
DATA GOVERNANCE APPLIED TO INFORMATION SCIENCE: ANALYSIS OF A SCIENTIFIC DATA SYSTEM FOR THE HEALTH AREA

GOVERNANÇA DE DADOS APLICADA À CIÊNCIA DA INFORMAÇÃO:
ANÁLISE DE UM SISTEMA DE DADOS CIENTÍFICOS PARA A ÁREA DA SAÚDE

GOBERNANZA DE DATOS APLICADA A LA CIENCIA DE LA INFORMACIÓN: ANÁLISIS DE
UN SISTEMA DE DATOS CIENTÍFICOS PARA EL ÁREA DE LA SALUD

Priscilla Lüdtke Espíndola¹


José Francisco Salm Junior¹

Francisco Rosa¹

Jordan Paulesky Juliani¹

¹Universidade do Estado de Santa Catarina

Correspondência

¹Priscilla Lüdtke Espíndola 

Universidade do Estado de Santa Catarina

Florianópolis, SC

E-mail: priscillaludtke@gmail.com



JITA: LN. Data base management systems.

Submitted on: 27/11/2017

Accepted on: 01/08/2018

Published on: 23/08/2018

ABSTRACT: Data are valuable inputs to organizations and to scientific studies. In order for scientific data to be used in studies related to Motor Assessment, and to produce reliable results, good collection, storage and retrieval practices are required. This research aims to apply "Data Life Cycle Model" data governance tools to identify opportunities for improvements to the Motor Assessment System, especially related to data quality. Regarding to the methodological aspects, the study is characterized as a applied research with an exploratory feature, data collection performed through research-action, and data analysis through qualitative methods. As results, it was verified the need to redesign the studied system, including mechanism for data treatment to avoid duplicity and guarantee homogeneity and completeness. Likewise, it was found necessary to create and implement a policy to restrict that only able health and education professionals could enter data in the System. It is understood that good data governance practices, "Data Life Cycle Model" principles and other tools adopted in this study contributed to diagnose failures and identify opportunities for improvement in the Motor Assessment System.

KEYWORDS: Data governance. Data Life Cycle. Scientific Data. Health data. Data quality.

RESUMO: Os dados são insumos valiosos para as organizações e para estudos científicos. Para que os dados científicos possam ser empregados em estudos relacionados à Avaliação Motora, e produzirem resultados confiáveis, é necessária a adoção de boas práticas de coleta, armazenamento e recuperação. Esta pesquisa tem como objetivo aplicar o "Modelo de ciclo de vida dos dados" e as ferramentas de governança de dados para identificar oportunidades de melhorias para o Sistema de Avaliação Motora, especialmente com relação a qualidade de dados. Quanto aos aspectos metodológicos, o estudo é caracterizado como uma pesquisa aplicada com caráter exploratório, coleta de dados realizada por meio de pesquisa-ação, e análise dos dados por meio de métodos qualitativos. Como resultados, constatou-se a necessidade de redesenhar o sistema estudado, incluindo mecanismo para tratamento dos dados para evitar duplicidade e garantir homogeneidade e completude. Igualmente, verificou-se ser necessário criar e implementar uma política para restringir que somente profissionais da saúde e da educação aptos possam inserir dados no Sistema. Compreende-se que as boas práticas da governança de dados, princípios do "Modelo de ciclo de vida dos dados" e demais ferramentas adotadas neste estudo contribuíram para diagnosticar falhas e identificar oportunidades de melhoria no Sistema de Avaliação Motora.

PALAVRAS-CHAVE: Governança de dados. Ciclo de Vida dos Dados. Dados Científicos. Dados em saúde. Qualidade de dados.

RESUMEN: Los datos son aportaciones valiosas para las organizaciones y para los estudios científicos. Para que los datos científicos se utilicen en estudios relacionados con la evaluación motora y para producir resultados confiables, se requieren buenas prácticas de recolección, almacenamiento y recuperación. Esta investigación tiene como objetivo aplicar herramientas de gobernanza de datos del "Modelo de ciclo de vida de los datos" para identificar oportunidades de mejoras en el Sistema de Evaluación Motor, especialmente en relación con la calidad de los datos. En cuanto a los aspectos metodológicos, el estudio se caracteriza como una investigación aplicada con una característica exploratoria, la recolección de datos realizada a través de la investigación-acción, y el análisis de datos a través de métodos cualitativos. Como resultado, se verificó la necesidad de rediseñar el sistema estudiado, incluido el mecanismo de tratamiento de datos para evitar la duplicidad y garantizar la homogeneidad y la integridad. Del mismo modo, se consideró necesario crear e implementar una política para restringir que solo los profesionales de salud y educación capaces pudieran ingresar datos en el Sistema. Se entiende que las buenas prácticas de gobernanza de datos, los principios del "Modelo de ciclo de vida de los datos" y otras herramientas adoptadas en este estudio contribuyeron a diagnosticar fallas e identificar oportunidades de mejora en el Sistema de Evaluación Motor.

PALABRAS CLAVE: Gobernanza de datos. Ciclo de vida de los datos. Datos científicos. Datos de salud. Calidad de datos.

1 INTRODUCTION

Nowadays, we are surrounded by data and information in a variety of everyday activities. The data and information have a fundamental role for living in Society, because is through them that we can maintain our decisions, know better the world around us, as well as acquiring knowledge.

In the corporate context, the way data is managed can impact the agility, productivity, costs and required time to meet both internal and external customer requests. In addition, data holds an important role in accountability in organizations with their stakeholders and with society in general.

Data governance is understood by Santos (2010) as a multidisciplinary action whose goal is to treat data as active and manageable inputs in organizations. For this, data governance determines policies, standards, processes, roles, responsibilities and technologies to better follow and monitor the data generated, stored, used and eliminated in the organization.

Still following the author's line of thought, the lack in data's quality is an inhibitor of success in organizations. Quality of data control in the business environment is one of the processes that belongs to the data government. Through data governance, metrics, procedures and requirements are defined to help the organization achieve the data quality necessary to meet its demands and reach its objectives (BARATA, 2015).

It is understood that qualified data are also essential for the success of scientific research, since the scientific data are in need the adoption of good practices of collection, storage and retrieval.

In Corrêa's view (2016) scientific data have a prominent role in the advance of scientific production, and for this reason, it is now demanding new actions to guarantee its preservation and recovery.

In this context, it is believed that the concepts and good practices determined by data governance in organizations can also be applied to information systems that perform the collection, storage and retrieval of scientific data.

Considering the importance of the data's life cycle process in scientific research, mainly in Information Science (IC), Santana (2013) performed an analysis of Life Cycle Data Models developed by several institutions and authors. These models presented the identification of four phases that represent the behavior and flow of information within the context of IC: collecting, storage, retrieval and disposal, as well as six objectives that work around those phases: privacy, quality, copyright, integration, dissemination and preservation.

The combination of these four phases, with the six objectives, was named by the author as Data's Life Cycle Model for Information Science (DLC-IC), which aims to be a data life cycle model that considers the characteristics and specifications of IC.

Accordingly, with Santana and Bonini (2014), through studies and research on the data life cycle, IC can broaden the democratization of data use, contribute to scientific dissemination, enable access to data and, consequently, knowledge generated through the application of this data.

Thus, given the importance of qualified data in scientific research and the need to adopt a life cycle process with defined steps, which may provide more security for the researcher in data management, we have the following research question: **how can DLC-IC and diagnostic data governance tools be used in the development of a proposal for improvements to the Motor Assessment System (MAS)?**

The MAS is an online system for inclusion of motor and child motor skills data, which is used by Health and Education Professionals (HEP) to assist in the application of the motor assessment method based on the Motor Development Scale (MDS). Once using the MAS, the HEP can record the motor evaluation data of each patient, consult the history of the performed evaluations, and be aware of the intervention's evolution performed in their patients (SYSTEM..., [201-?]). It is therefore a system with an open database, focused on the development of scientific research, definition of public policies and treatment of patients, for the academic, political and HEP publics respectively.

At this point, this study has as a general objective: to apply the DLC-IC and the data governance tools to identify opportunities for improvements to MAS, especially regarding to data quality. And as specific objectives: a) to perform diagnosis and identification of opportunities for improvements in the data flow of the MAS, through the Situation Analysis Canvas; b) propose adaptations and improvements to the data flow of the MAS, from the use of the Process Model Canvas tool and the alignment of the System characteristics to the DLC-IC model, c) map responsibilities about the quality of the data inserted in the MAS, from the proposed improvements, through the RACI Matrix².

It is noteworthy that the choice of DLC-IC model has been used as the basis for this research is justified by:

a) by the fact that it was constructed from the study and analysis of a series of models: the model proposed by the Data Documentarian Initiative (DDI), the model proposed by the United Kingdom Data Archive (UKDA) and the Economic and Social Data Service (ESDS),

² Acronym for Responsible, Accountable, Consulted and Informed (WENDE, OTTO, 2007).

the model proposed by the Data Management Association (DAMA) of data curators proposed by the Digital Curation Center (DCC);

b) being conceived to meet the demands of IC; and

c) by the fact that, in addition to the phases of the cycle, the objective of data quality, one of the central focuses of data governance and of this research.

Through this research, our goal is to contribute to the development of the IC area, above all about DLC-IC and the governance applied to scientific data, sharing the motivations of its execution, the methodology used, and the results achieved. In addition, it aims to influence other studies about the use of data governance tools in the diagnosis, analysis and proposal of improvement of real problems.

2 DATA'S LIFE CICLE

Davenport (1998, p. 19, our translation) defines data as "observations on the state of the world". They are the primary elements (raw material of information), untreated and alone have no meaning, such as a number, a word, or an image. It will be an amount of data, endowed with context, that will generate information. But it must be clear that when there is a lack of quality, it will be a loss for those who use the information, especially who depends on the information for decision making.

As an example, we can see an electoral survey, in which each participant provides the researcher with an opinion (or a set of opinions) about the election. The individualized data does not provide a significant outcome, but through the amount of several interviewees we were able to extract information about the possible result of the election.

Data is an important input to generate information and knowledge (BARBIERI, 2011), both in organizations and in scientific studies. Therefore, it is understood that it should be a constant concern of system managers to make data access available for professionals and researchers interested in these, to contribute with the dissemination of information and knowledge on a large scale.

To guarantee data access, Santana (2013) states that it is necessary to study and bring improvements to all phases related to data access through Information and Communication Technologies, from the planning of data collection and / or creation, until its visualization by the interested ones.

Yet following the author's line of thought, IC is an area that can contribute to the construction of new theoretical frameworks related to the process of access data to meet the different informational needs. Since the IC's objective is to study the production, storage, dissemination and reception of information (context data), especially with a focus on the technologies used during this process (CAPURRO; HJORLAND, 2007).

Through an analysis of several data life cycle models developed in studies of correlated areas, Santana (2013) developed a data life model proposal that considers the characteristics and specificities of IC. This model was named by the author of DLC-IC and can be visualized in Figure 1.

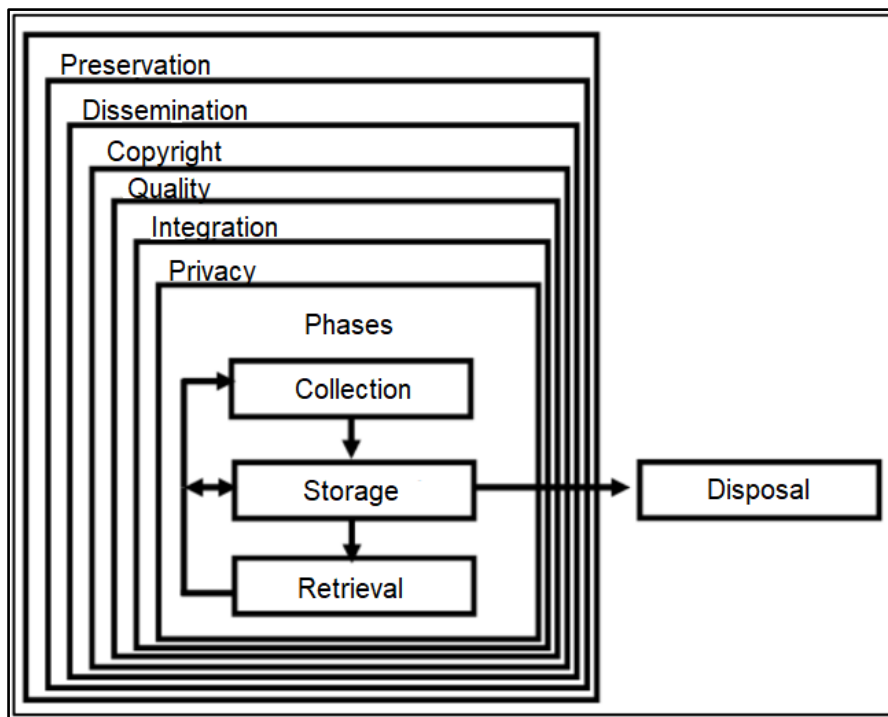


Figure 1. DLC-IC's Representative Diagram
Source: Santana (2013, p. 17, our translation)

The DLC-IC model consists of a four-step process: data collection, storage, retrieval, and disposal. These steps are related by six objectives: privacy, integration, quality, copyright, dissemination and preservation.

It should be noted that the phases of the mentioned Model served as a basis for the identification and characterization of the processes present in the MAS data flow, as will be presented in the methodology of this article.

According to the DLC-IC Model (SANTANA, 2013), in the collection phase, the initial planning of the data occurs, as well as its description through metadata, its evaluation and its selection. In the storage phase we have the activities of processing, transformation, insertion, modification, transmission, and other actions related to the digital data preservation. In the recovery, there is effective access to the data by professionals and researchers, as well as the activities of consultation and data visualization, along with the actions of structuring, filtering, treatment, representation, refinement and interactivity, which

are performed after obtaining the data. Disposal occurs when previously planned goals for the data are reached, or when the expected storage time limits are reached.

The author explains that "[...] the collection phase allows the storage phase to be initiated, which eases the execution of the recovery phase and can still generate new data by taking actions from the collection phase" (SANTANA, 2013, p. 17, our translation). In this way, the execution of the recovery phase can generate data that feed the cycle, which can resume actions of the collection phase or the storage phase.

3 DATA GOVERNANCE

Data governance can be defined as "the exercise of decision-making and authority for data-related matters." (DGI, c2017a). So, data governance determines policies, agreements, roles, and responsibilities with regards to the data generated in the organization, as well as defines which methods should be used in the creation, storage, evaluation, use and deletion of the data.

As Loftis (2014) data governance is a framework that guides and establishes strategies, policies and objectives in order to manage the data, like any other resource of an organization. So, for data governance, data are important resources for organizations, which must be managed, monitored and tracked as any other input that assists in fulfilling the organizational mission and the achievement of its objectives.

Also, according to the author, to implement data governance, it needs to be known how the organizational culture and the way decisions are made in the organization, so that the manager can extract advantages from the way your company communicates.

Among the objectives of data governance, the DGI (c2017b) highlights seven: improve decision making; reduce operational conflicts; protect the needs of those interested in the data generated; guide employees to adopt common solutions to the same data questions; build standards for processes; reduce costs and increase efficiency by coordinating efforts and ensuring process transparency.

Data governance also has an important role monitoring data management in the organization, to ensure that the data generated in the processes are aligned with the organizational objectives. In addition, it monitors and track the data so that it can be used effectively and efficiently in meeting the organizational goals set by the high management.

To achieve these objectives, data governance involves a set of processes, policies, standards, organization and technologies necessary to manipulate and ensure the availability, accessibility, quality, consistency, auditability and data security (SANTOS, 2010).

This way, by adopting good practices of data governance, it is possible to plan and implement guidelines, as well as internal standardizations that enable data manipulation.

The good practices of data governance ensure that the data generated in the institutional processes are available to those who need them and can be accessed quickly. Data governance also aims to guarantee that the data generated is secure, consistent and qualified, and can be audited for different purposes.

For Carvalho (2012) the practice of data governance, besides allowing the control of processes and methods used in data manipulation, allows to prevent adverse situations that may compromise the quality of the data generated in the organization. In accord to the author, through data governance it is possible to increase the security, confidentiality and quality of the data, as well as the speed and efficiency in its treatment.

Given the above, about data governance, it is understood that its principles and good practices can also be applied in processes that involve scientific data, and therefore, are not directly related to corporate data and to the achievement of organizational objectives.

Governance applied to scientific data is related to planning guidelines, create policies, determine roles and responsibilities, develop strategies and standardized processes aiming the data to have quality and are available for those who need them.

By applying the principles and good data governance practices in the MAS, the goal is that the data entered, stored and made available by said information system have more quality and more effectively meet the needs of HEP who used it.

Data quality can be considered one of the main focuses of data governance (BARBIERI, 2011). Given the importance of providing reliable and qualified data to the stakeholders of an organization, as well as to professionals and researchers using scientific data, the data quality theme will be addressed.

3.1 Data Quality

To assist in understanding what data quality is, we first resort to Davok and Garcia's (2014) view of quality. According to the authors, something, such as a service, has quality when it exhibits value and merit. Value is related to how much service is needed for the organization's stakeholders, so the more stakeholders need the service, the more value it has. The merit is associated with the use of resources efficiently and efficiently to provide the service in a way that meets the established quality standards.

So, in the context of data quality in organizations, it is understood that data has quality when it meets the needs of the institution's stakeholders from where it is produced, stored and

used. And also, when resources (technological, personal and financial) are used efficiently and effectively for their production, management and compliance with quality internal and external standards.

Wang and Strong (1996) affirm that qualified data are those that are fit for the use of their consumer. Mezzanzanica et al. (2015, p. 148) explain that this makes the data's quality a concept that depends on the context in which it is being addressed, since "[...] a dataset can be considered appropriate for one use while may not be suitable for a different one."

By using Davok and Garcia's (2014) definition of data governance, it is understood, for example, that separate departments that produce identical data, causing duplication of the data stored in the organization, make these data meritless. Since such data is produced through failed processes and that waste the resources of the organization (inefficient).

By duplicating the data, the organization will store data that is not needed for its employees, managers, shareholders, and other stakeholders, and therefore has no value. So, duplicate data, besides being meritless, also do not present value, because they do not meet the needs of the stakeholders who need them. That is, duplicate data does not have quality.

Likewise, it is understood that duplicate data is detrimental to information systems that collect, store and retrieve scientific data, since duplicate data has no value to researchers and practitioners who will use the system. As well, duplicate data in information systems are not meritorious, since they were generated through inefficient processes.

In addition to data duplicity, other aspects are related to poor data quality such as outdated data, inaccurate data, incomplete data, and so on.

Jesiļevska (2017) presents some of the consequences of poor data quality, such as: increased operating costs employed to detect, and correct errors associated with the data generated, and low confidence in data, which may imply a lack of acceptance of users in initiatives that use such data.

To evaluate the data quality, it is necessary to determine attributes, which are called by Jesiļevska (2017) as dimensions of it. Such dimensions can be used to measure the quality of data in different contexts.

Among such dimensions, we have, for example, accuracy, temporality, consistency and completeness of the data. Accuracy refers to the fact that the data are representative of reality; the temporality is related to the actuality of the data and the fact that it is accessible and available when it is necessary; the consistency refers to the perseverance of the data, to the fact that they do not undergo changes, and the completeness is associated with the data being complete, without null or missing values.

The dimensions of data quality are important parameters to diagnose the current situation related to the data and to identify the level of data quality that is sought to achieve.

Therefore, one of this research's focuses was related to diagnosing the data quality available in MAS, based on aspects related mainly to completeness, consistency and accuracy. For this, data governance tools were used, which will be described next.

3.2 Tools of Data Governance

It is understood that an important factor to assist in the control of data quality is the determination of roles and responsibilities about the data. Among the tools used for this, we have the RACI Matrix, which presents the roles involved in the process and their respective attributions of responsibility.

Through RACI Matrix, the internal documentation of the roles and responsibilities of the team members takes place, allowing researches to the tool to solve doubts and to reduce the conflicts among the teams (GRECO, 2014).

The RACI Matrix is used by governance frameworks such as the Control Objectives for Information and Related Technology (COBIT) and the Information Technology Infrastructure Library (ITIL) (FREITAS, 2013).

By using the RACI Matrix, benefits are provided such as: contributing to the clear division of tasks between the different roles involved; avoid to the forget key people for the process; improve understanding of the roles and responsibilities associated with the process; to serve as internal communication, and to systematize in a formal document the responsibilities of those involved in the process (SMITH; ERWIN, [2005]; RIVERO NETO, [201-?]; GRECO, 2014).

By this way, the RACI Matrix is an important tool that can be applied in data governance, helping to determine the roles and responsibilities of those involved in the activities, processes and sectors of organizations.

Another tool that can be used in data governance is the Situation Analysis Canvas. The mentioned tool can be applied individually or in conjunction with other tools, such as SWOT Analysis (COLOMBI, 2015).

Still according to the author, through the Situation Analysis Canvas it is possible to give an overview of the situation to be studied, presenting a current diagnosis and listing the priority actions (which must be accomplished in order reach the desired scenario) and the desired results.

For this, the Situation Analysis Canvas presents a framework composed of seven components responsible for assisting in the diagnosis and analysis of the situation to be studied: situation analysis; challenges and consequences; pivotal questions; recommended approach; reasons to believe; how to make it happen, and next steps (COLOMBI, 2015).

It is understood that the Situation Analysis Canvas allows for data governance a complete diagnosis of the situation studied, providing the analysis of the challenges and their consequences, and helping to plan what actions should be taken to improve the current data flow, aiming at ensuring that the data is available with quality to all stakeholders.

A third tool that can be used in data governance is the Process Model Canvas, which was developed by Marco Bijl, James P. Devlin and David Ruting (PROCESS..., c2014).

The Process Model Canvas was created as a complement to the Business Model Canvas, a template developed by Alexander Osterwalder as a strategic management tool. The Business Model Canvas provides a visual schema that can be employed to understand the operation of an existing business or to plan a business to be developed (MOURA, 2014).

By applying the Process Model Canvas in data governance it is possible to systematize in a single document: the reasons that motivate someone to use the data; the value proposition assigned to the data; what needs to be done for the desired value to be given to the data subject; what flow of information is needed for this to happen; what is the value received at the end of the process by the interested party in the data, and what is their reaction to receiving this value (PROCESS..., c2014).

Therefore, the mentioned tool makes it possible to connect the strategic management to the operational management (ibid), allowing to align what is planned by the high management with what must be executed so that those interested in the data are satisfied at the end of the process.

It should be highlighted that in this subsection we presented only some of the existing tools that can be applied in data governance. The approach was related to the tools used in this study, being how they were applied by their authors described below.

4 METHODOLOGY

As for the methodological aspects, it is understood that this research can be summarized as practical applied, with an exploratory characterization, data collection through active research, and data analysis through qualitative methods.

According to Silva and Menezes (2005), applied research aims to generate knowledge for practical application and for specific problems solving. Thus, this research is applied, as it seeks to solve a practical problem, related to the diagnosis, analysis and improvement proposal for MAS, to provide more quality to data submitted, stored and made available in the System.

Regarding the objectives, this research can be classified as exploratory, because it aims to provide greater familiarity with the object of study, seeking to make it more explicit (VIEIRA, 2002). Through this research, it was hoped to gather information about the processes of submission, storage and visualization of MAS motor data, as a way to explore the problems faced in these processes and to propose improvements, in order to provide more quality to the data available in the System.

To reach the specific objectives proposed, the methodology active research approach was used, which is characterized by Lima (2007) as the search for an innovative solution of a real problem, but that, unlike the Case Study, requires active involvement of the researcher and subjects in the context of the object of study.

According to the author, active research has an empirical basis, and is directed to the practical application, being composed by a cyclical process, called action-investigation, which basically consists of: finding a problem; reasoning about a solution to it; plan their execution; execute it and evaluate the results found to feed back the system, aiming to improve the process.

The principle of the active research cycle, described by Lima (2007), can also be observed in the study by Lorences and Ávila (2013), in which the authors propose a procedure for evaluation and improvement of information technology (IT) organizations. This procedure was developed with a focus on IT governance, however, it is understood that it can be applied equally in data governance, since it is an extension of the corporate governance and IT governance concepts (BARBIERI, 2011).

The Lorences and Ávila's procedure (2013) consists of four phases: evaluation, design, implementation and control. The evaluation phase is responsible, among other activities, for the current diagnosis of governance in the organization, for the characterization of the organization and for the analysis of the alignment of IT resources with the objectives of the organization. In the design phase the modeling and analysis of the IT processes is performed, allowing the identification of opportunities for improvement in the processes, from the results of the previous phase. In the implementation phase the execution of the modeled process occurs, and in the control phase there is the evaluation of the implemented process with a view to improving it continuously.

Considering the procedure proposed by the authors, as well as the steps of an active research described by Lima (2007), first, the situational diagnosis of MAS was done, through information raised by the experience of one of the authors' study. which was responsible for the development of the System and for its on-line availability. For this first phase, we used the Situation Analysis Canvas³ diagram.

Through the Situation Analysis Canvas, it was possible to realize an evaluation of the difficulties regarding MAS, outlining the challenges to be faced and the consequences of these challenges, determining the central problems to be solved in the System, what to do to solve them, and the next steps that must be followed for this to happen.

In the second phase, through the Situation Analysis Canvas and the DLC-IC (SANTANA, 2013), three MAS processes were identified that could be improved to contribute, more effectively, to their objective of disseminating qualified data related to the motor evaluation of children and the elderly.

The identified processes were: submission, storage and visualization of the data. However, it was decided to adapt the nomenclature of the collection stage described in the DLC-IC for submission, and the recovery step for visualization of the data, since it is understood that such nomenclatures are better suited to the context of the MAS.

In addition, the discarding step of DLC-IC was not included in this research, since it is understood that it is the purpose of the MAS to have a historical series of data to aid in the diagnosis and follow-up of the motor development of the patients attended by the professionals who use it the system.

In the third phase, a study was developed on each of the identified processes, aiming to better understand their deficiencies and what actions could be taken to solve such problems.

For this purpose the Process Model Canvas was used, it was determined for each MAS process identifying: the reasons that motivate the HEP to use this process; the amount to be delivered to these professionals; what it is necessary to do so that the desired value is reached; what flow of information necessary for the mentioned professionals to receive the desired value; what the value delivered to them at the end of the process, and what the reaction of these professionals when receiving the data of motor evaluation at the end of the process.

³ It should be noted that the diagrams used in this study and presented in the results section were developed using the online editing tool Piktochart, which is available at: <https://piktochart.com/>.

Finally, a RACI Matrix proposal was created on the responsibilities of qualifying the data inserted in the MAS, from different roles such as: the HEP, and the Information Technology Center (ITC) responsible for the system's management.

To obtain the results, we used qualitative data analysis procedures, which according to Martins and Theóphilo (2009) include the description, understanding and interpretation of the data. The use of qualitative procedures was due to the fact that the data collected are predominantly descriptive and require analysis that cannot be expressed in numbers.

5 RESULTS

The MAS's development was already started, and is still under construction, with two main objectives:

- a) collecting data generated from the motor evaluation tests performed using MDS protocol informational form;
- b) to expose the information collected to the society as a unique environment (public power, academic / scientific community and private sector).

Achieving both objectives is to bring efficiency, safety and reliability (through computerization) in the conduct of tests, and to present to society a tool for analysis that will allow the direction of public policies, data for new scientific research, and metrics for measuring the advancement of patient health.

In order to understand the operation of the MAS, particularly the flow of his data, first, there was an analysis focusing on the current situation of the entries of the HEP that use it, the mechanisms used to assess the quality of data entered, and the availability of data. The diagram resulting from this analysis can be seen in Figure 2.



Figure 2. Situation Analysis Canvas from MAS
 Source: From authors (2018)

From the situational analysis performed through the Situation Analysis Canvas, shown in Figure 2, it was possible to verify that currently to use the MAS it is necessary to request registration for the System moderators. After the analysis of the data by the moderators, the professional has his request for access denied or authorized. However, the current registration process for using MAS does not consider their ability to enter the data. So that the professional can end up inserting data with low quality, for ignoring some particularity of the System or the MDS motor evaluation method.

Thus, the need to create and implement a data entry policy was detected, restricting the inclusion of data only to qualified professionals through cross-validation of Recorded Professionals in partnership Professional Councils of health and education. The proof of aptitude to enter data and use MAS will also be through the approval of the online "Distance Learning Course" offered by Sociedade Brasileira de Motricidade Humana (SBMH)⁴.

Still through the Situation Analysis Canvas, the low participation of the HEP in the use of MAS was observed. Since the System is not the only tool available for recording motor data. In this way, the need to carry out MAS dissemination campaigns among its potential target public was identified, as well as to seek partnerships with class organs. In addition, it

⁴ For more information about: <https://ead.motricidade.com.br/curso/desenvolvimento-infantil>

was found necessary to add value to the System usability, so that HEP feel motivated to use it, rather than the other tools available.

It found to be necessary to redesign the MAS, to open the possibility of creating an open access portal to the motor evaluation data of children and the elderly, as currently only the professionals registered in the system can view the data. Each professional has access only to the data inserted by him, which makes it impossible for professionals interested in statistical analysis for the development of research and scientific studies have access to all data recorded in the System.

Finally, it was determined the absence of mechanisms to evaluate the completeness of the data inserted in the MAS. Consequently, the chance of inserting data with null or blank values in the system increases, which can affect the quality of the stored data. Therefore, it is understood that in redesigning the MAS, must be planned, developed and implemented suitable tools to verify and control the completeness of the data.

Further, in assessing the full flow of data, they identified three major processes, distinct from each other (in the proposed value, target audience and platform operation). Therefore, three Process Model Canvas were generated: Data submission (Figure 3); data storage (Figure 4); data visualization (Figure 5).

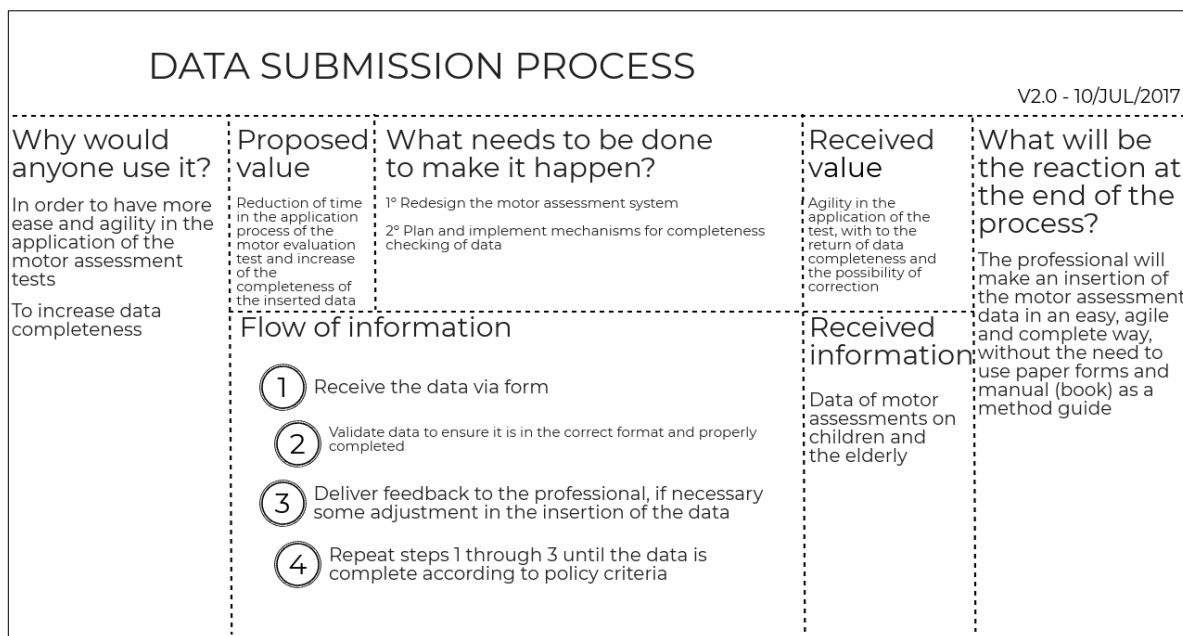


Figure 3. Process Model Canvas Data Submission
 Source: From authors (2018)

Through the Process Model Canvas data submission (Figure 3), it is understood that the proposed value of MAS is related to facilitate and speed up the application of the motor evaluation tests to the HEP that seek to insert data, besides increasing the completeness of the

inserted data, reducing the time of the test’s application. This will require redesigning the system and planning / implementing data completeness checking mechanisms.

Figure 3 also aided in visualizing the flow of information through which the data will pass after the redesign of the MAS, being: to receive the data through a form filled by the professional; validate the data, through the computerized system, so that they are in correct format and properly filled; present feedback, also through the computerized system, on the data inserted; repeat steps 1 to 3 until the data is properly adjusted according to predefined criteria.

In Figure 4, Process Model Canvas data storage, the value presented to the professional to use this service relies on the confidence that their data will be stored safely, through digital, and will undergo treatment that will avoid duplicity and guarantee homogeneity. It is understood that, for this to happen, MAS needs to be changed, that is, it becomes necessary to redesign its operation, developing and deploying the data storage and processing system

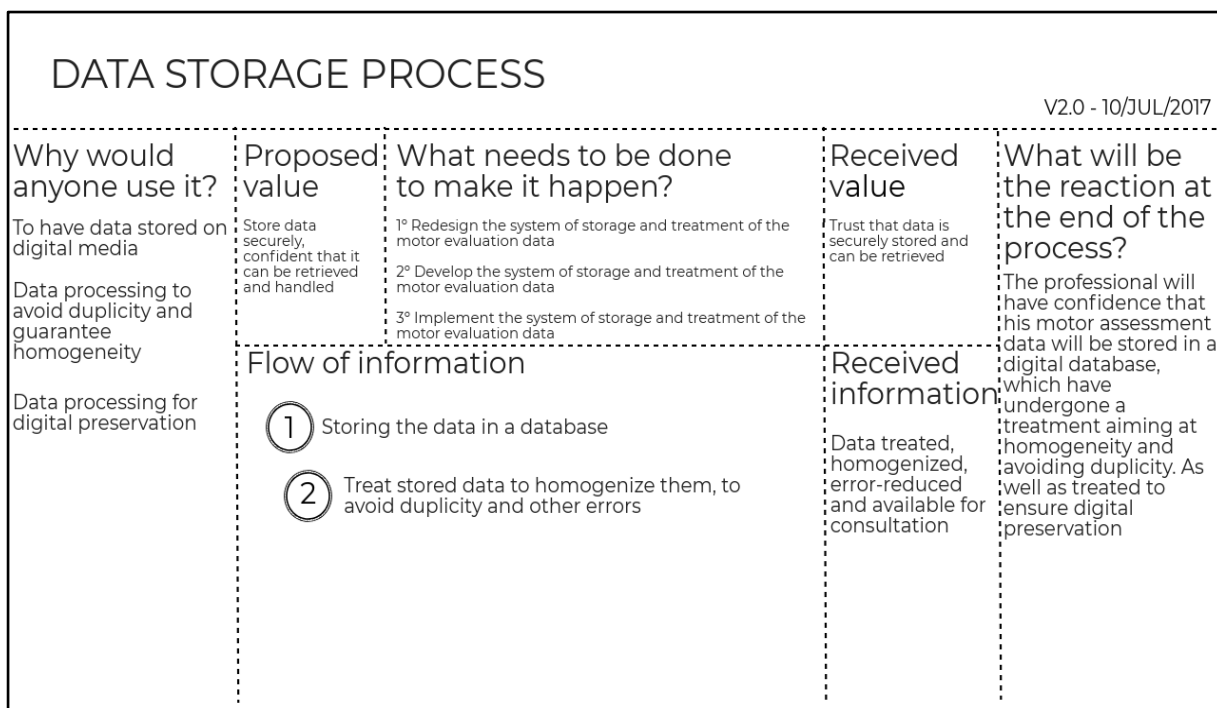


Figure 4. Process Model Canvas Data Storage
 Source: From authors (2018)

As an information flow it was possible to initially visualize the storage of the data in a database, later to be treated by means of the computerized system that will homogenize them, avoiding errors and eliminating duplications.

At end, in the last mapped process (Figure 5) referring to data visualization, it is understood that the value delivered to HEP using MAS is related to the visualization of reliable data on motor evaluation. Such data would assist in carrying out scientific research in

the area, as well as support the planning of public policies / actions that promote the treatment of human development disorders, as well as impact on the decision-making process on subjects related to the population's mobility.

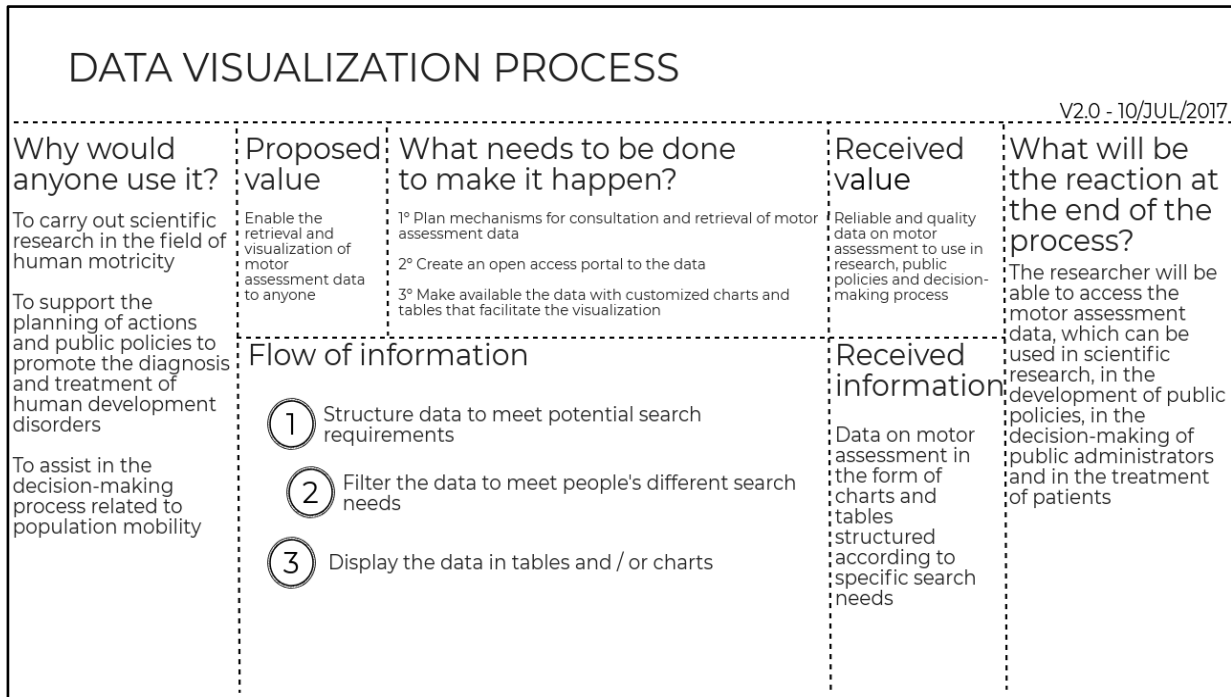


Figure 5. Process Model Canvas Data Visualization
 Source: From authors (2018)

The information flow presented in Figure 5 starts with the structuring of the data, followed by filtering based on the needs of the professional, and finally the presentation of this data in table format and / or graphics.

Figure 6 contains the proposed RACI Matrix developed for MAS.

LIABILITY MATRIX				
RACI MATRIX	Health and Education Professionals (HEP)	Information Technology Center (ITC)	Training Center (TC)	Professional Support Center (PSC)
Quality of data on the information sources.	R	S	P	P
<i>Completeness</i>	P	R	P	P
<i>Consistency</i>	R	S	P	P
<i>Accuracy</i>	R	S	P	P

Figure 6. Proposed RACI Matrix for MAS

Source: From authors (2018)

Legend: R= Role responsible for execution / control; A= Role responsible for approving; S= Role responsible for supervising; P= Role responsible for participating if necessary.

In the creation of the RACI matrix we focused on the evaluation of the distribution of responsibilities regarding data quality, subdividing this item into three that subsequently influenced the result of the principal. In order to evaluate each subitem, we identified the professionals involved in the data quality, being: the HEP, who insert the data in the MAS; the CTI, creator of MAS; the Training Center (TC), which trains those who will enter data into the MAS; the Professional Support Center (PSC), which is responsible for supporting professionals who use MAS.

Completeness was the first subitem evaluated. The CTI was defined as responsible for this item because it should prevent, through the technology used in the MAS, that any incomplete data be inserted into the system. While the other three involved are only participants.

When evaluating consistency and accuracy, we understand that HEP is responsible for both. It is the only one that has access to the source of the data, so the only one that can insert this data consistently and as close to the real value. We qualified the CTI as supervisor in both items, to follow possible signs of deviations of consistency and accuracy of the data, and TC and PSC as participants.

Finally, the responsibilities distribution of qualified data was performed based on the items that appeared the most for each one involved in the sub-items completeness, consistency and accuracy.

6 FINAL CONSIDERATIONS

At the end of this article, it is pointed out that the proposed objectives were accomplished, since it was possible to diagnose and identify opportunities for improvement and propose adequations to the data flow of the MAS, besides mapping responsibilities about data quality.

Through data governance, using tools such as the Situation Analysis Canvas and the Process Model Canvas, in addition to the RACI Matrix, it was possible to show flaws in the current security format of the MAS data, responsibilities erroneously allocated among those involved, and mainly concept errors in the planning and development of the tool.

These problems, which are now visible, can be solved, while those responsible for MAS will have subsidies to execute a new version of the platform, which will aim to add value:

- a) facilitating the including of motor evaluations data for the professionals, providing more agility to the process and greater completeness to the stored data;
- b) helping the professional in the elimination of duplicate data, besides providing tools that guarantee the homogeneity of these;
- c) ensuring the security and digital preservation of information entered by the professional in the system;
- d) allowing an intuitive research of data, for the professionals involved in the insertion of the system and for the society as a unique environment, assisting them in the development of scientific research, public policies, and better treatment of patients.

So, we conclude by means of this research, the importance and the value that is added to an information system when working with specific tools of data governance, allowing those who use them to create more efficient systems, and / or assisting in the evolution of existing systems that did not have the opportunity to use these tools during their creation, as is the case of MAS.

It is believed that by presenting this action research to the scientific community, with a practical example, we are contributing to the formation of the knowledge of the IC.

Likewise, it is understood that, through this work, it can create subsidies and even encourage other research on the search for solutions to real problems that can apply the principles and tools of data governance, as well as the concepts of the DLC-IC.

Finally, it is left as a suggestion of future study the analysis, by means of a case study, of MAS, in order to collect new data, and to evaluate how is the performance of the new platform, in relation to the old system. In addition to checking the benefits obtained with the

changes made, potential problems generated and new analyzes with suggestions for improvement.

REFERENCES

BARATA, André Montoia. **Governança de dados em organizações brasileiras: uma avaliação comparativa entre os benefícios previstos na literatura e os obtidos pelas organizações.** 2015. Dissertação (Mestrado em Ciências) - Programa de Pós-Graduação em Sistemas de Informação, Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2015. 155 f. Disponível em:

<<http://www.teses.usp.br/teses/disponiveis/100/100131/tde-28072015-215618/pt-br.php>>.

Acesso em: 06 ago. 2017.

BARBIERI, Carlos. **BI2: Business Intelligence: modelagem e qualidade.** Rio de Janeiro: Elsevier, 2011.

CAPURRO, Rafael; HJORLAND, Birger. O conceito de informação. **Perspec. Ci. Inf.**, Belo Horizonte, v. 12, n. 1, p. 148-207, jan./abr. 2007. Disponível em:

<<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/54/47>>. Acesso em: 08 ago. 2017.

2017.

CARVALHO, Mónica Isabel Machado. **Data Governance: estudo e aplicação na EDP distribuição.** 2012. Relatório de Estágio (Mestrado em Gestão) – Faculdade de Economia, Universidade de Évora, Évora, 2012. 57 p. Disponível em:

<<http://hdl.handle.net/10316/21346>>. Acesso em: 20 jul 2017.

COLOMBI, Luca. **Marketing performance measurement: how to build a situation analysis.** London: Conduit Data Services, 2015. Disponível em:

<<http://www.conduithub.com/marketing-performance-measurement-using-the-situation-analysis-canvas>>. Acesso em: 09 set. 2017.

CORRÊA, Fabiano Couto. O papel dos bibliotecários na gestão de dados científicos. **R. Digit. Bibliotecon. Ci. Inf.**, Campinas, v.14, n.3, p.387-406, set/dez. 2016. Disponível em:

<<http://dx.doi.org/10.20396/rdbci.v14i3.8646333>>. Acesso em: 29 jul. 2017.

DAVENPORT, Thomas. H. **Ecologia da informação: por que só a tecnologia não basta para o sucesso na era da informação.** São Paulo: Futura, 1998.

DAVOK, Delsi Fries; GARCIA, Renata Ivone. Modelo de avaliação de valor e mérito de estoques de informação de bibliotecas universitárias. **Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf.**, Florianópolis, v. 19, n. 39, p. 19-42, jan./abr., 2014. Disponível em:

<<http://dx.doi.org/10.5007/1518-2924.2014v19n39p19>>. Acesso em: 04 ago. 2017.

DATA GOVERNANCE INSTITUTE (DGI). **Definitions of Data Governance.** [S.l.]: DGI, c2017a. Disponível em:

<http://www.datagovernance.com/adg_data_governance_definition/>. Acesso em: 20 jul. 2017.

DATA GOVERNANCE INSTITUTE (DGI). **Goals and Principles for Data Governance**. [S.l.]: DGI, c2017b. Disponível em: <http://www.datagovernance.com/adg_data_governance_goals/>. Acesso em: 20 jul. 2017.

FREITAS, Marcos André dos Santos. **Fundamentos do gerenciamento de serviços de TI**. 2. ed. Rio de Janeiro: Brasport, 2013.

GRECO, Rafael Escaleão. **Governança de TI aplicada em Prefeitura: modelagem e melhoria dos processos de negócio utilizando boas práticas do Cobit**. 2014. 66 p. Monografia (Bacharelado em Sistemas de Informação) - Fundação de Ensino Eurípides Soares da Rocha, Centro Universitário Eurípides de Marília. Marília: UNIVEM, 2014. Disponível em: <<http://hdl.handle.net/11077/1010>>. Acesso em: 08 set. 2017.

JESILEVSKA, Svetlana. Data Quality Dimensions to Ensure Optimal Data Quality. **The Romanian Economic Journal**. v. 20, n. 63, mar. 2017. p. 89-103. Disponível em: <<http://www.rejournal.eu/sites/rejournal.versatech.ro/files/articole/2017-04-02/3443/6jesilevska.pdf>>. Acesso em: 13 set. 2017.

LIMA, João Alberto de Oliveira. Pesquisa-ação em Ciência da Informação. In: MUELLER, Suzana Pinheiro Machado (Org.). **Métodos para a pesquisa em Ciência da Informação**. Brasília: Thesaurus, 2007. p. 63-82. (Série Ciência da Informação e da Comunicação).

LOFTIS, Lisa. **O que é governança de dados?** [S.l.]: Youtube, 2014. Disponível em: <<https://www.youtube.com/watch?v=icIgpsqFbFo&t=18s>>. Acesso em: 20 jul. 2017.

LORENCES, Patricia Pérez; ÁVILA, Lourdes Francisca García. The evaluation and improvement of it governance. **JISTEM: J. Inf. Syst. Technol. Manag.**, São Paulo, v. 10, n. 2, p. 219-234, ago. 2013. Disponível em: <<http://dx.doi.org/10.4301/S1807-17752013000200002>>. Acesso em: 07 set. 2017.

MARTINS, Gilberto de Andrade; THEÓPHILO, Carlos Renato. **Metodologia da investigação científica para ciências sociais aplicadas**. 2. ed. São Paulo: Atlas, 2009.

MEZZANZANICA, Mario. et al. A model-based evaluation of data quality activities in KDD. **Information Processing & Management**, v. 51, n. 2, mar. 2015, p. 144–166. Disponível em: <<http://dx.doi.org/10.1016/j.ipm.2014.07.007>>. Acesso em: 06 ago. 2017.

MOURA, Luiz Antonio Rolim de. **Memória Seminário sobre Canvas Model com Alexander Osterwlder**. [S.l.]: SEBRAE Paraná, 2014. Disponível em: <<https://pt.slideshare.net/luizrolimdemoura/memoria-seminario-sobre-canvas-model-com-alexander-osterwlder-by-luiz-rolim>>. Acesso em: 09 set. 2017.

PROCESS Model Canvas. [S.l.: s.n], c2014. Disponível em: <<http://www.processmodelcanvas.com/>>. Acesso em: 09 set. 2017.

RIVERO NETO, Miguel. **Matriz RACI: entenda o que é e como usar**. Palhoça: VEC, [201-?]. Disponível em: <<http://valorecompetencia.com.br/gestao-de-operacoes/matriz-raci-entenda-o-que-e-e-como-usar>>. Acesso em: 06 ago. 2017.

SANTANA, Ricardo Cesar Gonçalves. Ciclo de vida dos dados e o papel da ciência da informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO. 14., 2013. Florianópolis. [Paper...]. Florianópolis: UFSC, 2013.

Disponível em:

<<http://enancib2013.ufsc.br/index.php/enancib2013/%20XIVenancib/paper/viewFile/284/319>

>. Acesso em: 13 jul. 2017.

SANTANA, Ricardo César Gonçalves; BONINI, Alfredo. Análise de dados sobre produção de leite: uma perspectiva da Ciência da Informação. **Inf. & Tecn.** v. 1, n. 1, 2014.

Disponível em: <<http://periodicos.ufpb.br/index.php/itec/article/view/19493/10863>>. Acesso em: 21 jul. 2017.

SANTOS, Isabel Maria Francisca dos. **Uma proposta de governança de dados baseada em um método de desenvolvimento de arquitetura empresarial.** 2010. 140 p. Dissertação (Mestrado em Informática) – Centro de Ciências Exatas e Tecnologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2010. Disponível em:

<<http://www2.uniriotec.br/ppgi/banco-de-dissertacoes-ppgi-unirio/ano-2010/uma-proposta-de-governanca-de-dados-baseada-em-um-metodo-de-desenvolvimento-de-arquitetura-empresarial/view>>. Acesso em: 21 jul. 2017.

SILVA, Edna Lúcia da; MENEZES, Estera Muszkat. **Metodologia da Pesquisa e Elaboração de Dissertação.** 4. ed. Florianópolis: UFSC, 2005. Disponível em:

<https://projetos.inf.ufsc.br/arquivos/Metodologia_de_pesquisa_e_elaboracao_de_teses_e_dissertacoes_4ed.pdf>. Acesso em: 20 jul. 2017.

SISTEMA de Avaliação Motora. [S.l. : s.n., 201-?]. Disponível em:

<<http://sam.motricidade.com.br/>>. Acesso em: 27 jun. 2017.

SMITH, Michael L.; ERWIN, James. Role & responsibility charting (RACI). In: **Project Management Forum (PMForum)**. [2005]. Disponível em:

<https://pmcie.org/images/downloads/raci_r_web3_1.pdf>. Acesso em: 06 ago. 2017.

VIEIRA, Valter Afonso. As tipologias, variações e características da pesquisa de marketing. **R. FAE**, Curitiba, v. 5, n.1, p. 61-70, jan./abr. 2002. Disponível em:

<<https://revistafae.fae.edu/revistafae/article/view/449>>. Acesso em: 24 jul. 2017.

WANG, Richard Y.; STRONG, Diane M. Beyond Accuracy: What Data Quality Means to Data Consumers. **Journal of Management Information Systems**, v. 12, n. 4, Spring 1996, p. 5-33. Disponível em:

<https://www.jstor.org/stable/40398176?seq=1#page_scan_tab_contents>. Acesso em: 06 ago. 2017.

WENDE, Kristin; OTTO, Boris. A contingency approach to data governance. In: INTERNATIONAL CONFERENCE ON INFORMATION QUALITY (IQ-2007), 12, 2007. Cambridge, USA. **Research Paper...** Cambridge, USA: Alexandria, 2007. Disponível em:

<<https://www.alexandria.unisg.ch/publications/213308>>. Acesso em: 06 ago. 2017.

