

**MEASURING OPEN DATA'S IMPACT OF  
BRAZILIAN NATIONAL AND SUB-NATIONAL  
BUDGET TRANSPARENCY WEBSITES  
AND ITS IMPACTS ON PEOPLE'S RIGHTS**

# MEASURING OPEN DATA'S IMPACT OF BRAZILIAN NATIONAL AND SUB-NATIONAL BUDGET TRANSPARENCY WEBSITES AND ITS IMPACTS ON PEOPLE'S RIGHTS

1<sup>st</sup> Edition

Nathalie Beghin and Carmela Zigoni

(Editors)

Brasília, 2014

**Research Coordinator**

Nathalie Beghin – Inesc

**Researchers**

Carmela Zigoni – Inesc

Gisele Craveiro – GPOPAI/USP

Marcelo Tavares – GPOPAI/USP

**Editors**

Nathalie Beghin – Inesc

Carmela Zigoni – Inesc

**Translate and review**

Luiz Marcos B. L. de Vasconcelos and Master Language Traduções e Interpretação Ltda.

**Graphic Project**

Lia Magalhães Graça Silva

The funding for this work has been provided through the World Wide Web Foundation ‘Exploring the Emerging Impacts of Open Data in Developing Countries’ research project, supported by grant 107075 from Canada’s International Development Research Centre ([web.idrc.ca](http://web.idrc.ca)). Find out more at [www.opendataresearch.org/emergingimpacts](http://www.opendataresearch.org/emergingimpacts).

---

Measuring open data’s impact of Brazilian national and sub-national budget transparency websites and its impacts on people’s rights. Brasília, Inesc: 2014.

117 pages

ISBN 978-85-87386-30-4

---

## **INESC's Team**

### **Board**

Adriana de Carvalho B. Ramos Barreto  
Caetano Ernesto Pereira Araújo  
Guacira Cesar de Oliveira  
Márcia Anita Sprandel  
Sérgio Haddad

### **Financial Council**

Silvia Ramos  
Armando Martinho Bardou Raggio  
Iliana Alves Canoff

### **Collegiate Management**

Iara Pietricovsky de Oliveira  
José Antonio Moroni

### **Head of Policy**

Nathalie Beghin

### **Financial and Human Resources Manager**

Maria Lúcia Jaime

### **Assistant to the Collegiate Management**

Ana Paula Felipe

### **Policy Advisers**

Alessandra Cardoso  
Carmela Zigoni  
Cleomar Manhas  
Márcia Hora Acioli

### **Administrative Assistants**

Adalberto Vieira dos Santos  
Eugênia Christina Alves Santana  
Isabela Mara dos Santos da Silva

### **Assistant for General Services**

Josemar Vieira dos Santos

### **Accounting Assistants**

Miria Thereza Brandão Consíglío  
Ricardo Santana da Silva

### **Interns**

Ana Júlia Barros Farias Zacks

### **Hired Services:**

#### **Accounting**

LC Mangueira Contabilidade  
(Rosa Diná Gomes Ferreira, accountant in charge)

#### **Media**

Empresa Vértice G  
Gisliene Hesse (journalist in charge)

#### **IT Support**

Leal Tecnologia  
(Leandro Rodrigues)

### **Institutional Support:**

Action Aid, Charles Stewart Mott Foundation, Christian Aid, Embaixada do Reino dos Países Baixos, Fastenopfer, Fundação Avina, Fundação Ford, International Budget Partnership (IBP), Instituto C&A, Instituto Heinrich Böll, KNH - Kinder Not Hilfe (KNH), Norwegian Church Aid, Oxfam, Pão para o Mundo: Serviço Protestante para o Desenvolvimento, Institute for Research in economics and Business Administration (SNF), Unicef e World Wide Web Foundation.

This is a publication of INESC — Institute for Socioeconomic Studies. All opinions expressed herein do not necessarily reflect the opinion of the ODDC network, the World Wide Web Foundation or Canada's International Development Research Centre. Its contents may be freely used, provided that the source is acknowledged. For accessing the databases of the quantitative research in an open format, kindly contact [inesc@inesc.org.br](mailto:inesc@inesc.org.br).

Inesc — Institute for Socioeconomic Studies  
SCS, QD 01 Bloco L, Nº 17, cobertura, Ed. Márcia,  
Brasília/DF – CEP: 70.307-900  
E-mail: [inesc.org.br](mailto:inesc.org.br) – Phone: +55 (61) 3212-0200;  
[www.inesc.org.br](http://www.inesc.org.br)



# TABLE OF CONTENTS

List of Acronyms .....	8
Executive Summary	
An Unprecedented Research in Brazil .....	9
A lot of additional efforts must still be made for budget information to be actually available in open data format .....	12
1. Introduction .....	17
2. Quantitative research	
Open data and public budget: assessing the performance of governments .....	28
2.1 Objectives .....	28
2.2 Methodology .....	29
2.3 Results .....	35
2.3.1. The information provided is complete .....	36
2.3.2. In general, the data are not <i>primary</i> .....	38
2.3.3. Most governments do not provide <i>timely</i> information .....	41
2.3.4. Access to information is not ensured to all individuals .....	43
2.3.5. Advances were observed in terms of the availability of <i>machine-processable data</i> .....	44

2.3.6. Access is free, <i>without discriminating</i> users .....	46
2.3.7. The minority presents data in non-proprietary format .....	47
2.3.8. It is not known whether the data are under <i>license</i> ...	48
2.4. Ranking of the Websites .....	48
2.5. Conclusions .....	50
3. Qualitative research	
Open data and public budget: impacts on the perception of intermediaries .....	53
3.1. Objectives .....	53
3.2. Methodology .....	55
3.2.1. Interview Guide.....	58
3.2.2. Mapping of Respondents .....	59
3.2.3. Profile of the respondents, institutions and organizations .....	61
3.3. Results .....	66
3.3.1. What is open data? .....	66
3.3.2. How are open data accessed? .....	74
3.3.3. What is done with open data that are accessed? .....	76
3.3.4. What does data <i>use</i> and <i>reuse</i> mean? .....	79

3.3.5. What is the impact of open data? .....	80
3.4. Conclusions .....	85
4. Final considerations .....	88
5. References .....	91
6. Digital references .....	94
Annex 1.	
Questionnaire — quantitative research .....	97
Annex 2.	
Interview guide — qualitative research .....	105

## LIST OF ACRONYMS

**CGU** Office of the Comptroller General

**CFEMEA** Feminist Center for Studies and Advisory

**CSV** Comma Separated Values

**DGA** Dados Governamentais Abertos

**eMAG** Accessibility Model of Electronic Government

**GPOPAI/USP** Group on Public Policies for Access to Information of the University of São Paulo

**IBP** International Budget Partnership

**INESC** Institute for Socioeconomic Studies

**IPEA** Institute for Applied Economic Research

**ITGP-M** Municipal Public Management Transparency Index

**LAI** Access to Information Act

**OBI** Open Budget Index

**ODDC** Open Data for Developing Countries

**OFAT** Federal Budget Within Everyone's Reach

**OGD** Open Government Data

**OGP** Open Government Partnership

**OKF Brasil** Brazil Open Knowledge Foundation

**PDF** Portable Document Format

**SIAFI** Integrated Financial Administration System of the Federal Government

**SISTN** Collection System of Accounting Data

**SIOP** Integrated Planning and Budget System

**SOF** Federal Budget Secretariat

**UFRPE** Federal Rural University of Pernambuco

**UFPE** Federal University of Pernambuco



# EXECUTIVE SUMMARY

## An Unprecedented Research in Brazil<sup>1</sup>

Having to access to budget information is critical for ensuring transparency in the public sector, thus enhancing its effectiveness and accountability. Recently, a movement was launched for data to be open, that is, freely available, in a timely fashion and to any citizen. This is a new process that is still under way and whose features vary according to the country where it is taking place.

In Brazil, the availability of data on public management has increased since the Access to Information Act (LAI, in its Portuguese acronym)<sup>2</sup> was passed in 2011, which governs the procedures to be followed by the federated entities to ensure access to information to society.

---

<sup>1</sup> We would like to express our special thanks to Tim Davis, from the ODDC team, for his invaluable contributions and key comments throughout the survey. We would also like to thank all those who granted us interviews and shared their experiences in the field of open data in Brazil.

<sup>2</sup> Law n°. 12,527 of November 18, 2011.

For nearly two decades, INESC has been actively engaged in public budget reviews based on the *tax justice* concept and on the development of the *Budget and Rights*<sup>3</sup> methodology. It monitors and decodes federal budgeting processes and their financial execution, carrying out technical analyses of public spending in general but also of specific aspects related to the so-called “thematic budgets”, i.e. budgets focused on socio-environmental issues, racial equality, food and nutrition security, children and adolescents, indigenous peoples, and public safety.

The organization only began to work on the specific topic of open data more recently. In 2011, Inesc published the research *Budget Transparency in Brazilian Capitals*<sup>4</sup>, based on the methodology adopted by the International Budget Partnership (IBP). In 2013, it launched the platform called *O Orçamento ao Seu Alcance*<sup>5</sup> (Budget at your Fingertips) in partnership with the Brazil Open Knowledge Foundation (OKF Brazil), which was prepared using open source tools such as that of OpenSpending.org and open data of the Brazilian federal budget, making it possible for a broader and more diverse number of users to have greater access to them with simpler visualization. Inesc also participates in the Open

---

<sup>3</sup> <http://www.inesc.org.br/biblioteca/publicacoes/textos/cartilhas-e-manuais-1/orcamento-e-direitos-construindo-um-metodo-de-analise-do-orcamento-a-luz-dos-direitos-humanos>.

<sup>4</sup> Barbosa, Lucídio (coordinator). *Pesquisa Inesc: Transparência Orçamentária nas Capitais do Brasil*. 2011. In: Inesc’s Digital Library: <http://www.inesc.org.br/noticias/biblioteca/textos/Transparencia%20Orcamentaria%20-%20Original>

<sup>5</sup> Web page of the tool: <http://orcamento.inesc.org.br/2013>

Government Partnership (OGP) initiative by taking part in its steering committee as representative of global civil society.

In this context, Inesc undertook to conduct the research *Measuring open data's impact of Brazilian national and sub-national budget transparency websites and its impacts on people's rights* carried out under the Open Data for Developing Countries (ODDC) initiative of the World Wide Web Foundation (WWWF).

This is an exploratory research designed to contribute to increase knowledge about open data in Brazil. The research is divided into two stages, namely, a quantitative and a qualitative stage. The quantitative research was designed to measure the scope of rules imposed by new Brazilian legislation for budget transparency in open data format. To this end, 27 websites of capitals<sup>6</sup>, one site of the federal government and one site of the Senate were analyzed. In addition, a ranking of these 29 national and subnational entities was developed. The qualitative research was designed to complement the data collected during the quantitative survey with the aim of capturing possible impacts of open data on expanding human rights in Brazil based on the perception of more qualified users on open data access and analysis, the so-called “intermediaries.”

---

<sup>6</sup> The Brazilian Republic is a Federation composed of Central Government; 26 States; a Federal District; and 5.570 Municipalities. The present research analyzes the transparency websites of 26 state capitals and Brasilia, the Federal capital, located in the Federal District.

The research as a whole, in its two stages, was guided by the Eight Open Data Principles, as defined in the Open Data Handbook<sup>7</sup>, for developing the tools that were used for collecting information. With the aim of deepening the reflection, a partnership was established with the Research Group on Public Policies for Access to Information of the University of São Paulo (Gpopai-Usp).

**A lot of additional efforts must still be made for budget information to be actually available in open data format.**

The quantitative research showed that the information provided is *complete*, that the data *are not primary* usually, that most governments *do not* provide *timely* information, that *access* to information *is not* guaranteed for all, that progress has been made in relation to the availability of *machine-processable* data, that access to them *does not discriminate* users, that the minority of websites presents data in non-proprietary format and that it *is not* known whether the data are under *license*. Table 1 shows the huge gap that exists between reality and legal requirements according to which budget information is to be published in open format. No capital city, and neither the Federal Government nor the Senate, got the highest score, which would be 10. The best positions were those of the municipalities of Rio de

---

<sup>7</sup> <http://opendatahandbook.org/>

Janeiro, Sao Luis and Joao Pessoa, although their scores were much below the desired ones. It should be mentioned that the performance of 13 capital cities was seen to be low, with scores of 2 or less - as they only met the requirements of providing complete and non-discriminatory data. It can be seen that the geographical region of a capital city does not explain its good performance, as opposed to what one could imagine. Economically better-off city halls, such as that of Sao Paulo, were given low scores. On the other hand, municipalities located in Brazil's poorest regions, such as those of Sao Luis and Teresina, are among the top-ranking ones.

The results of the qualitative survey confirm the findings of the quantitative research in relation to the need to improve the process of opening up data in Brazil. However, despite these limitations, the interviewed intermediaries believe that progress has been made in the field of human rights as a result of opening up data in the country. They also believe that access to data empowers civil society. Intermediaries reuse information in different ways (apps, analytical reports, media articles) through complex processes (hiring technicians, guidance to students, development of software for processing data, etc.) and carry out different dissemination activities (meetings with local social movements, local social media on the Web, press, meetings with government). They made a very positive assessment of the Access to Information Act, but they complained that the information they access is usually not *complete, timely and reliable*.

**Table 1. Ranking of Transparency Portals**

Position	Federative entity	TOTAL	Complete	Primary	Timely	Accessible	Machine-processable	Non-discriminatory	Non-proprietary	License-free
1st	City Hall of Rio de Janeiro	6	1	-	-	-	3	1	1	-
	City Hall of Sao Luis	6	1	-	-	1	2	1	1	-
	City Hall of Joao Pessoa	6	1	-	-	-	3	1	1	-
2nd	Office of the Comptroller General – Federal Executive	5	1	-	-	-	2	1	1	-
	City Hall of Teresina	5	1	-	-	-	2	1	1	-
	Federal Senate	5	1	-	-	-	2	1	1	-
3rd	City Hall of Palmas	4	1	-	1	-	1	1	-	-
	City Hall of Maceio	4	1	-	1	-	1	1	-	-
	City Hall of Curitiba	4	1	-	-	1	1	1	-	-
4th	City Hall of Salvador	3	1	-	-	-	1	1	-	-
	City Hall of Sao Paulo	3	1	-	-	-	1	1	-	-
	City Hall of Boa Vista	3	1	-	1	-	-	1	-	-
	City Hall of Recife	3	1	-	-	-	1	1	-	-
	City Hall of Belem	3	1	-	-	-	1	1	-	-
	City Hall of Florianopolis	3	1	-	1	-	-	1	-	-
	City Hall of Macapa	3	1	-	1	-	-	1	-	-
5th	City Hall of Fortaleza	2	1	-	-	-	-	1	-	-
	City Hall of Porto Alegre	2	1	-	-	-	-	1	-	-
	City Hall of Cuiaba	2	1	-	-	-	-	1	-	-
	Government of the Federal District	2	1	-	-	-	-	1	-	-
	City Hall of Vitoria	2	1	-	-	-	-	1	-	-
	City Hall of Goiania	2	1	-	-	-	-	1	-	-
	City Hall of Porto Velho	2	1	-	-	-	-	1	-	-
	City Hall of Belo Horizonte	2	1	-	-	-	-	1	-	-
	City Hall of Natal	2	1	-	-	-	-	1	-	-
	City Hall of Aracaju	2	1	-	-	-	-	1	-	-
	City Hall of Campo Grande	2	1	-	-	-	-	1	-	-
6th	City Hall of Rio Branco	2	1	-	-	-	-	1	-	-
	City Hall of Manaus	1	-	-	-	-	-	1	-	-

Source and Prepared by: Gopai/Usp, 2014.

A common understanding was observed that data “reuse” corresponds to more complex forms of processing data than just reading them, such as developing an app or writing an analytical text, among others.

Another result of the qualitative study was that of showing that the respondents were more aware of some principles than of others. The intermediaries see the data that they access as *complete*, *not very timely* and *non-discriminatory*. As regards whether the data are *primary* and *machine-processable*, the intermediaries only think that these principles are important when they use the data through software and apps - for those intermediaries who use them only for reading and writing purposes, these principles are of secondary importance. As for *accessibility*, it was not described as a problem, perhaps because there were no people with disabilities among the respondents. As for the *non-discriminatory* principle, this is not a relevant principle for the interviewed intermediaries, which reveals the weakness of the debate on data security and privacy and Internet regulation. The same can be said in relation to the *non-proprietary* principle, which is not an object of structured reflection on the part of the majority of the intermediaries either. Finally, with regard to the *license-free* principle, although the group of interviewed intermediaries included a few free-software activists, none of them knew whether the sources of the consulted open data used license-free digital structures - which coincides with the result of the quantitative research,

in which it was seen that the surveyed government portals simply disregard this information.

This research opens up avenues for further investigations, as the field of open data in Brazil is in full development and changes in access and impacts were seen to be rapid and complex. First, the ranking of the websites must be reproduced considering the same scoring criteria for the purpose of measuring whether or not governments made efforts to improve the openness of their data. With this in mind, the results presented here will be used in the stage of disseminating the results of this research to provoke public debate and advocacy, as there is a law that regulates the opening of data and there are spheres in charge of inspecting governmental structures (public prosecutors, ombudsman's offices, etc.). Second, based on the realization that users are not appropriately aware of important issues such as licensing, ownership and accessibility, as well as of difficulties to deal with complex databases, the need was observed to conduct studies designed to measure the extent to which Brazilian governments are actually promoting access to data and, consequently, social participation and control. Key research to be carried out would be one designed to map out initiatives and investment of government in technology development, organization of information and design to make their data more accessible and user-friendly, such as the establishment of interactive websites, contests for apps and training activities, among others.



# 1. INTRODUCTION

Having to access to budget information is critical for ensuring transparency in the public sector, thus enhancing its effectiveness and accountability<sup>8</sup>. Recently, a movement was launched for data to be open, that is, freely available, in a timely fashion and to any citizen. This is a new process that is still under way and whose features vary according to the country where it is taking place.

In Brazil, the availability of data on public management has increased since the Access to Information Act (LAI, in its Portuguese acronym)<sup>9</sup> was passed in 2011, which governs the procedures to be followed by the federated entities (Union, states and municipalities) to ensure access to information to society, as well as other legislative instruments that accompany it<sup>10</sup>.

---

<sup>8</sup> “Accountability” usually refers to fiscal responsibility.

<sup>9</sup> Law no. 12,527 of November 18, 2011.

<sup>10</sup> For example, Decree no. 7,724 of May 16, 2012, which “regulates Law no. 12,527 of November 18, 2011, which provides for access to information as set forth in section XXXIII of the caput of article 5, in section II of paragraph 3 of article 37 and in paragraph 2 of article 216 of the Constitution.” In other words, this decree regulates procedures

Data on public budgets are disseminated through digital tools and nowadays budget transparency websites are available at both the federal and state and municipal levels. At the federal level, in compliance with the LAI, public agencies gradually developed ways of giving access to data and today one can find web pages by budgetary unit (i.e. for each agency, meaning that each Ministry, Public Agency, State-Owned Company, etc. has its own website where open data can be accessed), but also tools that concentrate budget data from several units on a single portal, such as the *Siga Brasil* (Follow Brazil) Portal<sup>11</sup>. We must also consider that initiatives to “open up” data began a decade ago and the launching of the Transparency Portal in November 2004 was a landmark in this regard<sup>12</sup>. Therefore, even if we still don’t have an “open data culture” in Brazil, this is a process that has been going on for ten years now and its effects are already being felt in society.

---

for ensuring access to information and for classifying information under access restrictions, taking into account the degree and term of secrecy. Analytically, we can say that this decree was inspired by the Eight Open Data Principles defined by the Open Government Working Group (and adopted by Dataprev - Social Security Technology and Information Company, for example: <http://portal.dataprev.gov.br/tag/principios-de-dados-abertos/>).

<sup>11</sup> The *Siga Brasil* Portal is an initiative of the Federal Senate to make data on the Union budget available. In recent years, Inesc has been in charge of preparing the thematic budgets that are made available on this website (<http://www12.senado.gov.br/orcamento/tematico>).

<sup>12</sup> Regulated by Decree no. 5,482 of June 2005, which provides for the dissemination of data and information by agencies and entities of the Federal Administration through the World Wide Web - the Internet.

In Brazil, the main tools for making Union budget data available in concentrated form are the *Transparency Portal of the Federal Government*<sup>13</sup>, an initiative of the Office of the Comptroller General (CGU) launched in November 2004, which had a strong impact on Brazilian society for being the precursor of similar initiatives; the *Siga Brasil* portal<sup>14</sup>, an initiative of the Federal Senate, which is an information system on public budget through which the SIAFI<sup>15</sup> database can be accessed, as well as other databases of public plans and budgets using a single query tool; and the Federal Budget Within Everyone's Reach (*Orçamento Federal ao Alcance de Todos - OFAT*), of the Federal Budget Secretariat (SOF)<sup>16</sup>, which is a simplified document designed to improve accessibility to complex budget data<sup>17</sup>.

For nearly two decades, INESC has been actively engaged in public budget reviews based on the *tax justice* concept and on the development of the *Budget and*

---

<sup>13</sup> <http://www.portaltransparencia.gov.br>

<sup>14</sup> <http://www12.senado.gov.br/orcamento>

<sup>15</sup> Integrated Financial Administration System of the Federal Government, which is the main tool used for recording, monitoring and controlling the budgetary, financial and patrimonial execution of Brazil's Federal Administration.

<sup>16</sup> The Federal Budget Secretariat (SOF) is the agency in charge of coordinating and monitoring the budget policy. According to Article 17 of the Decree no. 7,675 of January 20, 2012, the Federal Budget Secretariat is in charge of: I - coordinating, consolidating and overseeing the drafting of the budget guidelines law and of the Union's budget bill, including the tax and social security budgets; II - establishing necessary rules for the preparation and implementation of federal budgets under its responsibility; III - without prejudice to the powers assigned to other agencies, monitoring the budget execution.

<sup>17</sup> <http://www.orcamentofederal.gov.br/educacao-orcamentaria/ofat/ofat>

*Rights*<sup>18</sup> methodology. It monitors and decodes federal budgeting processes and their financial execution, carrying out technical analyses of public spending in general but also of specific aspects related to the so-called “thematic budgets,” hosted in the *Siga Brasil* portal, i.e. budgets focused on socio-environmental issues, racial equality, indigenous peoples, policies for children and adolescents, public safety, food and nutrition security, among others<sup>19</sup>. Fiscal and budgetary issues are also key elements of the advocacy and knowledge formation and building actions carried out by the institution.

It should be mentioned that although Inesc has many years of expertise in monitoring and analyzing the budget and in carrying out advocacy actions for that purpose, its work in the specific area of open data is more recent. In 2011, Inesc published the research *Budget Transparency in Brazilian Capitals*<sup>20</sup>, based on the methodology adopted by the International Budget Partnership (IBP). In 2013, it launched the platform called *O Orçamento ao Seu Alcance* (Budget at your Fingertips)<sup>21</sup> in partnership with the Brazil Open Knowledge Foundation (OKF

<sup>18</sup> <http://www.inesc.org.br/biblioteca/publicacoes/textos/cartilhas-e-manuais-1/orcamento-e-direitos-construindo-um-metodo-de-analise-do-orcamento-a-luz-dos-direitos-humanos>.

<sup>19</sup> <http://www12.senado.gov.br/orcamento/tematico>.

<sup>20</sup> Barbosa, Lucidio (coordinators). *Inesc research: Budget Transparency in Brazilian Capitals*. 2011. In: Inesc Digital Library: <http://www.inesc.org.br/biblioteca/publicacoes/outras-publicacoes/publicacoes-em-ingles/budget-transparency-in-brazilian-capitals-2011>

<sup>21</sup> Web page of the tool: <http://orcamento.inesc.org.br/2013>.

Brazil), which was prepared using free structures such as that of OpenSpending.org and open data of the Brazilian federal budget, making it possible for a broader and more diverse number of users to have greater access to them with simpler visualization. Moreover, Inesc also participates in the Open Government Partnership (OGP) by taking part in its steering committee as representative of global civil society.

In this context, Inesc undertook to conduct the research *Measuring open data's impact of Brazilian national and sub-national budget transparency websites and its impacts on people's rights* carried out under the Open Data for Developing Countries (ODDC) initiative of the World Wide Web Foundation.

This is an exploratory research designed to contribute to increase knowledge about open data in Brazil and from an international comparative perspective. The research is divided into two stages, namely, a quantitative and a qualitative stage. The objective of the quantitative stage is to measure the scope of rules imposed by the new legislation based on an evaluation and production of a ranking of 27 transparency websites of capital cities and two similar federal websites. The qualitative stage was designed to complement the data collected in the quantitative stage with the aim of capturing the impacts that open data can have on expanding human rights in Brazil based on the perception of more qualified users on access to and analysis of open data. The research as a whole,

in its two stages, was guided by the Eight Open Data Principles, as defined in the Open Data Handbook, for developing the research tools, as will be explained in detail in the respective sections about the methodology.

With the aim of deepening the reflection, a partnership was established with the Research Group on Public Policies for Access to Information of the University of São Paulo (Gpopai-Usp). The GPOPAI is a research group recognized for its approach to the open data topic in its academic production and development of digital applications, together with training projects, with a focus on promoting citizenship, such as the *Cuidando do Meu Bairro* (Caring for my Neighborhood) tool<sup>22</sup>. For this purpose, the design of the quantitative stage was jointly developed with the Gpopai/Usp - *Grupo de Pesquisa em Políticas Públicas para o Acesso à Informação* (Working Group on Public Policies to Access Information from São Paulo University - Usp), so as to rely on its expertise in quantitative research into open data, and with Inesc researchers, to take advantage of their expertise in the budget and public transparency topics.

Considering the bibliographical survey carried out by Gpopai-Usp, the quantitative stage falls under the category of studies focused on the quality of the dissemination of government information, especially budget information, through Internet portals. Groff & Pitman (2004) studied the 100 largest U.S. counties.

---

<sup>22</sup> <http://www.gpopai.usp.br/cuidando/>.

Rodríguez & Bolívar (2006) proposed an index that evaluates financial information websites based on local Spanish public accounting standards.

Similar issues were observed in initiatives that studied how such information was disseminated at municipal level in Spain. Bolívar, Pérez & Hernández (2007) proposed an index for the dissemination of budget information and suggested an index to compare financial reports in hard copy and online, considering characteristics related to the quality of the information provided.

Another study (Pina, Torres & Royo, 2007) checked five local government websites (the website of the capital city and of the four next largest cities) of 15 European countries. This research included two groups of financial and non-financial information items: items related to financial reporting published on the Internet (performance, economic-financial, social and environmental information) and items related to transparency, interactivity, usability and maturity.

The municipal level was also studied by Jorge et al. (2011) in a research effort that combines issues related to budget and financial information to be disseminated with availability options, i.e. access/visibility, format and delivery method and proposed a dissemination index. The researchers evaluated a sample of 94 municipalities through institutional websites, 49 of which from Italy and 45 from Portugal.

In the Brazilian context, Cruz et al. (2012) studied the level of public administration information published on the websites of 96 municipalities included in

the list of the 100 most populous ones. They tried to determine which socio-economic characteristics and indicators of the municipalities can contribute to explain the observed level of transparency. The level of transparency in public administration was established based on a research model called Municipal Public Management Transparency Index (*Índice de Transparência de Gestão Pública Municipal* - ITGP-M) that was built based on international codes of good governance and transparency; Brazilian law; and experiences of similar previous studies carried out in Brazil and abroad.

In a context of availability of government data, the term Open Government Data (OGD) was born and became popular in early 2008 after a set of principles was published by activists in the U.S. in December 2007 (Craveiro, 2013). The term OGD denotes any data produced by public sectors for use by any individual for any purpose (Robinson et al., 2009), and the principles that ensure the right to information are the following ones<sup>23</sup>:

- 1) *Data Must Be Complete*: All public data are made available. Data are electronically stored information or recordings, including but not limited to documents, databases, transcripts, and audio/visual recordings. Public data are data that are not subject to valid privacy, security or privilege limitations, as governed by other statutes.

---

<sup>23</sup> Source: Laboratório Brasileiro de Cultura Digital & NICBr, 2011. .



- 2) *Data Must Be Primary*: Data are published as collected at the source, with the finest possible level of granularity, not in aggregate or modified forms.
- 3) *Data Must Be Timely*: Data are made available as quickly as necessary to preserve the value of the data.
- 4) *Data Must be Accessible*: Data are available to the widest range of users for the widest range of purposes.
- 5) *Data Must Be Machine-Processable*: Data are reasonably structured to allow automated processing of it.
- 6) *Access Must Be Non-Discriminatory*: Data are available to anyone, with no requirement of registration.
- 7) *Data Formats Must Be Non-Proprietary*: Data are available in a format over which no entity has exclusive control.
- 8) *Data Must Be License-Free*: Data are not subject to any copyright, patent, trademark or trade secret regulation. Reasonable privacy, security and privilege restrictions may be allowed as governed by other statutes.

Open data allow for increased reuse of budget data. However, one can see in the scientific literature on the evaluation of transparency portals and also in assessments carried out by civil society organizations, such as in the Open Budget Index (OBI), an IBP – International Budget Partnership initiative - conducted in various countries, that they have not fully, or in some cases not even partially,

addressed the intersection between making budget information available in open data.

Another point worth mentioning is that international assessments are often more focused on central government portals (in the case of Brazil, on federal portals) and it is a must to understand the heterogeneous realities of sub-national levels of government. Aware of this fact, this study developed a specific and unique methodology for producing a ranking of capital cities and two federal bodies in relation to their transparency websites that will be presented below.

The purpose of qualitative research is that of capturing, through narratives structured in a script, similarities and differences between agents in the contexts in which they act and their own knowledge and discourse background. Unlike quantitative methods, the qualitative method has the specific feature of being applicable to small groups of people and organizations without the main purpose of building representative statistical information for a given population. It should be noted, however, that although it is not statistically representative, the qualitative method provides a good example of a social group that one wants to know better, allowing for some generalizations to be made. The subjective dimension of the narrative is controlled for by means of a semi-structured interview script, apart from being delimited by the socio-cultural context shared by the respondents, by the fact that it addresses the open data topic, and by the

common characteristics (profile) of the members of the group to be surveyed. Because it can capture other levels of social reality, we believe that the qualitative study will be complementary to the quantitative research, allowing for new assumptions to be raised and addressed in future studies, as appropriate.

We emphasize that the open data topic is extremely important for Inesc, as it has an impact on the issue of human rights and citizenship. We believe that this first survey carried out under the Open Data for Developing Countries - ODDC initiative provides relevant results that can be immediately appropriated by the community of its stakeholders and be used as a basis for further comparative or deeper research into certain topics.

This report is structured as follows: initially, the quantitative research will be presented, as well as its objectives, methodology, results, especially the ranking of websites based on the principles, and a general sum. Next, the qualitative research will be presented, also considering its objectives, methodology and analytical results. Finally, concluding remarks will be made, a scoring of intersections between investigations will be presented, conclusive aspects will be mentioned and open issues for further investigations will be suggested.

## 2. QUANTITATIVE RESEARCH

### OPEN DATA AND PUBLIC BUDGET: ASSESSING THE PERFORMANCE OF GOVERNMENTS

#### 2.1 Objectives

The overall goal of building a “ranking of transparency in capital cities” was that of producing consistent information about open data and government transparency in Brazil based on an evaluation of the country’s main transparency portals, specifically the websites of the executive branch (Federal, Federal District and municipal executive branch in 26 capitals) and the website of the federal legislature (Senate). This evaluation is meant to equip society with useful tools for discussing the effectiveness of government actions to open up data, specifically budget data, as well as to measure the effectiveness of recent legislation on open data nationwide.

The research project also defined two specific objectives:

- Developing and conducting a survey designed to check the quality of budget data made available through open data government portals.
- Making a quantitative comparison based on the eight open data principles, which were interpreted in the specific context of budget data in the light of applicable codes of practice, regulations and laws, namely:

- ✓ Code of Good Practices on Fiscal Transparency (FMI, 2007);
- ✓ Constitution of the Federative Republic of Brazil;
- ✓ Complementary Law n°. 101/2000;
- ✓ Complementary Law n°. 131/2009;
- ✓ Law n°. 4,320/1964;
- ✓ Decree n°. 7,185/2010;
- ✓ Technical manual for the 2013 budget;
- ✓ Access to Information Act (Federal Law n°. 12,527/2011).

## 2.2 Methodology

The quantitative survey was intended to assess whether data transparency portals providing information on revenues and spending of the Federal Government, the Federal District Government and the governments of capital cities complied with the OGD principles systematized by Craveiro (2013). Basically, the methodology is based on the eight open data principles mentioned in the previous section.

For budget purposes, two large information sets are planned annually: revenues and spending. For a portal to comply with the first principle, “data must be complete,” it must present these two groups for the year 2013.

Both for revenue and spending, the Brazilian budget uses a numerical classification system. For example, according to Decree 7,185/2010, the levels to be published for spending are the following ones: budgetary unit (1), function (2), sub-function (3), nature of spending (4) and source of funds (5). In this numerical system, some numbers form a hierarchy where a sub-level is a breakdown of the immediately higher level, i.e. total function value (2) is the sum of all expenditures in sub-function (3). Thus, level 1 is the most aggregate level and level 3 is the most primary one. This logic is not repeated for nature of spending (4) and source of funds (5), as these are classification elements established for another purpose, such as that of identifying the origin of public funds for the spending in question. This classification provides the level of disaggregation of the information and, therefore, it allows for one to know to what extent a site is complying with the principle of being “primary.” Decree 7,185/2010 establishes three revenue information fields that must be reported to society and 12 spending fields, totaling 15 fields that must be reported and which were checked in this research.

For the third principle, “Data Must be Timely,” the date of the last revenue or expense published within the deadline set out in Decree 7,185/2010, i.e. one business day, was considered.

For the fourth principle, “Data Must be Accessible”, it was considered that the site should comply with the accessibility standards recommended for Brazilian

government websites, the so-called e-MAG<sup>24</sup>. This recommendation is based on other international models and classifies website development practices in “P1” actions (those that, if not taken, prevent a site from being accessed), “P2” actions (which do not prevent access to a site, but make it more difficult) and “P3” actions (which facilitate access to a site). For example, providing a text description for images is a P1 action, as without this description a screen reader program cannot read the meaning of an image for a visually impaired individual, undermining its understanding.

Possible technologies that could affect the “machine-processable” requirement, thus contradicting the fourth principle, such as the use of the CAPTCHA technology<sup>25</sup>, which depends on human intervention for accessing data, were considered. It should be noted here that the eight principles intersect and/or depend on each other to a certain extent, such as in the case of the two first principles, as one would not expect the data to be primary if they are not complete or that they are machine-processable if their format is not accessible.

The fifth principle, “Data Must be Machine-processable,” was addressed in three observations as it depends on the data being machine-processable, downloaded and available in non-proprietary formats for the information to be read

---

<sup>24</sup> <http://www.governoeletronico.gov.br/acoes-e-projetos/e-MAG>

<sup>25</sup> <http://pt.wikipedia.org/wiki/CAPTCHA>

and processed. It is also important to mention that, within the machine-processable category, this work will adopt levels included in the 5-star model proposed by Tim Berners-Lee (2010).

For the sixth principle, “Access Must Be Non-Discriminatory,” the need for any special access or registration was considered. When it was seen that no information had to be provided for accessing the site, it was deemed non-discriminatory. In this study, only the supply data in the version for the general public will be considered, not alternatives made available to experts upon registration.

The seventh principle, “Data Formats Must Be Non-Proprietary,” was addressed by observing some known format options (PDF, CSV, XLS, XML) and an open field for unforeseen cases.

The eighth principle, “Data Must Be License-Free,” was also addressed by searching for data use licenses and checking whether they provided for any restriction.

To check compliance of each site with the framework, a questionnaire was prepared (see appendix) with 20 questions for identifying the site and evaluating each principle. This identification includes email address, possible site redirection, address of the redirected site, state, region, name of the federative entity and type of federative entity.

To observe the first OGD principle, a multiple-choice question was answered to mark the publication of revenues and spending. Only websites that presented



**Table 1 summarizes explanations about our framework**

<b>Principle</b>	<b>Framework and ranking</b>
1. Data Must be Complete	Availability of information on income and spending for the year 2013 (0 or 1 score).
2. Data Must Be Primary	Observation of the categories and stages of revenues and spending (0 or 1 score; for the latter, the 15 information fields for revenues and spending provided for in the Decree 7,185/2010 must be published).
3. Data Must Be Timely	The date of the last update for 2013 should be 24 hours. (0 or 1 score)
4. Data Must be Accessible	There should be no P1 type bugs, which are those that prevent access to information, score 1; according to the e-MAG standards evaluated by the ASES automatic validation program for e-government (0 or 1 score).
5. Data Must Be Machine-processable	No score for PDF, score 1 for XLS, score 2 for CSV or XML. Score 3 if CSV and other more sophisticated alternatives are available, equivalent to 4 and 5 stars of the Tim Berners-Lee model.
6. Access Must Be Non-Discriminatory	It was checked whether any special access or registration was required (0 or 1 score).
7. Data Formats Must Be Non-Proprietary	PDF and XLS (no score) and CSV (score 1).
8. Data Must Be License-Free	It was checked whether there were licenses for using the data and whether they provide for any restriction (0 or 1 score).

Source and Prepared by: Gpopai/Usp, 2014.

the two budget execution groups were considered complete, regardless of the data being primary or not.

For the “primary” principle, two questions were answered, one for revenues and another one for spending, where the existence of each field of publication set out in Decree 7,185/2010, i.e. fifteen fields, was marked<sup>26</sup>. For data on a site to be considered primary, they should contain all the information provided for in the law.

To check the timeliness of the data, the date of the last published revenue or expenditure was taken note of, when possible, and it was checked

<sup>26</sup> Revenues: Anticipated, Entered and Collected; Spending: Committed, Settled, Paid, Budgetary Unit, Function, Sub-Function, Nature of Spending, Source of Funds, Type of Tender, Number of the Process, Beneficiary, Good/Service Provided.

whether it was within the deadline of one business day for publication, pursuant to Decree 7,185/2010.

To check the accessibility of the site, which is the fourth OGD principle for a user experience, the ASES program<sup>27</sup> provided by the Federal Government was used. It is an e-MAG recommendation evaluator for sites that identifies if there are any P1, P2 and P3 type bugs; for example, if an image has no description, it is considered that a P1 type bug exists. A “0” score was only given to sites with P1 type bugs. Besides the ASES, for the e-MAG the DaSilva online evaluator can also be used, but the ASES is the most complete program, according to the developers of the DaSilva evaluator themselves<sup>28</sup>.

The observation of the sample is based on the experience of users with the site, but to check compliance with the fifth principle it was verified whether data download in any file format is available. The quality of downloaded data was not checked, only the possibility of obtaining the data in other file formats.

In relation to the sixth principle, the need for users to identify themselves or to provide any personal data was checked. If users had to provide any personal information for data to be provided, a “0” score was given to the site.

---

<sup>27</sup> <http://www.governoeletronico.gov.br/aco-es-e-projetos/e-MAG/ases-avaliador-e-simulador-de-acessibilidade-sitios>

<sup>28</sup> <http://www.acessobrasil.org.br/software/>

For the seventh principle, the types of formats available for download in the user interface were checked. It was not checked whether there was any formatting error.

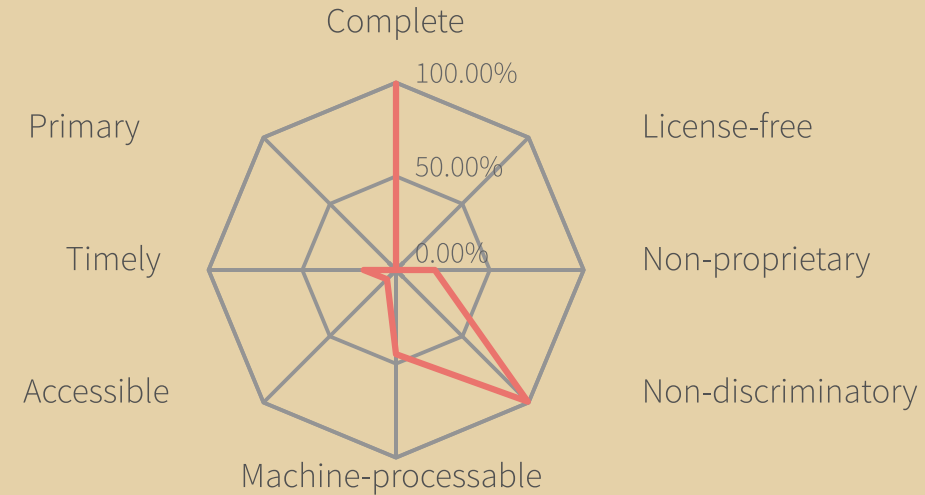
Despite being provided for in Brazilian law, to check whether data on budget execution could be used freely without a license, only the existence of an explicit statement of freedom to use the data in the evaluated sites was considered. If there was no such explicit statement, it was considered that the data were not license-free.

## 2.3 Results

The observations of the transparency pages of the sites evaluated were organized in a spreadsheet and scored in the light of the eight OGD principles as summarized in Table 1.

Figure 1 shows a percentage overview of compliance of the sites with the OGD Principles, with the external line corresponding to 100%. The figure shows that only the ‘data must be complete’ and ‘access must be non-discriminatory’ principles are being complied with by most transparency pages, while the other six principles are not being met by at least half of the transparency pages.

**Figure 1:** Compliance of the governmental sites with Open Government Data (OGD) Principles



Source and Prepared by: Gpapai/Usp, 2014.

To provide a more detailed overview of Brazilian reality, remarks and considerations about the scores given for compliance with each principle will be presented below. The order of portals included in each table is the same of the ranking presented at the end of this section.

### 2.3.1. The information provided is complete

Of the 29 sites that were evaluated, 28 provided information on revenues and spending (96.55%), except for the site of the city of Manaus (state of Amazonas), where no information about revenues was found.

None of the sites provided all the information set out in Decree 7,185/2010. For example, no information about revenues was found in the transparency website of the Federal Administration, hosted by the Office of the Comptroller General. None of the sites provided all the required information about revenues and only the sites of the cities of Salvador (state of Bahia) and Porto Alegre (state of Rio Grande do Sul) contained the twelve spending fields specified in the decree.

Based on the OGD Principles, pages with differences in the amount of information about revenues and spending had the same ranking. Table 2 provides a list of federative entities where those marked with a '✓' (tick) were seen to be in compliance with the 'data must be complete' principle. The 'revenues' and 'spending' columns contain the number of fields published in accordance with Decree 7,185/2010. For the 'complete' principle to be met, information about revenues and spending must be provided. The exception was the site of the City of Manaus, which did not provide data about revenues on its transparency page. Two cases, those of the municipalities of the cities of Belem and Natal, provided information about revenues, but not as provided for in the Decree 7,185/2010. However, because they did provide figures for revenues, it was considered that they were in compliance with the 'complete' principle.

Table 2. Evaluation of the 'Data Must be Complete' principle, revenues and spending			
Entity of the federation	Complete	Revenues (maximum score 3)	Spending (maximum 12)
City Hall of Rio de Janeiro	✓	2	11
City Hall of Sao Luis	T	1	10
City Hall of Joao Pessoa	✓	1	8
Office of the Comptroller General – Federal Executive	✓	2	11
City Hall of Teresina	✓	2	9
Federal Senate	✓	2	6
City Hall of Palmas	✓	2	9
City Hall of Maceio	✓	1	9
City Hall of Curitiba	✓	1	6
City Hall of Salvador	✓	2	12
City Hall of Sao Paulo	✓	2	11
City Hall of Boa Vista	✓	2	10
City Hall of Recife	✓	2	7
City Hall of Belem	✓	0	9
City Hall of Florianopolis	✓	1	8
City Hall of Macapa	✓	2	6
City Hall of Fortaleza	✓	2	11
City Hall of Porto Alegre	✓	1	12
City Hall of Cuiaba	✓	1	11
Government of the Federal District	✓	2	9
City Hall of Vitoria	✓	1	8
City Hall of Goiania	✓	1	8
City Hall of Porto Velho	✓	1	8
City Hall of Belo Horizonte	✓	2	6
City Hall of Natal	✓	0	7
City Hall of Aracaju	✓	1	6
City Hall of Campo Grande	✓	1	2
City Hall of Rio Branco	✓	1	2
City Hall of Manaus		0	9

Source and Prepared by: Gpopai/Usp, 2014.

### 2.3.2. In general, the data are not *primary*

Table 3 contains the revenue phases to be published in accordance with Decree 7,185/2010, namely, anticipated, entered and collected revenues. Publishing the three phases is a requirement to meet the 'primary' principle in relation to revenues. None of the transparency pages that were analyzed provided the three information items. Revenue collection is the information most often found on the pages and only the Office of the Comptroller General publishes entered revenues. The gaps in Table 3 suggest lack of observance of the law by federative entities as regards the requirement to publish revenue

Table 3. Primary nature of published revenues			
Entity of the federation	Anticipated	Entered	Collected
City Hall of Rio de Janeiro	✓		✓
City Hall of Sao Luis			✓
City Hall of Joao Pessoa			✓
Office of the Comptroller General – Federal Executive	✓	✓	
City Hall of Teresina	✓		✓
Federal Senate	✓		✓
City Hall of Palmas	✓		✓
City Hall of Maceio			✓
City Hall of Curitiba			✓
City Hall of Salvador	✓		✓
City Hall of Sao Paulo	✓		✓
City Hall of Boa Vista	✓		✓
City Hall of Recife	✓		✓
City Hall of Belem			
City Hall of Florianopolis			✓
City Hall of Macapa	✓		✓
City Hall of Fortaleza	✓		✓
City Hall of Porto Alegre	✓		
City Hall of Cuiaba			✓
Government of the Federal District	✓		✓
City Hall of Vitoria	✓		
City Hall of Goiania			✓
City Hall of Porto Velho			✓
City Hall of Belo Horizonte	✓		✓
City Hall of Natal			
City Hall of Aracaju			✓
City Hall of Campo Grande			✓
City Hall of Rio Branco	✓		
City Hall of Manaus			

Source and Prepared by: Gopai/Usp, 2014.

information, resulting in missing data for consumers who use transparency pages for research purposes.

Similarly, Table 4 presents a list of federative entities and information that must be compulsory published by them on public budget execution, totaling twelve fields. In this case, according to Decree 7,185/2010, they are required to publish information for three phases of spending (committed, settled and paid), for six classification elements (budgetary unit, function, sub-function, nature of spending, source of funds and type of tender) and for three elements detailing the tender corresponding to the expenditure (no. of the process, beneficiary and good/service provided).

**Table 4. Primariness of published spending**

Entity of the federation	committed	settled	paid	budgetary unit	function	sub-function	nature of spending	source of funds	type of tender	no. of the process	beneficiary	good/service provided
City Hall of Rio de Janeiro	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓
City Hall of Sao Luis	✓	✓	✓	✓	✓	✓	✓	✓		✓		✓
City Hall of Joao Pessoa	✓	✓	✓		✓			✓	✓	✓		✓
Office of the Comptroller General – Federal Executive	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
City Hall of Teresina	✓	✓	✓	✓	✓	✓			✓	✓		✓
Federal Senate	✓	✓	✓	✓	✓	✓						
City Hall of Palmas	✓			✓	✓	✓	✓	✓	✓	✓		✓
City Hall of Maceio	✓	✓	✓	✓	✓	✓		✓		✓		✓
City Hall of Curitiba	✓	✓	✓	✓	✓	✓						
City Hall of Salvador	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
City Hall of Sao Paulo	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓
City Hall of Boa Vista	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
City Hall of Recife	✓	✓	✓	✓	✓					✓		✓
City Hall of Belem	✓			✓	✓	✓	✓	✓	✓	✓		✓
City Hall of Florianopolis	✓	✓	✓	✓	✓	✓				✓	✓	
City Hall of Macapa	✓	✓	✓	✓	✓	✓						
City Hall of Fortaleza	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓
City Hall of Porto Alegre	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
City Hall of Cuiaba	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓
Government of the Federal District	✓			✓	✓	✓	✓	✓	✓	✓		✓
City Hall of Vitoria	✓	✓		✓	✓		✓		✓	✓	✓	
City Hall of Goiania	✓	✓	✓	✓	✓	✓				✓	✓	
City Hall of Porto Velho	✓	✓	✓	✓	✓			✓	✓	✓		
City Hall of Belo Horizonte	✓	✓	✓	✓	✓						✓	
City Hall of Natal	✓	✓	✓	✓	✓					✓		✓
City Hall of Aracaju	✓	✓		✓	✓	✓		✓				
City Hall of Campo Grande		✓			✓							
City Hall of Rio Branco				✓	✓							
City Hall of Manaus	✓			✓	✓	✓	✓	✓	✓	✓		✓

Source and Prepared by: Gpapai/Usp, 2014.



Decree 7,185/2010, which is used as a parameter of primariness, does not provide for any rules for publishing installment payments related to the same tender, meaning that aggregate figures for tenders can be published. However, apart from constituting non-compliance with the decree, failure to publish details about bidding processes means publishing data that are more aggregated than the law provides for. Because the tables follow the ranking presented at the end of this section, the City Hall of Manaus was ranked in the last position, even though it published nine of the twelve required information fields.

### **2.3.3. Most governments do not provide *timely* information**

With the aim of confirming the timeliness of published data, the last date of publication of revenue and spending information was checked on each transparency page. To get a score in relation to the ‘data must be timely’ principle, the difference between the last date of publication of revenue or spending data and the date of the observation must be of *one business day*. Thus, a page observed on Sunday should provide information for Thursday, as in Brazil Saturday is not a working day, so revenues and expenditures executed on Friday can be published on Monday.

Five sites (17.24%) provided information within the one business day deadline set forth in the law: Boa Vista (state of Roraima), Florianopolis (state of Santa Catarina),

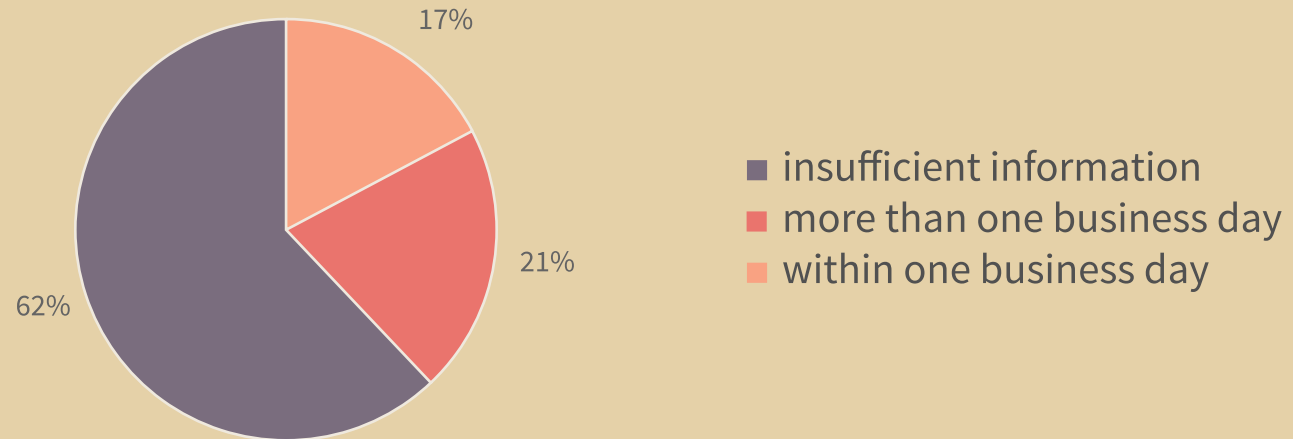
Table 5. Evaluation of the 'Data Must be Timely' principle and date observed				
Entity of the federation	Last publication	Day	Month	Year
City Hall of Rio de Janeiro	without sufficient information			2013
City Hall of Sao Luis	without sufficient information		11	2013
City Hall of Joao Pessoa	without sufficient information		10	2013
Office of the Comptroller General	without sufficient information			2013
City Hall of Teresina	without sufficient information		10	2013
Federal Senate	more than one business day	9	11	2013
City Hall of Palmas	within one business day	14	11	2013
City Hall of Maceio	within one business day	14	11	2013
City Hall of Curitiba	without sufficient information			2013
City Hall of Salvador	without sufficient information		11	2013
City Hall of Sao Paulo	without sufficient information		10	2013
City Hall of Boa Vista	within one business day	19	11	2013
City Hall of Recife	without sufficient information			
City Hall of Belem	without sufficient information			2013
City Hall of Florianopolis	within one business day	29	10	2013
City Hall of Macapa	within one business day	16	11	2013
City Hall of Fortaleza	more than one business day	14	11	2013
City Hall of Porto Alegre	without sufficient information		8	2013
City Hall of Cuiaba	without sufficient information		11	2013
Government of the Federal District	without sufficient information			2013
City Hall of Vitoria	more than one business day	8	11	2013
City Hall of Goiania	without sufficient information		11	2013
City Hall of Porto Velho	more than one business day	14	11	2013
City Hall of Belo Horizonte	more than one business day	31	10	2013
City Hall of Natal	without sufficient information		7	2013
City Hall of Aracaju	without sufficient information		10	2013
City Hall of Campo Grande	without sufficient information			2013
City Hall of Rio Branco	without sufficient information		11	2013
City Hall of Manaus	more than one business day	1	11	2013

Source and Prepared by: Gopai/Usp, 2014.

Macapa (state of Amapa), Palmas (state of Tocantins) and Maceio (state of Alagoas). Table 5 shows a list of federative entities, the situation of the last publication and data from the last date of publication found. Pages that do not indicate the day and/or month of publication of its data make it impossible for the timeliness of the data to be evaluated and they were thus described as 'without sufficient information' for scoring in the analysis of the 'timeliness' principle. In such cases, they were scored 0.

Figure 2 shows the percentage of pages and the publishing situation: 62% failed to publish sufficient information for evaluation purposes and 17% of the observed pages were seen to have complied with the deadline set forth in Decree 7,185/2010.

**Figure 2:** Percentage of evaluation of the ‘Data Must be Timely’ principle



Source and Prepared by: Gpopai/Usp, 2014.

#### 2.3.4. Access to information is not ensured to all individuals

To check compliance of the transparency pages with the ‘data must be accessible’ principle, the ASES program of the Federal Government was used, which analyzes the HTML page and checks whether the code complies with the e-MAG accessibility standard. P1 type bugs consist in page design problems that prevent people with special needs from accessing information, such as, for example, the lack of a description for an image that prevents screen reader programs to describe them to visually impaired individuals, something that is equivalent to the lack of subtitles in movies produced in another language for those

<b>Entity of the federation</b>	<b>P1 bugs</b>	<b>P2 bugs</b>	<b>P3 bugs</b>
City Hall of Rio de Janeiro	19	34	1
City Hall of Sao Luis	0	8	0
City Hall of Joao Pessoa	22	5	0
Office of the Comptroller General	3	21	0
City Hall of Teresina	5	0	0
Federal Senate	12	25	0
City Hall of Palmas	1	1	1
City Hall of Maceio	57	0	0
City Hall of Curitiba	0	0	0
City Hall of Salvador	26	16	3
City Hall of Sao Paulo	6	22	2
City Hall of Boa Vista	10	2	0
City Hall of Recife	20	3	0
City Hall of Belem	37	55	9
City Hall of Florianopolis	46	2	0
City Hall of Macapa	62	51	0
City Hall of Fortaleza	27	2	0
City Hall of Porto Alegre	8	0	0
City Hall of Cuiaba	32	4	0
Government of the Federal District	4	18	6
City Hall of Vitoria	19	9	7
City Hall of Goiania	10	1	0
City Hall of Porto Velho	12	1	0
City Hall of Belo Horizonte	112	56	1
City Hall of Natal	18	2	0
City Hall of Aracaju	10	3	1
City Hall of Campo Grande	25	2	0
City Hall of Rio Branco	32	11	0
City Hall of Manaus	103	24	1

Source and Prepared by: Gpopai/Usp, 2014.

who are not familiar with it. Table 6 shows the bug count by transparency page.

Two sites (6.90%) did not present any P1 bugs when analyzed using the ASES program: Curitiba (state of Parana) and Sao Luis (state of Maranhão). It can be observed that P1 bugs are the ones most often found on the transparency pages evaluated by the ASES program. Only in two portals no bugs were found, i.e. only two of them were scored in relation to the ‘accessible’ principle.

### **2.3.5. Advances were observed in terms of the availability of *machine-processable data***

Thirteen sites were provided enough data to be scored for the ‘machine-processable’ principle (44.83%), in seven of which downloads were only available in the XLS format, four in CSV and two in XML. All the transparency pages observed were developed with Web technology, otherwise it would not have been possible to view them using web browsers. The language used for developing these pages is called

Hypertext Markup Language, more known by the acronym HTML. This language is also the file format that web browsers recognize for presenting transparency pages to users.

HTML is not exactly a format for publishing data, but one for presenting information to visitors to web pages. This means that this format meets the visualization needs of human beings, but not machine-processing needs. For a system to process information from an HTML page, a manual programming effort must be made for the data to be correctly extracted. CSV (Comma Separated Values) is an example of a format that allows for easier processing, as it can be opened using a spreadsheet program such as LibreOffice Calc and Microsoft Excel. In the CSV format, data is arranged in rows and columns and the first line provides the name of the fields related to the data. Thus, the effort to process this information is lower than the one required on an HTML page. In fact, a user who is more experienced with spreadsheets can use some formulas and create charts from data in the CSV format.

The need for more or less human effort for data to be machine-processable is directly linked to how information is made available. PDF (Portable Document Format) is another example of a format from which it is difficult to extract data, as the simple act of copying and pasting text in this format often requires user effort to organize the copied text. Thus, HTML and PDF are formats that present

information to humans but whose machine-processing is difficult. CSV is an example of a format that is used by humans in spreadsheet programs, as the organization of data in columns and rows allows for tools of such applications to be used, meaning that the data are machine-processable.

Table 7 shows all the formats found in the observation of transparency pages. Because these pages were developed in HTML, this format is marked for all Brazilian federative entities. In addition to the HTML, CVS and PDF formats, data was also found in XLS/X, which is a format controlled by Microsoft, and in the SQL (Structured Query Language) and XML (eXtensible Markup Language) open formats, both of which are machine-processable. In this table, the formats that are open and machine-processable were marked with a star.

It was not the purpose of this research to evaluate the quality of downloaded data, as its focus was on the user's experience on the transparency pages. However, the existence of data in downloaded files was assessed. In the case of the City Hall of Joao Pessoa, although the possibility of downloading data in the CSV and XML formats is available, it was seen that, due to a problem on the page, no data existed. However, SQL data for download were also found and, in this case, data were present. This is also the only federative entity that provides SQL data.

<b>Name of the entity of the federation</b>	<b>HTML</b>	<b>PDF</b>	<b>XLS/X</b>	<b>CSV</b>	<b>XML</b>	<b>SQL</b>
City Hall of Rio de Janeiro	✓		✓	★	★	
City Hall of Sao Luis	✓			★		
City Hall of Joao Pessoa	✓			★	★	★
Office of the Comptroller General	✓			★		
City Hall of Teresina	✓	✓		★		
Federal Senate	✓	✓	✓	★		
City Hall of Palmas	✓	✓	✓			
City Hall of Maceio	✓	✓	✓			
City Hall of Curitiba	✓		✓			
City Hall of Salvador	✓	✓	✓			
City Hall of Sao Paulo	✓	✓	✓			
City Hall of Boa Vista	✓	✓				
City Hall of Recife	✓	✓	✓			
City Hall of Belem	✓		✓			
City Hall of Florianopolis	✓					
City Hall of Macapa	✓	✓				
City Hall of Fortaleza	✓					
City Hall of Porto Alegre	✓	✓				
City Hall of Cuiaba	✓	✓				
Government of the Federal District	✓	✓				
City Hall of Vitoria	✓					
City Hall of Goiania	✓	✓				
City Hall of Porto Velho	✓	✓				
City Hall of Belo Horizonte	✓	✓				
City Hall of Natal	✓	✓				
City Hall of Aracaju	✓	✓				
City Hall of Campo Grande	✓	✓				
City Hall of Rio Branco	✓	✓				
City Hall of Manaus	✓	✓				

Source and Prepared by: Gopai/Usp, 2014.

### 2.3.6. Access is free, *without discriminating users*

No site discriminates users for accessing data, but the tool of the Federal Senate on the *Siga Brasil* site provides two types of access, namely, ‘Open Access’ and ‘Expert Access.’ In the observation, the ‘Open Access’ option was used, because it did not require registration for data to be accessed.

### 2.3.7. The minority presents data in non-proprietary format

As noted above, for the seventh principle, the types of formats available for download on the user interface were checked. Because HTML is an open standard, meaning that it is not

exclusively controlled by a company, it could be said that every transparency page meets the ‘non-proprietary’ principle if only the language used in developing the transparency pages is considered. However, considering the machine-processable principle, in connection with which thirteen sites were scored, on seven, four and two of which data could only be downloaded in the XLS, CSV and XML format, respectively, it can be said that six of the analyzed sites (20.69%) provide files in ‘non-proprietary’ formats.

#### **2.3.8. It is not known whether the data are under *license***

No site explicitly stated that a license existed for using the data, either free or restrictive. Therefore, no state received a score for the ‘license-free’ principle.

### **2.4. Ranking of the Websites**

Table 8 shows the ranking of transparency pages and their score for each of the eight OGD principles. The ‘0’ (zero) score was replaced by a ‘-’ (dash) to improve the visualization of the scores. Three federative entities were tied in the first, second and third position. Seven were ranked fourth, twelve fifth and one sixth. Among those ranked first, there were differences in the composition of the score, as the city halls



**Table 8. Ranking of Transparency Portals**

Position	Federative entity	Score	Complete	Primary	Timely	Accessible	Machine-processable	Non-discriminatory	Non-proprietary	License-free
1st	City Hall of Rio de Janeiro	6	1	-	-	-	3	1	1	-
	City Hall of Sao Luis	6	1	-	-	1	2	1	1	-
	City Hall of Joao Pessoa	6	1	-	-	-	3	1	1	-
2nd	Office of the Comptroller General – Federal Executive	5	1	-	-	-	2	1	1	-
	City Hall of Teresina	5	1	-	-	-	2	1	1	-
	Federal Senate	5	1	-	-	-	2	1	1	-
3rd	City Hall of Palmas	4	1	-	1	-	1	1	-	-
	City Hall of Maceio	4	1	-	1	-	1	1	-	-
	City Hall of Curitiba	4	1	-	-	1	1	1	-	-
4th	City Hall of Salvador	3	1	-	-	-	1	1	-	-
	City Hall of Sao Paulo	3	1	-	-	-	1	1	-	-
	City Hall of Boa Vista	3	1	-	1	-	-	1	-	-
	City Hall of Recife	3	1	-	-	-	1	1	-	-
	City Hall of Belem	3	1	-	-	-	1	1	-	-
	City Hall of Florianopolis	3	1	-	1	-	-	1	-	-
	City Hall of Macapa	3	1	-	1	-	-	1	-	-
5th	City Hall of Fortaleza	2	1	-	-	-	-	1	-	-
	City Hall of Porto Alegre	2	1	-	-	-	-	1	-	-
	City Hall of Cuiaba	2	1	-	-	-	-	1	-	-
	Government of the Federal District	2	1	-	-	-	-	1	-	-
	City Hall of Vitoria	2	1	-	-	-	-	1	-	-
	City Hall of Goiania	2	1	-	-	-	-	1	-	-
	City Hall of Porto Velho	2	1	-	-	-	-	1	-	-
	City Hall of Belo Horizonte	2	1	-	-	-	-	1	-	-
	City Hall of Natal	2	1	-	-	-	-	1	-	-
	City Hall of Aracaju	2	1	-	-	-	-	1	-	-
	City Hall of Campo Grande	2	1	-	-	-	-	1	-	-
6th	City Hall of Rio Branco	2	1	-	-	-	-	1	-	-
	City Hall of Manaus	1	-	-	-	-	-	1	-	-

Source and Prepared by: Gpapai/Usp, 2014.

of Rio de Janeiro and Joao Pessoa were given a score of three for ‘processable’ and the city hall of Sao Luis was given a score of two, but the latter was the only one of the three to be given a score for the ‘accessible’ principle. However, in the group of the highest-ranking city halls, Rio de Janeiro was ranked higher than Sao Luis and just below Joao Pessoa. This order was determined by the amount of primary data that each of them publishes. Tables 2, 3 and 4 show that the city hall of Rio de Janeiro publishes 13 of the mandatory fields according to Decree 7,185/2010, while that of Sao Luis publishes 11 and that of Joao Pessoa 9 of them.

Thus, we can say that in the groups of municipalities that were ranked fifth, the city hall of Rio Branco is the one closest to the sixth position and that of Fortaleza is the one closest to the fourth position.

## 2.5. Conclusions

The aim of this research was that of analyzing the provision of data and information on budget execution on transparency portals of Brazilian capital cities, as well as on the websites of the Office of the Comptroller General and of the Federal Senate, the *Siga Brasil* site.

Its purpose was that of providing an overview of information offered to citizens interested in exercising social control over public budget mainly at the sub-national

level, that of the largest municipalities in each state. It should be noted that, in Brazil, many major public policies are under the responsibility of the municipal executive branch and, moreover, the general public is more interested in keeping track of decisions made at the local level, where it is more affected by them.

The unprecedented nature of this survey lies in the fact that it assessed the supply of data in relation to open data principles, as we believe that reuse of the government data made available on the analyzed sites can raise social control, citizen empowerment and the fight against corruption to a higher level.

In our analysis, we saw that all the entities surveyed keep a transparency portal in compliance with Complementary Law 131, but it was disappointing to see that none of them offers the detailed and primary data provided for in Decree 7,185/2010, which regulates Complementary Law 131.

To be useful for effective social control, data must be timely and the Brazilian legal framework is quite advanced in this regard, as it sets forth a deadline of 24 working hours for data to be published after an expenditure is made. It was seen that 17% of the surveyed websites complied with this requirement.

We were hoping that the portals were accessible and, in this research, we chose to assess whether they were developed in such a way as to not exclude people with limitations. Unfortunately, only two of the portals evaluated were considered accessible. On the other hand, it should be highlighted that none of

the surveyed sites required login, a password or any justification for accessing data and, therefore, they were considered as non-discriminatory.

A major contribution of this research was that of checking the forms and formats in which data are being made available to citizens, taking into account the ‘machine-processable’ and ‘non-proprietary formats’ dimensions. All portals provide data on HTML pages and most of them also offer PDF documents, but although these are suitable for a person to read and interpret them, they make it extremely difficult for the data to be reused if, for example, a citizen wishes to use a spreadsheet software for making calculations.

It is interesting to note that the open government data movement is already influencing some governments and this fact was observed in our sample, where data are also provided in a semi-structured (XLS and CSV) and structured fashion, with proprietary and open formats.

Finally, the absence of explicit licenses for reusing data made available on transparency portals suggests that a truly comprehensive discussion on a data licensing policy in Brazil is yet to be held.

## 3. QUALITATIVE RESEARCH

### OPEN DATA AND PUBLIC BUDGET: IMPACTS ON THE PERCEPTION OF INTERMEDIARIES

#### 3.1. Objectives

The purpose of the qualitative stage of the research was that of evaluating/measuring, through interviews with open data *intermediaries* (who process budget data, but not only such data), the impact of these data on expanding human rights and citizenship rights in Brazil. Open data intermediaries are those who “*get hold of data to transform and recombine them to then present them to the public,*” since a lot of data will not directly reach end users<sup>29</sup>, i.e. organizations or individual agents who somehow process, analyze and/or reuse existing data to improve access to information available to society.

Several products can be prepared with open data by these intermediaries (use and reuse): such products range from newspaper articles and technical-scientific papers to digital technologies capable of processing the data and presenting them in other formats. It is in such diverse scenario of access, use and reuse of data that

<sup>29</sup> Source: [governoeletronico.gov.br](http://governoeletronico.gov.br), accessed in September 2013.

we based the initial assumption that intermediaries could have different levels of understanding of transparency policies, of the consensual principles applied to open data, of the laws that regulate access to data and of practical ways to assess whether or not data is truly accessible on a digital platform that makes it available. This assumption was confirmed in the interviews, as will be seen below.

From this starting point, an effort was made to understand the *access, use and reuse* produced by open data intermediaries in connection with the public budget, i.e. their motivations to “work on” data related to budget cycles, how they work, who they want to reach, the types of information they are producing that are more user-friendly and what their suggestions are to improve official data for the purpose of ensuring greater transparency.

Based on data on access, use and reuse, we tried to identify the effects and impacts of open data both on the routines of organizations and on promoting human and citizenship rights, as can be observed, for instance, in examples of effective political advocacy with the legislative and executive branches, in the promotion of public debate generated by media articles or in training projects with local social movements.

For this purpose, the period determined by the publication of the Access to Information Act, i.e. from 2011 to today, was considered in the interviews. However, due to the existence of previous laws, such as those providing for the opening of data on the Transparency Portal of the Federal Government in 2004,

the respondents considered a longer period in their reflections on the process of opening data in Brazil.

The research project also had specific objectives, namely:

- Identifying the motivations of intermediaries to work with budget transparency;
- Checking to what extent these intermediaries *know* what open data are;
- Checking whether intermediaries *use* and also *reuse* open data;
- Checking whether in their *reuse* of data the intermediaries make them available in an open format;
- Checking the perceptions of intermediaries on the impact of their work on expanding human rights in Brazil;
- Identifying possible evidence to support these perceptions;
- Collecting suggestions.

In the section on results, we will see the extent to which the objectives were achieved and also some reports that may signal the need for further investigations.

### 3.2. Methodology

According to the idea of this research project, the objective of this investigation is “to measure the perception of budget transparency and accessibility, and

its impact on people's life." By referring to studies designed to evaluate public policies, it is possible to better define the contours and limits of this research exercise, which is intended to capture "perceptions" and measure "impacts." Evaluation studies can be designed to address the following dimensions<sup>30</sup>:

- a) *Processes*: studies on aspects that favor or hinder the process of implementing a policy or program, including their design and organizational and institutional dimensions.
- b) *Results*, which comprise:
  - i) Performance (or results in their strict sense, i.e. outcomes), which refers to the "products" of a program, as defined in its goals;
  - ii) Impact: changes in the situation of beneficiaries directly caused by a program;
  - iii) Effects: other results of a program, whether social or institutional, either expected or not, that end up being brought about as a result of the program.

The basic assumption was that *impact assessment research* is intended to establish causal relationships by identifying positive, negative or non-existent changes in a given social context as a result of a new element. In this study, the causes of change - a specific change in the routines of organizations that access public budget

<sup>30</sup> Draibe, 2001 apud Vaitsman, Rodrigues & Paes-Sousa. MDS, Unesco, 2006.



data; and a cultural change brought about by changes in the law and in how public budget data is accessed - would be the availability of budget data in digital tools.

Subsequently, based on the network for sharing methodologies provided by the project coordinator - the Web Foundation – the concepts of “outcomes” and “impacts” were incorporated, which were defined, according to the materials that were made available, as follows: “*Outcomes: The results of what has been done (Is anyone better off as a result? Positive and negative outcomes; Unexpected outcomes); Impact: Wider influence, the demonstrable contribution to society.*”<sup>31</sup>

From this perspective, it should be highlighted that the initial methodology has been somewhat improved as a result of the methodological dynamic of the Web Foundation: meetings held in the network of researchers, as well as the fact that texts and presentations were made available in a single shared platform, provided inputs for reflections that changed the final approach of the research. For example, the interview guide was designed according to the “follow the data” model<sup>32</sup>, with the aim of leading the interviewees to build their own arguments and concepts based on the data, sources and products.

---

<sup>31</sup> Source: Clarifying terms. Shared drive of the ODDC research. Accessed in 2014.

<sup>32</sup> The “follow the data” methodology was suggested under the research support provided by the World Wide Web Foundation and consists of “*methods that can be used to understand how data is being used in practice, and to trace the different actors involved in the use of data*” Source: Follow the data: understanding the social life of data. Shared drive of the ODDC research. Accessed in 2014.

### 3.2.1. Interview Guide

The interview guide (attached herewith) was designed according to the “follow the data” model, with the aim of collecting perceptions on the 8 Open Data Principles without asking direct and immediate questions on concepts. First, respondents were provoked to quote their data sources and describe the process of using the data (access, problems, benefits), as well as concrete outputs (articles, software, web pages, etc.). They were then asked to speak about changes in the routines of the organization (NGO, newspaper) or of people (journalist, activist) - i.e. which queries were incorporated into their routine work, which products (use and reuse) they prepare and how these products are disseminated. Subsequently, they were asked direct questions about concepts (What are open data? How do you define use and reuse? How do you assess the 8 principles?). Finally, the guide addressed the *perception of intermediary agents of the impact of open data on the rights and citizenship of the population*, i.e. how they assessed observable social changes from their perspective.

The guide was pre-tested in November of 2013 with members of two NGOs (Cfemea and OKF Brazil), and it was deemed appropriate and not requiring significant adjustments in its structure and content. After the pre-test, other interviews were conducted in November and December of 2013 and in January and February of

2014. The interviews were conducted in person or using digital communication tools (Skype, Google Hangout, etc.) and recorded for subsequent transcription and analysis — respondents were told previously that their answers would be recorded.

It should be stressed that the guide instructs interviewers on how to engage in a dialogue-based interaction to get consistent answers. Thus, the interviews allow for comparability, but it is not a matter of tabulating answers and their correspondence in a survey (quantitative tool), but rather of finding patterns of meaning and mapping out the diversity of answers to make it possible to prepare summaries.

### 3.2.2. Mapping of Respondents

The mapping of respondents was based on some specific criteria. In the research project, five categories of “intermediaries” were defined - potential users of open data operating nationwide to make up a research sample capable of capturing different types of actions on open budget data: NGOs, journalists, universities, hackers<sup>33</sup> and government itself. The Inesc researcher attended the 2nd National Meeting on Open Data in

---

<sup>33</sup> There are several definitions for “hacker”, but all of them are similar in the general sense. According to the *Manual dos Dados Abertos: Governo* (2011, p. 7), translated from the Open Data Manual and edited in Brazil, “a hacker is a person who enjoys having a deep understanding of the internal processes of a system, of a computer or of a computer network in particular. The term is often used pejoratively, where ‘cracker’ would be the correct name.”

2013<sup>34</sup>, where it was possible to become acquainted with the field formed by groups with an interest in open data in Brazil, and mapped some people for interviews.

A second moment, after initial contacts were established with some organizations and research was done on the Internet, consisted in selecting individuals in these categories who stood out in the field of open data in Brazil either for their data dissemination activities or for their innovative development of applications for re-using government data<sup>35</sup>. It should be mentioned that the interviews lasted 1 hour in average and that they were conducted by the advisor to Inesc Carmela Zigoni.

A first result of the research deserves immediate mention: the interviews actually showed that these are hybrid categories - for example, the professor and researcher Kellyton Brito can also be regarded as a hacker: he was awarded a prize in a contest on data reuse applications in 2013 and works independently, not necessarily supported by any university. IPEA is another example, as it is both a research institute and a government body, as well as OKF Brazil, an NGO that has members with a “hacker” profile. A brief description of the profiles of the institutions, organizations and persons who were interviewed will be provided below.

---

<sup>34</sup> Event jointly organized by W3C and the Federal Government (<http://2.encontro.dados.gov.br/>).

<sup>35</sup> The Federal University of Pernambuco, for example, emerged as a field in full development beyond the Rio-Sao Paulo area: the states of Pernambuco and Ceara have evolved as IT and innovation development zones in Brazil and, as a result, academic research into open data has become fertile ground. Source: ComputerWorld: O Porta Voz do Mercado de TI e Comunicação, accessed in 2014. (<http://computerworld.com.br/negocios/2013/02/01/bahia-ceara-e-pernambuco-sao-os-que-mais-investem-em-ti-no-brasil/>).

Box 1. Profile of the Respondents					
Organization	Category	Main action with open data	Respondent	Date of the interview	Means
CFEMEA	NGO	Analysis of the federal budget for policies for women.	Priscilla Brito	11/06/2013	Face to face
OKF Brazil	NGO	“Data School”; Development of Data Reuse Tools.	Everton Zanella	11/19/2013	Skype
O Estado de Sao Paulo newspaper	Journalist	“Basometro”	Daniel Bramatti	12/02/2013	GoogleHangout
Public Investigative Journalism Agency	Journalist	News stories about public investment, especially on impacts of large projects on the Amazon region and on World Cup-related projects.	Natalia Viana	01/23/2014	Skype
UFRPE	Academia	“meucongressonacional.com”	Kellyton Britto	12/13/2013	Skype
UFPE	Academia	Research interests: Data integration, Dataspaces, PDMS, Semantic Web, Linked data, XML, Information quality and Decision support systems.	Bernadette Farias Loscio	01/08/2013	Skype
IPEA	Research Institute	Research into the effectiveness of the Access to Information Act.	Rodrigo Orair and Jose Mauricio	12/19/2013	Skype
General Secretariat of the Presidency of the Republic (General Coordinating Board for New Media and Other Participation Languages).	Government	Platforms focused on social participation.	Ricardo Poppi	02/19/2014	Skype

Source and Prepared by: Inesc, 2014.

### 3.2.3. Profile of the respondents, institutions and organizations

#### a) NGOs:

- ✓ Cfemea (NGO): Brazilian non-governmental, non-profit organization that has been working for 23 years in the following theoretical and policy fields: feminism, human rights, democracy and racial equality (source: <http://cfemea.org.br/>). Cfemea uses open data to analyze the Brazilian federal

government budget and is responsible for the “Thematic Women’s Budget,” hosted on the budget transparency website of the Federal Senate, Siga Brasil (<http://www12.senado.gov.br/orcamento/tematicomulher>).

- ✓ Priscilla Brito is a political scientist who has been working with budget-related issues at Cfemea since 2011. Her job consists in analyzing open data of the federal public budget related to public policies for women, carrying out advocacy with the executive and legislative branches and promoting training activities designed to empower women’s social movements (rural women’s movements, for example) in relation to the information that is collected and analyzed by the organization. The methodology of the Thematic Women’s Budget is developed through dialogue with women’s movements and with the support from external technical consultants.
- ✓ OKF Brazil (NGO): non-governmental organization that is the Brazilian version of a global network for open knowledge, defined as “*any good - whether content, data or information in general - that can be freely used, reused and distributed by anyone without any restriction.*” (source: <http://br.okfn.org/>). It is engaged in several open data projects, such as one for developing a tool for monitoring the federal budget in partnership with INESC. Another product of this organization deserves special mention, namely, the Data School, which is an initiative focused on

training citizens for reading open data that also relies on a network of educational exchanges and training in programming, among other activities (source: <http://escoladedados.org/>).

- ✓ Everton Zanella Alvarenga: he is currently the executive director of the Open Knowledge Foundation in Brazil, which in Portuguese is referred to as “*A network for open knowledge*”.

b) Journalism:

- ✓ O Estado de Sao Paulo newspaper: Traditional Brazilian daily newspaper of large circulation (<http://www.estadao.com.br/>).
- ✓ Daniel Bramatti (Journalist): reporter of the Political Editorial Staff and Data Center of the O Estado de Sao Paulo newspaper (“Estadao Dados”: <http://blog.estadaodados.com/>) who is in charge of developing and maintaining the *Basometro*, an application that was developed to monitor the activities of Congress in relation to the behavior of its members — their presence in plenary sessions, voting sessions, coalitions, benches, with the aim of mapping out political trends in Congress. Since 2009, he has been working with statistical data produced by academia and open data, mainly in the context of elections.
- ✓ Publica Agency for Investigative Journalism: the organization bets on a non-profit journalistic approach to remain independent. All its reports are

freely reproduced by several media outlets under the creative commons license. Its mission is to produce reports based on the public interest on major issues for Brazil from the standpoint of its population – with the aim of strengthening the right to information, improving the democratic debate and promoting human rights. The agency’s investigative work is focused on the preparations for the 2014 World Cup; mega investments in the Amazon region; and military dictatorship. Its partners include independent journalism centers in Latin America, the US and Europe and traditional and new media outlets. Besides producing reports, the agency promotes independent investigative journalism through training programs for young journalists and grants. (Source: <http://www.apublica.org/>)

- ✓ Natalia Viana: Award-winning journalist and director of the Publica Agency for Investigative Journalism. She worked in Brazilian media outlets such as the Caros Amigos magazine and collaborated with the BBC, The Guardian, The Independent, The Sunday Times, The Nation and the WikiLeaks, coordinating leaks of Brazilian documents.

c) Research:

- ✓ Kellyton Brito: Professor at the IT Department of the Federal Rural University of Pernambuco (<http://www.ufrpe.br/>). Winner of the Hackathon prize of the Brazilian House of Representatives held in



2013 with the application *Meu Congresso Nacional* (my national congress) (<http://meucongressonacional.com/>), which reuses open budget data provided by offices of representatives and senators.

- ✓ Bernadette Farias Loscio is an assistant professor at the IT Center (CIn) of the Federal University of Pernambuco - UFPE, Brazil. She holds a PhD in computer science from the Federal University of Pernambuco and an MS from the Federal University of Ceara. Recently, she coordinated the opening of data of the City Hall of Recife and of the city's Transparency Portal, in addition to providing consultancy for opening data in the Government of Costa Rica.
- ✓ Institute for Applied Economic Research (Ipea): this is a federal public foundation linked to the Strategic Affairs Secretariat of the Presidency. It provides technical and institutional support to government for the formulation and reformulation of public policies and development programs in Brazil (<http://www.ipea.gov.br/portal/>).
- ✓ Rodrigo Orair and José Maurício: Researchers of the Public Finance Coordinating Board of Ipea's Directorate for Studies on Macroeconomic Policies.

d) Government:

- ✓ General Secretariat of the Presidency of the Republic: Agency in charge of coordinating the dialogue between the Presidency and social

movements. It mediates the relationship with and promotes the institutional strengthening of civil society organizations or joint bodies such as the National Youth Council.

- ✓ Ricardo Poppi: member of the General Coordinating Board for New Media and Other Participation Languages.

### 3.3. Results

#### 3.3.1. What is open data?

During the interviews, after “following the open data”, i.e. after asking questions to the respondents about data sources and uses, we deliberately asked them more directly to provide a conceptualization of the topic: “*in just a few words, how would you define open data?*.” Respondents are alerted to the fact that they should provide a discursive elaboration, without any consultations, meaning that they should feel comfortable and try and convey their understanding of this concept.

We can say that the result is similar in all answers and we identified a few patterns. Regarding the open data themselves, these would be data made available quickly, broadly, to the general public, efficiently, periodically, safe data that comply with the institutional standards set for making them available, which meet the

needs of users looking for them. The testimony of one of the journalists illustrates this understanding: *“It is a data management policy focused on promoting transparency and on encouraging community use and innovation in relation to the data.”*

There is another recurring element in how open data are defined, namely, the *information* that such data are intended to translate. We know that data and information can be conceptualized in different ways (data as a smaller unit that makes up a piece of information), but in all cases these dimensions are intertwined, as intermediaries are interested in data that express some meaning. The Access to Information Act defines *information* as synonymous with data: *“Information: data, processed or not, that can be used for production and transmission of knowledge, contained in any medium, platform or format.”*

In this regard, the respondents understand that the information should be simple, direct, clearly presented, transparent, free (*“with no payment required to access them”*) and usable at the same time. In this connection, criticism of the government was observed for making lots of information available but presenting them in ways that are “inconsistent”, “not comparable”, “with different levels of disaggregation” or “organized under different categories.” According to some respondents, especially those from civil society organizations, detailing — i.e. the primary aspect of the data — can jeopardize effective access to the data. Another issue that was brought up both in relation to the definition

and to the promotion of rights was the BigData issue, i.e. it is not enough to open up the data, as it is also necessary to understand what is important and to know how to structure the data in a way that makes them useful and effectively accessible.

One respondent pointed out an interesting dimension of data and information: creating the concept of “data life cycle,” this respondent, who was from the “academic” category, highlighted the existence of “zombie data”, i.e. data that look like they are alive but are not, since they are static in the sense of only allowing for queries and not for processing, for being in PDF format or for not being up to date.

Overall, the statements of the respondents reflect knowledge of the first five open data principles (data must be complete, primary, timely, accessible and machine-processable). As for the “non-discriminatory” principle, all of them understand what it is, but some of them only understood what it meant when the interviewer explained it with examples. With regard to the “non-proprietary” principle, most of them didn’t know what it meant, including respondents falling under the categories of research and journalism.

In relation to the “license-free” principle, most of them know what it means, but they don’t see it as a relevant piece of information in their searches on open data portals and could not tell whether the searched web pages were license-free

or not. This may reflect two issues: the debate on free knowledge is not widespread in society, not even among the most skilled users of open data; with the exception of IT academics and hackers, other intermediaries outsource the technological element of using and reusing data to other professionals who more actively engaged in discussions on software development, etc. However, when the licensing policy of the organizations to which they belong was checked, the following scenario emerged:

<b>Box 2. License information of the Websites of the Organizations</b>	
<b>Organization</b>	<b>Licensing of the website of the organization to which it belongs</b>
CFEMEA	There is copyleft for content, but no information is provided about the site's licensing policy.
OKF Brazil	All the contents of this site licensed under a Creative Commons Attribution 3.0 Unported license (all jurisdictions). —See more at: <a href="http://br.okfn.org/sobre/#sthash.EuapAWrx.dpuf">http://br.okfn.org/sobre/#sthash.EuapAWrx.dpuf</a>
O Estado de Sao Paulo newspaper – Estadao Dados	Copyright 2013 - Estado Group - Source Code under AGPL license — Contents under license Creative Commons By-SA.
Publica Agency for Investigative Journalism	There is copyleft for content, but no information is provided about the site's licensing policy.
MeuCongressoNacional.com	License-free.
IPEA	Creative Commons Attribution 2.5 Brazil license.
City Hall of Recife	Private License.
General Secretariat of the Presidency of the Republic	No information about the licensing policy on the official page.

Source and Prepared by: Inesc, 2014.

As can be seen, most of them adopt a license-free policy or at least an open knowledge license related to contents. In the interviews, we saw that some of the reuse products are open (application and database codes) while others are not, such as texts in PDF format - however, as there is sensitivity to copyleft, we can say that, in general, they make data available in an open format. The fact that the organizations adopt the practice of having websites with relatively open licenses and copyleft policies contrasts with the unawareness of such conduct of official sources of open data that they browse, if they are license-free or not, and we should consider that, in the Brazilian context, there is still no regulatory framework for the Internet and the debate on license-free formats advances at times and loses momentum at others in the political agenda of the executive branch: in practice, the license to be adopted for official portals is defined at the discretion of the manager of the ministry or area in question.

Regarding the licensing policy of the websites that they access, the researcher checked the portals one by one and the reality that she identified is the following one:

<b>Box 3. License Information of the Official Websites Accessed by the Organizations</b>	
<b>Portal</b>	<b>License</b>
House of Representatives (legislative activity)	No information provided.
Federal Senate (legislative activity)	No information provided.
House of Representatives (budget data of the house)	No information provided.
Federal Senate (budget data of the senate)	No information provided.
Transparency Portal	No information provided.
Siga Brasil	No information provided.
SIOP	The information provided on the site is not clear. Developed in partnership with the University of Brasilia (a federal university) and Dell.
System for Recording Votes in Plenary Sessions of the House of Representatives	No information provided.
Receita Federal (Brazil's Internal Revenue Service)	No information provided.
SISTN (Caixa Economica Federal, the Brazilian federal savings bank)	No information provided.

Source and Prepared by: Inesc, 2014.

In this regard, the unawareness of intermediaries of this principle might have been influenced by the fact that none of the sites provides information about its license, even if a restrictive one.

We also asked the respondents to assess the Access to Information Act and all of them assessed it positively both in relation to its content, which was described as politically correct and complete, and to its social impact, as that Act regulates access to data and promotes a culture of public transparency. All of them believe that the Act promotes the empowerment of society. They pointed out some

positive aspects of the Act, such as the importance of this regulation and the fact that its text sets out clear and detailed rules for data availability, as well as the fact that it sets standards for the three branches of power (federal, state, municipal branches). With regard to data availability standards, we can say that the Access to Information Act reflects all of the 8 open data principles to some extent, especially in relation to accessibility (which is even regulated by another specific law, the e-MAG law), the need for data to be primary, digital availability, integrity (reliability), open format, non-proprietary requirements, among others. In this context, the very existence of the Act leads data intermediaries to reflect on some important aspects related to open data and their definition. For the group of journalists, the Access to Information Act has a special positive feature:

“We journalists, for example, are often asked ‘what do you want these data for?’ and, in this sense, civil servants, who are sort of the guardians of those data, cannot interfere in the final use that will be made of the data. So the Act strengthens the rights of citizens, as the more tools citizens have to explore the wealth of information produced by government, the more value they add to the public debate.” (Journalist).

But there were also criticisms associated with the need for improvements in the effectiveness of the Act — the respondents believe that not all levels



of government (state and municipal levels) are aware of the need to comply with the Access to Information Act and that penalties can be applied for non-compliance with it.<sup>36</sup> Another issue is the fact that the venue for filing appeals for refusal to open up data is the State itself, which makes it difficult for them to be promptly and impartially judged: one of the respondents mentioned the possibility of, for example, setting up a joint council with civil society and government representatives to monitor compliance with the Access to Information Act.

The IPEA researchers made a more precise and informed assessment of the Access to Information Act, due to the fact that they are conducting a research into its effectiveness in relation to users who are “ordinary citizens”, whose final report will be issued in 2014. The researchers reported that ordinary citizens face difficulties to access data when they invoke the Act to get data that are unavailable. Reinforcing this perception, one of the journalists who were interviewed made a similar criticism in relation to cases where a user of the data who needs information that is not available on the Web or who cannot understand it has to deal with a lot of red tape to request the

---

<sup>36</sup> “Warning; fine; termination of any employment relationship with government; temporary suspension from participating in public tenders and prohibition from entering into contracts with the public administration for no more than two (2) years; and declaration of unfitness to bid or contract with the government until rehabilitation is promoted before the very authority that imposed the fine.” (Source: Access to Information Act).

**Professor Kellyton Britto**, developer of the award-winning Meu Congresso Nacional tool (<http://meucongressonacional.com/>), uses the portals of the House of Representatives and of the Senate, plus data from Brazil's Internal Revenue Service. According to him, the key point is that the data are not complete. For example, in the data for resources of the National Congress, there are representatives for whom “zero spending” is registered, which is not true, as by consulting other sources of the House of Representatives itself (hard-copy invoices, for example) one sees that there are records of these expenditures. What this means is that congresspersons are not feeding the database that is made available to the public. According to him, the data available on these portals are up to date, as ideally the updating is daily. However, he made the following remark: “But because they are not complete, the only ones that are up to date are those of MPs who actually feed the databases.”

information or clarifications with the relevant government agency, which can jeopardize the effectiveness of the Act: a journalist (intermediary) must request the information to the communication department of the agency, which has a staff of journalists and advertising experts who are not familiar with the requested information and have to relay the request to a public manager of the policy in question who is actually in charge of managing such information. We can therefore question whether the communication departments of these agencies are the best units to process requests for information and, if yes, whether it would not be the case of reshaping these departments by bringing in additional expertise.

### 3.3.2. How are open data accessed?

As mentioned above, respondents were not generally aware of the eight open data principles as defined internationally, but it was seen that they were especially aware of the first five when they were asked about the practice of accessing data in both their practical (technical) and political contours. The websites accessed by the respondents included in our sample are generally the same, making it possible for us to systematize this perception in the box below.

**Box 4. User perception about the Open Data Websites that were Accessed**

Principle Portal	Complete	Primary	Timely	Accessible	Machine-processable	Non-discriminatory	Non-proprietary	License-free
House of Representatives (legislative activity)	Yes	No	Yes	Yes	No	Yes	Yes	Yes
Federal Senate (legislative activity)	Yes	No	Yes	Yes	No	Yes	Yes	Yes
House of Representatives (budget data of the house)	No	Yes	Yes	Yes	Yes	Yes	Unknown	Unknown
Federal Senate (budget data of the senate)	No	Yes	Yes	Yes	Yes	Yes	Unknown	Unknown
Transparency Portal	Yes/doesn't know	Yes/doesn't know	Yes	Yes/no	Yes	Yes	Unknown	Not made explicit on the site
Siga Brasil	Yes	Yes	Yes	Yes	Yes	Yes/no (registration for advanced users)	Yes	Don't know.
SIOP	Yes	Yes	Yes	Yes	Yes	Yes	Unknown	Unknown
System for Recording Votes in Plenary Sessions of the House of Representatives	Yes	Yes	Yes	Yes	Yes	Unknown	Unknown	Unknown
Receita Federal (Brazil's Internal Revenue Service)	Yes	Yes	Yes	Yes	Yes	No (CNPJ - corporate income tax ID number - and captcha required)	Unknown	Unknown
SISTN (Caixa Economica Federal, the Brazilian federal savings bank)	No	Yes	No	Yes	No	Yes	Yes	Unknown

Source/Prepared by: Inesc, 2014.

### 3.3.3. What is done with open data that are accessed?

The respondents produce different kinds of products, according to the goals of their organizations and institutions they belong to. Box 5 illustrates purposes for which intermediaries use the data:

Box 5. Use of open data						
Organization	Category	Main area of work with open data	Collection, reading, processing.	Products Use and reuse	Information Dissemination	Measurable impacts
CFEMEA	NGO	Analysis of the federal budget for policies for women.	<ul style="list-style-type: none"> <li>Hiring of consultants specializing in public budgeting;</li> <li>Extracting data about public budgeting and policies;</li> <li>Analysis and Production of Report.</li> </ul>	<ul style="list-style-type: none"> <li>Report;</li> <li>Academic articles;</li> <li>Technical Note;</li> <li>Methodology for the Siga Brasil website;</li> <li>Letters to MPs;</li> <li>Semáforo da Execução Orçamentária.</li> </ul>	<ul style="list-style-type: none"> <li>Meetings with public managers;</li> <li>Hearings in Congress;</li> <li>Workshops with social movements; social media (facebook, twitter);</li> <li>Contents for media.</li> </ul>	<ul style="list-style-type: none"> <li>Women's movements established;</li> <li>Parliamentary amendments incorporated into the budget;</li> <li>Public policies improved.</li> </ul>
OKF Brazil	NGO	"Data School"; Development of Reuse Tools.	<ul style="list-style-type: none"> <li>Collection and technical interpretation, Development of applications and digital tools, Establishment of hacker community, NGOs and social movements.</li> </ul>	<ul style="list-style-type: none"> <li>Reuse Tools (Budget at Your Fingertips - Inesc);</li> <li>Online courses;</li> <li>Tutorial videos.</li> </ul>	<ul style="list-style-type: none"> <li>OKF Brazil website, interactive website (data school);</li> <li>Distance and classroom learning;</li> <li>Knowledge sharing networks on the Internet;</li> <li>Social media (facebook, twitter, youtube channel).</li> </ul>	<ul style="list-style-type: none"> <li>NGOs and movements established around the open data topic capable of managing digital tools;</li> <li>Hackers capable of developing programming.</li> </ul>

O Estado de São Paulo newspaper	Journalist	Basometro	<ul style="list-style-type: none"> <li>• Data collection and reading, feeding of the tool;</li> <li>• Political analysis of data and production of news content;</li> <li>• Dialogue with academia.</li> </ul>	<ul style="list-style-type: none"> <li>• Basometro;</li> <li>• Newspaper articles;</li> <li>• Book.</li> </ul>	<ul style="list-style-type: none"> <li>• Digital newspaper;</li> <li>• Social media (facebook, twitter).</li> </ul>	<ul style="list-style-type: none"> <li>• Academics wrote a book using data generated by the application;</li> <li>• Access to the tool and sharing of stories.</li> </ul>
Publica Agency for Investigative Journalism	Journalist	News stories about public investment, with an emphasis on the impacts of large undertakings in the Amazon region and of the World Cup; Weak Leaks in Brazil.	<ul style="list-style-type: none"> <li>• Collection of official websites;</li> <li>• Reading by journalists, assisted by partnerships and with the participation of readers or not;</li> <li>• Scholarships for young journalists.</li> </ul>	<ul style="list-style-type: none"> <li>• Articles;</li> <li>• Videos;</li> <li>• Public debates.</li> </ul>	<ul style="list-style-type: none"> <li>• Website;</li> <li>• Social media (facebook, twitter);</li> <li>• Thematic contests.</li> </ul>	<ul style="list-style-type: none"> <li>• Access to and sharing of stories;</li> <li>• Dialogue with the Government and social movements;</li> <li>• New journalistic investigations based on society's needs.</li> </ul>
UFRPE	Academia	"meucongressonacional.com"	<ul style="list-style-type: none"> <li>• Data collection, evaluation, cleaning;</li> <li>• Development of the application;</li> <li>• Publication in digital media.</li> </ul>	<ul style="list-style-type: none"> <li>• Application.</li> </ul>	<ul style="list-style-type: none"> <li>• Website of the application;</li> <li>• Social media (facebook, twitter);</li> <li>• Interviews for the media.</li> </ul>	<ul style="list-style-type: none"> <li>• Prize awarding for the application;</li> <li>• High number of accesses to the tool.</li> </ul>
UFPE	Academia	Research (data integration, Dataspaces, PDMS, Semantic Web, Linked data, XML, Information Quality and Decision Support Systems); consultancies for governments.	<ul style="list-style-type: none"> <li>• Review of databases and data opening for governments (City Hall of Recife and Government of Costa Rica).</li> </ul>	<ul style="list-style-type: none"> <li>• Structured transparency websites;</li> <li>• Academic contents;</li> <li>• Guidance for university students.</li> </ul>	<ul style="list-style-type: none"> <li>• Transparency websites on the Internet;</li> <li>• Academic production published.</li> <li>• No. of students guided.</li> </ul>	Not monitored by the intermediary.

IPEA	Research Institute	Research into the effectiveness of the Access to Information Act.	<ul style="list-style-type: none"> <li>• Evaluation of databases;</li> <li>• Development of softwares for automatic breaking of PDF files;</li> <li>• Organization of the data in a new database;</li> <li>• Analysis and production of technical notes and reports;</li> <li>• Dialogues with government sectors; development of evaluative research.</li> </ul>	<ul style="list-style-type: none"> <li>• Research Projects;</li> <li>• Research Reports;</li> <li>• Technical Notes;</li> </ul>	<ul style="list-style-type: none"> <li>• Meetings with public managers;</li> <li>• Interviews for journalists.</li> </ul>	Improved public policies
General Secretariat of the Presidency of the Republic.	Government	Platforms focused on social participation, social control and citizen interactivity (Consocial, Web Cidadania Xingu, Social Participation System of the Federal Government; Digital Public Consultations, among others); Linking of policies designed to promote open data in Brazil (2nd Action Plan of the Open Government Partnership, National Data Infrastructure, Internet Civil Landmark, among others).	<ul style="list-style-type: none"> <li>• Establishment of partnerships for building websites (UNB, social movements, NGOs);</li> <li>• Meetings with managers of several Ministries for transferring knowledge;</li> <li>• Meetings with information technicians of various Ministries with the aim of developing expertise in apps.</li> </ul>	<ul style="list-style-type: none"> <li>• Online websites;</li> <li>• Social networks of government bodies active (visitors, shares, comments, accesses).</li> </ul>	<ul style="list-style-type: none"> <li>• Website;</li> <li>• Social media (facebook, twitter);</li> <li>• Analog spaces (meetings, public conferences);</li> <li>• News media.</li> </ul>	<ul style="list-style-type: none"> <li>• Number of accesses to participation platforms (websites, social media);</li> <li>• Number of suggestions recorded on digital queries;</li> </ul>

**In the case of the Basometro** (<http://estadadados.com/basometro/>) tool, the portals that are used are mainly those of the House of Representatives and of the Federal Senate. The most important data of this tool are those related to plenary voting sessions in Congress. However, as pointed out by Daniel Bramatti, data are not made available on both portals in the same way. After manual collection, as collection is not automatic, it is necessary to clean the data (human review), which is a factor of difficulty, as data cannot be updated in real time, causing a delay between what happened and what is available in the Basometro. The respondent stressed the collaboration of technicians of the two houses, who assist the newspaper when its journalists need clarifications.

### 3.3.4. What does data *use* and *reuse* mean?

In the field of open data, the terms data “use” and “reuse” are often heard. However, as we carried out this research project, we noticed that there is a conceptual void in the literature in relation to such terms. After all, what defines and differentiates use and reuse?

Based on the research results, we can say that use and reuse are terms that users have not reflected upon sufficiently. After stimulating a rationalization of the terms in the interview, the common understanding emerged that this difference is related to the type of use one makes of the data. Let us take a look at some of the definitions provided by the respondents:

“I would say that the Basometro reuses data of the Legislative Chamber, as they are presented in another form, i.e. you need to capture, rework on and republish them. What characterizes reuse is the informative dimension that is added in the process.” (Journalist Category, 2013).

“Using, accessing the data and making them available to third parties. Reusing the data would be using them for another type of application, carrying out a data mining operation on them, crossing them with other databases to generate new information, that is, generating information and services secondarily from the raw data, crossing them, updating them.” (Academic Category).

In the interviewees perception, therefore, use would be related to simple access to available data (accessing, reading) and reuse would be associated with different forms of data processing, such as the development of an application or the drafting of a text, among others.

### **3.3.5. What is the impact of open data?**

With regard to the impacts of open data, one can say that it is already possible to visualize their configuration based on the products prepared with the data, as can be seen in Box 5. Let us consider here some of the impacts identified in the comments of the intermediates of open data. The impact of open data can be metaphorically described as a ripple or domino effect, as explained below.

One of their significant impacts is educational and cultural in nature: open data have become a topic of interest for Brazilian society. It is something that is being debated and around which training courses and seminars are being held, research is being conducted, prizes are being awarded, newspaper stories are being published, meetings of social movements are being held and public policies are being improved. While open data are not synonymous with “open budget data,” knowledge about them has a direct impact on public transparency, since issues related to government spending, especially problems such as squandering



of funds, corruption and the inability of governments to promote effective and efficient public policies, have aroused the interest of society. When qualified information is available, this interest spurs social control by organized civil society or individuals. And this research shows that NGOs, universities and the media are mobilized to develop a better understanding of the open data topic, data available on the Internet, so as to monitor competent agencies appropriately and politicize this field of dispute with the aim of promoting rights and citizenship.

This leads to other demands: intermediaries are unanimous in reporting difficulties in accessing data for different reasons. Some of them are not very familiar with the design of public policies, with the structure of spending and with the laws governing them. Others are not familiar with technological tools, including the most basic ones for structuring databases, such as Excel, and more sophisticated tools, such as *Siga Brasil*, among others. The intermediaries also reported difficulties such as databases that are overly differentiated or with poorly standardized categories, forcing them to build new databases for analysis purposes. All of these problems generate demand for a more reliable presentation of data for intermediaries to play their role of “mediating” complex data for society appropriately by making them more user-friendly. All of them agree that the government should invest in this process of promoting access to data by investing in technological development to make data available

through interactive and accessible applications (e.g. for mobile phones), improving the quality of the data and teaching the population how to access and use them. According to the government representative who was interviewed, there have been improvements in the process under way, but like in any cultural process the government itself needs time to change and develop the ability to not only open up data, but also to train the people who will be managing this process, making the data available to society (metadata) and keeping track of the impacts of such information. In this context, the use of open data is an act of social participation in itself, one that generates debate and spurs government to promote change.

A specific aspect deserves special mention, namely, its impact on the praxis of knowledge in different fields. The process of opening up data in Brazil has changed professions and we will highlight two examples. The first one is that of journalists, who in an era in which virtually everybody produces content and information have reinvented their craft in several ways: today in Brazil some groups are already discussing “data journalism” and trying to practice it, as is the case of the respondents involved in this research. The 2nd National Meeting on Open Data (Brasilia, 2013) was attended by a journalist who was invited to make a presentation on this topic and who translated the *Data Journalism Handbook* into Portuguese.

The other example is the change observed in routine research in applied economics, which must now deal with a new data standard for financial analysis. According to one of our respondents:

“The profile of applied research into public finance has changed a little, as you had no information before, so it was necessary to build databases, do primary research. Today the situation is different, as lots of information is available. You now have the SISTN, 27 transparency portals of the states, besides the portals of the state audit courts and those of the municipalities. So the problem in the field of applied research now is that of structuring data and extracting content that is consistent with what we need to carry out studies. A trend has been observed in the field of statistical economics around ‘big data’, i.e. how to work with this ‘informational broth.’” (Economist, Research Institute category).

Over a longer period, it will be possible to conduct specific research into this change in the world of labor, in which technology has an impact on professional practice and also on one’s conception of the world (ontologies), besides generating new forms of knowledge production (epistemologies).

Another impact that should be mentioned is how civil society organizations and movements are becoming more skilled in monitoring public policies and spending, a process that contributes to the expansion of rights. Open data,

especially through the work of intermediaries who process data and train people, have contributed to the empowerment of social movements and to progressive cultural change regarding access to public information. This is a complex phenomenon, as it implies processes for training social movements in how to carry out qualified advocacy. It also requires the establishment of a permanent network of hackers and developers for sharing codes and structures and also political issues intrinsic to the debate on open data. That is, we can foresee a situation where data in this field will be fed in a networked fashion, with a focus on producing knowledge and promoting rights.

Finally, we would like to highlight a last point of impact: all the respondents made a positive assessment of the Access to Information Act and of the processing of opening up data in Brazil and believe that this process contributes to promoting human rights, citizenship, the empowerment of society and, therefore, that it has a positive impact on Brazilian society. According to these mediators of information, the LAI lays down the foundations for regulating the process of opening up public data in Brazil and it is up to the relevant managers to enforce the Act broadly, with society's participation and under the supervision of the judiciary.

It is necessary to continue to monitor this process with the aim of recording changes and measuring impacts more precisely over longer periods of time (longitudinal research, for example).

### 3.4. Conclusions

Based on the survey and on our analysis of the respondents' perception, some conclusions can be drawn. In terms of the motivations of the intermediaries to work with budget transparency, we reached the conclusion that the selected intermediaries have both a political motivation - working with open data is part of their agenda of promoting public transparency and rights - and a professional motivation, namely, enabling their organizations to process and analyze open data enhances their legitimacy in relation to public dialogues and to intellectual production and media stories on public policies and spending, as well as on rights and citizenship.

As for the extent to which these intermediaries *know* what open data are, we can say that the respondents have this knowledge and, with respect to the eight principles as internationally defined, we saw that their knowledge of them is fairly good, especially about the first five ones, perhaps because these are the most relevant ones for their daily practice of processing and analyzing data. The respondents hesitated at some points to provide conceptualizations, which shows how new this field is from the academic point of view and also from the standpoint of social practice.

The objective of checking how intermediaries *use* and *reuse* open data was achieved and we can conclude that their use of them is varied, complex and very

intense. We also saw that they conceptualize use and reuse of data intuitively, and not yet in a systematized way.

In our checking of whether the intermediaries make the data available in an open format upon *reusing* them, we saw, based on Table 2, that they usually make information of their organizations available on free platforms or allow copyleft of its contents. This practice contrasts with the finding that the ‘license-free’ principle is not relevant when the respondents search for open data on official platforms on the Web, as most of them reported that they “don’t know” or “did not pay attention to this.”

The perceptions of the intermediaries on the impact of their work on expanding human rights in Brazil were checked, and it was seen that all of them perceive a general change in the culture of public transparency as a result of the recent process of opening up data in the country. As for their products, they reported impacts on the responses of society, both of social movements and of public institutions and official political spheres, to the products that they produce (Table 5).

Our final conclusion is that, despite the fact that they work in different institutions and that their roles and political views vary, the respondents tend to have very similar perceptions on open data in Brazil: advances in legislation have led to significant social change and the participation of civil society organizations in processing, analyzing and disseminating data is critical to democratize this process and enable it to promote rights and citizenship.

The research into impacts, considering the measurement of transparency at national and subnational levels, as well as in the fields of human rights and citizenship, revealed several issues both in relation to the process of opening up data in Brazil and to weaknesses and limitations that are yet to be overcome.

The quantitative research showed that the information provided is *complete*, that the data are not *primary* usually, that most governments do not provide *timely* information, that *access* to information is not guaranteed for all, that progress has been made in relation to the availability of *machine-processable* data, that access to them *does not discriminate* users, that data are presented in formats that can be described as fairly *non-proprietary* and that it is not known whether the data are under *license*.

The qualitative research revealed that the intermediaries perceive that progress has been made in relation to the rights brought about by open data in Brazil, that access to such data empowers civil society, that data reuse is systematic, creative and quite diverse, but that there is a demand among users for more complete, timely and reliable information.

It was also observed that the intermediaries are more aware of some principles than of others and that this phenomenon might have different causes. In this regard, some intersections between the two surveys should be highlighted: according to the intermediaries, the data that they access are *complete*, not very *timely* and *non-discriminatory*. Regarding whether the data are *primary*

## 4. FINAL CONSIDERATIONS

and *machine-processable*, principles that were seen to be unmet in the quantitative research, the intermediaries regard these principles as important only if the data is used through software and apps - for those intermediaries who use them only for reading and writing purposes, these principles are of secondary importance. As for *accessibility*, the quantitative research showed that the portals were not designed to meet the needs of people with disabilities, an aspect that was not captured in the qualitative investigation perhaps because none of the intermediaries who were interviewed needed to rely on alternative means of access. As for the *non-discriminatory* principle, although all the sites surveyed in the quantitative research provide free access to data, this is not a relevant principle for the interviewed intermediaries, which reveals the weakness of the debate on data security and privacy and Internet regulation among users. The same can be said in relation to the *non-proprietary* principle, which is not an object of structured reflection on the part of the majority of the intermediaries either. Finally, with regard to the *license-free* principle, although the group of



interviewed intermediaries included a few free-software activists, none of them knew whether the sources of the consulted open data used license-free digital structures - which is not surprising, as none of the surveyed government portals provides this information.

Finally, we can say that a change is under way in Brazil in the debate on open data and, consequently, on different and related fields: the field of public transparency, especially budget transparency; the field of digital technologies and new media; and the field of rights, which crosscuts the previous two. Paradigms are being broken as a result of this intersection, such as that of Brazilian patronage of public affairs. And there are also other elements being reconfigured and built, as well as tensions and negotiations - civil society and governments are key actors in these new processes, each taking on their responsibilities and playing their respective roles in society.

BERNERS-LEE, T. “Is your Linked Open Data 5 Star?” (2010). In: <http://www.w3.org/DesignIssues/LinkedData.html>. Accessed on: 01/10/2014.

BOLÍVAR, M. P. Rodríguez, PÉREZ, C. Caba and HERNÁNDEZ, A. M. López. “Cultural contexts and governmental digital reporting. International Review of Administrative Sciences,” 2006, 72(2), pp. 269-290.

BOLÍVAR, M. P. Rodríguez, PÉREZ, C. Caba and HERNÁNDEZ, A. M. López. “E-Government and Public Financial Reporting: The Case of Spanish Regional Governments,” The American Review of Public Administration, 2007, 37(2), pp. 142-177.

CRAVEIRO, G.S., ALBUQUERQUE, J.P and SANTANA, M. S. “Assessing Open Government Budgetary Data in Brazil” International Conference on Digital Society, 2013, Nice, France.

CRUZ, C. F., FERREIRA, A. S., SILVA, L. M. and MACEDO, M. S. “Transparência da gestão pública municipal: um estudo a partir dos portais eletrônicos dos maiores municípios brasileiros”, Revista de Administração Pública, Jan./Feb., 2012, 46(1), pp. 153-76.

GROFF, J. E. and PITMAN, M. K.. “Municipal Financial Reporting on the World

## 5. REFERENCES

Wide Web: A Survey of Financial Data Displayed on the Official Websites of the 100 Largest U.S. Municipalities,” *Journal of Government Financial Management*, 2004, 53(2), pp. 20-30.

JORGE, S., SÁ, P. M., PATTARO, A. F. and LOURENÇO, R. P. “Local Government financial transparency in Portugal and Italy: a comparative exploratory study on its determinants,” 13th Biennial CIGAR Conference, Bridging Public Sector and Non-Profit Sector Accounting 9-10 June 2011, Ghent (Belgium).

PINA, V., TORRES L. and ROYO, S. “Are ICTs Improving Transparency and Accountability in the EU Regional and Local Governments? An Empirical Study,” *Public Administration Review*, 2007, 85(2), pp. 449–472.

ROBINSON, D., YU, H., ZELLER, W. P., and FELTEN, E. W. “Government Data and the Invisible Hand,” *Yale Journal of Law and Technology*. 2009, 11(1).

Open Knowledge Foundation, “Beyond Access: The Right to (Re)Use Public Information,” 2011, pp. 8-9.

UNESCO, MDS. The System for Evaluating and Monitoring Social Development Programs and Policies: the case of the Ministry of Social Development and the Fight against Hunger in Brazil Policy Papers /17. Jeni Vaitsman, Roberto W. S. Rodrigues & Rômulo Paes-Sousa, Management of Social Transformations. Brazil: 2006.

<http://opendatahandbook.org>

<http://dados.gov.br/dados-abertos/>

<http://www.od4d.org/2013/01/17/usoereuso/>

<http://br.okfn.org/2013/08/28/dados-meio-abertos-sobre-o-uso-e-reu-so-dos-dados-governamentais-brasileiros/>

[www.gpopai.usp.br](http://www.gpopai.usp.br)

[http://www.planalto.gov.br/ccivil\\_03/\\_ato2007-2010/2010/decreto/d7185.htm](http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2010/decreto/d7185.htm)

<http://www.governoeletronico.gov.br/acoes-e-projetos/e-MAG>

<http://edemocracia.camara.gov.br/>

<http://www.consocial.prefeitura.sp.gov.br/>

<http://www.cidadedemocratica.org.br/>



## 6. DIGITAL REFERENCES

competitions/4-qual-seu-sonho-para-o-xingu

<http://www.apublica.org/>

<http://www12.senado.gov.br/orcamento/sigabrasil>

<http://www.portaltransparencia.gov.br/>

<https://www1.siop.planejamento.gov.br/acessopublico/?pp=acessopublico>

<http://www.spm.gov.br/>

<http://www.pac.gov.br/>

<http://www.planejamento.gov.br/>

<http://www2.camara.leg.br/transparencia/dados-abertos>

<http://dadosabertos.senado.gov.br/>

[http://www.receita.fazenda.gov.br/pessoajuridica/cnpj/cnpjreva/cnpjreva\\_solicitacao.asp](http://www.receita.fazenda.gov.br/pessoajuridica/cnpj/cnpjreva/cnpjreva_solicitacao.asp)

[https://www.contaspublicas.caixa.gov.br/sistncon\\_internet/index.jsp](https://www.contaspublicas.caixa.gov.br/sistncon_internet/index.jsp)

<http://portaltransparencia.recife.pe.gov.br/codigos/web/geral/home.php>

## **INESC: 2013 budget transparency portals**

This questionnaire can be answered several times for the same URL/API for each year. It includes 20 questions.

### **Browsing and searching for information**

This set of questions should be answered from the standpoint of those who browse the transparency page. Only online tools should be considered; if you need to download any file, even a simple text, that depends on a program other than the browser, this file should not be considered in this set of questions.

#### **Electronic address (url): \***

Kindly put your answer here: \_\_\_\_\_

#### **Budget data were found: \***

Kindly choose the options that apply:

( ) on the visited domain

( ) On another domain redirected from the visited page

The domain of a page is the main part of an Internet address, such as 'brasil.gov.br' or 'tce.sp.gov.br'. In most cases, there is an initial identification such as 'www' or 'portal' that can change during navigation, but the one responsible for the data is the same; in this case, it is considered as being in the same domain.

**In you were redirected, the redirected electronic address (url) is: \_\_\_\_\_**



## ANNEX 1.

### QUESTIONNAIRE — QUANTITATIVE RESEARCH

#### **Only answer this question under the following conditions:**

The answer was ‘on another domain redirected from the visited page’ in question ‘2 [a2]’ (Budget data were found :)

Kindly put your answer here: \_\_\_\_\_

#### **The budget transparency page is maintained by: \***

Kindly choose only one of the following options:

- City Hall of the Municipality
- State Government
- Federal Government
- Audit Court

#### **The budget data refer to: \***

Kindly choose only one of the following options:

- Municipality
- State Capital
- State/Federative Unit
- Union
- Audit Court

In some situations, the unit responsible for the page only indicates its own expenses, in which case you should choose the ‘unit responsible for the page’ option.

**Federative Unit: \***

Kindly choose only one of the following options:

- AC, Acre
- AL, Alagoas
- AP, Amapa
- AM, Amazonas
- BA, Bahia
- CE, Ceara
- DF, Federal District
- ES, Espirito Santo
- GO, Goias
- MA, Maranhao
- MT, Mato Grosso
- MS, Mato Grosso do Sul
- MG, Minas Gerais
- PA, Para
- PB, Paraiba
- PR, Parana
- PE, Pernambuco
- PI, Piaui

- RJ, Rio de Janeiro
- RN, Rio Grande do Norte
- RS, Rio Grande do Sul
- RO, Rondonia
- RR, Roraima
- SC, Santa Catarina
- SP, Sao Paulo
- SE, Sergipe
- TO, Tocantins
- Others

Kindly choose the options that apply:  
 general information about the budget  
 revenues  
 expenditures

**In relation to revenues, the following was found on the site (for humans): \***

**Only answer this question under the following conditions:**

**Name of the federative entity: \***

Kindly put your answer here: \_\_\_\_\_ The answer was in question '8 [p1] e.g.: State of Sao Paulo, Audit Court of the Municipality of Sao Paulo.

**The page provides information on: \***

Kindly choose the appropriate answer

for each item:

	Yes	I don't know	No
--	-----	--------------	----

Anticipated revenue			
Recorded revenue			
Recorded revenue			
Collected revenue			
Economic category			
Origin			
Type			
Heading			
Item			
Subitem			
Details below a subitem			

**Observations about revenues:**

**Only answer this question under the following conditions:**

The answer was in question '8 [p1] '(The page provides information on:)

Kindly put your answer here: \_\_\_\_\_

**In relation to expenditures, the following was found on the site (for humans): \***

**Only answer this question under the following conditions:**

The answer was in question '8 [p1] '(The page provides information on:)

Kindly choose the appropriate answer for each item:

	Yes	I don't know	No
--	-----	--------------	----

Budgeted expenditure			
Committed expenditure			
Settled expenditure			
Paid expenditure			
Agency			
Budgetary/managing unit			
Function			
Subfunction			
Program			
Action			
Nature of the expenditure			
Source of funds			
Type of tender			
No. of the tender process			
Beneficiary			
Good or service provided			

**Observations about expenditures:** \_\_\_\_\_

**Only answer this question under the following conditions:**

The answer was in question '8 [p1] '(The page provides information on:)

Kindly put your answer here: \_\_\_\_\_

**There are time series data: \***

Kindly choose the options that apply:

- ( ) for the current year
- ( ) for the last 2-5 years
- ( ) for the last 6-10 years

some years are missing

**What is the last date of recorded expenditure or revenue**

Kindly put your answer(s) here:

day

month

year

**An obstacle/technical requirement was found for accessing the data or site:**

Kindly choose the options that apply:

CAPTCHA

Required Flash plugin

Required Java plugin

Missing audio for DV

Others: \_\_\_\_\_

Consider anything that needs to be installed beyond the browser and obstacles that must be dealt with by humans, such as CAPTCHA. If there is any mechanism that makes it difficult for visually impaired individuals (VI) to access the data for lack of audio, choose the option 'audio missing for VI.'

**According to the test carried out with the accessibility validator, does the site satisfy the e-MAG?\***

Kindly choose only one of the following options:

- yes
- no
- partially
- It was not possible to use the validator

Comment your choice here: \_\_\_\_\_

**Data download is available in the formats:**

Kindly choose the options that apply:

- PDF
- XLS/X
- CSV, ODC
- JSON
- XML
- XHTML/RDFa
- RDF
- Broken link

Others: \_\_\_\_\_

**Have you observed any restrictions for accessing data or other elements? \***

Kindly choose only one of the following options:

- prior user registration is required for accessing the data

- part of the data can only be accessed upon user registration
- registration is optional
- no access restriction

**As for data use licensing: \***

Kindly choose only one of the following options:

- there are no visible licenses for the data
- there are licenses and they restrict the use of data
- there are licenses that allow free use of the data

**Observations about licensing**

Kindly put your answer here: \_\_\_\_\_



## ANNEX 2.

### INTERVIEW GUIDE – QUALITATIVE RESEARCH

#### Measuring the impacts of Open Data –Interview Guide

##### 1. Heading

- 1.1. Date:
- 1.2. Start time of the interview:
- 1.3. End time of the interview:
- 1.4. Researcher in charge:
- 1.5. Medium (phone, Skype, face to face):

##### 2. Characterization of the interviewee

Mark the category to which you belong (NGO, Media, Academia, Hacker, Government).

- 2.1. Name:
- 2.2. Position:
- 2.3. Organization:
- 2.4. Since when have you been working with public budget/budget cycle? (in years);
- 2.5. Briefly describe the work of your organization (or individual work) in the areas of public budget, open data, human rights.
- 2.6 Does the wok of your organization help to promote human rights?

##### 3. Motivations

- 3.1. What are your motivations for working with public budget/open data/human rights?

3.2. What is the audience you want to reach?

#### **4. Open data**

##### 4.1. Products

4.1.1. Tell me a little about the products that you or your organization have been producing with budget information. (texts, articles, software, games, videos, training activities, advocacy with government).

4.1.2. On use of the data: how were the data collected at their source turned into another product? Briefly describe the process through which the data are used, what worked what didn't (narrative).

##### 4.2. Access to open data

4.2.1. What data are you accessing and which data suit your purposes? What sources have you been using to get budget information (websites)? Do you search for information directly on transparency websites or do you search for information that has already been processed by other sources? Ask the respondent to specify the addresses of the sites he or she consults.

4.2.2. What software do you use to access the data? Do you or your organization use free software (Linux, Ubuntu) or private software (Windows, Apple)?

4.2.3. How do you know that you have permission to turn the data you collect into other products?

Now I'd like to talk a little about open data.

4.2.4. Do you know what open data are? (This is not a test, if the respondent doesn't know it's not a problem).

4.2.5. If you know, how would you define open data? (If the respondent doesn't know, you should make a record that he or she doesn't know what open data are).

4.2.6. Do you know what use and reuse of open data mean? How would you define data use and reuse?

I will now ask you questions about the information you access on the main website where you work with budget-related topics (remind the respondent about the site he or she mentioned at the beginning of the interview)

4.2.7. Primary and complete: do you find the data you need? Are they disaggregated as you would like them to be? What level of disaggregation do you use? Have you ever looked for data and didn't find it? If yes, what kind of data?

4.2.8. Timely: The periodicity at which the information is provided is adequate?

4.2.9. Accessible: Do you find it difficult to access the data? If yes, what kind of difficulty do you face?

4.2.10. Machine-processable: are the data available for automatic machine-based reuse (excel, csv, xlm)? Or do you find data in formats that require processing that is almost manual (pdf)?

Non-discriminatory: can you access data directly on the sites with which you work or is some form of registration required to access them?

License-free: can you tell if you have explicit permission to analyze/reuse the data that you access? (this information is usually provided at the bottom of the home page in small print)

Non-proprietary: when trying to download data from websites, did you notice whether you had to use a specific, private software both to download and save the information (Windows, Office, Java, another one)? Did you have to buy a specific software to work with this information?

## **5. Use of open data:**

What is most important for you when working with the budget data that you collect (the answers have to do with the characteristics of open data - try and prioritize):

- 5.1. Complete (not important at all, important, very important)
- 5.2. Timely (not important at all, important, very important)
- 5.3. Accessible, i.e. that can be accessed without any requirement such as registration (not important at all, important, very important)
- 5.4. Machine-processable (not important at all, important, very important)
- 5.5. Non-proprietary (not important at all, important, very important)
- 5.6. Non-discriminatory (not important at all, important, very important)
- 5.7. License-free (not important at all, important, very important)
- 5.8. The availability of data in an open format, i.e. in a format that meets the characteristics listed above, has changed your way of working? How? Does it

allow you to produce new information in ways that were not possible before?  
5.9. The way you make data available to your audience is also in open format?  
If not, how do you present the data to your audience?

## **6. Impact of open data**

6.1. Has any product been widely covered by the media or discussed in social networks or communities with which you work (radio, open meetings, schools, services)?

6.2. How do you assess the use of data you produce overall? Do you think it has empowered you, your organization, the communities with which you work?

6.3. How would you describe the impact generated by the information you produce in the budget area in general? With regard to human rights?

6.4. What is this perception based on?

6.5. How do you assess the Access to Information Act in relation to human rights?

## **7. Suggestions**

7.1. If you had to make suggestions to those in charge of the websites you use (in the budget area), what would you suggest?

7.2. If you could improve your work in this field, what would you do?

7.3. Any additional comments?

## **8. Identifying new intermediaries**

8.1. Can you mention up to three (3) persons or organizations that reuse budget data made available by the Union?

Support:

