

BRAZILIAN GOVERNMENT OPEN DATA PROJECT: A CASE STUDY

Carolina de Oliveira¹

Universidade Federal do Estado do Rio de Janeiro – Unirio

Introduction

This research is a case study about Brazil's Government Open Data Project, which is part of the Electronic Government Program (GOV.BR). Among the goals of this research is to demonstrate the importance of digital records management with the Archival Science theory foundation for maintaining the quality of data, such as the data reliability, authenticity, usability, traceability and accessibility, especially in an open initiative. The body of information for the research includes the Ministry of Justice's datasets (Excel spreadsheets from Microsoft Office) that are available for use and reuse by Brazilian citizens.

In this paper we communicate our first impression of the foundation, related studies and methodology for the development of the Brazilian Government Open Data Project research. On the Open Government Data issue we will discuss data quality, specifically data reliability and authenticity for use and reuse by data consumers.

The open initiative, especially in the context of government, is an important opportunity for democratic advancement and citizen empowerment, but some queries should be made: Does the Brazilian Government Open Data Project note that digital records management should be based on Archival Science theory foundation? What does open data mean? What are the qualities of the open government data? How does the Ministry of Justice maintain authenticity and reliability of government data in Excel spreadsheets (xls format)?

The Open Government Data and data quality themes have been studied by a diversity of research groups, primarily from Computer Science and Information Systems, but this is also an opportunity for Archival Science researchers to work in an interdisciplinary approach. Collaborative efforts between different disciplines can bring new insight from the theoretical and practical experiences of each specific area of knowledge.

Thus, in this paper we propose to contribute to the Open Government Data and data quality debate in the Brazilian scenario. The rest of this paper is structured as follows: Section 2, Brazilian Government Open Data Project; Section 3, Open Government Data and data quality; Section 4, Records management, Archival Science and data quality standing for Open Government Data; Section 5, Methodology; and Section 6, Conclusion.

2 Brazilian Government Open Data Project

In Brazil, the Federal Administration approved its e-Government (e-gov) program in 2000 to make information available to civil society through information and communication technologies (ICTs). Among such technologies, computer based systems are being used by

1

Currently studying for a master's degree in Mestrado profissional em Gestão de Documentos e Arquivos.

federal public organizations to carry out management and to promote citizens' access to information resulting from the functions and activities performed by federal public organizations.

In recent years, we have seen an increasing movement of "openness": Open Government, Open Data, Open Archives and Open Innovation. To cope with this scenario, Brazilian's e-gov defined in 2011 the Government Open Data Project (Projeto Dados Abertos Governamentais).

The Brazilian Government Open Data Project follows the principles of an open data initiative, so the decision to make available the data produced by the government is a proactive initiative for transparency and for civil society's participation in government decisions.

In Brazil, there is an act that obligates all levels of government to make their financial records, such as account reports, plans, management reports and other similar records, available using an electronic environment, such as the Internet. This act is called the Revenue Responsibility Act (Lei Complementar nº 101, de 4 de maio de 2000). Among the initiatives for public transparency is the *Portal da Transparência*² (Transparency Portal), which provides information about public service transactions and employee salaries. Thus, it was the first step for Brazil to become internationally recognized for its open government initiatives.

In 2011, as a result of Brazil's federal government actions for transparency and access to information, the country was one of the co-founders of the Open Government Partnership (OGP) with Indonesia, Mexico, Norway, the Philippines, South Africa, the United Kingdom and the United States³. One of Brazil's commitments was to elaborate an action plan that would cover a two-year period. The coordination of the Brazilian Open Government Plan was organized by an Inter-Ministerial Committee on Open Government (CIGA) (see Table 1).

Table 1 – CIGA members

Inter-Ministerial Committee on Open Government (CIGA)
I – <i>Casa Civil da Presidência da República</i> (coordination);
II – <i>Secretaria-Geral da Presidência da República</i> ;
III – <i>Controladoria-Geral da União</i> ;
IV – Ministério da Justiça;
V – <i>Ministério da Fazenda</i> ;
VI – <i>Ministério do Planejamento, Orçamento e Gestão</i> ;
VII – Ministério da Ciência, Tecnologia e Inovação;
VIII – Ministério das Comunicações;
IX – <i>Ministério das Relações Exteriores</i> ;
X – Ministério da Educação;
XI – Ministério da Saúde;
XII – Ministério do Esporte;
XIII – Ministério do Meio Ambiente;
XIV – Ministério da Integração Nacional;

2 Available at <<http://transparencia.gov.br/>>.

3 See more at: <<http://www.opengovpartnership.org/about#sthash.j55uM8vc.dpuf>>.

XV – Ministério da Previdência Social;
 XVI – Ministério do Desenvolvimento Social e Combate à Fome;
 XVII – Secretaria de Direitos Humanos da Presidência da República; e
 XVIII – Secretaria de Comunicação Social da Presidência da República.

Another action by federal government was to define an infrastructure whose purpose is to define, to structure and to coordinate the open data policy. This infrastructure is called National Open Data Infrastructure (INDA).

The INDA was defined by Instrução Normativa nº 4 (IN nº 4), 13/4/2012, from the Logistics and Information Technology Secretary (SLTI). Its composition is shown in Table 2.

Table 2 – INDA’s Manager Committee

Comitê Gestor da INDA
a– Ministério do Planejamento, Orçamento e Gestão, represented by Secretaria de Logística e Tecnologia da Informação – SLTI/MP (coordination); b – Casa Civil da Presidência da República; c – Controladoria Geral da União – CGU; d – Ministério da Ciência, Tecnologia e Inovação; e – Ministério do Desenvolvimento Social; f – Ministério da Educação; g – Ministério da Saúde; h – Secretaria-Geral da Presidência da República; i – Instituto Brasileiro de Geografia e Estatística – IBGE; j – civil society representative; k – IT academic sector representative.

For this paper, at Figure 1, we had designed a timeline of Brazil's federal government actions for an open government initiative.

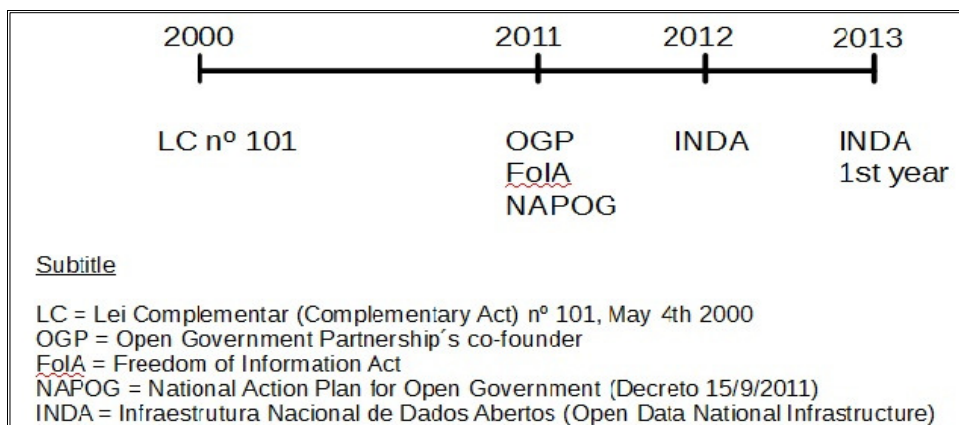


Figure 1 – Federal government action’s timeline for an open government

An issue that should be disclosed here is that in the scope of the Brazilian federal government's actions in coping with the open data initiative, neither the Archival Science and digital records management nor the National Archives of Brazil were mentioned. The latter could be invited as member of INDA's Manager Committee as said in IN nº 4, § 5º. However, until this phase of the research, we had not had any converse evidence of it.

3 Open Government Data and data quality

This paper addresses the relevance of Open Government Data and studies about data quality. If government bodies want to make available their data for citizens' free use, then they will need to ensure the quality of the data. However, what is Open Government Data and what is data quality?

According to the Open Knowledge Foundation (OKF), open government data means "Data produced or commissioned by government or government controlled entities (...) that can be freely used, reused and redistributed by anyone"⁴. By this definition we observe the merging of open data and open government notions.

It is relevant to reveal here the eight principles of Open Government Data: "Government data shall be considered open if it is made public in a way that complies with the principles: complete, primary, timely, accessible, machine processable, non-discriminatory, non-proprietary, license-free."⁵ (Open Government Data, 2007).

Any data available by the government on its websites should have the qualities that presume they are good to be used, reused and redistributed. Despite the idea that government data are of high quality and trustworthy (Shadbolt et al, 2012, p. 3), we should consider which principles should comprise the Open Government Data qualities.

Data quality is relevant to precise decision-making. If inaccurate or outdated data are available for use, business losses could happen as a result (Martins, 2009). With the purpose of defining data quality requirements, an international standard had been suggested by the International Organization for Standardization, which is *ISO/TS 8000-110:2008 specifies general, syntax, semantic encoding and data specification requirements for master data messages between organizations and systems*. Other parts of ISO/TS 8000 include part 100 (introduction), part 120 (provenance), part 130 (accuracy) and part 140 (completeness) (Benson, 2009).

In the last 20 years, research regarding data quality has advanced (Sadiq et al, 2011). For instance, a framework of data quality attributes from the point of view of data consumers developed by Wang and Strong, in 1996, remains a classical one. According to those authors, Information System theory accuracy is the most relevant attribute for data, but for data consumers other attributes are relevant (Wang et al, 1996), as shown in Figure 2.

4

See more at: <<http://opengovernmentdata.org/#sthash.6yt9HzqD.dpuf>>.

5 See more at: <<http://opengovdata.org>>.

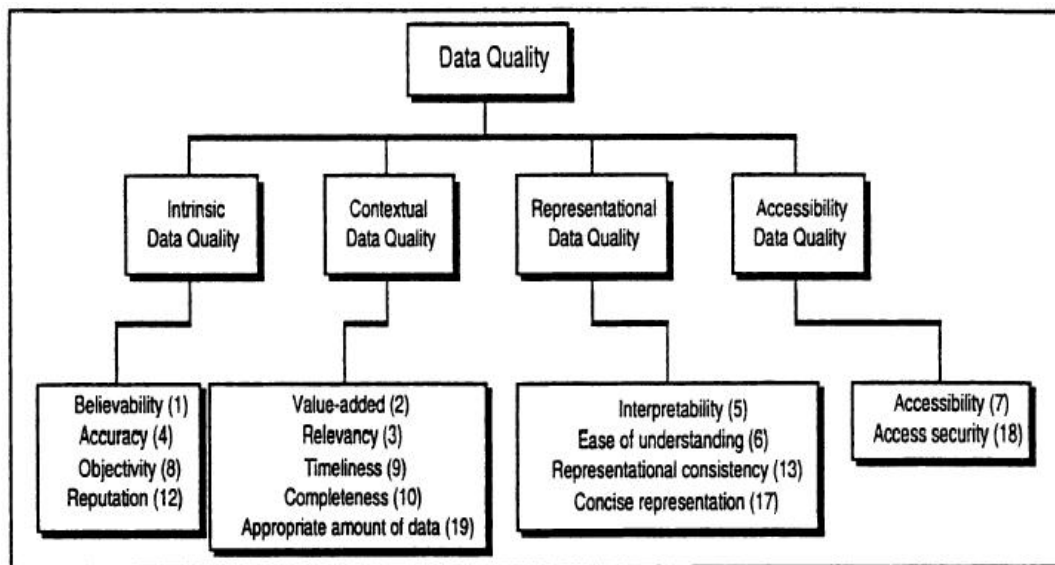


Figure 2 – A conceptual framework of data quality (WANG; STRONG, 1996)

Recent research by Sadiq, Yeganeh and Indulska (2011) to identify new studies on data quality presents a taxonomy of keywords that refer to data quality. From this study we selected “Data Quality Dimensions”, which are presented in Table 3.

Table 3 – Taxonomy of keywords referable to data quality dimensions (SADIQ, KHODABANDEHLOO and INDULSKA, 2011)

Data quality dimensions	Usability
	Information understandability
	Information relevance
	Fitness for use
	Data completeness
	Data reliability
	Data accuracy
	Information usefulness
	Data currency
	Data consistency
	Data uncertainty > probabilistic databases
	Data privacy

Data quality dimensions (or factors/attributes for some researchers) are especially important in an open initiative: government, data, archives, innovation, access, etc. In this paper, we note two dimensions: reliability and authenticity, focusing on a possible relationship between Open Government Data and digital records.

It is known that maintaining the quality attributes of digital records is a difficult task because

of IT upgrades, changes in business models and the fact that users' needs are dynamic and fast. However, digital records management founded on Archival Science could help government officials to cope with this scenario.

However, what about the quality of the data? Are they authentic, reliable, available or traceable? What are the sources of those data? What is the National Archives of Brazil's role in open data initiatives? The purpose of this research is an attempt to find the answers for these questions within a case study methodology.

4 Records management, Archival Science and data quality standing for Open Government Data

Archival Science theory can be defined as “a body of concepts and methods directed toward the study of records in terms of their documentary and functional relationships and the ways in which they are controlled and communicated” (Duranti and MacNeil, 1996, p. 47).

At public or private offices, records must be controlled for good governance, accountability and transparency of the business process, especially within government agencies. Thus, records management based on Archival Science theory is important because it defines processes for the production, use, maintenance, classification, appraisal, disposal, preservation and accessibility of records, particularly in a digital environment.

It is preponderantly records managers' and IT professionals' awareness of the importance of digital records and their values--administrative, informative, probative and historical—as is the awareness of their context (legal, administrative, provenance, proceedings, documentary, technological) because without the analysis of the context records management becomes impaired (Rondinelli, 2013, p. 239; Duranti, 1994, p.59).

The concern about how to coordinate activities to plan, control, make available and retrieve records, information or data is a point shared in Archival Science and Computer Science/Information Systems essays. Since the 20th century we can conclude that record and data quality themes have also been an issue for interdisciplinary studies.

The Brazilian archivist Rondinelli (2004; 2013) defines records qualities as having authenticity, naturalness, inter-relationships, uniqueness and impartiality. The “Creator Guidelines”⁶, a publication from InterPARES Project (2010), also defines these aspects in records qualities: fixity (fixed form and stable content), bounded variability, reliability, identity and integrity constitutive of authenticity.

In Section 3, we introduced data qualities research developed by Wang and Strong (1996) and Sadiq et al (2011); in the discussion above we have introduced records qualities. Thus, it is possible to identify that some qualities of data and records are equivalent. If considering data as part of a record does it mean both data and records could be studied based on the same foundation? If yes, which one: Archival Science or Computer Science/Information

6 See at:
<[http://www.interpares.org/ip2/display_file.cfm?doc=ip2\(pub\)creator_guidelines_booklet.pdf](http://www.interpares.org/ip2/display_file.cfm?doc=ip2(pub)creator_guidelines_booklet.pdf)>.

System? We support the idea of an interdisciplinary knowledge construction.

For an interdisciplinary knowledge construction, an extensive effort should be made by Archival Science and Computer Science/Information Systems scholars, especially when the debate is around vocabulary, terms, definitions and even jargon of each area (O'Toole, 2000, p. 475-476). One of the reason that this is necessary is the object of study: information records is a field for partnership but also for dispute (Marques, 2011).

Open Government Data is a positive change that allows archivists and records managers to work together with IT professionals. The more that partnership can be achieved, the more findings for better records management proceedings will be possible. In turn, good government data for use and reuse of Brazilian citizens will be available and accessible.

5 Methodology

For this research, the Ministry of Justice was selected as the test-bed because it is one of the Brazilian federal government agencies making government data available on its site. Furthermore, the Ministry is responsible for citizen, justice and public safety assurance and promotion. However, how are its goals being achieved? One possible answer could be Open Government Data, as presented in the Section 3 and Section 4.

It is remarkable that citizen, justice and public safety are essential components of civil society and that effective access to them is a human right (Cesar, 2002, p. 46), as is the access to records, information or data. Overall, the Ministry of Justice is a government agency and according to Brazilian freedom of information Act (Lei nº 12.527, November 18th, 2011), one of the items that must be available to citizens is the number of employees working at the Ministry, compared with public employees or outsourcing. The purpose is to control financial expenses involving staff.

The case study method chosen for this research is based on Robert K. Yin's (2004) studies. According to the author "[...] Compared to other methods, the strength of the case study method is its ability to examine, in-depth, a 'case' within its 'real-life' context" (p. 1). Another reason for this approach is the explanatory question for the essay's problem, which is how the Ministry of Justice maintain the authenticity and reliability qualities of Excel spreadsheets available for use and reuse on its website.

The research follows a conceptual analysis approach with a bibliographic references revision about data quality, digital records management in an Archival Science context and Open Government Data, comprising the theoretical basis. The main purpose is to identify which are the open government data qualities and compare them with the digital records' qualities, e.g., authenticity and reliability.

For the scope of this study, we selected 13 Excel spreadsheets available on the Ministry of Justice website at <<http://www.justica.gov.br/dados-abertos>>. These datasets refer to the outsourced employees who worked at the Ministry of Justice departments from December 2012 to December 2013.

Each of the thirteen spreadsheets was analyzed from the point of view of a data consumer, to assess whether the spreadsheets kept their authenticity and reliability qualities when downloaded to a personal computer. A first approach was to verify the records properties. A second approach was to distribute a questionnaire to the people responsible for Brazil's Government Open Data Project with the aim of discovering what the guidelines are for making the spreadsheets available. How does the Project IT staff maintain authenticity and reliability qualities of government data? Have they heard about Archival Science theory?

In an attempt to express the similarities between open government data quality and digital records qualities, we constructed a conceptual framework and, based on the framework, a benchmarking tool for maintaining authenticity and reliability qualities of government data regardless of ICTs.

6 Conclusion

The democracy in a country in its different levels of government – national, regional or local – is sufficient when civil society is acting directly upon the governments' actions, which is possible with the government opening up its data. Developed countries and emerging ones are looking forward to achieving this goal in developing their policies, projects, ICT background and crew qualification.

The hypothesis we wish to advance in this paper is that digital data without appropriate qualities could cause damage to government business processes, resulting in harm and misunderstanding to the civil society, which in turn means a broken democracy. Looking to avoid this scenario, Archival Science researchers have been developing studies about digital records management. Whether considering data as part of a record, records qualities also should support maintaining good data quality, such as reliability and authenticity.

Considering that the research is underway, we expect to obtain the answers for the queries in this paper and contribution for the Archival Science scenario.

For the Brazilian Archival Science scenario, research about the relationship between open data initiatives and digital records management is in its beginning stages. Furthermore, records managers and archivists should be prepared to cope with data issues. Thus, a scientific contribution of this research to Brazilian Archival Science is to identify the similarities between records qualities and open government data qualities, and moreover to identify how to maintain the authenticity and reliability of data, regardless of ICTs.

Bibliography

- BENSON, Peter (2009). "ISO 8000: a new international standard for data quality". En: *Data quality PRO*. <<http://dataqualitypro.com/data-quality-pro-blog/iso-8000-new-international-standard-data-quality>>. [Last accessed: 25/6/2014].
- CESAR, Alexandre (2002). "Acesso à justiça e cidadania". Cuiabá: Ed. UFMT.
- DURANTI, Luciana (1994). "Registros documentais contemporâneos". En: *Revista Estudos Históricos*.

- DURANTI, Luciana; MACNEIL, Heather (1996). "The protection of the integrity of electronic records: an overview of the UBC-MAS research project". En: *Archivaria*, nº 42. <<http://journals.sfu.ca/archivar/index.php/archivaria/article/download/12153/13158>>. [Last accessed: 17/8/2014].
- InterPARES 2 Project (2010). *Creator guidelines: making and maintaining digital materials: guidelines for individuals*. School of Library, Archival and Information Studies University of British Columbia Vancouver, Canada. <[http://www.interpares.org/ip2/display_file.cfm?doc=ip2\(pub\)creator_guidelines_booklet.pdf](http://www.interpares.org/ip2/display_file.cfm?doc=ip2(pub)creator_guidelines_booklet.pdf)>. [Last accessed: 17/8/2014].
- MARQUES, Angelica Alves da Cunha (2011). *Interloquções entre a Arquivologia nacional e a internacional no delineamento da disciplina no Brasil*. Brasília: FCI/UnB. **Tese de Doutorado**. <<http://hdl.handle.net/10482/8730>>. [Last accessed: 17/8/2014].
- MARTINS, Daniel Barbosa (2009). *Extração personalizada e incremental de dados em ambientes de BI em tempo real*. Dissertação (Mestrado) - Programa de pós-graduação em Informática, Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro.
- OPEN GOVERNMENT DATA (2007). *Principles*. <https://public.resource.org/8_principles.html>. [Last accessed: 17/8/2014].
- O'TOOLE, J. M. (2000). "On the idea of permanence". En: JIMERSON, R. C. (edited by). *American archival studies: readings in theory and practice*. Chicago, USA: The Society of American Archivist, p. 475-494.
- RONDINELLI, Rosely Curi (2004). *Gerenciamento arquivístico de documentos eletrônicos: uma abordagem teórica da diplomática arquivística contemporânea*. 2 ed. Rio de Janeiro: Editora FGV.
- RONDINELLI, Rosely Curi (2013). *O documento arquivístico ante a realidade digital: uma revisitación conceitual necessária*. Rio de Janeiro: Editora FGV.
- SADIQ, Shazia et al (2011). "20 Years of Data Quality Research: Themes, Trends and Synergies". Australasian Database Conference. 22nd (ADC 2011), Perth, Australia, January 2011. En: *Conferences in Research and Practice in Information Technology (CRPIT)*, Vol. 115, Heng Tao Shen and Yanchun Zhang, Ed. <http://www.researchgate.net/publication/49515275_20_years_of_data_quality_research_Themes_trends_and_synergies/file/3deec5179ddf947a29.pdf>. [Last accessed: 26/5/2014].
- SHADBOLT, Nigel et al (2012). "Linked Open Government Data: lessons from Data.gov.uk". En: *IEEE Intelligent Systems*. <<http://www.computer.org/intelligent>>. [Last accessed: 17/8/2014].
- WANG, Richard Y et al (1996). "Beyond accuracy: What data quality means to data consumers". En: *Journal of Management Information Systems*. vol. 12, número 4 (Spring 1996), p.5-34. <http://courses.washington.edu/geog482/resource/14_Beyond_Accuracy.pdf>. [Last accessed: 7/4/2014].
- YIN, Robert K. (2004). *Case study methods*. Cosmos Corporation (Revised draft). January, 20th. <<http://www.cosmoscorp.com/Docs/AERAdraft.pdf>>. [Last accessed: 17/8/2014].