

Open Government Data in Brazil

A Systematic Review of its Uses and Issues

Emilio Feliciano de Oliveira

School of Technology - Pontifícia Universidade Católica do
Rio Grande do Sul
Porto Alegre, RS, Brazil
emilio.oliveira@live.com

Milene Selbach Silveira

School of Technology - Pontifícia Universidade Católica do
Rio Grande do Sul
Porto Alegre, RS, Brazil
milene.silveira@pucrs.br

ABSTRACT

Context: With the advancement of the internet and information systems, more and more data is available to the public. Governments around the world are trying to find the best way to present their citizens with relevant information and transparency of public actions, expenditures, and investments. But how is this being approached in Brazil, in the federal, state, and municipal spheres of government?

Objective: To find what initiatives are being conducted in Brazil, how open government data is being used, what types of data are most used in studies, and what are the challenges to implement open government data across the country through a systematic review of the literature. At last, based on the difficulties related to open government data, is suggested the first actions that should be taken to solve it.

Method: A search on the main repositories to find literature that pertains to Open Government Data initiatives in Brazil, followed by a systematic review and classification of said literature.

Results: There are not many research documents available that use open government data in the context of Brazil. Of those that exist, the main area that has studies on involve politicians and their expenses, and also education investments and results of said investments. Other subjects and uses of open government data are shown in this paper as a means to show that the interest in said data is spread across several areas.

Conclusion: Many steps need to be done before an effective use of the government open datasets can be made. In most studies analyzed, researchers needed to take a first step of treatment of the data available before effectively using it. This one difficulty, if solved, would make a great impact on the use of government data. The conclusion of this paper is that there is much ground work that needs to be done yet to give companies and researchers a footing to make discoveries from the datasets.

CCS CONCEPTS

• **General and reference** → **Surveys and overviews**; • **Information systems** → *Semantic web description languages*;

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

dg.o '18, May 30-June 1, 2018, Delft, Netherlands

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-6526-0/18/05...\$15.00

<https://doi.org/10.1145/3209281.3209339>

KEYWORDS

Brazil, Open Government Data, Open Data, Electronic Government, Open Government Partnership, Linked Open Data.

ACM Reference Format:

Emilio Feliciano de Oliveira and Milene Selbach Silveira. 2018. Open Government Data in Brazil A Systematic Review of its Uses and Issues. In *dg.o '18: dg.o 2018: Proceedings of the 19th Annual International Conference on Digital Government Research, May 30-June 1, 2018, Delft, Netherlands*. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3209281.3209339>

1 INTRODUCTION

Information is everywhere. In today's world, only one thing is more powerful than having information in our hands: having knowledge. To extract knowledge, one needs to make sense of all the information one receives, and from that extract reason and comprehension that allows her to make use of this information. With governments, it is no different. Governments have an overload of information accumulated, spread in many systems, that usually do not talk to each other and do not save information in the same way. This makes the use of information to create knowledge one of the greatest challenges for public administrations nowadays to be able to effectively manage their assets and services.

Open data is the name given to datasets that are free to use by anyone, without restrictions. Datasets may refer to any type of data, like medical study results, an index of movie titles, image galleries, sound samples, etc. When the source of this open data is a government, we call this data Open Government Data. In Brazil, with the Information Access Law (Law 12.527/2011), government entities have been pressed to not only deal with all their data, but to release it for public use, following the precepts established on the letter of the law. This precepts are listed in section 4.1.

In this systematic literature review, the goal is to identify the state-of-the-art of initiatives to produce Open Government Data (OGD) in Brazil. Also, to determine what initiatives there are, this review aims to demonstrate the common challenges found when implementing OGD in Brazil, the difficulties found when implementing these strategies. By starting with a systematic review, the researchers intend to find what are the areas where they can act and deliver most impact to improve the usage of OGD in Brazil, be it in the consumption of the datasets to produce knowledge or in the construction and distribution of the datasets themselves.

The 3 spheres of government in Brazil - Federal, State and Municipal - will be taken into account, as there are different levels of commitment to an open government in each sphere, as well as diverse initiatives. Also considered and, when possible, clustered, are initiatives by the subject area of the datasets. For example, health

studies will be grouped and compared to other health studies. This way, it will be possible to identify areas that are not being studied, or that are studied in one sphere of government but not others, and through this, it will be possible to identify areas for improvement or investment.

2 CONTEXT OF THIS STUDY

This study is part of a larger research project called "Agregando valor ao cidadão: triangulação e visualização de dados abertos governamentais como apoio à tomada de decisão numa gestão resiliente"¹, which in a free-form translation means "Adding value to the citizen: Triangulation and visualization of open government data as a support to decision-making in a resilient administration", that focuses on how to integrate and use open government data and visualization tools in order to help public officials on their decision-making processes and to anticipate events that might disrupt the day-to-day activities of their citizens. Although the scope of the larger research study is not limited to Brazil, the focus of this systematic review is, as all the researchers involved in the larger research project are located in Brazil and the end result of this study is more prone to cause an impact to society in a local scale.

2.1 Brazil

In order to understand open government initiatives in Brazil, some introduction to how the country is structured is needed. Brazil is one of the largest countries in the world, with a federalist structure of government. As such, the government is divided in 3 spheres (federal, state and municipal), each responsible for basic services to the population, including education, public safety, health care, among others. Each sphere is then divided in 3 independent powers: executive, legislative and judiciary.

That being so, is common in a Brazilian city, especially the state capitals, to have services provided by all three spheres of government. For example, having federal, state and municipal schools in the same city. Although at first sight this might seem as just a separation of where the funding comes from for each service, this represents a challenge for collecting data on each of the services. Returning to our example of the school system, to get all the data on funding of education in one city, 3 different datasets (one for each sphere of government) will need to be combined, each with different formats, data, availability, etc. Since every sphere is granted a great level of independence, each of them is responsible for the publishing (or not) of their data. There are national data that is collected from the states, which on their turn collected it from the municipalities, but that results in most of the datasets being aggregates of that and not raw data (more on why this is important in section 4.1).

3 RESEARCH METHOD

A systematic review of the literature was performed for the study presented here. The use of a systematic approach, as described in Kitchenham [25], allows for future reproduction of this study, and assures that the same criteria was used independently of the publisher the article reviewed came from. Aside from that, with a well defined methodology, it is less likely that a bias from the author

ends up tarnishing the final result of the study. Since, as it will be shown in the next sessions, there's not that many studies, it was decided to use a hybrid approach to the evaluation of the results. It will be mostly a qualitative evaluation of the studies returned in the research, but some quantitative data will also be presented, as they can show a beginning of trend being set (although the sampling is too scarce to be statistically relevant).

3.1 Research Questions

Based on the focus of this systematic review being which open government data initiatives for use and publication of Brazilian sets of data, the following generic research question was defined:

- RQ - What is the state-of-art of Open Government Data research in Brazil?

Based on this question, 3 other more specific question were derived. They are:

- RQ1 - What types of Open Government Data from Brazil have researchers worked on (E.g. Health, Government Expenditure, etc)?
- RQ2 - What studies can be found in each sphere of the Brazilian Government?
- RQ3 - What challenges the implementation of Open Government Data initiatives in Brazil face?

3.2 Search Strategy

The following databases were used for this review. They were chosen for being the most significant repositories for studies that involve Computer Science, but not limited to:

- ACM Digital Library
- IEEE Xplore Digital Library
- Science Direct
- Springer Link

Scopus was considered to be utilized as well, but since its results were also returned in the other databases, especially Science Direct, it was not included. From the research questions, the following terms where utilized in the search of the databases:

- Brazil
- Open Government Data
- Open Government
- Open Data
- e-gov
- Electronic Government
- Linked Open Data

This last item, linked open data, was added after the first results showed that this was a term that was gaining traction with open government data. Although adding it did not return many significant studies, it helped in bringing more technical ones to be evaluated. The results of the original search can be seen in Table 1.

To set the research strings for each database, it was determined that, due to the scope of the project, Brazil was a mandatory term (via use of an "AND" operator with the other terms). All other terms where combined using an "OR" operator and encapsulated in parenthesis. Two of the databases required a little extra effort. For ACM, it was necessary to combine Brazil with each of the other terms individually, and then combine the results in a single result

¹Edital BPA/2017 N° 05/2017 (PRAIAS), PUCRS

Table 1: Initial search results

Database	Number of Articles Returned	Oldest	Newest
ACM Digital Library	110	2008	2017
IEEE Xplore	76	1994	2017
Science Direct	471	1995	2017
Springer	371	1985	2017

Table 2: Research queries per database

Database	Research Query
ACM Digital Library	(+brazil +"open government data") (+brazil +"open data") (+brazil +"government data") (+brazil +"electronic government") (+brazil +"e-gov") (+brazil +"open linked data")
IEEE Xplore	("open data" OR "open government" OR "open government data" OR "government data" OR "e-gov" OR "electronic government" OR "open linked data") AND (Brazil)
Science Direct	brazil AND open data OR open government OR open government data OR government data OR e gov OR electronic government OR open linked data
Springer	("open data" OR "open government" OR "open government data" OR "electronic government" OR "e-gov" OR "open linked data") AND Brazil

set, removing duplicates. That is because for some reason, the use of the OR operator was being ignored, which cause the search to return hundreds of results. After this compromise of combining 6 searches, the ACM search resulted in a much more realistic set of 110 results. The other database that required some tweaking was the Science Direct database. For this database, the removal of the hyphen in e-gov was necessary for the research string to work. The research queries used are shown in Table 2.

The search was done in the full articles, and not only on the titles or abstracts of the papers. This is because this is a fairly recent area of research, and the terminology for it still requires some consolidation.

3.3 Study Selection

For this study, as stated before, the focus was on not only Open Government Data from Brazil, but also on what types of study, data and difficulties are being found by researches. For that matter, the following criteria was used to select which studies would be considered.

- Inclusion Criteria
 - Use of open data that comes from a government in Brazil or government agency.
 - Data that is easily accessible by anyone.
 - Data that is reusable.
 - Technological implementations of open government data in Brazil.
 - Strategies to make open government data in Brazil available.

- Strategies to stimulate the creation of tools and applications that make use of said data.

- Exclusion Criteria
 - Studies that only marginally mention Brazil, citing it as an example.
 - Studies from events that happened in Brazil but that don't talk about open government data from Brazil.
 - Studies that evaluate accessibility or usability only.
 - Studies that use data from Brazilian satellites but are not from Brazil.
 - Studies that talk about electronic government initiatives, but not the data involved.
 - Papers that are part of a larger study, where the larger paper was selected instead.
 - Papers that make use of geographic reference information made available by the Brazilian government, without associating it with other data.
 - Studies that are focused on public participation only.

Based on this criteria, 2 reviews of the articles were made to decide which ones would effectively be selected. On the first passing, the focus was on finding the search terms and on a fast scan of the articles to see if the article was about open government data and Brazil. This resulted in a large number of articles excluded, and others marked either as approved or in need of review, as shown in Table 3. On the second passing, the articles were read in detail, which determined the final number of approved ones, and resolving all the articles marked for review. On this second pass, the articles approved were also classified as to the type of study,

the sphere of government the data came from, and the type of data. This classification will be explained in section 5. The final table of approved articles is shown in Table 4.

3.4 Overview of Included Studies

A quantitative analysis of the 40 studies that made into this review reveals that the majority of the studies focused on federal datasets, 28 of them, to be precise. Of those 28, 7 of those also have state datasets (8 studies with state datasets in total), and 5 have municipal datasets too (16 studies with municipal datasets in total). One of the reasons for this is that federal datasets are easier to find. Most of them are available at the open data website of the central government of Brazil². Also, federal data is firmly required to adhere to the precepts of open data as stated in the Brazilian Information Access Law and on the Open Government Partnership. Table 5 describes the articles included in this study.

4 TERMINOLOGY

In this section, an explanation of the terminology used in open government data is given, as a way to facilitate the understanding of the results that follow it.

4.1 Open Government Data

By open government data, this review uses the definition stated by the 8 principles of open government data³. As shown in Corrêa [13], these principles can be almost in all of its totality mapped to what the Brazilian Information Access Law states. The 8 principles are as follows:

- **Complete:** Meaning that all public data is made available.
- **Primary:** Data is made available as it was collected, without transformation.
- **Timely:** Data is made available as quickly as possible.
- **Accessible:** Data is made available to the maximum of users possible.
- **Machine Processable:** Data should be formatted so a machine can process it.
- **Non-discriminatory:** Anyone can access the data, with no need to identify himself.
- **Non-proprietary:** The format the data is made available is open.
- **License-free:** The data is not subject to copyright, patent, trademark or trade secret, but minimal privacy and security privileges are observed.

4.2 Linked data

Linked data⁴ is a set of best practices on how to publish data, so that data that comes from different sources can be hyper-linked, through the use of URIs, forming, as time passes, a single global graph of all linked data. Observing this principle when modelling data for publication, makes following up on the data easier and machine friendly, allowing the development of data crawlers and machine learning. Many linked datasets were created openly to enable the standardization and interoperability of data among datasets [36].

²<http://www.dados.gov.br>

³<https://opengovdata.org/>

⁴<https://www.w3.org/wiki/LinkedData>

For instance, the GeoNames dataset, which contains geographical information on all countries of the world⁵.

4.3 Ontology

Gruber [23] defines an ontology as a mechanism to describe the objects and their relationships, which form the vocabulary that formalizes the representation of knowledge. In open data, ontologies are used to describe the universe of discourse for a determined area. For instance, an ontology for the federal budget as presented in Araújo [5], where in fine granularity, every expense predicted in the Brazilian annual budget is described. Information such as the department who made the acquisition request, to what end, who made the request, are all part of the ontology defined, making it easy to identify each expense.

4.4 Open government partnership

The Open Government Partnership⁶ is a partnership signed by various countries in which the countries commit to promote transparency and fight corruption through the use of open data technologies. Countries that decide to join the Open Government Partnership must commit to have an Open Government, and present an action plan to make true on their commitment. On its part, the Open Government Partnership provides support and guidance to the countries to help them complete their actual plan. Brazil became a member of the Open Government Partnership in 2011 and the country is committed first and foremost with public transparency and with securing the publication of open government data, as noted in Breitman [9].

4.5 Technologies and Standards Used

In regards to the technologies used to make open government data happen, two deserve to be looked into more detail, as they seem to be the technological future of said implementations.

- **RDF:** Also known as Resource Description Framework, it is a standard defined by the World Wide Web Consortium⁷, and that is a model for data interchange on the web. It allows data from different sources and with different schemas to be combined, by using URIs and triples. It is the technical basis for implementing Linked Data.
- **SparQL:** Is a query language used to extract information from RDF structures. With it, it is possible to query different data sources, being them in RDF natively or viewed as RDF through the use of a middleware⁸.

5 RESULTS

In this section, the results of the research will be presented in two manners, looking at what data each article makes use of, both in terms of subject and sphere of government. With this, RQ1 and RQ2 will be answered. RQ3 will be analyzed in the next section.

⁵<http://www.geonames.org>

⁶<https://www.opengovpartnership.org/>

⁷<https://www.w3.org/RDF/>

⁸<https://www.w3.org/2001/sw/wiki/SPARQL>

Table 3: First screening

Database	Number of Articles	Approved Articles	Discarded Articles	Articles Marked for Review
ACM Digital Library	110	21	86	3
IEEE Xplore	76	13	58	5
Science Direct	471	8	458	5
Springer	371	14	357	0
Total	1028	56	959	13

Table 4: Second screening

Database	Number of Articles	Approved Articles	Discarded Articles	Articles Marked for Review
ACM Digital Library	110	15	95	0
IEEE Xplore	76	11	65	0
Science Direct	471	6	465	0
Springer	371	8	363	0
Total	1028	40	988	0

5.1 Transparency

Transparency is one of the most common subject area to appear in this research. This seems to be a reflex of the recent turmoils on the Brazilian political scenario. In Dos Santos Brito [17] [20] [19], the authors show how their implementation of an application called "Meu Congresso Nacional" positively influenced the public during the 2014 election cycle in Brazil. Through a survey with users, they found that their implementation of the open government data on the candidates available was perceived as more reliable than official government channels, and a good resource for knowing who was running in the election. In Sandoval-Almazan [37], they propose a model to compare the transparency portals of Mexico and Brazil. Through studies of the legal framework, open data, collaboration, co-production and institutional arrangements made by each country, they evaluated several transparency portals, concluding that both countries are on a similar level of transparency, although not a good level.

Corrêa [13] in his work compares the data present in transparency portals for 20 municipalities with the precepts for open government data present in the Information Access Law. The results are staggering, as most of the municipalities are far from compliant with the law. In accordance to this study, another study by Coelho [12] analyses how transparency compares through each sphere of government in Brazil. It clearly shows that municipalities have the lower level of transparency, followed by states, and by the central government, who is the most transparent. It also shows an interesting statistic, that transparency and social inequality seems to have a correlation. The more inequality in a municipality, the lower it scored in transparency. Lastly on this subject, Matheus [27] presents a study of the 27 state court of accounts in Brazil, and proposes tools to help the courts combat corruption, discussing how an increase in public transparency allows for more mechanisms to fight.

Although not directly referring to transparency, another study can be added here: Ghedini [22] proposes a multi-agent data mining system to combat corruption through the identification of patterns that reveal the formation of cartels in government procurement processes. With their proposed implementation, they were able to reach a 90% accuracy rate in cartel identification, proving that their approach can become an important tool for improving transparency.

5.2 Ontology

Four studies talk of constructing the ontologies that will serve as the basis for the open government initiative. The aforementioned Araújo [5], where the elaboration of the ontology for the public budget is described. That study has a counterpart in Da Silva [15], where to the ontology being developed was added an extension to put geolocation information on the budget.

With a focus on the software behind open data, two papers, one by Tosin [41] and another by Monteiro [30], discuss how to implement data integration and data retrieval in linked datasets using ontologies to facilitate the process. The latter use the technological solution presented to also propose an ontology to determine social vulnerability indexes.

5.3 Crime

In De Melo [16] the use of open government data is made to evaluate if patterns can be determined in criminal occurrences. It shows in a very clear way that for one type of crime (passerby robbery) there's an axis of the city where they occur, around one of the main streets. Using the law of crime concentrations at places developed for the north american reality, they found that it could be applied to a Brazilian reality just as well.

Table 5: Studies used in this review

Reference	Authors	Publication	Year	Sphere of government	Type of data
Albano, 2015 [1]	Cláudio Sonáglio Albano, Leonardo Bidese de Pinho	ICEGOV	2016	Federal	Education expenditure
Alves, 2016 [2]	Ronaldo Fernandes Santos Alves, Eduardo Faerstein	International Journal for Equity in Health	2016	Federal	Health
Andre, 2016 [3]	Luiz André P. Paes Leme, Chiara Renso, et al	WISE	2016	Municipal (Rio de Janeiro, RJ)	Route calculation
Andrews, 2013 [4]	Pierre Andrews, Flávio Soares Corrêa da Silva	ICEGOV	2013	Federal, State (SP), Municipal (São Paulo, SP)	Law data
Araújo, 2015 [5]	Luís Sérgio de Oliveira Araújo, Mauro Tapajós Santos, et al	MEDES	2015	Federal	Budget ontology
Attard, 2015 [6]	Judie Attard, Fabrizio Orlandi, et al	Government Information Quarterly	2015	Federal	Transparency
Azevedo, 2015 [7]	Patricia Carolina Neves Azevedo, Guilherme Sousa Bastos, et al	GISTAM	2015	Federal, State (MG)	Flood data
Azevedo, 2016 [8]	Patricia Carolina Neves Azevedo, Vitor Afonso Pinto, et al	GISTAM	2016	Federal, State (MG)	Flood data
Breitman, 2012 [9]	Karin Breitman, Percy Salas, et al	IEEE Intelligent Systems	2012	Federal	Multiple datasets
Buregio, 2015 [10]	Vanilson Burégio, Kellyton Brito, et al	WWW	2015	Federal	Multiple datasets
Cacho, 2016 [11]	Nelio Cacho, Frederico Lopes, et al	ISC2	2016	Municipal (Natal, RN)	Multiple datasets
Coelho, 2016a [12]	Taiane Ritta Coelho, Thomaz Anderson Barbosa da Silva, et al	HICSS	2016	Federal, State e Municipal	Transparency
Corrêa, 2014 [13]	Andreiwid Sheffer Corrêa, Pedro Luiz Pizzigatti Corrêa, et al	dg.o	2014	Municipal (20 cities in SP)	Transparency
Craveiro, 2014 [14]	Gisele S. Craveiro, Jose P. Alcazar, et al	Lecture Notes in Computer Science	2014	Municipal (São Paulo, SP)	Official Gazzette
Da Silva, 2014 [15]	Daniel A. da Silva, Rafael Timóteo de Sousa Júnior, et al	Iberian CISTI	2014	Federal	Health and Education
De Melo, 2015 [16]	Silas Nogueira de Melo, Lindon Fonseca Matias, et al	Applied Geography	2015	Municipal (Campinas, SP)	Criminality
Dos Santos Brito, 2014 [17]	Kellyton dos Santos Brito, Marcos Antônio da Silva Costa, et al	COMPSAC	2014	Federal, Municipal (Recife, PE)	Multiple datasets
Dos Santos Brito, 2014a [18]	Kellyton dos Santos Brito, Marcos Antônio da Silva Costa, et al	dg.o	2014	Municipal (Rio de Janeiro, RJ, Recife, PE)	Health
Dos Santos Brito, 2015 [19]	Kellyton dos Santos Brito, Marcos Antônio da Silva Costa, et al	dg.o	2015	Federal	Transparency
Dos Santos Brito, 2014b [20]	Kellyton dos Santos Brito, Misaél dos Santos Neto, et al	dg.o	2014	Federal	Multiple datasets
Gama, 2017 [21]	Kiev Gama	CSI-SE	2017	Municipal (Recife, PE, Olinda, PE)	Multiple datasets
Ghedini, 2012 [22]	Célia Ghedini Ralha, Carlos Vinicius Sarmiento Silva	Expert Systems with Applications	2012	Federal	Crimes in acquisitions
Heaton, 2016 [24]	Lorna Heaton, Patrícia Dias da Silva	IFIP	2016	Federal	Biodiversity
Matheus, 2012a [26]	Ricardo Matheus, Manuella M. Ribeiro, et al	ICEGOV	2012	Federal, State (SP), Municipal (São Paulo, SP)	Multiple datasets
Matheus, 2012 [27]	Ricardo Matheus, Manuella M. Ribeiro, et al	ICEGOV	2012	State (MG, PA, PB, PI)	Transparency
Matheus, 2014 [28]	Ricardo Matheus, Manuella M. Ribeiro, et al	ICEGOV	2014	Municipal (Rio de Janeiro, RJ)	Multiple datasets
Michener, 2015 [29]	Gregory Michener	World Development	2015	Federal	Multiple datasets
Monteiro, 2012 [30]	Adriana Costa Monteiro, Luis Enrique Zárate Gálvez	CLEI	2012	Federal	Social vulnerability
Oliveira, 2016 [31]	Marcelo Iury S. Oliveira, Hélio Rodrigues de Oliveira, et al	dg.o	2016	Federal, State (6 states), Municipal (6 capitals)	Multiple datasets
Pena, 2015 [32]	Karen Isabel Cabrera Peña	RUSC	2015	Federal	Multiple datasets
Penteadó, 2016 [33]	Bruno Elias Penteadó	WWW	2016	Municipal (All)	Education
Pereira, 2016 [34]	Gabriela Viale Pereira, Marie Anne Macadar, et al	Information Systems Frontiers	2016	Municipal (Rio de Janeiro, RJ)	Socioeconomic Info
Pinheiro, 2016 [35]	Emerson B. Pinheiro, Emanuel F. Coutinho, et al	EATIS	2016	Federal	Car fleet
Ribeiro, 2013 [36]	Cristiano E. Ribeiro, Adriana S. Vivacqua	ICSC	2013	Municipal (Rio de Janeiro, RJ)	Meteorological
Sandoval-Almazan, 2013 [37]	Rodrigo Sandoval-Almazan, Fabro Steibel	ICEGOV	2013	Federal	Transparency
Schimit, 2014 [38]	P.H.T. Schimit, L.H.A. Monteiro, et al	CNSNS	2014	Federal	Socioeconomic Info
Shkabatur, 2016 [39]	Jennifer Shkabatur, Alon Peled	CeDEM	2016	Federal	Multiple datasets
Terra, 2017 [40]	Rafael Terra, Enlison Mattos	Journal of Urban Economics	2017	Federal, Municipal	Education expenditure
Tosin, 2017 [41]	Thyago Tosin, Sandro J. Rigo, et al	SCCC	2017	Federal	Acquisition
Ueti, 2016 [42]	Roberto da Mota Ueti, Daniela Fernandez Espinosa, et al	Big Data Applications and Use Cases	2016	Federal, State (GO, SP, RJ)	Multiple datasets

5.4 Biodiversity

One of the more interesting uses of open government data returned in this literature review is presented in Heaton [24]. This study discusses the use of open government data for an international collaboration to create a global index of biodiversity. To try and remove the hurdles in implementing this strategy, a partnership with Canadian researchers was established, which brought its own difficulties along.

5.5 Other Subjects Worth Noting

In this section are shown other subjects that although not research in extent in the papers selected in the systematic review, are still worthy of mentioning:

- **Health:** In Dos Santos Brito [18], the authors use public information to make a mobile app that shows the nearest health units to the user.
- **Education:** Penteadó [33] makes an interesting study, correlating what the education indexes for a city are, with what is the schools performance, making use of linked open data.
- **Meteorologic:** Two studies by the same author, Azevedo [7] [8], show how to use linked data, RDF and SparQL to treat flood information and use it in an application.

Table 6: Studies by type of study

Type of study	Total articles
Initiative analysis	6
Data analysis	5
Applications	5
Website evaluation	9
Implementation strategy	5
Technical solution description	6
Other	4

6 CHALLENGES

The usage of open government data in the studies selected for the systematic review raised many challenges. The majority of them have to do with the quality and the format of the datasets. In this section, they are presented. Although they may not be the main subject of the studies included in this review, these are the problems faced by the researchers and the areas where they needed to spend time and effort on just to be able to deliver the research they first intended. If not for those, more researches, and deeper researches could have been conducted.

6.1 Data availability and timeliness

Many of the datasets, although they have been publicly published, they are not updated in a timely fashion [6]. In some cases, it has been observed that for data to be published, it needed a catalyst factor. Either an inspection from the Open Government Partnership, or a meeting going to happen in the country, were observed to trigger a flood of new data being published, and that after these facts, the rhythm of publishing diminished to a crawl.

6.2 Data set quality

Many of the datasets that are available present one of the following issues to their quality [31]:

- **Data is aggregated:** This is a very common issue. A good deal of the datasets available from public agents in Brazil are not the raw data, but only aggregates that follow a logic used by the governing body. With this, many of the insights that could be retrieved from the data are lost.
- **Data has a gatekeeper:** Even though one of the tenets of open government data says that data has to be publicly available, many datasets still demand that a request is made to an government agency who will determine if the data can or cannot be released, and if released, if all of the data or just part of it will be made available, as described in Ghedini [22].

6.3 Data access and format

Aside from the quality of the datasets themselves, another issue commonly associated with open government data is the issue of proprietary data formats [31]. By that we mean that many of the datasets are only available as PDF files, or as HTML pages. These formats are far from being easily scannable by a machine to retrieve the data, making the use of the data very laborious and prone to error (since it will need much more human interaction to clean up the data).

Also, it was observed that in none of the studies was mentioned any data repository that allowed the retrieval of the data via web services or any other newer technology. Data always need to be downloaded and fed to the researchers system before being used. There is simply no way of doing on demand retrieval of the data.

6.4 Lack of standardization

This is perhaps the biggest issue of all. Since there is no standardization between the different data sources, it is very difficult to combine different sources for a study. As reported in most of the papers selected, many of the researchers invested a good deal of time and effort just adjusting the datasets so they could relate to each other [17]. This is a much bigger and difficult issue to tackle, either because there are no public ontologies for the subject that one is trying to correlate, or because the data is collected from several different information systems, that are owned by different spheres of government, each with its own secretariats, departments, agencies, etc.

7 CONCLUSION

As we can see from this study, open government data initiatives in Brazil are still on its infancy. Considering that the Information Access law has been signed only in 2011, the same year Brazil joined the Open Government Partnership, it is understandable that there are still very few studies on the subject, as mentioned in Attard [6]. Interesting to know is that, although on the research made on the databases there were papers returned that dated all the way back to 1994, the papers approved after the inclusion and exclusion criteria were applied only date as far back as 2012. This shows that, although some research was made on this subjects before, only after the Information Access law was signed the subject of

open government data started to be formally studied, as shown in Breitman [9].

From the studies as well, we could see two researchers that are on the forefront of this area in Brazil: Kellyton dos Santos Brito, who focusses in civic hackatons and in how to stimulate a larger public to make use of open government data in their applications, and Ricardo Matheus, who focuses more on studies of transparency and of what investments are happening in states and municipalities in Brazil. The one systematic review included in this study, Attard [6], was included here for the purpose of ratification of this affirmation. This paper by Attard is a systematic review of open government data initiatives, without focusing solely in Brazil, and the authors cited here were the two Brazilian authors that also shown up on Attards review.

Observing Table 6, we can see that most studies in Brazil for open government data are still initial studies and tests with the data sets that are starting to become available. Most of the studies refer to website evaluations, where the easiness and options available to retrieve the data are the focus of the study.

Following this type of study, comes initiative analysis, that analyze what is to be done in the future, and technical solution descriptions, that can also be seen as Minimum Valuable Products to test if the data being made available can be used and how that can be done.

Considering the amount of issues with the data sets reported on the studies, we can conclude that improving the data, before investing in solutions that make use of it is a good strategy that should be observed by the three spheres of government in Brazil.

Lastly, and as an example of this kind of investment, I would like to appraise the study done by Araújo [5]. This is a study conducted only by civil servants that had the need to come up with a standardization to be able to conduct their jobs, and that were kind enough to produce an academic paper with their findings. What makes this a rare example, is that no university was involved in the study, and having a study publish that does not comes from academia is a rare feat in Brazil.

ACKNOWLEDGEMENT

This systematic review of the literature was achieved in cooperation with Hewlett Packard Brasil LTDA. using incentives of Brazilian Informatics Law (Law nº 8.2.48 of 1991).

Also, we would like to thank professor Sabrina Marczak for all the guidance on how to author a systematic review of the literature.

REFERENCES

- [1] Cláudio Sonáglio Albano and Leonardo Bidese De Pinho. 2015. Publishing Data in Open Format : Formalizing the Process in a Brazilian Federal University. *International Conference on Theory and Practice of Electronic Governance (ICEGOV)* (2015), 382–383. <https://doi.org/10.1145/2910019.2910059>
- [2] Ronaldo Fernandes Santos Alves and Eduardo Faerstein. 2016. Educational inequalities in hypertension: complex patterns in intersections with gender and race in Brazil. *International Journal for Equity in Health* 15, 1 (2016), 146. <https://doi.org/10.1186/s12939-016-0441-6>
- [3] Luiz Andre, P Paes Leme B, Chiara Renso, Bernardo P Nunes, Giseli Rabello Lopes, and Marco A Casanova. 2016. Searching for Data Sources for the Semantic Enrichment of Trajectories. *19th International Conference on Web Information Systems Engineering (WISE)* 10042 (2016), 238–246. <https://doi.org/10.1007/978-3-319-48743-4>
- [4] Pierre Andrews and Flávio Soares Corrêa da Silva. 2013. Using parliamentary open data to improve participation. *International Conference on Theory and Practice of Electronic Governance (ICEGOV)* (2013), 242–249. <https://doi.org/10.1145/2591888.2591933>
- [5] Luis Sérgio Oliveira Araújo, Mauro Tapajós Santos, and Daniel Aguiar Silva. 2015. The Brazilian Federal Budget Ontology – A Semantic Web Case of Public Open Data. *7th International Conference on Management of computational and collective intelligence in Digital EcoSystems* October (2015), 85–89. <https://doi.org/10.1145/2857218.2857232>
- [6] Judie Attard, Fabrizio Orlandi, Simon Scerri, and S?ren Auer. 2015. A systematic review of open government data initiatives. *Government Information Quarterly* 32, 4 (2015). <https://doi.org/10.1016/j.giq.2015.07.006>
- [7] Patricia Carolina Neves Azevedo, Guilherme Sousa Bastos, and Fernando Silva Parreiras. 2015. A linked open data approach for visualizing flood information a case study of the Rio Doce Basin in Brazil. In *Geographical Information Systems Theory, Applications and Management. First International Conference (GISTAM)*. 227–232.
- [8] Patricia Carolina Neves Azevedo, Vitor Afonso Pinto, Guilherme Sousa Bastos, and Fernando Silva Parreiras. 2016. Using Linked Open Data in Geographical Information Systems. *Geographical Information Systems Theory, Applications and Management. First International Conference (GISTAM)* 582 (2016), 22–35. <https://doi.org/10.1007/978-3-319-29589-3>
- [9] Karin Breitman, José Viterbo, Percy Salas, Daniel Saraiva, Regis Pires Magalhães, Vinicius Gama, Marco Antonio Casanova, Miriam Chaves, and Ednylton Franzosi. 2012. Open government data in Brazil. *IEEE Intelligent Systems* 27, 3 (2012), 45–49. <https://doi.org/10.1109/MIS.2012.25>
- [10] Vanilson Burégio, Kellyton Brito, Nelson Rosa, Misael Neto, Vinicius Garcia, and Silvio Meira. 2015. Towards Government as a Social Machine. *24th International World Wide Web Conference* (2015), 1131–1136. <https://doi.org/10.1145/2740908.2743976>
- [11] Nelio Cacho, Frederico Lopes, Everton Cavalcante, and Irani Santos. 2016. A smart city initiative: The case of Natal. *IEEE International Smart Cities Conference (ISC2)* (2016), 1–7. <https://doi.org/10.1109/ISC2.2016.7580774>
- [12] Taiane Ritta Coelho, Thomaz Anderson Barbosa Da Silva, Maria Alexandra Cunha, and Marco Antonio Carvalho Teixeira. 2016. Uncovering governmental transparency in federative states: Diverse government spheres, heterogeneous outcomes. *49th Hawaii International Conference on System Sciences (HICSS)* 2016-March (2016), 2809–2818. <https://doi.org/10.1109/HICSS.2016.352>
- [13] Andreiweid Sheffer Corrêa, Pedro Luiz Pizzigatti Corrêa, and Flávio Soares Corrêa da Silva. 2014. Transparency Portals Versus Open Government Data. An Assessment of Openness in Brazilian Municipalities. *dg.o '14* (2014), 178–185. <https://doi.org/10.1145/2612733.2612760>
- [14] Gisele Craveiro, Jose Alcazar, and Andres Martano. 2014. Collaborative Construction of an Open Official Gazette. In *Transactions on Large-Scale Data- and Knowledge-Centered Systems XXVII*. Vol. 8800. 111–128. https://doi.org/10.1007/978-3-662-53416-8_7
- [15] Daniel A. Da Silva, Rafael Timoteo De Sousa, Valterio A. Martins, Evandro Nogueira Exposto, Fabio L. Mendonca, and Carlos E L Veiga. 2014. Extension of the Brazilian Federal Government Budget Ontology to Support the Representation of Geolocated Human Development Indicators. *9th Iberian Conference on Information Systems and Technologies (CISTI)* (2014). <https://doi.org/10.1109/CISTI.2014.6876938>
- [16] Silas Nogueira de Melo, Lindon Fonseca Matias, and Martin A. Andresen. 2015. Crime concentrations and similarities in spatial crime patterns in a Brazilian context. *Applied Geography* 62 (2015), 314–324. <https://doi.org/10.1016/j.apgeog.2015.05.012>
- [17] Kellyton Dos Santos Brito, Marcos Antonio Silva Costa, Vinicius Cardoso Garcia, and Silvio Romero De Lemos Meira. 2014. Experiences Integrating Heterogeneous Government Open Data Sources to Deliver Services and Promote Transparency in Brazil. *38th Annual Computer Software and Applications Conference (COMPSAC)* 475743 (2014), 606–607. <https://doi.org/10.1109/COMPSAC.2014.87>
- [18] Kellyton dos Santos Brito, Marcos Antônio da Silva Costa, Vinicius Cardoso Garcia, and Silvio Romero de Lemos Meira. 2014. Brazilian Government Open Data: Implementation, Challenges, and Potential Opportunities. *dg.o '14* (2014), 11–16. <https://doi.org/10.1145/2612733.2612770>
- [19] Kellyton dos Santos Brito, Marcos Antônio da Silva Costa, Vinicius Cardoso Garcia, and Silvio Romero de Lemos Meira. 2015. Assessing the Benefits of Open Government Data: The Case of Meu Congresso Nacional in Brazilian Elections 2014. *dg.o 2015* (2015), 89–96. <https://doi.org/10.1145/2757401.2757422>
- [20] Kellyton dos Santos Brito, Misael dos Santos Neto, Marcos Antônio da Silva Costa, Vinicius Cardoso Garcia, and Silvio Romero de Lemos Meira. 2014. Using Parliamentary Brazilian Open Data to Improve Transparency and Public Participation in Brazil. *dg.o 2014* (2014), 171–177. <https://doi.org/10.1145/2612733.2612769>
- [21] Kiev Gama. 2017. Preliminary findings on Software Engineering Practices in Civic Hackathons. *IEEE/ACM 4th International Workshop on CrowdSourcing in Software Engineering (CSI-SE)* (2017). <https://doi.org/10.1109/CSI-SE.2017.5>
- [22] Célia Ghedini Ralha and Carlos Vinicius Sarmento Silva. 2012. A multi-agent data mining system for cartel detection in Brazilian government procurement. *Expert Systems with Applications* 39 (2012), 11642–11656. <https://doi.org/10.1016/j.eswa.2012.04.037>

- [23] Thomas R. Gruber. 1995. Toward principles for the design of ontologies used for knowledge sharing? *International Journal of Human-Computer Studies* 43, 5-6 (1995), 907–928. <https://doi.org/10.1006/ijhc.1995.1081> arXiv:0701907v3
- [24] Lorna Heaton and Patricia Dias da Silva. 2016. Cultures of Science and Technology in the Trading Zone: Biodiversity and Open Source Development. *IFIP International Federation for Information Processing* 490 (2016), 19–31. <https://doi.org/10.1007/978-3-319-50109-3>
- [25] Barbara Kitchenham and Stuart Charters. 2007. Guidelines for performing Systematic Literature reviews in Software Engineering Version 2.3. *28th International Conference on Software Engineering (ICSE)* 45, 4ve (2007), 1051. <https://doi.org/10.1145/1134285.1134500>
- [26] Ricardo Matheus, Manuella M Ribeiro, and José Carlos Vaz. 2012. New Perspectives for Electronic Government in Brazil: The Adoption of Open Government Data in National and Subnational Governments of Brazil. *International Conference on Theory and Practice of Electronic Governance (ICEGOV)* (2012), 22–29. <https://doi.org/10.1145/2463728.2463734>
- [27] Ricardo Matheus, Manuella M Ribeiro, José Carlos Vaz, and Cesar A de Souza. 2012. Anti-corruption Online Monitoring Systems in Brazil. *International Conference on Theory and Practice of Electronic Governance (ICEGOV)* (2012), 419–425. <https://doi.org/10.1145/2463728.2463809>
- [28] Ricardo Matheus, José Carlos Vaz, and Manuella Maia Ribeiro. 2014. Open government data and the data usage for improvement of public services in the Rio de Janeiro City. *International Conference on Theory and Practice of Electronic Governance (ICEGOV)* (2014), 338–341. <https://doi.org/10.1145/2691195.2691240>
- [29] Gregory Michener. 2015. Policy Evaluation via Composite Indexes: Qualitative Lessons from International Transparency Policy Indexes. *World Development* 74 (2015), 184–196. <https://doi.org/10.1016/j.worlddev.2015.04.016>
- [30] Adriana Costa Monteiro and Luis Enrique Zárate Galvez. 2012. Prospects and limitations in the context of knowledge discovery in database for manipulation of domains through ontologies to support the modeling of data warehouse -Case study in social databases. *XXXVIII Conferencia Latinoamericana En Informatica (CLEI)* (2012), 0–8. <https://doi.org/10.1109/CLEI.2012.6427213>
- [31] Marcelo Iury S. Oliveira, Hélio Rodrigues de Oliveira, Lairson Alencar Oliveira, and Bernadette Farias Lóscio. 2016. Open Government Data Portals Analysis: The Brazilian Case. *dg.o '16* (2016), 415–424. <https://doi.org/10.1145/2912160.2912163>
- [32] Karen Isabel Cabrera Peña. 2015. Comparative analysis of public policies in open access models in Latin America. Brazil and Argentina cases. *International Journal of Educational Technology in Higher Education* 12, 1 (2015), 15–24. <https://doi.org/10.7238/rusc.v12i1.1947>
- [33] Bruno Elias Penteadó. 2016. Correlational Analysis Between School Performance and Municipal Indicators in Brazil Supported by Linked Open data. *25th International World Wide Web Conference* (2016), 507–512.
- [34] Gabriela Viale Pereira, Marie Anne Macadar, Edimara M. Luciano, and Mauricio Gregianin Testa. 2016. Delivering public value through open government data initiatives in a Smart City context. In *Information Systems Frontiers*. 1–17. <https://doi.org/10.1007/s10796-016-9673-7>
- [35] Emerson B. Pinheiro, Emanuel F. Coutinho, Leonardo O. Moreira, Carla I. M. Bezerra, and Gabriel A. L. Paillard. [n. d.]. ([n. d.]).
- [36] Cristiano E. Ribeiro and Adriana S. Vivacqua. 2013. A framework for composition and reuse on the linked open data. *IEEE Seventh International Conference on Semantic Computing (ICSC)* (2013), 17–24. <https://doi.org/10.1109/ICSC.2013.14>
- [37] Rodrigo Sandoval-Almazan and Fabro Steibel. 2013. Benchmarking Mexico & Brazil open government websites: model and metrics. *International Conference on Theory and Practice of Electronic Governance (ICEGOV)* 2013, October (2013), 372–373. <https://doi.org/10.1145/2591888.2591965>
- [38] Pedro Henrique Triguís Schimit, Luiz Henrique Alves Monteiro, and Nizam Omar. 2014. Cash transfer program and education investment: A model for social evolution. In *Communications in Nonlinear Science and Numerical Simulation*. Vol. 19. 570–577. <https://doi.org/10.1016/j.cnsns.2013.07.018>
- [39] Jennifer Shkabatur and Alon Peled. 2016. Sustaining the Open Government Data Movement Worldwide: Insights from Developing Countries. *Conference for E-Democracy and Open Government (CeDEM)* (2016), 172–179. <https://doi.org/10.1109/CeDEM.2016.10>
- [40] Rafael Terra and Enlison Mattos. 2017. Accountability and yardstick competition in the public provision of education. *Journal of Urban Economics* 99 (2017), 15–30. <https://doi.org/10.1016/j.jue.2016.12.001>
- [41] Thyago Tosin, Sandro J. Rigo, Jorge L.V. Barbosa, and Clarissa Rodrigues. 2017. A model for data integration in open and linked databases with the use of ontologies. *35th International Conference of the Chilean Computer Science Society (SCCC)* Part F1262 (2017). <https://doi.org/10.1109/SCCC.2016.7835994>
- [42] Mota Ueti, Daniela Fernandez Espinosa, Laura Rafferty, and Patrick C K Hung. 2016. Case Studies of Government Use of Big Data in Latin America: Brazil and Mexico. In *International Series on Computer Entertainment and Media Technology*. 197–214. <https://doi.org/10.1007/978-3-319-30146-4>