Data Collect Requirements Model

I. Tikito [1]
[1] Mohammed V University - Rabat
EMI-SIWEB Team Rabat
Morocco
tikito.iman@gmail.com

N. Souissi [1,2]
[2] Ecole Nationale Supérieure des Mines de Rabat
Computer Sciences Department Rabat
Morocco
souissi@enim.ac.ma

ABSTRACT

In Big data era, managing data requires sufficient tools, last computer science evolution and developed methodologies. To be able to satisfy customer and the big need of information, multiple methods are developed to handle the complexity as well as the huge amount of data in different phases of data lifecycle. We notice for each complicated situation in data lifecycle we focus more particularly to develop storage or Analysis processes. For this reason in this paper, we try to have a different approach to resolve basic issues on targeting the first phase of data lifecycle, which is data collect. We present it as a System of systems, since the complexity of each phase of data lifecycle. In this research, we are interested by the collect system and particularly the process of Creation/Reception of data for which we model the requirements in order to manage smart data at the first level of the cycle. To build this model, we follow a methodology that required three major steps. Starting with requirement identification to defining criterion for each requirement, and in the last step will provide requirement modeling. This research highlight the importance of managing data collect to identify and restrict the issues of big data era.

CCS CONCEPTS

• **Information systems** → Data management systems, Information systems applications.

KEYWORDS

Big data; Data lifecycle; System of Systems; BPMN; Seven views; Requirement Model; Collect system; Creation/Reception Process.

BDCA'17, March 29-30, 2017, Tetouan, Morocco

ACM ISBN 978-1-4503-4852-2/17/03...$15.00
http://dx.doi.org/10.1145/3090354.3090358

1 INTRODUCTION

In the past, the one who held the information had the power, but now with the technological evolution, having the information is only the first step towards glory. Due to computer science growth we have more data coming using several tools and ways. Social media is the famous one to provide various data as well as mobile apps or cookies. Indeed, organizations are producing and storing vast amount of data to manage and understand clients' needs, in [15] this paradigm is called as Big Data. Referring to [16] Amir Gandomi and Murtaza Haiderhe collect a definition of big data based on an online survey of 154 global executives in April 2012 finding 28% define big data as Massive growth of transaction data, including data from customers and the supply chain. 24% presume new technologies designed to address the volume, variety, and velocity challenges of Big Data. 19% see the Big data as a requirement to store and archive data for regulatory and compliance. 18% describe it as explosion of new data sources (social media, mobile device, and machine-generated devices) and 11% give some other definition.

In a world where competition is ruleless and time is money, building knowledge from the collected data is a tough task needs a good management and a processes to answer client requirements. Traditional systems, and the data management techniques associated with them, have failed to scale to Big Data [17]. Otherwise, multiple researches try to find ways to optimize Big data as [18]. In a Big data context, the data lifecycle cannot be seen anymore as elementary system but as a System of Systems (SoS). The classical phases that can be found in elementary lifecycle become increasingly complex, and single phase now is seen as a complete system.

To identify the main phases of data lifecycle in big data context, we studied several researches like [40, 41, 42, 43, 44] which was taken into account in our data lifecycle model.

In this paper we focus specifically on Creation/Reception process present under Collect system in data lifecycle. In order to make the process smart a good management is mandatory, thus many actions can be much easier if they have been managed and optimized in advance [37, 38, 39]. Accordingly we defined some requirements using collect strategy and the 7Vs present in Big Data context. Afterwards we will describe depending on definitions given a set of criteria for each requirement, and we will model them using seven views approach.

This paper is organized as follow. We begin the paper by defining Data lifecycle as SoS. We highlight the fact that Data lifecycle is not an elementary system due to Big data as well as SoS definition. We then give the methodology followed in the paper. In order to meet our objective it's important to take into account a number of requirements established by the customer as well as experts in the trade. So we choose to schematize our processes according to the BPMN approach to model the creation/reception process. Hence, when we discuss requirement identification, we will focus on V's given in Big data definitions, to be the key for our conditions. We then expand the discussion by providing the criteria for each requirement with some examples. Given that we model the requirements using seven views approach. We conclude by highlighting the expected developments to realize in the near future to manage and give more importance to the collect data.

2 DATA LIFECYCLE AS SOS

Big data can transform the entire business process as mention in [12] displaying the big impact for the world economy. Finding data nowadays is simple, subsequently extract the right information on time become more complicated due to the Big data era. Multiple research as [13, 14, 15] evokes definitions of Big data to surround the context for a better understanding and be able to give answers and solutions to manage these amount of data that become wild.

In this context we consider the data lifecycle as a complex system and we present it as a SoS.

2.1 System of systems definition

The "system of systems" term has been in use for several years now, even if there is no universally accepted definition. Keating and his co-authors define in [1] their view of systems of systems based on several definitions present in literature; "The design, deployment, operation, and transformation of metasystems that must function as an integrated complex system to produce desirable results. These metasystems are themselves comprised of multiple autonomous embedded complex systems that can be diverse in technology, context, operation, geography, and conceptual frame."

Several researchers have also developed their own definitions for a system of systems like Popper 2004 define SoS as "a collection of task-oriented or dedicated systems that pool their resources and capabilities together to obtain a new, more complex 'meta-system' which offers more functionality and performance than simply the sum of the constituent systems". However, multiple researches [3, 4, 5] refer to Maier definition as a group of components which individually can be viewed as systems, and which had five properties [6]:

Evolutionary development: This means that SoS are not developed in a single project but progress over time from their constituent systems.

Emergence: SoS has emergent characteristics that only become apparent after the SoS has been created.

Geographical distribution of elements: The elements of a SoS are often geographically distributed across different organizations.

Operational Independence of the Components: If we divide system-of-systems into its component, each component systems can usefully operate independently. That is, the components fulfill customer-operator purposes on their own.

Managerial Independence of the Components: Each element is managed, at least in part, for its own purposes rather than the purposes of the collective.

The last two characteristics are specified as the major ones to make the difference between SOS and system in many researches. Two further characteristics are added to Maier's list in [7]:

Data intensive: A software SoS typically relies on and manages a very large volume of data.

Heterogeneity: The different systems in a software SoS are unlikely to have been developed using the same programming languages and design methods.

2.2 Mapping Cycle/SoS

We consider a data lifecycle as a system of systems. Indeed, Data intensive is present in all the cycle due to Big Data era. Different devices in diverse location interact to gather data using several systems to analyze and manage them provided by different resources in different area. Although, we can notice in Storage phase the use of different devices cloud, flash memory, Blu-Ray and others explain heterogeneity concept. Furthermore, each component in data lifecycle is managed by different parties, due to the separate purpose and the geographical situation for each part. Relying on the definitions given in the literature, also the comparison between SoS and regular system present in [8, 9] based on: Stakeholder Involvement, Governance, Operational Focus, Acquisition, Test and Evaluation, Boundaries and Interfaces, Performance and Behavior we can present a data lifecycle as SoS.

In this further paper we present a mapping Cycle/SoS in table 1, to discard any misunderstanding of terminologies. We consider a cycle as a set of phases, having a task for each phase. Respectively a system of systems is a set of systems, composed on multiple processes. Which is a structured and homogeneous set of activities that respond to a business objective.

Table 1 : Mapping Cycle Vs SoS

Cycle	System of Systems
Set of phases	System
Phase	Process
Task	Activity

2.3 Data lifecycle model

The presentation of data lifecycle as SoS display five major systems, we present in Fig 1 an overall view of data lifecycle.

One-way or two-way arrows indicate that the data moves between different stages. "Management system" is ubiquitous due to the importance to have a goal and rules before starting any process for optimizing useless processes and gain time, we can also modify these data from "Storage system" in some specific cases that must be already defined and approved like modifying or saving some existing or new methodologies that will be estimated valuable. "Management system" needs to be handled by business analyst as well as some developed process to offer the best solution for each client requirement. This step is the key of all the process because it gives the strategy to follow. In all phases in the cycle, data will go through many process leading to the omnipresence of "Storage system" to keep tracking those changes.

"Analysis system" is the processes allowing to change a data to an information that can be used by the client. Multiple researches as well as tools are present to improve this system.

Once Analysis system done, "Processing system" define the making the information and the results can be used and shared.

"Collect system" present all the processes before Analysis systems. This system will be detailed further in this paper.

We consider "Storage system" as a data center where data will be saved. Also we are focusing only on "Collect system", so we are not giving further detailed for the other systems.

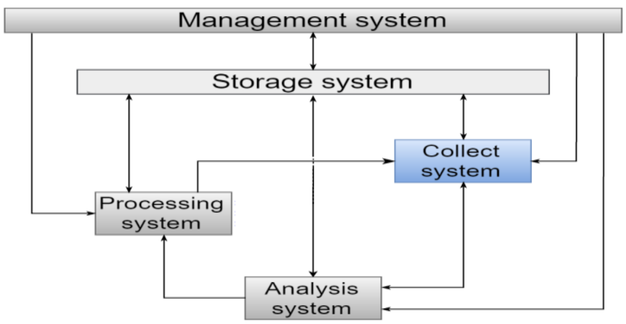

Figure 1 : Data lifecycle

3 METHODOLOGY

To model the requirements related to data collect, we will follow three steps present in Fig. 2.

1st step - Requirements Identification: to identify the requirements we need two inputs. The first one is from the management system providing client requirement. The second one related to big data context.

2nd step - Defining criterion per requirement: consist to define the criteria for each requirement provided by the output of the first step, which are the 7V's of Big data.

3rd step - Requirements Modeling: to model requirements and provide our model, we should have the output present in second step providing the set of criteria for each requirement.

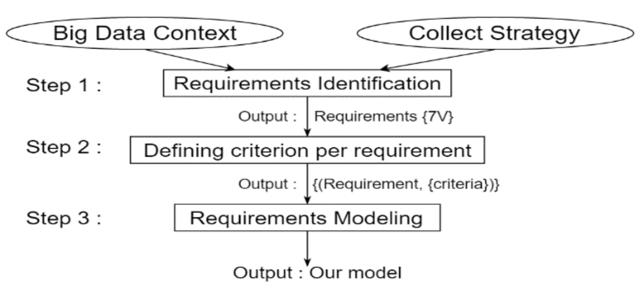

Figure 2 : Building methodology of requirements model

4 IDENTIFY REQUIREMENTS

In this section we are identifying the collect data requirements based on 7Vs provided in Big data context, consequently an analysis of the collect system is required. Thus, a schema representing the interaction between Collect system and the other systems of cycle are present to give a general view.

4.1 Collect system

For a better understanding how the Collect system works, we will model the processes of this system from the point of view of the user. Even if SysML is a modeling language specific to system engineering and offers several improvements over UML, which is more focused on the software. SysML does not allow to model the interaction between SoS components in a simple way to be understandable by all the users of the company and not only by developers. The nine diagrams that are covered in SysML are like mention in [10, 11]:

- New diagrams compared to UML: Requirement diagram and parametric diagram.
- Updated diagrams compared to UML: Activity diagram, Block definition diagram and Internal block diagram.
- Unchanged diagrams compared to UML: Sequence diagram, State machine diagram, Use case diagram and Package diagram.

Accordingly we will model our processes using BPMN (Business Process Model and Notation) having as primary goal to provide an easy way to understand notation for all involved stakeholders, from business users to technical developers. The graphical notation will simplify the understanding of the performance of the process.

We are schematizing 7 blocks in fig 3 and we describe below the flows numbered from (1) to (17).

Supplier: describes the ways to provide data from different resources. This system is important because it's used as the input of two processes "Creation/Reception" and "Enrichment" present respectively in (2) and (9).

Before data collect begins, a strategy that reflects the client's request need to be defined in Customer bloc (1). Even if in some cases clients can provide their own strategy, but it will not be accepted until we check it in "Strategy Assess" process.

Strategy assess prevent duplicated strategies as well as update and complete client proposal strategy for a better result using business analyst as well as template strategy.

Template strategy archive the strategies for different business areas that are well estimated in their fields.

Once the collect strategy is defined and approved by the customer, the management process allows the collect of return requirements in terms of data. In other words, the Creation/Reception process must meet the requirements defined by the client to collect raw data from different sources and in some cases to create data following a predefined request. The output of this process (3) will be the first level of smart data (L0).

Nowadays, many ways are possible to provide data like, archives, surveys or articles exposed in paper format only (BD paper), or stored in databases (BD). Moreover we can use the exchange of information and data (IoT) from devices connected to the Internet to receive new data. Furthermore, data can be created as a result of researches and observations made by qualified persons (Observer). Due to the diversity of methods, sources and heterogeneity of data, the phase “Integration” allows to structure these data (4) for a better use. The output of this process will be (5) the second level of smart data (L1).

Once data are integrated we can filter the data (6) to proceed with the analyze (13) or directly to analyze system (14) if we do have adequate data. Filtering according to criteria provided by the client reduce the amount of data and optimize the analysis process and provide (7) smart data (L1).

The Enrichment process can be accessed in three situations. The first is the most common after the archiving of data, the data will become after a delay and in some cases obsolete so an update will be necessary, the timing to trigger the action should be present in the strategy (10). The second (17) after Processing system, the customer can required some extra data. The last case after the filtering phase (8), resulting shows that the data does not meet the requirements. The business referent can solve this situation under certain conditions and in this case we go back to the filtering process (12). However, the business referent has the knowledge required in a specific field enabling it to enrich in some cases the data already collected. But if there is a lack of information that cannot be resolved by the business referrer, we will need to go back to the Creation/Reception process (15). The output of this process will be (11) the third level of smart data (L2).

After Analyzing System data should answer the main functionality to gathering data like Visualization or Sharing present in Processing System, or we restart the Collect (16).

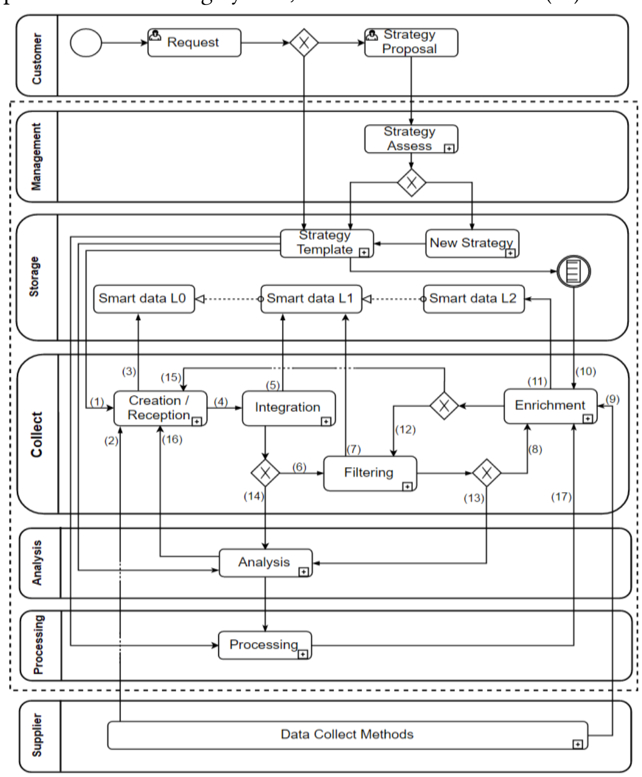

Figure 3 : Interactions between systems and collect system

4.2 Requirements description

Since the appearance of the term "Big Data" in 1997 by Michael Cox and David Ellsworth [19], many researchers start to be interested in data-related topics. The definition of "Big Data" in 3V appeared in 2001 by Doug Laney in [20], although the term itself was not quoted in his research. It explains the value of increasing the definition of data across the three dimensions: **Volume**, **Velocity** and **Variety**. This vision is known today as 3V.

The number of "V" has increased over the years from three to five to keep increasing nowadays. We find imperatively 5V: Volume, Velocity, Variety, Veracity and Value, but also other «V» have appeared as **Validity**, **Volatility**, **Variability** or **Visualization** [21, 22].

In a world where competition is increased, the one who holds the information has the power. As a result, very high growth in the volume of data collected to meet the need has become a fact [23, 24, 25]. Therefore, it’s necessary to collect, read, process and manage these data quickly. The velocity of analyzing and extracting information remains a great challenge [26, 27]. The data must be stored on arrival, otherwise there will be a risk to lose the information or a potential competitor can use them first.

In order to meet the demand and the need in terms of data, several tools are used for the collect, generating a Variety of data. Inconsistent data formats, non-aligned data structures, and inconsistent data semantics represent significant challenges [27, 28].

The Veracity is defined in [29] as the accuracy of data in relation to reality or, in other words, its ability to be without "lies". Thus, extracting the value of a data taking into account the need to enrich the raw data and untreated in order to achieve higher level of knowledge to use through different scenarios is the objective of the data [30].

In [21], they define the Validity as the accuracy and precision of the data regarding the intended use. However, Volatility is evoked to respect the policy of retention of the structured data and that once the period of conservation expires, the data must be destroyed.

The Variability is defined in [31] by the ability to interpret data for a client, given the changes in the data structure.

The notion of Visualization appeared in 2012 in [32] highlighting the importance and impact of visualizing the data in the big data era.

The number of "V" to define the term "Big data" reveal the importance and the difficulty of storing, managing and analyzing the great flow of data which from day to day and even from second to second increase. Various solutions have emerged to remedy this dilemma using the technological revolution. Hardware, software, but also very powerful storage solutions surfaced during these years [33, 34, 35]. However, these efforts remain limited and insufficient to resolve the issue. Hence the importance, we propose to put the V’s of big data as requirements managing data collect.

We consider that **Big Data's 7V (Volume, Velocity, Variety, Veracity, Value, Variability and Visualization)** are good requirements in order to measure the effectiveness of our approach. We do not retain Validity because it is a criterion managed in the planning and analysis phases in spite of the Big Data context. Similarly for Volatility, this is a concept that must be managed during the planning phase for each customer and for each stored data.

5 CRITERIA DEFINITION

In this section we will present the set of criteria for each requirement. Table 2 provides the information with a short description for each criteria for a better understanding.

Table 2 : Requirement per Criteria

Requirement	Criterion	Description / Example
Volume	Capacity	Bandwidth capacity
	Storage	Devices, Cloud, …
	Transactions	Time dimension to describe the data
Velocity	Speed	The rapidity to gather data
	Real time	The data should be collected once it's needed and available.
	Frequency	The occurrence to gather data
	Data flow	The number of data able to move from one point to another
Variety	Structured	Formalized data using a clear data model
	Unstructured	Data without a predefined data model nor predefined organized
	Different formats	Html, jpeg, PDF, video, …
	Different sources	Connected objects, databases, Social network, Cookies, GPS, Paper, …
Veracity	Trust	Belief in the reliability
	Authenticity	Data received is original and received without variation
	Root	The source of data
	Accuracy	The results of collecting data should be exact
	Integrity	The data records are real and were not faked or modified
Value	Statistics	Characteristic or measure obtained from a snip of data
	Events	Data actions
	Dependencies	Define the dependencies between data
	Issues	The value shouldn't generate any problem and be exploitable
	Interpretability	The value of information should provide a meaning
Variability	Customer requirements	Customers request
	Appropriate amount	Suitable quantity to have the required data.
	Quality	Good quality to provide a high level of information
	Interpretable	To provide a comprehensible information
Visualization	Security	Protect sensitive information and
	Understandable	Data should be comprehensible.
	Availability	Data should be disposable
	Access	Information should be reachable
	Facility	It should be easy to reach it

6 REQUIREMENTS MODEL

In this section we are modeling using the seven views approach, the result of our research presented by set of criteria for each requirement. We explain as well the choice of seven views.

6.1 Modeling approaches

Requirement view is present only on seven views as well as SysML, but for this last one as present in [10] the notion of requirement is defined informally and in textual form. Also, Relationships between requirements do not have precise semantics, causing confusion. For these reasons we are modeling requirements using seven views.

The seven views approach is a meta-model technique for modeling business processes providing basis for analysis and discussion, as mention in [36].

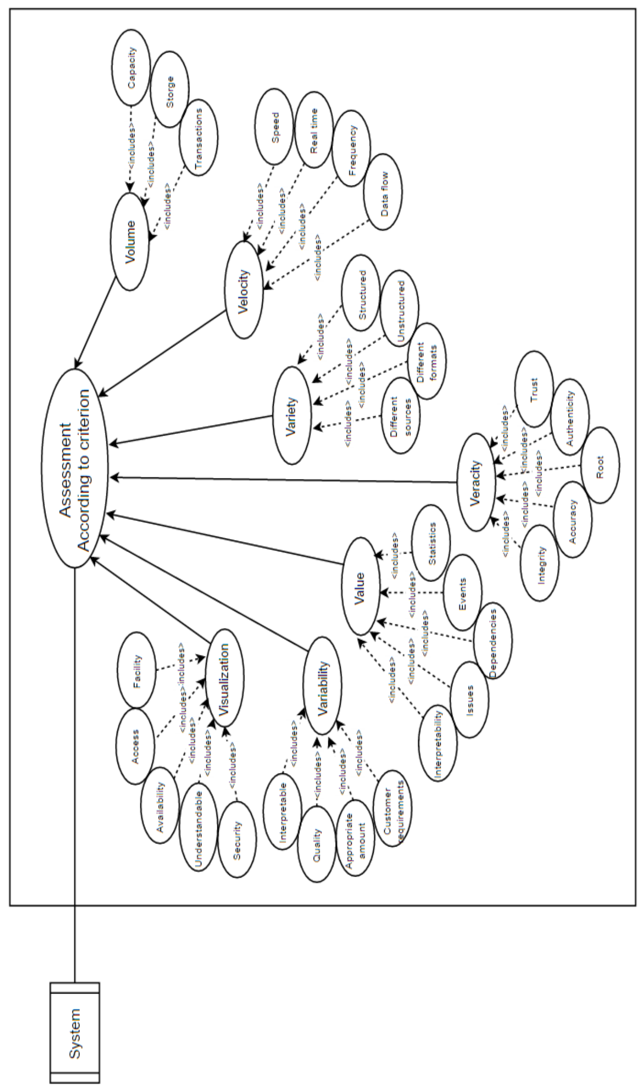

Figure 4 : Requirements Model of data Collect

6.2 Model

We present in Fig 4, the requirements of creation/reception process using seven views approach. For each requirement we find a set of criteria present in Fig 4.

The actor in this model is the system that evaluate the data based on 7Vs requirements. Moreover, the criteria present are the definition key for each requirement, that are present using the dashed arrow with an open arrowhead and labeled with <include> keyword. Those criteria are not all mandatory but it depends on client request, nevertheless some elementary ones will be obligatory like for example authenticity, quality and security. The gain of this model is to provide the first template for the user to decide regarding the basic criteria to follow in all data lifecycle.

7 CONCLUSIONS

In this paper we put under spotlight the first process of data lifecycle "Creation/Reception". Since the amount of data and the ways to provide them increased, we tried to optimize and simplify the followed system for a better use of data.

A set of criteria based on requirements are proposed using big data 7Vs context. Volume, Velocity, Variety, Veracity, Value, Variability and Visualization are the maintained requirements. Using this approach a different level of Smart data will be generated in different steps to optimize the subsequent systems of data lifecycle.

This research is the first step for a further work to have a smart data lifecycle using a smart collect data system. In view of this the methods used nowadays will be more powerful and efficient.

REFERENCES

[1] KEATING, Charles, ROGERS, Ralph, UNAL, Resit, et al. System of systems engineering. Engineering Management Journal, 2003, vol. 15, no 3, p. 36-45.

[2] POPPER, Steven W., BANKES, Steven C., CALLAWAY, Robert, et al. System of systems symposium: Report on a summer conversation. In : Proceedings of the 1st System of Systems Symposium. 2004.

[3] SITTON, Miri et REICH, Yoram. Enterprise Systems Engineering for Better Operational Interoperability. Systems Engineering, 2015, vol. 18, no 6, p. 625-638.

[4] IEEE-Reliability Society. Technical Committee on 'Systems of Systems'-WHITE PAPER. Oct. 15, 2014. p 1-5.

[5] SAGE, Andrew P. et CUPPAN, Christopher D. On the systems engineering and management of systems of systems and federations of systems. Information knowledge systems management, 2001, vol. 2, no 4, p. 325-345.

[6] MAIER, Mark W. Architecting principles for systems-of-systems. In : INCOSE International Symposium. 1996. p. 565-573.

[7] Ian Sommerville. Software engineering book - Software Engineering 10th Edition, 2015.

[8] Office of the Director, Defense Research and Engineering Systems Engineering Guide for Systems of Systems Essentials, 2010.

[9] DAHMANN, Judith S., REBOVICH JR, George, et LANE, Jo Ann. Systems engineering for capabilities. MITRE CORP MCLEAN VA, 2008.

[10] GNAHO, Christophe et SEMMAK, Farida. Une extension SysML pour l'ingénierie des exigences dirigée par les buts. In : INFORSID. 2010. p. 277-292.

[11] FRIEDENTHAL, Sanford, MOORE, Alan, et STEINER, Rick. A practical guide to SysML: the systems modeling language. Morgan Kaufmann, 2014.

[12] WAMBA, Samuel Fosso, AKTER, Shahriar, EDWARDS, Andrew, et al. How 'big data'can make big impact: Findings from a systematic review and a longitudinal case study. International Journal of Production Economics, 2015, vol. 165, p. 234-246.

[13] LETOUZÉ, Emmanuel et JÜTTING, Johannes. Official statistics, big data and human development. Tech. rep., Data-Pop Alliance (Harvard Humanitarian Initiative, MIT Media Lab and Overseas Development Institute) and Paris2, 2015.

[14] RIGGINS, Frederick J. et WAMBA, Samuel Fosso. Research directions on the adoption, usage, and impact of the internet of things through the use of big data analytics. In : System Sciences (HICSS), 2015 48th Hawaii International Conference on. IEEE, 2015. p. 1531-1540.

[15] ASSUNÇÃO, Marcos D., CALHEIROS, Rodrigo N., BIANCHI, Silvia, et al. Big Data computing and clouds: Trends and future directions. Journal of Parallel and Distributed Computing, 2015, vol. 79, p. 3-15.

[16] GANDOMI, Amir et HAIDER, Murtaza. Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 2015, vol. 35, no 2, p. 137-144.

[17] MARZ, Nathan et WARREN, James. Big Data: Principles and best practices of scalable realtime data systems. Manning Publications Co., 2015.

[18] RICHTÁRIK, Peter et TAKÁČ, Martin. Parallel coordinate descent methods for big data optimization. Mathematical Programming, 2016, vol. 156, no 1-2, p. 433-484.

[19] COX, Michael et ELLSWORTH, David. Application-controlled demand paging for out-of-core visualization. In : Proceedings of the 8th conference on Visualization'97. IEEE Computer Society Press, 1997. p. 235-ff.

[20] Laney D (2001) 3-d data management: controlling data volume, velocity and variety. META Group Research Note, 6 February.

[21] UDDIN, Muhammad Fahim, GUPTA, Navarun, et al. Seven V's of Big Data understanding Big Data to extract value. In : American Society for Engineering Education (ASEE Zone 1), 2014 Zone 1 Conference of the. IEEE, 2014. p. 1-5.

[22] YLIJOKI, Ossi et PORRAS, Jari. Perspectives to definition of big data: a mapping study and discussion. Journal of Innovation Management, 2016, vol. 4, no 1, p. 69-91.

[23] MCAFEE, Andrew, BRYNJOLFSSON, Erik, DAVENPORT, Thomas H., et al. Big data. The management revolution. Harvard Bus Rev, 2012, vol. 90, no 10, p. 61-67.

[24] AL-SALIM, Ali M., ALI, Howraa M. Mohammad, LAWEY, Ahmed Q., et al. Greening big data networks: Volume impact. In : Transparent Optical Networks (ICTON), 2016 18th International Conference on. IEEE, 2016. p. 1-6.

[25] SANGROYA, Amit et SINGHAL, Rekha. Performance Assurance Model for HiveQL on Large Data Volume. In : 2015 IEEE 22nd International Conference on High Performance Computing Workshops. IEEE, 2015. p. 26-33.

[26] MASTELIC, Toni et BRANDIC, Ivona. Data Velocity Scaling via Dynamic Monitoring Frequency on Ultrascale Infrastructures. In : 2015 IEEE 7th International Conference on Cloud Computing Technology and Science (CloudCom). IEEE, 2015. p. 422-425.

[27] KAISLER, Stephen, ARMOUR, Frank, ESPINOSA, J. Alberto, et al. Big data: issues and challenges moving forward. In : System Sciences (HICSS), 2013 46th Hawaii International Conference on. IEEE, 2013. p. 995-1004.

[28] CERQUEUS, Thomas, DE ALMEIDA, Eduardo Cunha, et SCHERZINGER, Stefanie. Safely managing data variety in big data software development. In : Proceedings of the First International Workshop on BIG Data Software Engineering. IEEE Press, 2015. p. 4-10.

[29] STARON, Miroslaw et SCANDARIATO, Riccardo. Data veracity in intelligent transportation systems: The slippery road warning scenario. In : Intelligent Vehicles Symposium (IV), 2016 IEEE. IEEE, 2016. p. 821-826.

[30] DEBATTISTA, Jeremy, LANGE, Christoph, SCERRI, Simon, et al. Linked'Big'Data: Towards a Manifold Increase in Big Data Value and Veracity. In : 2015 IEEE/ACM 2nd International Symposium on Big Data Computing (BDC). IEEE, 2015. p. 92-98.

[31] FAN, Wei et BIFET, Albert. Mining big data: current status, and forecast to the future. ACM sIGKDD Explorations Newsletter, 2013, vol. 14, no 2, p. 1-5.

[32] CHEN, Hsinchun, CHIANG, Roger HL, et STOREY, Veda C. Business Intelligence and Analytics: From Big Data to Big Impact. MIS quarterly, 2012, vol. 36, no 4, p. 1165-1188.

[33] HILBERT, Martin et LÓPEZ, Priscila. The world's technological capacity to store, communicate, and compute information. science, 2011, vol. 332, no 6025, p. 60-65.

[34] GUO, Yu-Guo, HU, Jin-Song, et WAN, Li-Jun. Nanostructured materials for electrochemical energy conversion and storage devices. Advanced Materials, 2008, vol. 20, no 15, p. 2878-2887.

[35] LOU, Mingji, BEKER, Orhan, et FENG, L. U. DATA STORAGE DEVICE GENERATING AN OPERATING SEEK TIME PROFILE AS A FUNCTION OF A BASE SEEK TIME PROFILE. U.S. Patent No 20,160,104,504, 14 avr. 2016.

[36] HOLT, Jon. A pragmatic guide to business process modelling. BCS, The Chartered Institute, 2009.

[37] PROVOST, Foster et FAWCETT, Tom. Data science and its relationship to big data and data-driven decision making. Big Data, 2013, vol. 1, no 1, p. 51-59.

[38] SCHAAF, Roseann C. Creating evidence for practice using data-driven decision making. American Journal of Occupational Therapy, 2015, vol. 69, no 2, p. 6902360010p1-6902360010p6.

[39] ZHANG, Yin, CHEN, Min, MAO, Shiwen, et al. CAP: Community activity prediction based on big data analysis. Ieee Network, 2014, vol. 28, no 4, p. 52-57.

[40] N. Khan, I. Yaqoob, I. A. T. Hashem, Z. Inayat, W. K. Mahmoud Ali, M. Alam, M. Shiraz, and A. Gani, "Big data : survey, technologies, opportunities, and challenges," The Scientific World Journal, vol. 2014, 2014.
[41] J. L. Faundeen, T. E. Burley, J. A. Carlino, D. L. Govoni, H. S. Henkel, S. L. Holl, V. B. Hutchison, E. Martín, E. T. Montgomery, C. Ladino, et al., "The united states geological survey science data lifecycle model," tech. rep., US Geological Survey, 2014.
[42] Y. Demchenko, C. De Laat, and P. Membrey, "Defining architecture components of the big data ecosystem," in Collaboration Technologies and Systems (CTS), 2014 International Conference on, pp. 104–112, IEEE, 2014.
[43] S. Allard, "Dataone : Facilitating escience through collaboration," Journal of eScience Librarianship, vol. 1, no. 1, p. 3, 2012.
[44] L. Lin, T. Liu, J. Hu, and J. Zhang, "A privacy-aware cloud service selection method toward data life-cycle," in Parallel and Distributed Systems (ICPADS), 2014 20th IEEE International Conference on, pp. 752–759, IEEE, 2014.