-added); 3) representational DQ describing that DQ is related to data representation (e.g. interpretability); and 4) accessibility DQ that emphasizes the importance of computer systems that provide access to data (e.g. accessibility).High DQ is explained by Wand and Wang [23]: "high quality data should be intrinsically good, contextually appropriate for the task, clearly represented, and accessible to the data consumer" (p. 22).

In order to unlock value from the data, organizations very often combine many datasets from various data sources whether internally or externally. Those datasets may have varieties of DQ which organizations should take into account when they process them. When DQ is low, often a process are started to deal with the low quality, before the data can be used. In this way, 'garbage in is garbage out' is avoided.

3.2 Data Processing

Due to the variability of DQ in datasets, there is no uniform way to process them. As such, which process should be followed depends on the DQ. Normally, data are processed sequentially in data lifecycles which encompass all facets of data generation to knowledge creation [10]. There are many models of data lifecycles in the literatures. Some prominent ones are Data Documentation Initiative (DDI) Combined Lifecycle Model [9], DataOne Data Lifecycle [10], and ANDS Data Sharing Verbs [11]. The DDI Combined Life Cycle Model has eight activities in a data lifecycle, namely 1) study concept; 2) data collection; 3) data processing; 4) data archiving; 5) data distribution; 6) data discovery; 7) data analysis; and 8) repurposing. Meanwhile, activities defined in the DataOne Data Lifecycle are 1) planning, 2) collecting, 3) assuring, 4) describing, 5) preserving, 6) discovering, 7) integrating, and 8) analyzing. The ANDS Data Sharing Verbs consist of: 1) create, 2) store, 3) describe, 4) identify, 5) register, 6) discover, 7) access, and 8) exploit.

Although these models use various terminologies, all models of the data lifecycles have common activities which reflect data provider's and data consumer's point of view. Our study focuses on data consumer's point of view because we focus on the process of unlocking value from many data which have varying DQ.

From data consumer's perspective, the first step of data processing is to discover relevant data from data providers. It could be conducted by using searchable interfaces to locate the data or by making agreements with data providers. This step may require user registration and signing in.

The next step is to access the data. Data can be accessed either through an automated system (i.e. using a Web link, perhaps passing through an authentication barrier and/or licensing agreement), or by an application to a data consumer.

Third, data need to be exploited. Data exploitation requires good technical metadata (fields, descriptions, metrics, etc.), which provide contextual information about the way the data were created. Cleansing, parsing, and other functions to prepare the data to be fit for analysis are also involved in this step. Moreover, it also includes the transformation of several different datasets into a common representation (format, coding scheme, and ontology), accounting for methodological and semantic differences while preserving a provenance trail. In addition, the dataset very frequently needs to be combined with other datasets so that more insights or knowledge could be obtained.

The final step is to analyze the data. We apply statistical and analytical models to the data in order to extract meaningful answers to the prior research questions.

3.3 Process Patterns

The aforementioned data lifecycle provides the bases for creating process patterns. Data processing may vary based on DQ of the data. For example, internal data which have a high DQ need not to be assessed as this is already known, but external data (e.g. Twitter data) should be assessed first and maybe cleansed prior to exploitation. The variation of data lifecycle for dealing with a dataset results in a process pattern.

The terminology process pattern is comprised of "process" and "pattern". According to Davenport [24], a process is "a specific ordering of work activities across time and place with a beginning and an end, and clearly identified inputs and outputs: a structure for action" (p. 21). Inline with this, Ambler [25] defines a process as "a series of action to produce one or more outputs from one or more inputs" (p. 2). He also defines a pattern as "a general solution to a common problem, one from which a specific solution may be derived" (p. 4). Patterns have been applied in various domains, e.g. architecture, economics, telecommunication, business, and software engineering [26]. Patterns in software engineering come in many flavors, including (but are not limited to) analysis patterns, design patterns, and process patterns. Hagen and Gruhn [27] define process patterns as "patterns that represent proven process which solves a frequently recurring problem in a pattern like way" (p. 1). Process patterns provide flexibility in their use since one can select and apply a suitable process pattern according to the situation under study.

In the literature, there is no consensus about what should be included in a process pattern. Buschmann, Meunier [26] mentions that a pattern must consist of contexts, problems, and solutions. A context of a pattern describes a design situation that gives rise to a design problem. The problem describes a concrete situation which may emerge in the contextual application. A pattern should mention internal and external forces, e.g. influences of customers, competitors, component vendors, time and money constraints and requirements. The solutions describe the process that consists of a set of activities that are supposed to solve the problem if they are executed. Process patterns of overcoming DQ challenges assist organizations in creating value from the data. They also serve as catalogs and repositories to the organizations for future use.

4 Case Study

The goal of the case study was to derive process patterns. For this reason, the case study involved multiple methods for collecting data. In the case study, the primary processes of PT Telekomunikasi Indonesia Tbk., a state-owned telecom company in Indonesia, were selected for analysis.. The primary processes of the CDMA marketing department were focused on, as the marketing department dealt with various sources of data having a variety of data qualities.

The traditional way of marketing based on intuition mostly resulted in ineffective targeting, segmenting, and positioning of products. As a result, the program often ended with unsatisfactory returns on marketing investments. The program did not have sufficient justification and it was hard to predict its success. Moreover, the marketing activities sometimes unexpectedly turned back to the rise of customer complaints, customer churn, and financial loss. The tight competition in the market and the increasing power of customers kept forcing them to respond competitor moves and customer voices with attractive programs. On the other hands, the programs should be designed very carefully not to impact their customers' satisfaction and long-term profitability, e.g. discounting cash-cow products for which customers were willing to pay or giving massive nationalwide promotion which would result in network overload and congestion in dense cities.

Many data generated by the telecom, e.g. transaction data, customer data, machine logs, network performance data, etc., and external data such as social media, crowdsourced maps, could potentially be used to target customers better. By combining those data, they designed an attractive discount program. As mentioned by Verhoef, Kooge [28], organizations can obtain value from the data in a bidirectional way, i.e. value to customer and value to organization. From the program, customers of the telecom benefited by obtaining budget communication solution and perceived quality. Meanwhile, as the impact of customer experience, the telecom benefited by increasing its market share, improved brand recognition, and high return on marketing cost. The way the telecom turned the data into value is illustrated in Fig. 1.

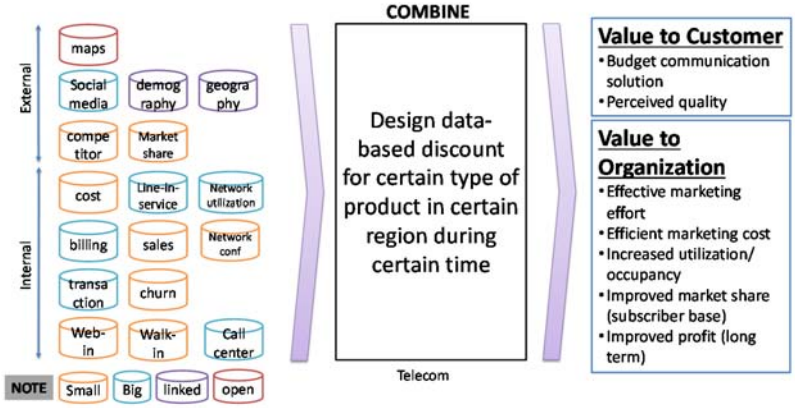

Fig. 1. Case Study: A telecom designed a data-based discount program

The company built the information system that had a number of functionalities to process the big data. Prior to running the program, an initial kick-off meeting that included data providers and related departments was held. The marketing and IT department proposed a model that described how to turn the data into decision. From the model, they listed all the required data and made agreements with the data providers on access, metadata, cut-off time, etc. The IT department built a data lake to pool data that had been retrieved with restricted/limited access and some data from machines with concurrency issue. They also employed a number of tools to cleanse the low quality data and parse the data that had unfit representation for the further process. Syncsort DMX-h Hadoop application was utilized to exploit the data. The application had extracting, transforming, aggregating, and loading functionalities. Many datasets were combined and transformed based on the task at hand. The processes involved a number of execution activities that include one or more datasets, e.g. joining, aggregating, then manipulating fields, rejoining, etc. Furthermore, the data were analyzed using trade-off analytics and visualized using Microsoft Excel. The program was then proposed to the board of executives for decision. Sometimes, iterations between the aforementioned processes occurred.

Initially, drawing value from the data seemed straightforward using the functions provided in the information system. However, it was found to be complex in terms of data quality variation. Since the telecom incorporated many datasets, the DQ of the datasets varied greatly. Internal big data and partners' data usually come with high intrinsic DQ because these data were self-managed (e.g. by periodic calibration of data-generating sensors, quality control, or using service-level agreements with partners). However, the datasets may have low accessibility DQ, contextual DQ and representational DQ. For example, call center recordings which were mostly unstructured caused difficulties for technical staffs to process (ease of operation); many data were just thrown to the data lake but never used (value-added); machine logs had varied representation depending on the machine's manufacturer (consistent representation).

Unlike internal big data, external big data such as social media very often had low intrinsic DQ. For example, Twitter data might have a data-biased issue (i.e. they represent only certain groups of people, e.g. young generation). The unbiased data could lead to inaccurate outcomes if employed to make a conclusion about the entire population. External big data were also reported to have low accessibility DQ (e.g. license/subscription fee which leads to no/limited access to the data), low representational DQ (e.g. no metadata which causes a problem in understanding and interpreting the data), and low contextual DQ (e.g. outdated statistical data which was not fit in the task).

As a result, there were many ways of data processing based on DQ of the data. Fortunately, some patterns were indicated in the case that showed specific solutions for particular problems of DQ. From these, process patterns could be derived. Recurring specific processes were found, namely data processing, for a particular DQ problem, from which a process pattern was derived. For example, accuracy problem was solved by cleansing the data before exploitation.

5 Discussion

The objective of this paper was to identify process patterns for dealing with different DQ. To do so, the organizations developed an information system having various functionalities. Data were processed in a data processing sequence, which followed the data lifecycle from a high level of abstraction. From the literature and the case study, we derived the following typical data lifecycle. We used similar steps as found in the literature (See Section 3.2), but we extended this by including a managing process that could occur in any step of data lifecycle. It consists of connecting, controlling, and integration functions so that the data processing sequences can be executed. Moreover, we listed all functionalities related to every step of nominal data lifecycle. The nominal data lifecycle together with the functionalities used to process the data is shown in Fig. 2.

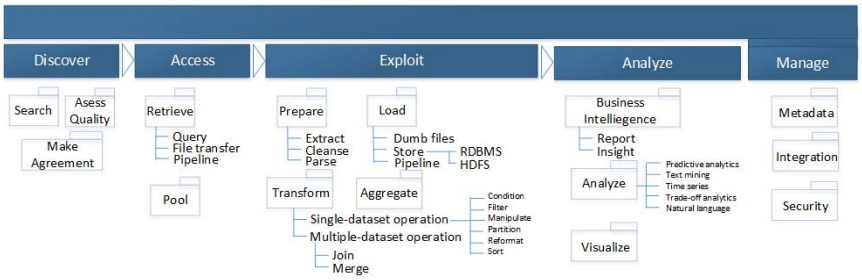

Fig. 2. Functions used in each step of a nominal data lifecycle

The first step in the nominal data lifecycle was the "discover data" step. In this step, some functions such as search, assess quality, and make agreement were employed. Search functions assisted them to quickly find relevant data from many datasets in the data lake. Assessing quality is important to determine whether actions to improve the quality are needed in the subsequent steps. In order to use the data properly, organizations make agreements with the data providers on: 1) what data should be included in the process?; 2) how to retrieve the data?; 3) when was the cut-off time or the retrieval time?; 4) how to read the data?; and 5) what if the data were not intrinsically good (e.g. corrupted)?

The "access data" step consists of retrieving and pooling the data. Retrieving the data is strongly related to accessibility. A number of activities were used, such as query, flat file transfer, or process pipeline. Sometimes organizations pool the data in the data lake for several reasons, such as limited/restricted access, concurrency issue, etc.

The third step, "exploit data" is one of the most challenging steps in terms of the application complexity. Because there is seldom a single application that has all the functionalities, various applications having separate functions are composed together to perform data exploitation. Interoperability and standardization are key success factors to get all applications working together. Some functions in this step are preparing,

transforming, aggregating, and loading the data. In the "preparing step", some data might need to be extracted because they are retrieved as compressed flat files, cleansed because they contain low intrinsic quality (e.g. low accuracy), or parsed because their original representation is not fit for the further process. The organizations transform the data using single-dataset and multi-dataset operations. Functions such as conditioning, filtering, manipulating, partitioning, reformatting, sorting, joining, and merging are selected based on the task in hand. The combination and iteration of those functions are found very often. Aggregating the data is supposed to reduce the data based on certain fields. The outputs are loaded either to dumb flat files, stored in the relational databases, passed to HDFS, or put into the pipeline to the next process.

The next step was to "analyze" the data. Functions included in this step are business intelligence, analyzing, and visualizing the data. Business intelligence has been used extensively to generate reports. Analyzing the data is the most difficult task because it creates the value of the data. The data were analyzed using various analytical methods such as predictive analytics, text mining, time series, trade-off analytics, and natural language, depending on the task in hand. In the case study, the telecom exhibited trade-off analytics between the projected revenue (from existing customer and new subscribers) and projected cost (from revenue opportunity loss and marketing campaign expense). Visualizing data is important to quickly grasp insights (e.g. trend, relationship) between datasets.

The step of "manage" data is not part of the sequential process, but manages all the aforementioned data processing steps. It ensures the data pressing sequence run smoothly. It is conducted thru metadata, integration, and security. Metadata is important in order to understand and interpret the data so that they could be reused. Integration ensures the involvement of many actors and the utilization of many applications could run smoothly.

From the case study, we found that every dataset had a variation of data processes depending on its DQ. Fig. 3 shows a data process pattern for the situation in which all datasets have high DQ.

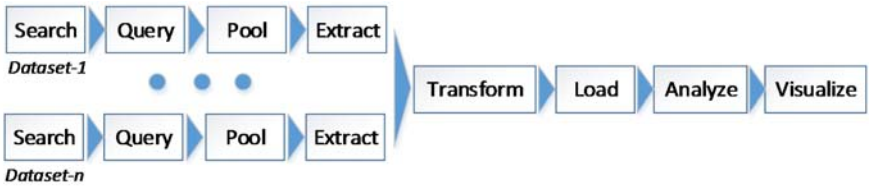

Fig. 3. Typical data process pattern when all datasets have high DQ

However, if any dataset has a low DQ, the data processing takes different paths, different from the typical process. Combining the concept of DQ from Wang and Strong [4] and the case study, we derived four process patterns as described in Table 1. The process patterns consist of DQ context, problem, and the solution that reflect the modification of the typical data process.

Table 1. Process patterns of DQ challenge: the context, problem, and solution (the red words indicates different patterns from the typical process)

DQ dimension	Dataset	Problems	Process Pattern
Intrinsic - Accuracy - Objectivity	Customer complaint from social media	Some data were from real customers, other data maybe from black campaigners	1 Search ➔ Assess accuracy ➔ Query ➔ Pool ➔ Extract ➔ Cleanse ➔ Transform ➔ Load ➔ Analyze ➔ Visualize
Representational - Interpretability - Consistent representation	Network performance data	Varied terminologies and data representation across vendors' machines	2 Search ➔ Make agreement ➔ Metadata ➔ Integration ➔ Query ➔ Pool ➔ Extract ➔ Transform: Manipulate ➔ Load ➔ Analyze ➔ Visualize
Accessibility - Accessibility - Security	Transaction data	- Existing machines were not capable of handling many concurrent access (scalability) - Very restricted access - Privacy issue	3 Search ➔ Access securely ➔ Query ➔ Pool: Data lake ➔ Extract ➔ Transform: Manipulate ➔ Load ➔ Analyze ➔ Visualize
Contextual - Value-added	Many datasets	Lack of knowledge of how to derive value	4 Search ➔ Metadata ➔ Query ➔ Pool ➔ Extract ➔ Transform ➔ Load ➔ Analyze: Model ➔ Visualize

Process pattern 1 represents the change of typical data processing to take low intrinsic DQ into account. The example of dataset from the case study is customer complaint from social media (e.g. Twitter). The data have low accuracy because some conversations were not generated by real customers, but maybe driven by fake accounts and black campaigners. Moreover, not the entire customers are represented in the social media, e.g. the old generation. Therefore, prior to exploitation, the data should be assessed for their accuracy. Because we are unable to improve their accuracy, cleansing is the only way to remove data with low accuracy.

Process pattern 2 considers the low representational DQ. From the case study, the network performance data generated by machines from many vendors are hard to interpret because varied terminologies are used across vendors' machines. Therefore, metadata are very important to understand so that they can be reused. The data also have inconsistent representation, because each vendor has different formulations of per-

formance indicators (e.g. drop call). Organizations should make agreements on performance indicators (e.g. standardization) that are applied across vendors' machines. In the exploitation step, the fields containing performance indicators need to be manipulated so that they represent consistently for the subsequent process. In a multi-software vendor environment, often different methods of access are used, e.g. direct query to databases, file transfer, query from the vendor's application, SNMP logs, etc. Therefore, integration ensures that the information system could handle multiple ways of access.

Low accessibility DQ is represented by *process pattern 3*. From the case, the transaction data were generated by the machines that were not designed to process many concurrent connections. Therefore, organizations create a data lake to store the data so that they access the source once but reuse the data many times from the data lake. Moreover, organizations may have strict regulations about access to the machines. Hence, accessing data from the data provider in a secure way is very important. Privacy issue could concern organizations. Therefore, in the exploitation process, they may manipulate the fields related to privacy to be anonymous.

Process pattern 4 addresses low contextual DQ. Most of datasets have unknown value prior to the use. Therefore, the model to use the data in the analysis step is important for the organizations to create value from the data. Metadata are also important so that the data could be put into a contextual use.

The process patterns describe the recurring problems of big data quality together with the solutions that consist of certain functions to solve the problem. The process patterns can be reused for any organization in order to create value from the data.

6 Conclusion

The objective of this paper is to derive process patterns for creating operational value from data in information-intensive organizations. Four patterns have been identified based on differences in data quality. A case study in a telecom was investigated. In the case, the telecom combined many data, including internal and external data, to design a big data marketing program in order to increase market share and profit.

The creation of value from the data depends heavily on the data quality. Various data need different data processing steps based on the quality they have. This led us to look for process patterns for every big data quality problems. Combining literature review and the case study, we proposed four process patterns that map big data quality problems in data processing, together with the following solutions. . Process pattern 1 deals with low intrinsic data quality, e.g. inaccurate and biased. Functionalities such as accessing accuracy and cleansing are added to the typical data processing pattern. Low representational data quality is encountered by process pattern 2. Challenges like interpretability and consistent representation are solved by the functionalities such as metadata, making agreements, integration, and manipulation. Process pattern 3 considers low accessibility data quality. Secure access, building a data lake, and data manip-

ulation are needed from dealing with restricted access, concurrency, and privacy. Process pattern 4 copes with low contextual data quality. To turn the data into value, models to use the data and metadata are two important elements in this pattern.

A limitation of this study is that the patterns were derived using a single case study in a particular field. It is recommended that more empirical research is conducted in other fields to test and refine the proposed process patterns as well as to evaluate the process patterns and the significance of their elements.

References

1. Mayer-Schönberger, V. and K. Cukier, *Big data: A revolution that will transform how we live, work, and think*. 2013: Houghton Mifflin Harcourt.
2. Bauer, F. and M. Kaltenbock, *Linked Open Data: The Essentials*. 2011.
3. Lee, Y.W., et al., *AIMQ: a methodology for information quality assessment.* Information & management, 2002. **40**(2): p. 133-146.
4. Wang, R.Y. and D.M. Strong, *Beyond accuracy: What data quality means to data consumers.* Journal of management information systems, 1996: p. 5-33.
5. Gantz, J. and D. Reinsel, *Extracting value from chaos.* IDC iview, 2011(1142): p. 9-10.
6. Manyika, J., et al., *Big data: The next frontier for innovation, competition, and productivity.* 2011.
7. Zikopoulos, P.C., et al., *Understanding big data.* New York et al: McGraw-Hill, 2012.
8. LaValle, S., et al., *Big data, analytics and the path from insights to value.* MIT sloan management review, 2013. **21**.
9. Green, A. and J.-P. Kent, *The Metadata Life Cycle.* MetaNet Work Package 1: Methodology and Tools, 2002: p. 29-34.
10. Michener, W.K. and M.B. Jones, *Ecoinformatics: supporting ecology as a data-intensive science.* Trends in ecology & evolution, 2012. **27**(2): p. 85-93.
11. Burton, A. and A. Treloar, *Designing for discovery and re-use: the 'ANDS data sharing verbs' approach to service decomposition.* International Journal of Digital Curation, 2009. **4**(3): p. 44-56.
12. Reid, C., et al., *Seizing the information advantage: How organizations can unlock value and insight from the information they hold.* 2015.
13. Yin, R.K., *Case study research: Design and methods*. 2013: Sage publications.
14. Nag, R., D.C. Hambrick, and M.J. Chen, *What is strategic management, really? Inductive derivation of a consensus definition of the field.* Strategic management journal, 2007. **28**(9): p. 935-955.
15. Zuiderwijk, A., et al., *Socio-technical impediments of open data.* Electronic Journal of e-Government, 2012. **10**(2): p. 156-172.
16. Chen, C.L.P. and C.-Y. Zhang, *Data-intensive applications, challenges, techniques and technologies: A survey on Big Data.* Information Science, 2014.

17. Fan, J.Q., F. Han, and H. Liu, *Challenges of Big Data analysis.* National Science Review, 2014. **1**(2): p. 293-314.
18. Marx, V., *THE BIG CHALLENGES OF BIG DATA.* Nature, 2013. **498**(7453): p. 255-260.
19. Millard, I., et al., *Consuming multiple linked data sources: Challenges and Experiences.* 2010.
20. Zhou, Z.H., et al., *Big Data Opportunities and Challenges: Discussions from Data Analytics Perspectives.* Ieee Computational Intelligence Magazine, 2014. **9**(4): p. 62-74.
21. Zicari, R.V., *Big data: Challenges and opportunities.* Big data computing, 2014: p. 103-128.
22. Redman, T.C., *The impact of poor data quality on the typical enterprise.* Communications of the ACM, 1998. **41**(2): p. 79-82.
23. Wand, Y. and R.Y. Wang, *Anchoring data quality dimensions in ontological foundations.* Communications of the ACM, 1996. **39**(11): p. 86-95.
24. Davenport, T.H., *Process innovation: reengineering work through information technology*. 2013: Harvard Business Press.
25. Ambler, S.W., *More Process Patterns: Delivering Large-Scale Systems Using Object Technology*. 1999: Cambridge University Press.
26. Buschmann, F., et al., *A system of patterns: Pattern-oriented software architecture.* 1996.
27. Hagen, M. and V. Gruhn. *Towards flexible software processes by using process patterns*. in *IASTED Conf. on Software Engineering and Applications*. 2004.
28. Verhoef, P.C., E. Kooge, and N. Walk, *Creating Value with Big Data Analytics: Making Smarter Marketing Decisions*. 2015: Routledge.