

The purpose of data governance in external control organizations



Ricardo Dantas Stumpf

is an employee of the Federal Court of Accounts – Brazil. He works at the Department of IT Solutions in the area of Technical Management of Data. He has a B.A. in Computer Science from the University of Brasilia (UnB).

ABSTRACT

There are different levels of governance in an organization, whether dealing with strategies, IT, data in general or open data. Data are manageable assets, whether for the healthy operation of the institution or for sophisticated data analysis. This paper aims at defining the utility of data governance in this context, as well as the basic elements for its implementation with a bias for External Control organizations.

Keywords: Data governance. Data management. Data Science.

1. INTRODUCTION

What is important in an organization? People, budget, equipments and real estate, data... It is difficult even to think of the operation of organizations if any of these elements is missing, which proves the crucial importance of all of them. Several of these assets are essential to most entities and, consequently, they must be well **managed and governed** in order to maximize their value. In summary, managing means to guarantee that we do things right, while governing means to guarantee that we do the right things (WODZINSKI et al, 2015).

Now that the Federal Court of Accounts (TCU) has been investing heavily in data analysis, taking care



of the data became even more critical, as we can infer from the excerpt from the quarterly report below:

A consensus has been reached on the three main challenges that the court must overcome in order to obtain good results with data analysis: the technical challenge, especially related to data quality; the regulatory challenge, regarding legal and normative restrictions; and the cultural one, concerning behavioral aspects of the people involved (TCU QUARTERLY REPORT - 4th QUARTER/2014, p. 101).

The second edition of the TCU Governance Guide (2015) brings a summary of interaction between management and governance:

“The functions of governance are:

- a. To define strategic direction;
- b. To oversee management;
- c. To involve stakeholders;

Figure 1:

TCU – Basic Governance Reference Guide applicable to the Public Administration bodies and entities, 2nd edition, p. 32



- d. To manage strategic risks;
- e. To manage internal conflicts;
- f. To audit and evaluate the management and audit system; and
- g. To promote accountability and transparency” (TCU, BASIC GOVERNANCE REFERENCE GUIDE APPLICABLE TO THE PUBLIC ADMINISTRATION BODIES AND ENTITIES, 2nd edition, p. 31).

This logic applies both to the strategic level and to the different levels and more focused facets of the institution. Governance and information technology management, among others, gravitate towards corporate governance and management. Governance and management of produced and custodial data management gravitate towards the latter. Even data governance and management subdivide in the most complex organizations.

2. DATA AND INFORMATION

It is common to use data and information as synonyms, as well as knowledge and skill. However, stric-

tly speaking, they represent different concepts according to professor Valdemar Setzer (2015), from the Universidade de São Paulo.

Data is “a sequence of quantified or quantifiable symbols”, while information means “an abstraction on one’s mind”.

As a simple example, the sequence of symbols **880**. It is a **data**, either digital or not. If one recognizes 880 in a list of salaries, we then have the **piece of information** of a numerical salary with 880 units.

On the other hand, **knowledge** demands a personal abstraction and experimentation (SETZER, 2015). A company’s employee knows, according to their experience, that it means BRL 880,00, minimum wage in October 2016 attributed to a just-hired employee. The number 880 changed from data into information and then into knowledge. Finally, an auditor has the **skill** to interact and act on such knowledge during an audit.

This process takes place every day in professional environments. “Process” is a document, a repeatable workflow, a legal instrument, a computer code in execution or a protrusion of bones? It depends.

That is when **Data Governance** comes in, by using people’s knowledge and skill to define policies, responsibilities, glossaries, metadata, workflows of data

Figure 2:

Different levels of governance “gravitating” simultaneously

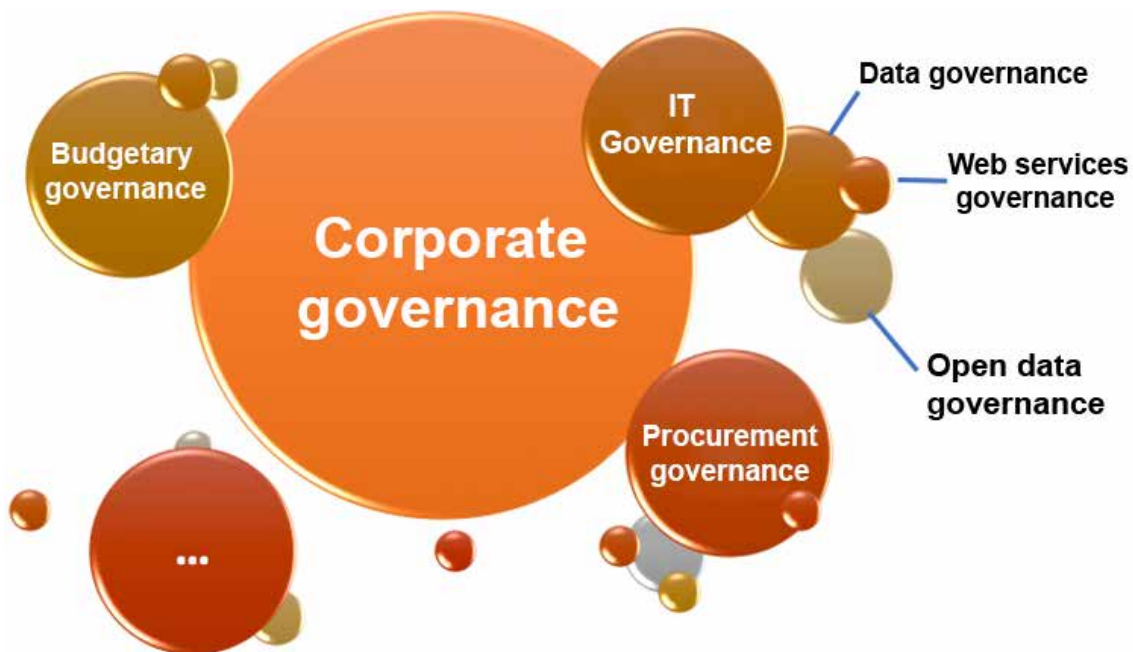
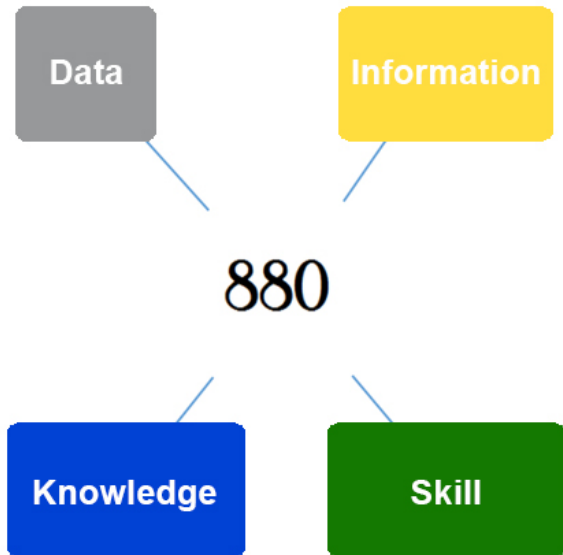
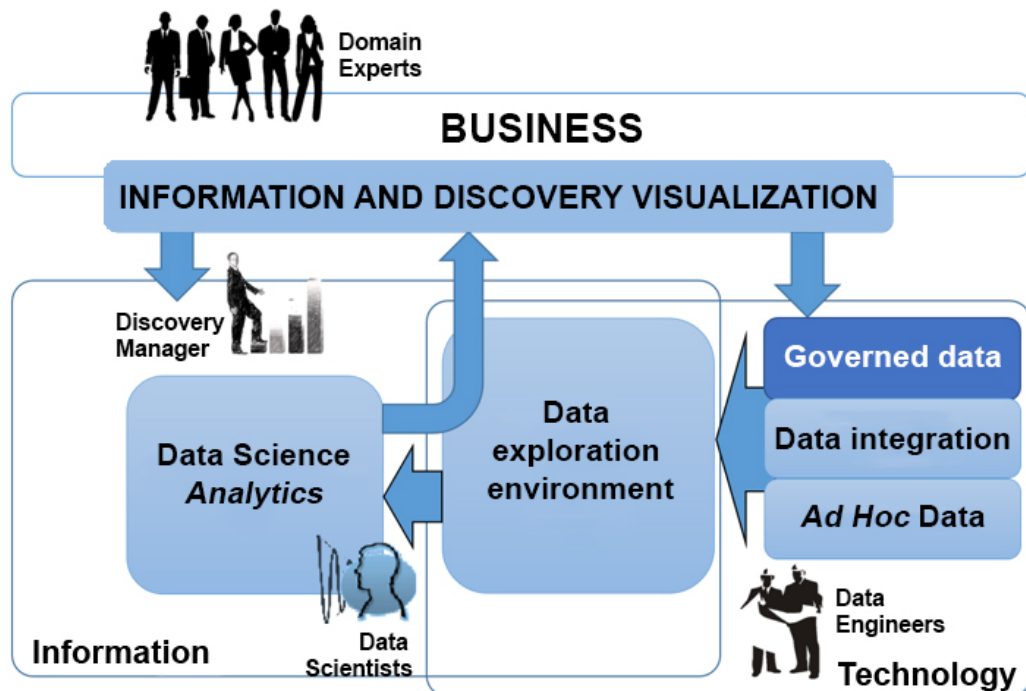


Figure 3:
Example of how symbols relate to several concepts



in movement, data monitoring and lineage, which allow the organization to improve their quality, discovery and understanding in order to simplify information and knowledge extraction, leading to better analysis and business decisions.

Figure 4:
Data cycle samples, adapted (HENDERYCKX, 2016)



3. NEW OIL IS DIGITAL

Data are being called “new oil” (VANIAN, 2016), but they have always been important and present, even before the digital format. However, our focus is on digital data, which have been growing in such a rhythm and variety that they have been associated to “tsunamis”, “landslides” and “storms”, or Big Data. Internal data, usually the most aligned to the corporate internal work processes are called Small Data (HENDERYCKX, 2016).

A trend among companies is to invest more and more in data analysis and visualization, in order to envision more business opportunities and to increase operational efficiency. Likewise, public bodies, especially oversight ones, have been collecting great quantities of data from their partners and auditees, in addition to the feedback from society in the social networks.

Jan Henderyckx, DAMA Benelux, Chairman, presented a generic model for the data cycle:

In this model, notice that the data coming from business areas must be well governed. That does not mean that custodial external data do not require any governance, but primary worries may be others. For example, the TCU has an environment used by the External Control as an exploration and external and internal data-crossing laboratory.

When necessary, the area responsible for the lab adopts the “Opportunity Principle”, which justifies the ad hoc inclusion of new bases, even before there are quality treatments and more detailed documentation. The reason is to avoid the window of opportunity to pass. From then on, however, such base becomes eligible for a more complete treatment process.

In addition, there is the **contextual** issue. Low quality external bases may be useful to discover audit findings in information systems, while good quality derived bases, with full documentation of metadata, facilitate the analysis of the effective use of transfers of financial funds, for instance.

When the model presented is low in quality and organization, problems occur in three moments (DYKES, 2016):

- a. *a priori* people do not trust the data;
- b. there are so many sources that people do not know which one to use; and
- c. there is analysis paralysis, when it is difficult to know when the analysis is good enough.

That said, we can turn to data management and governance.

4. DATA MANAGEMENT

Organizations usually already manage their data, especially the structured and semistructured data. Regarding the way it is done, it is common for management to follow good practices in their niches by force of necessity and knowledge shared by the internal peers and partners, such as the ones from the IT-Audit community, with around 20 oversight bodies that discuss better practices in IT. The systematization of the performance niches is displayed in frameworks as proposed in the DMBok (MOSLEY et al., 2012), of DAMA International.

In the new version of this framework, even techniques that are more modern, such as integration by web services and publish-subscribe, have been addressed as ways of data to circulate locally or in clouds among systems or systems modules (BRADLEY, 2016).

Nevertheless, the topic Data Governance (which is a kind of “glue” for all the processes related to the subject) is only now gaining energy, since com-

panies and organizations that depend on those data want to extract the maximum value from this more and more widespread and more complex asset. It is worth remembering that there might already be areas that govern their data well. In such cases, according to the English politician Lucius Cary, “where it is not necessary to change, it is necessary not to change” (In: SEINER, 2014).

Information is the raw material for the work of any oversight body. Their final products (determinations, recommendations) are primary sources of information to all of their auditees.

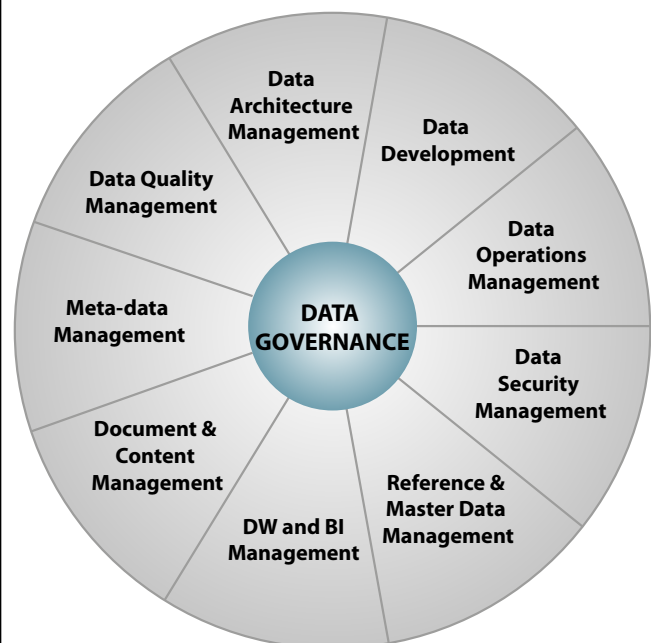
5. DATA GOVERNANCE

We finally reached a formal definition, according to John Ladley (2012):

Data governance is the organization and implementation of policies, procedures, structure, roles and responsibilities which outline and enforce the rules of engagement, decisions, rights and accountabilities for the effective management of information assets.

Figure 5:

Synthetic view of the data framework, according to DMBok



Regarding the definition, we will briefly discuss the need for principles, policies, data stewards and data classification, glossaries and metadata, leaderships and implementation approaches.

It is important to know the limits. Very bureaucratic data governance is an invitation to disobedience; on the other hand, excessive flexibility can lead to dis-governance, to a less efficient data management. That is why we should start by defining **principles**.

5.1 PRINCIPLES

Very specific, logical and feasible principles facilitate all the next steps. For instance, if the principle “data are corporate assets” is adopted by the organization, we must treat the entity “data” as we treat other assets: continuous improvement, definition of responsibilities etc. (LADLEY, 2012).

Each organization has its own principles. The following list is illustrative and is not exhaustive.

5.2 POLICIES

As John Ladley (2012) said, policy is what “give teeth to the principles”. It is the formal document that makes the organization adopt the discussed principles. It is also common for policy to define the responsibility for the data (data stewardship), as well as the organizational structure which will conduct and monitor the efforts of Data Governance.

As a good example, we suggest reading the Information Governance Policy of the Central Bank of

Brazil, published under Ordinance 47, of February 20, 2013, updated in 2016.

5.3 DATA STEWARDSHIP AND CLASSIFICATION

Data stewards are people or groups of people who have the responsibility of **taking care of data under their business sphere**. This is a fundamental change since it shares with IT the mission of taking care of the corporate data.

Suppose that the quality principle has been adopted. That means that business areas have to worry about the quality of data they generate for themselves and for others, and they must state quality problems in reports and any other sources that come from other business areas. The **Data Governance Office**, or its equivalent, is the department that will support data stewards in this duty.

Law 12.527/2011, known as the Information Access Law, formalized a guideline for the Public Administration, which says “publicity [of information] as a general rule and of secrecy as an exception”. Therefore, the data stewards must also actively classify information they produce and guard, in order to accelerate the availability of public data or to restrict them according to the law.

5.4 GLOSSARIES AND METADATA

Glossaries, or vocabularies, define and eliminate ambiguities regarding business terms, such as the term “process”, mentioned in section 2. The TCU

Table 1:

List of some of the principles for implementation of Data Governance

Principles	Description
Golden rule	All data are treated as corporative assets
Federation	There are defined patterns for data structures
Efficiency	Relevant data must be available to authorized users at the right moment, at the right place and in the right format
Quality	Corporative data are measured and managed to provide quality
Risk management	Comply with legislation, policies and internal rules regarding data
Collaboration	Corporative data are shared and disclosed resources
Contextualization	The context of use of data changes the way it is stored, treated and used
Innovation	New techniques are encouraged, following the other principles

Source: adapted from LADLEY (2012)

keeps the Vocabulary of External Control (VCE), created, according to the Minister-President Aroldo Cedraz, “to standardize the treatment of specialized information and to grant greater agility and precision to the recovery of contents in the TCU information systems” (VCE, p.5).

The classic definition of metadata is “data about the data”. However, we can extend such concept to two aspects: business and technical metadata.

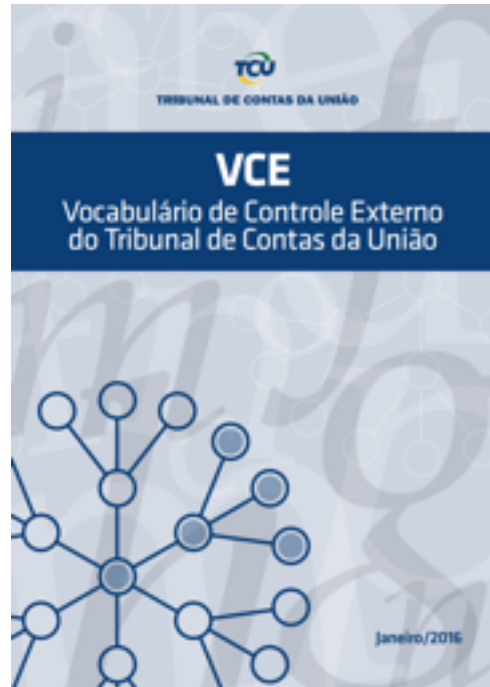
Business metadata are those that contextualize data that is stored or in movement. For instance, consider the WZV table of TTNN (hypothetical) database. Since the business area and the domain experts know this subject, they can **enrich the existing metadata** with specific information. Now, data scientists will know that the WZV table contains suppliers’ registration data and that the TTNN base brings audited organizations that, among others, have purchased from those suppliers.

On the other hand, technical metadata are those which bring IT information about the data, such as tables and columns names, load dates, version, sizes, types etc.

The “magic” happens when technical and business metadata are **integrated to discovery** and use for analysis and change management areas. At TCU, there are some actions in this direction. The VCE data are being slowly integrated to other information technology solutions, in order to optimize and standardize the use of concepts. Another one is a tool specialized in metadata that shows, in an integrated way, information originated from data modeling, integrations through ETL (*Extract Transform Load*) and corporate databases.

Figure 6:

VCE front cover of the January/2016 issue



5.5 STRCUTURES

Each Corporation has its own culture and needs, and it would not make sense to have only one type of organizational structure to govern data. The CDO (Chief Data Officer) figure is increasingly common. The following are some of the ways used, according to Bergson Rêgo (2013):

Figure 7:

Edited sample of integrated technical metadata

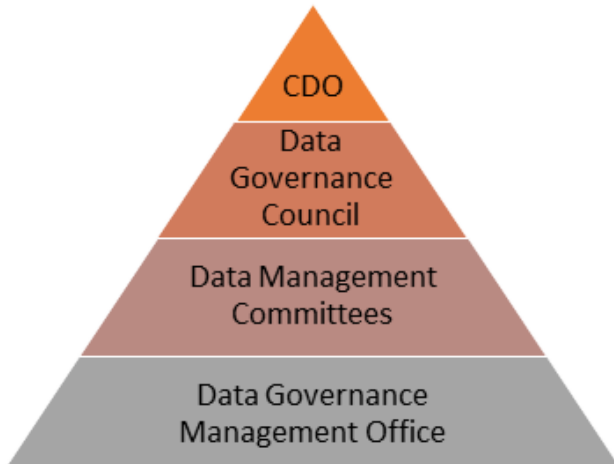


- a. CIO heading IT centralized team, which interacts with the business;
- b. CDO interacting with IT and a central data management area of the business;

- c. CDO interacting with IT and several data management areas of the business; or
- d. a CDO for each great line of business interacting with IT (hybrid).

Figure 8:

Simple hierarchic structure (RÉGO, 2013)



The hybrid structure is particularly interesting and used in practice in a great multinational bank (IPPOLITI, 2016). Due to the wide internal variety of available data, they have decided to divide them in great lines of business.

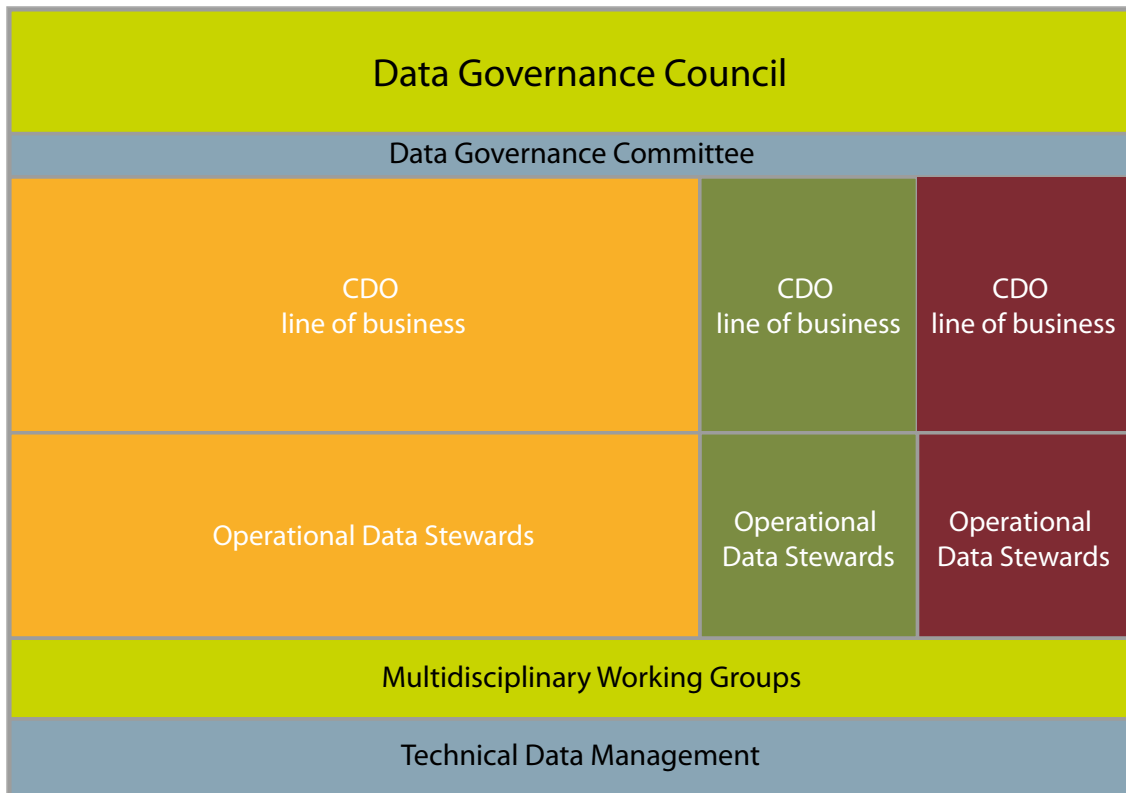
5.6 IMPLEMENTATION APPROACHES

There are three more common structured ways of implementing Data Governance (SEINER, 2016):

- a. Command and control;
- b. Traditional; and
- c. Non-invasive.

Figure 9:

Hybrid structure, adapted (IPPOLITI, 2016)



On the command and control approach, the highest-level committee defines the rules, determines data stewards, purchases tools and demands results.

The traditional one identifies the data stewards and guides them with more generic processes and existing tools, and measures results by analysing the involved data.

Finally, the non-invasive one formalizes the tacit data stewardship, focuses on applying the existing processes, encourages building of tools and the use of the current ones and measures results through the perceived increase in efficiency and effectiveness of analysis capabilities.

As we have already highlighted, none of them is necessarily better than the other one. It all depends on the culture and needs of each organization.

4. CONCLUSION

Despite the current importance of the theme Data Governance, it will be even greater in the future, when there will be great automation, even of intellectual activities, as long as they are repeatable. Data, whether originated from informatized systems or from sensors in day-by-day objects, will be a basic instrument for such automation.

That means that organizations should capture data, store them in an effective way and automate analysis with cognitive algorithms (HEN-

DERYKCX, 2016). According to Ian Rowlands (2016), business metadata will have their semantics, and not only their syntax, inferred by algorithms and crowdsourcing (big groups thinking and collaborating).

It is also an increasing challenge to identify and reduce dark data, which represent loss of opportunities, waste of resources and risk (ROWLANDS, 2016). Rowlands reiterates an excerpt from a statement of the Auditor General of Canada in which this is a problem to be fought:

One of the topics that unite many of our audits is that data collected from many governmental organizations are either unusable or are not usable, or are not used (Auditor General of Canada in ROWLANDS, 2016).

To make such reduction possible, the need to govern such data is pressing – in the right dosage for each case and context, always **focusing on the mission and vision of the institution**.

REFERENCES

BRADLEY, C. The new DMBOK 2 discipline of Data Integration. Disponível em: <<https://www.brighttalk.com/webcast/12405/186095/the-new-dmbok-2-discipline-of-data-integration>>. Acesso em: 12 abr. 2016.



BRASIL, BANCO CENTRAL DO BRASIL – BCB. Portaria nº 47, de 20 de fevereiro de 2013. Diário Oficial da União – DOU, 21 de fevereiro de 2013. Seção 1, p. 24.

BRASIL, Comunidade de Tecnologia da Informação Aplicada ao Controle – TI Controle. Sobre a Comunidade. Disponível em: <<http://www.ticontrole.gov.br/ticontrole/sobre-a-comunidade/sobre.htm>>. Acesso em: 20 out. 2016.

BRASIL. LEI nº 12.527, de 18 de novembro de 2011. Lei de Acesso à Informação. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm>. Acesso em: 20 out. 2016. 2011.

BRASIL, TCU – TRIBUNAL DE CONTAS DA UNIÃO. Referencial básico de governança aplicável a órgãos e entidades da administração pública. Versão 2 – Brasília: Tribunal de Contas da União. Secretaria de Planejamento, Governança e Gestão. 80p. Disponível em: <<http://www.tcu.gov.br/governanca>>. Acesso em: 20 out. 2016. 2014.

_____. Relatório de Atividades: 4º trimestre de 2014 – Brasília: Tribunal de Contas da União. Secretaria-Geral da Presidência, p.101. Disponível em: <<http://portal.tcu.gov.br/publicacoes-institucionais/relatorios/relatorios-de-atividades>>. Acesso em: 20 out. 2016. 2015.

_____. Vocabulário de Controle Externo do Tribunal de Contas da União – Brasília: Tribunal de Contas da União. Instituto Serzedello Corrêa. Disponível em: <<http://portal.tcu.gov.br/vocabulario-de-controle-externo/>>. Acesso em: 20 out. 2016. 2015.

CANADÁ, Office of the Auditor General of Canada – OAG. Spring Reports of the Auditor General of Canada. Auditor General's Opening Statement. Disponível em: <http://www.oag-bvg.gc.ca/internet/English/osm_20160503_e_41358.html>. Acesso em: 20 out. 2016. 2016.

DMBOK. MOSLEY, M.; BRACKETT, M.; EARLEY, S. HENDERSON, D. DAMA Guia para o corpo de conhecimento em gerenciamento de dados. Technics Publications, versão brasileira 2012.

DYKES, B. Big Data Paralyzing Your Business? Avoid These 3 Common Traps. Forbes, Entrepreneurs. Disponível em: <<http://www.forbes.com/sites/brentdykes/2016/09/28/big-data-paralyzing-your-business-avoid-these-3-common-traps/>>. Acesso em: 20 out. 2016.

HENDERYCKX, J. Sustaining Data Driven Innovation. In: DATA MANAGEMENT CONFERENCE LATIN AMERICA, 2016, São Paulo. Anais eletrônicos... São Paulo: DAMA Brasil, 2016. 42 slides. Disponível em: <<http://www.dmc-latam.com/palestrantes/jan-henderyckx/>>. Acesso em: 19 set. 2016.

IPPOLITI, J. Implementing Data Governance Strategies. In: ENTERPRISE DATA GOVERNANCE ONLINE, 2016. Anais eletrônicos... 20 slides. Disponível em: <<http://video.dataversity.net/video/keynote-implementing-data-governance-strategies/>>. Acesso em: 12 fev. 2016.

LADLEY J. Data Governance: How to Design, Deploy and Sustain an Effective Data Governance Program. The Morgan Kaufmann Series on Business Intelligence. Morgan Kaufmann. 2012.

RÊGO, B. L. Gestão e Governança de Dados: Promovendo Dados Como Ativo de Valor Nas Empresas. Rio de Janeiro: Brasport, p. 60-74. 2013.

ROWLANDS, I. Data Management in Motion. In: DATA MANAGEMENT CONFERENCE LATIN AMERICA, 2016, São Paulo. Anais eletrônicos. São Paulo: DAMA Brasil, 2016. 25 slides. Disponível em: <<http://www.dmc-latam.com/palestrantes/ian-rowlands/>>. Acesso em: 19 set. 2016.

SEINER, R. Non-Invasive Data Governance: The Path of Least Resistance and Greatest Success. Technics Publications. 2014.

_____. Comparing Approaches to Data Governance. Disponível em: <<http://tdan.com/comparing-approaches-to-data-governance/20386>>. Acesso em: 5 out. 2016.

SETZER, V. W. Dado, informação, conhecimento e competência. Journal Data & Information 1(1), DAMA Brasil: p. 38-55. 2015.

VANIAN, J. Why Data Is The New Oil. Fortune, Brainstorm Tech. Disponível em: <<http://fortune.com/2016/07/11/data-oil-brainstorm-tech/>>. Acesso em: 20 out. 2016. Julho de 2016.

WODZINSKI, M. et al. Building an Impactful Data Governance Program One Step at a Time. Disponível em: <<https://www.brighttalk.com/webcast/10477/142503/building-an-impactful-data-governance-program-one-step-at-a-time>>. Acesso em: 26 fev. 2015.