APPROACHES TO PROVIDING SECURITY IN DATA QUALITY MANAGEMENT

Marisia TOMA[1]
Ștefania LEUCA[2]

ABSTRACT:
TECHNOLOGICAL DEVELOPMENT HAS BEEN INCREASING IN VELOCITY AND THE COMMITMENT OF PROVIDING DATA QUALITY BECOMES PREREQUISITE ESPECIALLY FOR SECURITY PURPOSES. THUS, ORGANIZATIONS REQUIRE ONLY RELEVANT DATA IMPOSING AN INDISPENSABLE NEED FOR A COMPREHENSIVE COLLECTION, MEASUREMENT AND MANAGEMENT OF DATA. THIS ARTICLE IS A BRIEF OVERVIEW OF THE MAIN PRESUMPTIONS IN THE FIELD, IN ORDER TO PROVIDE A NOTIONAL UNDERSTANDING FOR DATA QUALITY MANAGEMENT AND ITS OPERATION TO INFORMATION SECURITY. THE CONCEPT THAT EVOLVED FROM A TREND INTO A REAL NECESSITY HASN'T BEEN EXHAUSTIVELY ASSERTED IN LITERATURE. THEREFORE OUR OUTLINE IS A CONSTRICTED PERSPECTIVE ABOUT ITS MEANING, PROCESSES AND APPLICATION ADDRESSED TO OUR FELLOW STUDENTS AND SCIENTIFIC RESEARCH COMMUNITY UNDERSTANDING. THE ORGANIZATIONAL SPECTRUM REQUIRES THE NEED OF SECURING DATA QUALITY ADMINISTRATION AS AN ESSENTIAL ASSET HENCE, THIS ASSUMPTION CALLS FOR A MORE DEEPENED ANALYSIS.

KEY WORDS: DATA QUALITY, DATA MANAGEMENT, INFORMATION SECURITY

INTRODUCTION

New technologies have opened up the world to us but at the same time have opened us up to the world as well. These high tech components have permitted us to create our own digital universe, operating in an always-on differential environment: we work online, we socialize online, we follow news, we bank online and everything we do leaves a digital footprint. Performing in such a climate requires us to share fundamental data conceiving a vast ocean of information. While the world has recognized an exponential data growth, the ability to look at it and analyzing it has been increasing. Hence, people and organizations encountered the necessity a prerequisite distinction between the

[1]BA, "Babeș Bolyai" University, Romania, marisia.toma@gmail.com.
[2]BA, "Babeș Bolyai" University, Romania, stefanialeuca02@gmail.com.

ISSN-P: 2247-4455 / ISSN-E: 2285-9632

relevant and irrelevant data. The process of distinction represents the data management and the quality lies in relevance vs. irrelevance of data.

Our presentation outlines the main presumptions in the area of intelligence, in order to provide a notional understanding for data quality management and its operation to information security. Therefore, our synthesis is a constricted perspective about its meaning, processes, application and a requirement of the need to securing data quality administration as an essential asset. Through a descriptive, qualitative and quantitative analysis of the most representative publications in the field highlighting the chosen methodology, our aim is to introduce our public to the process of data quality management, formulating a brief literature review to its discipline.

Firstly, we consider of high importance to emphasize the basic concept that stands as a core principle of data quality management process: data. Secondly we have reviewed the attributions that data quality implies, in providing the security to a flourishing business. Lastly, in section 3 we evaluated the approaches of the scholars in regard with the effectiveness and efficiency of data quality management.

WHAT IS DATA?

Along with the evolution of technology, the world has found itself drowned in an "ocean" of data. What is actually data? Data represents a set of characters or meaningful symbolic constructs, (alphabetic letters, numbers, genetic sequences etc. as a basis)[3] that has been gathered and translated for the purpose of analysis. The conducted analysis concluded in an indispensable demand and usage of information for the global development. Depending upon the nature of its use, data has various explications and can be utilized at different levels of filtration, as a result of the semiotic analysis.

Semiotics as a theoretical framework makes the transition from the physical world where signs are created, to the social world of norms where a meaning is attached to it. This leads to a comprehensive understanding of data, information and knowledge. The semiotic levels that can be attributed to the 4 concepts are: syntactic, semantic, pragmatic and empiric. The syntactic level conveys a set of signs structured and governed by formal rules together becoming data. The semantic level refers to the study of the meaning of signs which become useful, only when they indicate a certain action. When a meaning is attributed to the selected data, in a particular context it becomes information. Pragmatics is concerned with the relationship between signs and their behavior and involves information that has intentional use. Information becomes knowledge to the moment of its interpretation. Empirics is the physical property concerned with thee signals used to code and transmit the message. Communication channels and their characteristics stand as the core to correct transmission of content.

Semiotics analysis can be used to understand the technical, formal and informal systems of an organization. The technical level involves technology and system security measures, while formal rules and procedures address the information systems security issues arising at the formal level. At the informal level, pragmatic concerns are paramount towards the development of security culture and environment. As we have discovered above, when speaking about data, we refer to the syntactic level of the semiotic analysis. What is actually the study of syntactics? It comprises the

[3] Gurvirender Tejay, Gurpreet Dhillon, and Amita Goyal Chin, "Data Quality Dimensions for Information Systems Security: A Theoretical Exposition," in *Security Management, Integrity, and Internal Control in Information Systems* (Springer, 2005), 23.

logic and grammar with an emphasis on the physical form of the system analyzed. Syntactics studies the relation of signs to one another and how complex signs originate from simple ones. Therefore, the concern at a syntactic level is rather on the formal and structural aspect, meaning that the information is viewed as just a simple raw material. Considering these elements, data represents empirical facts that have been arranged and organized according to a certain structure, to a set of meaningful rules. Moving forward with the semiotic analysis, the next level is expressed by semantics. This level concentrates on the meaning of the data, in order to transform it into information. In this regard, data needs a particular context to become meaningful and useful to the recipient. "Meaning is appropriated when we draw from our knowledge and apply understanding to information. Ensuring good quality at a semantic level therefore moves from maintaining quality of data to that of information"[4].

Finally the pragmatic level, the last level in the semiotic analysis, is evaluating the relationship between the empirical facts and the meaning attributed to them. The beliefs, shared assumptions, expectations etc. are all part of the pragmatic structure. In this level, the information has an intentional use, thus becoming knowledge. Nonaka & Takeuchi (1995, pp. 58) suggest a relationship to knowledge: "...information is a flow of messages, while knowledge is created by that very flow of information anchored in the beliefs and commitment of its holder. This emphasizes that knowledge is essentially related to human action".

A semiotic framework for understanding data quality objectives in the context of the decision process is more appropriate than using data quality objectives alone, because it reflects which quality objectives are most significant at each stage of the production process.

PROVIDING DATA QUALITY

Data quality is becoming a prerequisite priority in everyday life, as an important asset of data management. The aftermath of data analysis endorses data quality issues that are causing significant lose in money, time and opportunities. An aspect of poor quality is the cost usually hidden and not obvious to those not looking for it. In the security field, the necessity of understanding the data quality dimensions becomes essential. As such, semiotics helps us in this endeavor by providing a notional analysis of the dimensions at the 4 levels: empiric, syntactic, semantic and pragmatic. Therefore, data quality dimension represents a set of data quality attributes that represent a single aspect or construct of data quality.

The empiric level is concerned with the establishment of the means of communication and data handling. Hence, the dimensions of data quality performing at this level include: accessibility, timeliness, security, portability and locality. These dimensions are concerned with the problems of medium of communication rather than data itself. However, they are o highly importance as unavailable channels would hinder the access to data and unauthorized access would lead to massive security breaches.

Syntactics represents the imperative (momentous) level in the analysis of the data quality dimensions. These dimensions cover the logic, grammar and structural aspects that compose data in order to ensure the best quality of that specific data. The research literature has found these aspects as being consequential: accuracy appearance, arrangement, clarity, coherence, comparability, compatibility, completeness, composition, conciseness, consistency, correctness,

[4] Tejay, Dhillon, and Chin, 25.

ISSN-P: 2247-4455 / ISSN-E: 2285-9632

ease of operation, ease of use, flexibility, format, freedom from bias, integrity, level-of-detail, objectivity, portability, presentation, readable, redundancy, robustness, uniqueness and usable. At a semantic and pragmatic level, the dimensions deal with the issues of information and knowledge rather than data. Nonetheless, facets associated with knowledge quality and information quality should be endorsed as part of any equation devised to address the problems of data quality. Thus, there is an inter-dependent nature between knowledge, information and data that should always be addressed when referring to data quality.

DEFINING THE CONCEPTS IN DATA QUALITY MANAGEMENT

Data quality and data management are part of a larger image when we address to a broader use of information. In order to have a proper understanding of the explanatory expansion of our article we will associate the concept of *data* to the concept of *information* which we will describe as "data that have been processed".[5] Manufacturing information is the development of data processing and it can be viewed as a processing arrangement that uses raw materials in order to create physical products. When producing information products managing for quality is particularly important because use of the data may expose broad impacts to the efficiency of an organization, to future scientific investigation, or to our society. Therefore, principles, guidelines and techniques are required for information products quality and knowledge has been created for data quality practice.

Information manufacturing comprises the evolution of data where the input is represented by raw data, the process is accomplished by the information systems and the output is represented by information products. For continuous investigation and advancement an organization would need guidelines in order to identify the critical issues and develop procedures to assess data quality projects. Information manufacturing system is referred to be a system that produces information products. The concept of information product asserts that the information products from the manufacturing information systems have transferable value to the consumers internal or external and we define three roles of the consumers in the process of information manufacturing systems:

1. Information suppliers are those who create and collect data for the information products.
2. Information designers are those who develop and maintain the data and the systems infrastructure.
3. Information consumers are those who use the information products in their work.

Also, we define information product managers as those who are responsible for managing the entire information product production process and the information product life cycle. Each of them is associated with a process or task:

- Suppliers are associated with data-production processes.
- Designers are dealing with data storage, maintenance, and security.
- Consumers are associated with data-utilization processes.

Poor data quality limits the usefulness of the information products establishment. Defining, measuring, analyzing, and repeatedly improving information quality is essential to ensuring high-quality information products. Therefore, a total data quality management is required. The definition of the total data quality cycle describes important data quality dimensions and the

[5] Bruce McNutt, *The Fractal Structure of Data Reference: Applications to the Memory Hierarchy*, The Kluwer International Series on Advances in Database Systems 22 (Boston, MA: Kluwer Academic, 2000), 2.

corresponding data quality requirements. The measurement component produces data quality metrics. The analysis component identifies foundation causes for data quality problems and calculates the impacts of poor quality information. Finally, the improvement component provides techniques for improving data quality. These components are applied along data quality dimensions according to requirements specified by the consumer.

DEVELOPING SOLUTIONS

The main aspect to discuss concerns the managerial approach, for example, the strategy that has to be adopted in an organization in order to take proper technical choices. The choices are in terms of data quality activities to be achieved, databases and flows to be treated, and techniques adopted. So, in the final stage of data quality management, the focus is moved from technical to managerial facet. The expansion of the steps which are: assessment, improvement and management of improvement solutions from an organizational perspective, provides evidence of the attention devoted to this concern. Specific tasks of the managerial perspectives involve:

1. Assessment of organization preparation in following data quality processes.
2. Analysis of customer satisfaction, in order to discover problems at the source, for example, directly from the service users.
3. Initial focus on a pilot project, in order to experiment with and tune the approach and avoid the risk of failure in the initial phase, which is typical of large-scale projects performed in one single phase. This principle is inspired by the well-known motto “think big, start small, scale fast.”
4. Definition of information stewardship, for example, the organizational units and their managers who, with respect to the laws and rules, have specific authority on data production and exchange.
5. Pursuing the results of the preparation assessment, analysis of the main barriers in the organization to the data quality management perspective in terms of protection, to change processes, control establishment, information sharing, and quality certification.
6. Founding a specific relationship with senior managers, in order to get their consensus and active participation in the process.

Before concluding, there are also a second set of major managerial principles we should take into consideration:

• “Principle 1. Since data are never what they are supposed to be, check and recheck schema constraints and business rules every time fresh data arrive. Immediately identify and send discrepancies to responsible parties.

• Principle 2. Maintain a good and strict relationship with the data owners and data creators, to keep up with changes and to ensure a quick response to problems.

• Principle 3. Involve senior management willing to intervene in the case of uncooperative partners.

• Principle 4. Data entry, as well as other data processes, should be fully automated in such a way that data be entered only once. Furthermore, data should only be entered and processed as per schema and business specifications.

• Principle 5. Perform continuous and end-to-end audits to immediately identify discrepancies; the audits should be a routine part of data processing.

• Principle 6. Maintain an updated and accurate view of the schema and business rules; use proper software and tools to enable this.

ISSN-P: 2247-4455 / ISSN-E: 2285-9632

• Principle 7. Appoint a data steward who owns the entire process and is accountable for the quality of data.
• Principle 8. Publish the data where it can be seen and used by as many users as possible, so that discrepancies are more likely to be reported".[6]

In order for organizations to have a significant competitive advantage, they must structure, evaluate and implement an effective and efficient data quality management program. This program requires indispensable tools, brand-new techniques and effective information systems to maintain the integrity of the data as a valuable asset. To achieve tenacity in the outcomes of the process, a data quality assurance framework is needed to mitigate the consequences of misconceived and misapplied strategies calling for a commitment to security.

CONCLUSION

A taxonomy or conceptualization cannot be empirically confirmed and is not intended to provide prescriptive guidance; its value comes from its applicability and establishes the theoretical foundations for quality management, as well as to identify and develop techniques for quality improvement based on the comprehensive body of knowledge. Understanding the nature of quality enables us to focus on the most applicable issues for future research and development.

Production process are unique in three ways: unlike total data quality management the production process surrounds the generation of raw data and products derived from those data; the lifetime of the data product stored in an archive will be longer than the life time of the instruments that generated the data and their quality must be evaluated subjectively as well as objectively; and the continuous improvement process must include the clarification of data products in an archive, and the integration of continuous learning into the production process to enable the clarification.

Issues can have a very long life. We believe that some of them can live forever. This leads to the need for an explicit treatment of them. It also calls for issues to be very accessible in their own collection of dada. Issue resolutions are often considered interruptive to the normal flow of work through departments that develop and deploy information technology. They will tend to get altered easily if not monitored and placed in front of management on a regular basis. These activities need to become the normal flow of work. Monitoring data quality and making corrections to improve it should not be considered an inconvenience, but should be considered a regular part of information systems operations. This article highlights the need to coordinate the activities of data quality assurance with the complete information systems agenda.

[6] Carlo Batini and Monica Scannapieca, *Data Quality: Concepts, Methodologies and Techniques*, Data-Centric Systems and Applications (Berlin ; New York: Springer, 2006), 177.

REFERENCES

1. **Abraham, Ajith; Grosan, Crina; Ramos, Vitorino**, Swarm intelligence in data mining, Studies in computational intelligence v. 34, Springer, Berlin ; New York, 2006.
2. **Batini, Carlo; Scannapieca, Monica**, Data quality: concepts, methodologies and techniques, Data-centric systems and applications, Springer, Berlin ; New York, 2006.
3. **Clarke, Roger**, „Quality assurance for security applications of Big Data", în Intelligence and Security Informatics Conference (EISIC), 2016 European, IEEE, 2016, pp. 1–8.
4. **Craig, Terence; Ludloff, Mary E**., Privacy and big data, O'Reilly Media, Inc, Beijing, 2011.
5. **Dasu, Tamraparni; Johnson, Theodore,** Exploratory data mining and data cleaning, Wiley series in probability and statistics, Wiley-Interscience, New York, 2003.
6. **Loshin, David**, Master data management, Elsevier/Morgan Kaufmann, Amsterdam ; Boston, 2009.
7. **Mcnutt, Bruce**, The fractal structure of data reference: applications to the memory hierarchy, The Kluwer international series on advances in database systems 22, Kluwer Academic, Boston, MA, 2000.
8. **Olson, Jack E.,** Data quality: the accuracy dimension, Morgan Kaufmann, San Francisco, 2003.
9. **Sadiq, Shazia**, Handbook of Data Quality, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, http://link.springer.com/10.1007/978-3-642-36257-6, data accesării 14 iunie 2018.
10. **Stackowiak, Robert; Rayman, Joseph; Greenwald, Rick**, Oracle data warehousing and business intelligence solutions, Wiley Pub, Hoboken, N.J, 2007.
11. **Tejay, Gurvirender; Dhillon, Gurpreet; Chin, Amita Goyal**, „Data quality dimensions for information systems security: A theoretical exposition", în Security management, integrity, and internal control in Information Systems, Springer, 2005, pp. 21–39.