

Teaching Open Data for Social Movements: a Research Strategy

1. Introduction

Since the year 2009, the release of public government data in open formats has been configured as one of the main actions taken by national states in order to respond to claims of transparency and participation by the civil society. The United States (<http://data.gov>) and the United Kingdom (<http://data.gov.uk>) were pioneers, and today over 46 countries have their own Open Government Data Portal¹, many of them fostered by the Open Government Partnership (OGP), an international agreement aimed at stimulating transparency.

The premise of these open data portals is that, by making data publicly available in reusable formats, society would take care of building applications and services, and gain value from this data (Huijboom & Broek, 2011). According to the same authors, the discourse around open data policies also includes increasing democratic control and participation and strengthening law enforcement.

Several recent works argue that the impact of open data policies, especially the release of open data portals, are still difficult to be assessed (Davies & Bawa, 2012; Huijboom & Broek, 2011; Zuiderwijk, Janssen, Choenni, Meijer, & Alibaks, 2012). One important consideration is that “The gap between the promise and reality of OGD [Open Government Data] re-use cannot be addressed by technological solutions alone” (Davies, 2012). Therefore, sociotechnical approaches (Mumford, 1987) are mandatory.

The target users of open government data lie over a wide range that includes journalists, non-governmental organizations (NGO), civil society organizations (CSO), enterprises, researchers and ordinary citizens who want to audit governments actions. Among them, the focus of our research is on social (or grassroots) movements. These are groups of organized citizens at local, national or international level who drive some political action, normally placing themselves in opposition to the established power relations and claiming rights for oppressed groups.

A literature definition gives a social movement as “collective social actions with a socio-political and cultural approach, which enables distinct forms of organizing the population and expressing their demands” (Gohn, 2011).

Social movements have been using data in their actions repertory (as can be seen in Table 1 and Listing 1) with several motivations. From our experience, an overview of

¹ According to <http://data.gov/open-gov/>, access in 26/12/2014.

several applications where social movements use open data reveals some of these motivations. A better understanding of reality, a more solid basis for their claims, or even building a counter-hegemonic discourse based on data produced inside the social movements were the main motivations detected.

An interesting example is the Citizen Public Depth Audit Movement (<http://www.auditoriacidada.org.br/>) which takes place in Brazil. This movement, which is part of an international network (<http://www.citizen-audit.net/>), claims that “significant amounts registered as public debt do not correspond to money collected through loans to the country” (Fattorelli, 2011), and thus origins of this debt should be proven. According to the movement, in 2014 Brazil paid 45% of its Federal spent in debt services.

Recently, a number of works tried to develop comparison schemes between open data strategies (Atz, Heath, & Fawcet, 2015; Caplan et al., 2014; Ubaldi, 2013; Zuiderwijk & Janssen, 2014). Huijboom & Broek (2011) listed four categories of instruments applied by the countries to implement their open data policies: voluntary approaches, such as general recommendations, economic instruments, legislation and control, and education and training. One of the conclusions is that the latter was used to a lesser extent.

Social movements, in general, are composed by people with less experience of informatics, either for lack of opportunities or interest. Although it is recognized that using data is important for the social movement objectives, the training aspect still hinders a wider use of it.

In order to address this issue, an open data course for social movements was designed. Besides building a strategy on open data education, the course also aims to be a research strategy to understand three aspects:

- The motivations of social movements for using open data;
- The impediments that block a wider and better use; and
- Possible actions to be taken for enhancing the use of open data by social movements.

This paper describes the methodology used for developing the open data course for social movements and its associated research, expecting that this approach can be useful for the Community Informatics field. In the sequel (Section 2), we picture the open data scene in Brazil, where this research was driven, contextualizing with some related social movements actions. Section 3 presents some open data education experiences, and

Section 4 lays out the methodological inspirations taken on the course and in the research. A description of the open data course is presented in Section 5, followed by the analysis of its presentation, in Section 6. Finally, we draw some conclusions in Section 7.

2. Open Data in Brazil and Social Movements

In this section, a general panorama of open data in Brazil is given, together with some perspectives on the use of open data by social movements.

2.1. Open Data in Brazil

The first concrete action of the Brazilian government regarding the publication of data on the Internet was the Federal Transparency Portal (<http://transparencia.gov.br>), released in 2004. Even without using the term “open data”, this initiative offered the citizens rich information about public expenditure. The main goal of the Federal Transparency Portal is to fight corruption and misuse of public money. Besides the Federal Transparency Portal, all Brazilian states also have their own versions, as well as many cities.

Between 2004 and 2011, several legal instruments were published in order to create a political infrastructure to support a wider open data policy. Finally, at the end of 2011, Brazil joined the Open Government Partnership (OGP) and published its Freedom of Information Act (FoIA), the so called Access to Information Law (Brasil, 2011). Recently, an evaluation of the two years under the law showed that only 69% of the demands received responses, and only 57% received satisfactory answers (Michener, Moncau, & Velasco, 2014).

Shortly after the publication of the FoIA, the Brazilian Open Data Portal (<http://www.dados.gov.br/>) was released. It is presented as a tool “for promoting dialogue between actors of the society and government, to establish the better use of data on behalf of a better society”. At the present time, it numbers almost 500 datasets, and more than 4000 resources.

2.2. Open Data and Social Movements

A closer look at the activities of social movements can easily detect a wide use of data in their discourses and forms of actions. Data is used to denounce specific situations, to attest the results of their actions and to understand the contexts where they are working.

In order to illustrate some examples, Table *Erro: Origem da referência não encontrada* shows some campaigns driven by social movements and examples of data used.

Table 1: Examples of use of data by social movements.

| Campaign | Data Elements |
|--------------------------------|--|
| Against Pesticides | Each Brazilian citizen is exposed to 5.2l of pesticides every year. (Brazilian Union of the Industry of Agricultural Defence Products, 2010) |
| For a land owning limit | In Brazil, 0.9% of the land owners control 45% of the arable area of the country. (Brazilian Agricultural Census, 2006) |
| Against the closing of schools | In the last 10 years, more than 24 thousand schools were closed in Brazil. (Brazil School Census, 2013) |
| Violence against Women | 70% of the women in the world have suffered or will suffer some kind of violence. (World Health Organization, 2012) |

Besides those examples, where civil society organizations use official data to ground their discourses and claims, there is another use which is worth mentioning: social movements which produce data by themselves. This is the case, for example, of Comissão Pastoral da Terra (a Christian service for rural areas), which built a database of violent events in rural areas (<http://cptnacional.org.br/>), and two maps of Environmental Justice, one covering the whole country (<http://www.conflitoambiental.icict.fiocruz.br/>), and another focused on the state of Minas Gerais (<http://conflitosambientaismg.lcc.ufmg.br/>). Moreover, the Solidarity Economy Movement promoted a mapping which identified almost 22,000 Solidarity Economy Enterprises in Brazil. This database was used to build an information system (Tygel & Alvear, 2011).

With this outlook, we can conclude that data is an important element in social movements' elaborations. But some elements still hinder its use, and a sound one is the lack of special training.

3. Education in Open Data

Maybe the most prominent initiative on teaching open data comes from the School of Data (<http://schoolofdata.org/>), an initiative by Open Knowledge and Peer 2 Peer University. They define themselves as working “to empower civil society organizations, journalists and citizens with the skills they need to use data effectively”, under the slogan “Evidence is Power”.

In 2014, School of Data organized 90 events taking place in 30 countries, reaching over 2000 participants. Besides Europe, where most of them happened, School of Data

reached places like Lebanon, Nigeria, Indonesia, Mexico, Brazil, Bosnia and Herzegovina, Tanzania and Philippines, training and exploring data about water, elections, and many other issues (School of Data, 2014). Open Knowledge offer courses in Germany, with a special focus on Data Journalism (<http://okfn.de/training/>).

Initiatives on open data education were also reported to be fostered by countries as the United States, the United Kingdom, Spain, Australia, and especially in Denmark, where the focus is on standardization of open data strategies between different government institutions (Huijboom & Broek, 2011).

Fioretti (2011) also remembered the importance of using open data in schools, emphasizing that it could help connect school curricula with real life and stimulate active citizenship in the students. The need of some skills to understand data, such as mathematics, was also mentioned. The author proposes two main lines of action: using open data, and producing open data as an official school policy.

4. Action Research and popular education as research method

Based on the previous discussion, we propose an open data course designed for social movement activists. The course has its pedagogic foundation on the Popular Education theory (Freire, 2005), which suggests an emancipatory practice, where several kinds of knowledge are considered valid, not only the academic, and the knowledge building is a joint approach between educators and educands². The Popular Education starting point is the educands' context, experiences, and knowledge. This theory is grounded mostly on Paulo Freire, a Brazilian educator and philosopher, who started his experience by teaching the alphabet to workers in the north-east of Brazil, one of the poorest regions of the country, in the 1960's.

The main work of Popular Education is the book *Pedagogy of the Oppressed* (Freire, 2005), written in 1968, when he was in exile, ideologically persecuted by the civil-military dictatorship in Brazil. It was only published in his own country in 1974. *Pedagogy of the oppressed* was translated into more than 40 languages, and has already has 50 editions in Brazil.

To understand the concept of dialogicity (dialogue between ideas instead of imposition of one idea over another), it is worth citing Freire's analogy of the banking education:

² Freire proposed the terms *educator* and *educand* in order to express the dialectic relationship between them in the knowledge building process, in opposition to the oppressing relation between teacher and student.

For the anti-dialogical banking educator, the question of content simply concerns the program about which he will discourse to his students; and he answers his own question, by organizing his own program.

For the dialogical, problem-posing teacher-student, the program content of education is neither a gift nor an imposition – bits of information to be deposited in the students – but rather the organized, systematized, and developed “re-presentation” to individuals of the things about which they want to know more. (Freire, 2005, p. 95)

In the case of open data teaching, it is important to offer the student-activists, who already use data in their work, a systematized view of the whole process, and to present some tools that allow them to enhance their data based analysis and synthesis capacity. By understanding the context where people already use open data in social actions, it is possible to teach based on students' interest in knowing more. Methodologies derived from Popular Education for open data teaching are explained in details in Section 5.

Another theoretical concept which inspires this work is Participatory Action-Research (PAR) (Morin, 2004). This methodological paradigm associates with every research a transforming action, whose planning and application should always be carried out with the researched subjects. Using the PAR perspective in community informatics research is a natural choice, since both approaches are centred in the community, both for practice and analysis.

PAR was used in the community informatics field for building a portal in a poor community in Rio de Janeiro, Brazil. In that case, research about community based organizations was the starting point to build a web portal run by those organizations (Alvear & Thiollent, 2011).

One of the main references in Brazil on PAR lies in Michel Thiollent, who suggests that action research

consists in coupling action and research in a process, in which the implied actors participate, together with the researchers, to interactively elucidate the reality in which they are inserted, identifying collective problems, seeking and experimenting solutions in real situations. Simultaneously, there is production and use of knowledge. (Thiollent, 2009, p. 14)

Fals Borda and Rahman (1991) also emphasize the importance of generating knowledge from practice in Latin American context. In this perspective, PAR is an appropriate methodology to generate knowledge for oppressed groups and social transformation.

Finally, Baskerville (2007), within a more pragmatic North American PAR perspective, presents some possibilities for developing information systems using this methodology. In Section 6, we discuss the research results gathered from the open data course according to a PAR perspective.

5. Course Description

Motivated by research on use and publication of open data by social movements and grounded on popular education and PAR principles, an open data course was developed. According to the dialogicity principle, the course objective is double: (i) to tackle the issue of open data education, indicated to be one of the factors hindering the use of open data; and (ii) to use the training stages to observe the activists using data and gather information for the research.

The course programme was elaborated seeking a balance between the social aspects of the use of data, which acted as motivators, and the technical issues, as tools for data manipulation. The methodology switches between expository stages and individual and collective activities by the students. It is expected that, at least, the students can achieve a critical view about data, understand the possibilities and limits of its use, be aware of the political questions involved in data production and publishing, and, finally, have a technical starting point for manipulating data.

The course is divided into four stages of four hours each, but can be adjusted to needs of the people involved. A website containing teaching materials, links to data sources, and a discussion forum was developed (<http://dados.eita.org.br/>), which in each presentation of the course is fed with more data.

Only two requirements are asked of people interested in attending the course: basic knowledge in informatics (web navigation) and a computer, which could also be offered by the organizers. Good quality internet access provided by the organizers is also highly desirable.

5.1. First Stage – Introduction

The first stage starts with a short description of the course, and the participants are informed they will also be contributing to a research project. This stage is intended to get people on the same level, by discussing sociotechnical and political aspects about data. The aim is to start from the educands' own experience, as suggested by the Popular

Education approach. For this, all participants are asked to present themselves, state their expectations and why he or she decided to take part.

Afterwards, a challenge is proposed: some socially relevant statistical results are presented (see Listing 1), and the educands are asked to find the data sources related to those figures. Doing the inverse data path (information to data), we expect to raise curiosity and show, in practice, the importance of knowing what is behind the statistics.

Listing 1: Example of data driven statements used to stimulate a critical view of data sources (based on Brazilian statistics agencies)

| | |
|---|---|
| 1 | 0.9% of the biggest landowners own 45% of arable land in Brazil |
| 2 | In Brazil, white men earn more than white women, who earn more than black men, who earn more than black women |
| 3 | 77% of young people killed in 2011 in Brazil were black |
| 4 | 46.7% of Brazilian exportation in 2013 were basic products, 12.6% were semi-manufactured, and 38.4% were manufactured |

In the sequel, several open data related topics are discussed:

- **How does data arise:** a data path is presented, from the occurrence of something, passing through its systematization until publication. Concepts such as facts, dimensions, and measures are discussed, together with the political motivations and consequences of those design choices. This topic is intended to put data neutrality in question, by showing that data produced by research is an outcome of several choices, made according to some goal.
- **Data visualization:** the same dataset can be observed in many ways, and the conclusions to which one may come heavily depend on this. Visualizing data as tables, graphics, networks (graphs), or maps may reveal different aspects and induce several kinds of conclusions.
- **Open Data:** In this topic, we motivate the understanding of open data using analogies (see Table 2). In the sequel, we define open data according to the David Eaves' three rules: data must be findable in the Web, published in machine readable formats, and cannot have licenses which prevent re-use (Eaves, 2009). A debate about linking and semantically marking data through the use of Linked Open Data (LOD) is also proposed with examples.
- **Transparency Policy:** At this point, we present the context of open data in Brazil and in the world, especially through transparency policies. It starts with

the Freedom of Information Act (FoIA), and goes up to Internet governance, with the Brazilian recent regulation based on three foundations: net neutrality, privacy and freedom of expression.³ International efforts on transparency, such as the Open Government Partnership (OGP) are also presented.

- **Synthesis:** After presenting all topics, educands are asked to discuss about how open data is related to their activism.

Table 2: Open and closed analogies to help understand what open data is.

| Open | Close |
|--|--|
| Text in digital format (txt, odt, doc) | Printed Text |
| Presentation in editable format (odp, ppt) | Presentation in PDF format |
| Source code | Executable software |
| Raw Data | Information (statistics, graphics, maps) |

5.2. Second Stage – Data Sources

The second stage of the course is dedicated to an overview of some important datasets on the Internet. It is worth noting that some of them are not “open” by the classical definition (Eaves, 2009), mainly because the raw data is not available for download. However, when an aggregate data querying system is offered, it makes data even more useful for common user than if raw data was available.

Different forms of accessing data are discussed. We recognize that, in respect to data access means, there is a trade-off between the ease of analysing data and the autonomy one can have in assessing their own conclusions. When a database is published in raw data, following all open data principles, this could still not help a citizen who want to know how much was spent on education in his city. Large amounts of data coded in specialized formats (e.g. R, SPSS, SAS, SQL, XML, RDF) allows a high analysis autonomy level, but special skills are needed to work with it. On the contrary, aggregate data, reports and charts allow people to have access to this information, but it has already passed through someone else's filter. Figure 1 depicts this debate.

³ An English version of the “Marco Civil da Internet” can be found at:
<https://www.publicknowledge.org/documents/marco-civil-english-version>.

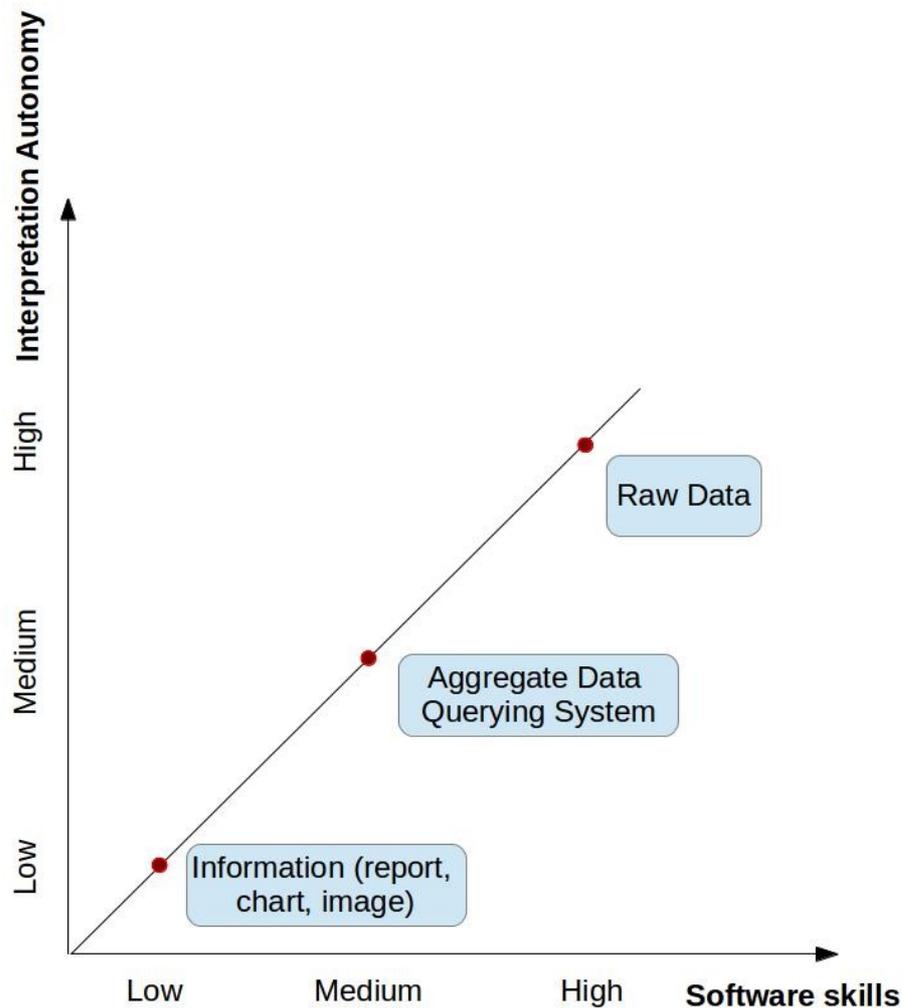


Figure 1: Trade-off between interpretation autonomy and software skills needed.

Besides the data access means, we propose a classification of data according to the type of provider:

- Data produced by the state:** This is the wider category, since the state has structural conditions and legal liability to produce data. In Brazil, the biggest data producer is the Brazilian Institute of Geography and Statistics (IBGE, in Portuguese), responsible for demographic, economic, geographic and many other sorts of data. The Unified Health System (SUS, in Portuguese) is also an important data generator, mostly about health and illnesses. Worldwide, the United Nations (UN) and the World Bank are also important data suppliers. Even though they are not governments, most of their data is compiled from

country data. It is important to emphasize that this kind of data carries with it the visions and ideologies of those who generated it. All the design choices made during the data production, including definition of facts, dimensions, and measures, in some degree follows the government intentions.

- Data produced by the state and shown by society:** In many cases, data produced by the state is not open, and when it is open, there are no tools for the citizens to easily analyse and take their own conclusions. Specialists are needed in order to translate data into useful information. In order to tackle this issue, many society-driven applications using official data have been recently released. In many cases, they help visualize data in a way that it leads to conclusions against the states' interest. One example is the Brazilian's "Congress Owners" application. Based on raw (and hard to analyse) donation data published by Electoral Justice, a civil society organization has developed an application where people can easily access and visualize the amount of donations received by politicians and parties, or paid by enterprises or economic sector (<http://donosdocongresso.com.br/>).
- Data produced by society:** The case where organized groups of the civil society produce their own data is interesting because: (i) as in the case of state data, data produced by the civil society contains its ideological influences in the design choices; (ii) it allows other perspectives on subjects already explained by the state. Data related stories can oppose well established hegemonic opinions. One example is the Brazilian Map of Environmental Injustice. Agribusiness is considered to be a good development alternative for the country, based on its relevant contribution to the gross domestic product. The map shows 82 occurrences of Environmental Injustice related to the agribusiness (from a total of 501), where activities of this sector cause damage to poor communities and/or to the nature. Table 3 shows a number of society driven databases. It is worth noting that, in some cases, the funding for building those databases comes from the Government. In principle, we consider that this does not hurt society autonomy to put their view in the design process.

Table 3: Examples of society driven databases, used by social movements with several purposes.

| Initiative | Address | Publisher |
|---------------------------------------|---|------------------|
| Environmental Injustice Maps (Brazil) | http://www.conflitoambiental.icic.t.fiocruz.br/ | Fiocruz and FASE |

| | | |
|---|---|--|
| Environmental Conflicts Maps (Minas Gerais, BR) | http://conflitosambientaismg.lcc.ufmg.br/ | Federal University of Minas Gerais, Brazil (in partnership with a number of social movements) |
| Atlas of Environmental Justice | http://www.ejolt.org/maps/ | 23 worldwide organizations (see http://www.ejolt.org/section/team for a complete list) |
| Agroecology initiatives | http://agroecologiaemrede.org.br | ANA/ABA (Brazilian national social movements related to agroecology) |
| Land Conflicts | http://cptnacional.org.br/ | Comissão Pastoral da Terra (CPT) |

In the final activity of this stage, educands are asked to add new data sources to the course web page, according to their interests. New sources can come from students' experiences, or be searched for during the class time. However, it is important to find the exact link, since this is reported to be a difficulty, as it will be seen later.

5.3. Third Stage – Tools

In the third stage, focus is given on tools for manipulating data. The goal is to present means for working with raw or aggregate data resulting from queries. It begins with an introduction to the Comma Separated Values (CSV) format, which is an open, universal and easy-to-use way of exchanging tabular data. Concepts such as primary and foreign keys are also discussed, in order to help comprehend how relationship between tables and databases can be made. Nevertheless, database design is beyond our scope.

This is an essentially practical stage. Several tools are presented, so that each student can choose which one he or she wants to work with, according to individual interests and ability with computers.

The first tool presented is a spreadsheet editor. The task consists in downloading a CSV sheet with a two dimensional table (production of food in tons, by type of food and year) and drawing a line chart. Students are also asked to plot percentage changes between first and last year production. The second part of the task consists in working with dynamic tables, which allows building analysis frameworks with more than two dimensions.

Other tools presented are related to map building and infographics drawing. Sometimes a mathematical background revision is necessary, since working with number variations requires some previous knowledge of percentages.

5.4. Fourth Stage – Final Work

The fourth and final stage is dedicated to a jointly decided activity. The goal is to develop some data based communication product, based on the three previous stages. Ideally, there should be more than one facilitator in the room, so that the work can be divided into groups, with each group being accompanied by one instructor.

Some suggested options are writing news text based on data, and building infographics and maps on specific subjects. The intentionality – what and why we want to communicate – is discussed first. Then, we evaluate the feasibility of the task – is there data about this subject? – and finally, the communication instrument is chosen. In the end, results are presented and an evaluation of the course is done.

The next section brings an analysis of presentations of this course, and draws some research results based on it.

6. Course Analysis

The course described above was presented five times in the second semester of 2014, in Brazil. While three presentations happened in Rio de Janeiro, one was held in Vitória (state of Espírito Santo) and another in Porto Alegre (state of Rio Grande do Sul). Two presentations were held in a workers union and three in universities, organized by groups who work with social movements in extension projects. A total of 52 students enrolled and participated in at least one stage. There were no fees pay, and the only requirements were basic informatics knowledge and a computer, sometimes provided by the organizers. Table 4 shows a summary of the presentations.

Table 4: Summary of the presentations of the open data course for social movements.

| N. | Kind of Place | City | Duration (h) and time distribution | Participants enrolled | Forms responded |
|--------------|---------------|----------------|------------------------------------|-----------------------|-----------------|
| 1 | Union | Rio de Janeiro | 16h (4 days at night) | 6 | 2 |
| 2 | University | Rio de Janeiro | 16h (2 full days) | 11 | 3 |
| 3 | University | Vitória | 16h (4 days at night) | 13 | 4 |
| 4 | University | Porto Alegre | 12h (1 half day/1 full day) | 12 | 3 |
| 5 | Union | Rio de Janeiro | 8h (two nights) | 10 | 3 |
| Total | | | 68 h | 52 | 15 |

The analysis will be based on two instruments: an evaluation questionnaire that all students were asked to fill in, and a participant observation driven during the presentations. The goal of the analysis is to respond to the research questions: (i) why

social movements use data (motivations); (ii) what are the mains problems (impediments); and (iii) what could be done to enhance the use (improvements). Also, the evaluations about the course can be used to improve it.

6.1. Questionnaire based analysis

All the participants were requested to answer a questionnaire after attending to the course. Thus, we assume that the opinions given are strongly influenced by the discussions held over the course. This decision was taken having in mind that: (i) open data is not a subject of the educands' everyday life; so, answering before the course could lead to meaningless results; (ii) according to the popular education methods, we expect each educand to be able to relate content unseen before with its experiences and synthesize in the end their own conclusions about the process. Table 5 shows the questionnaire and the mean, maximum, and minimum values for numerical questions.

Table 5: Questionnaire answered by course attendants. All the numerical results are in over a base of 15 (n = 15), and N/A means “not applicable”.

| # | Question | Mean (maximum - minimum) |
|---|---|--------------------------|
| 1 | Age | 31 (25–48) |
| 2 | Knowledge of informatics (1: poor knowledge – 5: good knowledge) | 2.7 (1–5) |
| 3 | Work/Profession/Activism | N/A |
| 4 | Why have you attended to the course? Why do you think open data is important? | N/A |
| 5 | Educator's performance (didactics, material, knowledge, punctuality) (1: poor – 5: very good) | 4.5 (3–5) |
| 6 | Self performance (participation, attention, punctuality) (1: poor – 5: very good) | 3.3 (1–4) |
| 7 | Was the subject according to your expectations? (1: totally distinct – 5: totally according) | 4.6 (2–5) |
| 8 | What is the main impediment perceived by using data? | N/A |
| 9 | How do you imagine that the use of data could be improved? | N/A |

The medium age of participants was 31 years, with the youngest being 25 years old and the oldest 48. They considered themselves to have medium knowledge of informatics. Before enrolling, participants were asked to have some informatics knowledge, but no admission tests were given.

Some participants were exclusively activists or academics, but most of them were activists with some academic involvement. There were journalists, lawyers and social scientists, all engaged with some social movement. No participant had formal informatics training, meaning that no one was an informatics expert.

The teacher's performance was well rated, but this was somehow expected in a free course. On the other hand, no one rated him or herself with very good participation performance. In Question 7, only one participant seemed to have very different expectations about the course content. All the others marked 4 or 5, indicating that open data is not so distant of non-informatics people's lives, at least for those who answered the questionnaire.

In order to analyse questions 4, 8 and 9, we will pick answer elements and classify them according to research goals: motivations, impediments, and improvements. Question 4 was aimed to catch motivations, but impediments and improvements were also cited. Question 8 raised only impediments, and Question 9 only improvements, as wished.

An effort was made in order to extract concrete elements from the discursive text. An equilibrium was sought between merging similar statements and not losing opinion diversity. These concrete elements extracted can be seen on the Appendix, in Tables 6, 7 and 8.

Sometimes, the separation between the classes is not very clear. All impediments (e.g. "Open Data Portal is hard to use") have implicit improvements (e.g. "Open Data Portal could improve usability"), as all improvements also have implicit impediments. Some motivations (e.g. "Use spending data to fight corruption") also could be interpreted as impediments (e.g. "Few spending data is available") or improvements (e.g. "More spending data must be made available"). We tried to classify according to the respondent's intention.

6.2. *Observation based analysis*

In this section, some remarks are made based on the 68 hours observation of the course. This observation was driven inspired on the ethnographic method of participant observation (Atkinson & Hammersley, 1994). Within this approach, the researcher plays an established participant role in the studied scene, in this case, as an educator, taking field notes during the class. Ethnography inspired methods are complementary of objective and quantitative evaluations, since, according to Atkinson & Hammersley (1994), ethnography deals with the "analysis of data that involves explicit interpretation of the meanings and functions of human actions", and "represents a uniquely humanistic, interpretive approach, as opposed to supposedly 'scientific' and 'positivist' positions." Since two of our research questions deal with human actions and feelings – what are the motivations of social movements for using open data and what are impediments that block a wider and better use – we considered the participant

observation an appropriate methodological direction. We aimed to comprehend the point of view of the educands, and this was done from the educator stance, which certainly influenced the analysis.

As described in Section 5.1, in the first stages of the course, participants are shown statistical statements (see Listing 1), and are asked to search for data that generated those figures. Below, we list some behaviours observed:

- The first impulse of users is to paste the phrases directly into a web search engine. Normally, the results are news commenting that statement, or reports containing that information, and never the actual data source;
- For some people, it is difficult to understand the difference between the statements and the data sources from which they were originated. One way to overcome this misunderstanding is to slightly rephrase the statement and ask what would be the new figures. For example, relating to statement 1 (Listing 1), we would ask: “how much land do the 0.1% of the biggest landowners possess?”;
- Overall, only few people reached the actual data source. This shows that one of the main problems of data sets and their query/download systems is that they are frequently hidden in the deep web, i.e., regular search machines cannot find them.

In Section 5.2, we described the second stage of the course, where some data sources are presented, divided in three categories. About this stage, we would like to remark:

- In general, although interested, users are unfamiliar or unaware of data sources. This ignorance is, as expected, worse for society driven OGD based applications, and for data produced by social movements, which usually have no official means of dissemination;
- Students were stimulated to add new data sources to the course website, according to their own interests or activism. In some cases, participants inserted already known data sources, but in most cases data sources were found during the activity.

The practical stage, described in Section 5.3, revealed one of the strongest difficulties in open data usage: the manipulation of software tools, particularly of spreadsheets. The knowledge about CSV tabular files, considered as a fundamental skill to use data on different systems, was practically absent. This problem got even worse because of the

inability of the most used proprietary spreadsheet application (MS Excel) to deal with such kind of file. LibreOffice, its open source counterpart, facilitates this task.

Another issue that was highlighted in this stage was the mathematical difficulty faced by most of the students. Dealing with statistical open data requires, most of the time, simple mathematical operations. Therefore, sometimes a small revision of percentage was necessary.

Unfortunately, the fourth stage, described in Section 5.4, did not work as expected. This stage was only reached in two of the five presentations described in Table 4. In the first one, students, mainly journalists, decided to individually write stories and impressions about open data. They were published in the course website (<http://www.dados.eita.org.br/category/reflexoes/>). The second experience reached closer to the goal: participants decided to investigate a local case of environmental injustice. Data about enterprises, population health, environmental licensing and other issues were gathered, but no final product was obtained. In the remaining three presentations, the time ran over twice, and once the students said they were tired, as this course was run on two full days, at a weekend.

One possible way to overcome this issue is to propose this work at the beginning of the course and organize tasks during the other stages. This has the advantage of motivating students with a concrete problem during the course. Nevertheless, the challenge remains on how to prepare the course without predefining the problem. Another option would be to increase the number of hours, what would depend on participants' availability.

6.3. *Synthesis*

As explained above, by a simple rephrasing, an impediment or a motivation can turn into an improvement. By doing a careful analysis of Tables 6, 7 and 8 (Appendix), an improvement classification tree was built. It is aimed at orienting actions for engagement of social movements in open data in the Brazilian context. The classification tree can be seen in Figure 2.

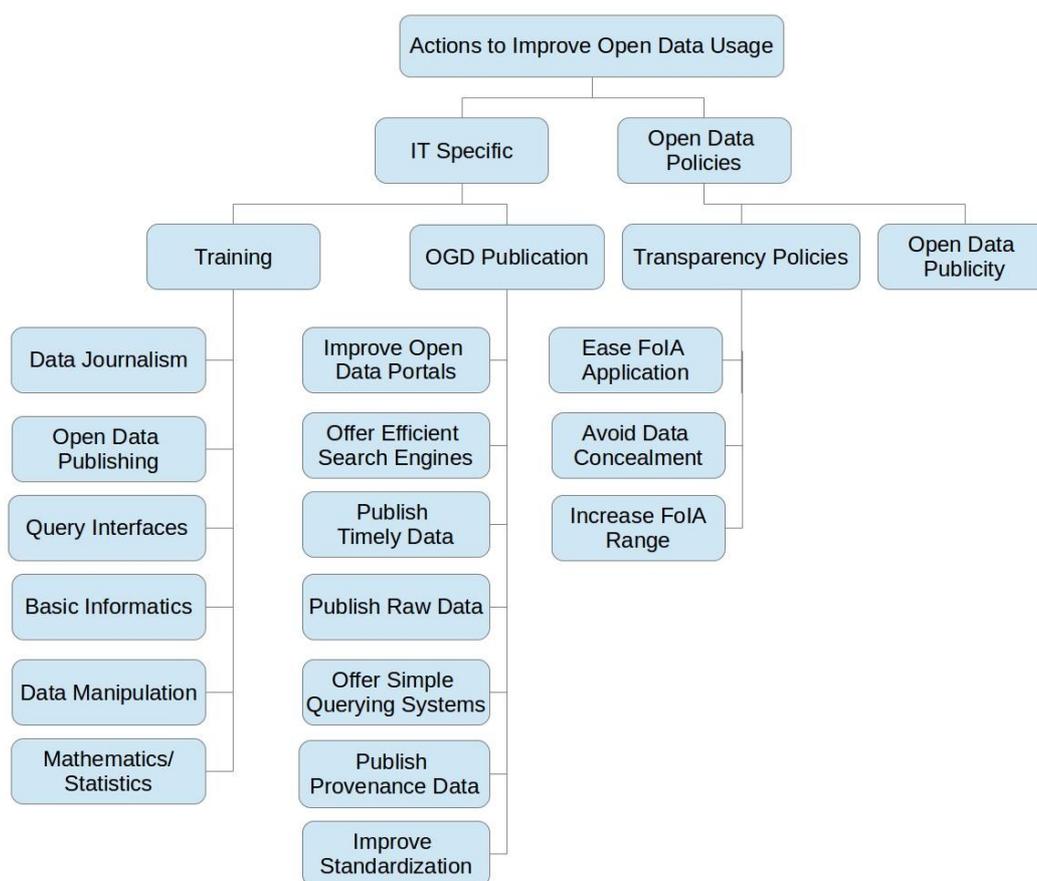


Figure 2: Classification tree for open data engagement actions, systematized from Tables 6, 7 and 8.

The first classification is a distinction between *Information Technologies (IT) Specific* and *Open Data Policies* related issues. There is no intention to imply a duality between social and technical issues, however, one can easily recognize that some elements are directly related to information technology, and others are not.

The *IT Specific* issues are divided into *Training* and *Open Government Data (OGD) Publication*. The first class encompasses cited impediments which could be approached with educational investments, and the second is related to actions to be taken by government data publishers. Our proposed course tackles all cited educational demands, except data publishing, since it still demands a higher level of informatics knowledge. As to *OGD Publication* related issues, the need for better search engines was the most cited enhancement.

The right side of the tree presents general issues related to *Transparency Policies* and *Open Data Publicity*. We can conclude that in order to improve open data usage, actions

must be taken far above data level. In this case, the whole structure for information access must be enhanced. Difficulties in claiming the FoIA within local government levels were reported, as well as accessing information on private foundations who run public money. Finally, many participants suggested that more publicity on open data already available would also improve usage.

7. Conclusions

This article described a methodological strategy used in research on Open Data and Social Movements. An open data course was proposed in order to help understand the motivations for the use of open data by social movements, the impediments to succeed on this, and the improvements that could help the engagement of society with open data.

The results of five presentations of the course were organized into a classification tree of possible open data engagement actions, shown in Figure 2. Training related actions should be fostered as part of a Transparency Policy, but social movements should also prioritize open data in their agenda.

Some improvements related to OGD publication could be addressed by using new technologies being developed under the Linked Open Data (LOD) framework. By semantically annotating data with commonly used vocabularies and ontologies, the LOD approach offers technical means to link different data sources and jointly query them. A solid set of tools to implement LOD is being developed (Auer, Bryl, & Tramp, 2014), but strong efforts must be driven in order to hide the complexity of the representation and to highlight its benefits so that it can be considered a viable option. Other improvements are only possible through effective political willingness of governments for being transparent.

As a methodological approach for research in informatics, the course has shown to be an efficient tool, since it accomplished its dialogical function indicated by the Popular Education theory. At the same time as they were subjects on an open data education action, the educands that participated on the course acted as objects in a research project. The dialogical approach resulted, for one side, in a set of appointments for open data publishers, and, on the other side, in a satisfactory educational experience, as shown by the good educator evaluation (Table 5) and by the rich answers collected (Tables 6, 7 and 8).

The open data engagement actions gathered here are not supposed to be complete. The methodology used to raise these actions allows only to affirm that they can improve the

use of a specific actor – social movements, who think about open data with a special intentionality. Nevertheless, our efforts sums to a number of other initiatives (Atz et al., 2015; Davies, Fernando, & José, 2013; Gurstein, 2011; Ubaldi, 2013; Veljković, Bogdanović-Dinić, & Stoimenov, 2014; Zuiderwijk & Janssen, 2014), encompassing other actors, aiming to picture a desired open data scenario in the future.

8. Acknowledgement

We would like to thank all the educands who took part in the courses, and especially to the organizers of the 5 presentations: Journalists' Union of Rio de Janeiro, Citizenship University/Federal University of Rio de Janeiro, ORGANON (Federal University of Espírito Santo), and GPACE (Federal University of Rio de Grande do Sul), all in Brazil.

9. Bibliography

- Alvear, C. A. S. de, & Thiollent, M. (2011). Participatory Development of Technologies as a Way to Increase Community Participation: the Cidade de Deus Web Portal Case. *The Journal of Community Informatics*, 7(1-2), 1–19.
- Atkinson, P., & Hammersley, M. (1994). Ethnography and Participant Observation. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 248–260). Thousand Oaks: Sage.
- Atz, U., Heath, T., & Fawcet, J. (2015). *Benchmarking Open Data Automatically*. London.
- Auer, S., Bryl, V., & Tramp, S. (2014). *Linked Open Data -- Creating Knowledge Out of Interlinked Data*. (S. Auer, V. Bryl, & S. Tramp, Eds.) (Vol. 8661). Cham: Springer International Publishing. doi:10.1007/978-3-319-09846-3
- Baskerville, R. (2007). Educing Theory from Practice. In *Information systems action research: an applied view of emerging concepts and methods*. Laredo, Texas, USA: Springer.
- Brasil. (2011). Lei de Acesso à Informação. *Lei N^o 12.527 de 18 de Novembro de 2011*.
- Caplan, R., Davies, T., Wadud, A., Verhulst, S., Alonso, J. M., & Farhan, H. (2014). *Towards common methods for assessing open data: workshop report & draft framework* (Vol. 31). New York.
- Davies, T. (2012). Supporting open data use through active engagement. In *Proceedings of the W3C Using Open Data Workshop* (pp. 1–5). Brussels.

- Davies, T., & Bawa, Z. A. (2012). The Promises and Perils of Open Government Data (OGD). *Journal Of Community Informatics*, 8(2), 1–6.
- Davies, T., Fernando, P., & José, A. (2013). *Researching the emerging impacts of open data ODDC conceptual framework*.
- Eaves, D. (2009). The Three Laws of Open Government Data. Retrieved January 22, 2015, from <http://eaves.ca/2009/09/30/three-law-of-open-government-data/>
- Fals Borda, O., & Rahman, M. A. (1991). *Action and knowledge: breaking the monopoly with participatory action-research*. Appex Press.
- Fattorelli, M. L. (2011). *AUDIT: an essential tool to prove the origins and causes of the current “debt” crises in the United States and Europe*.
- Fioretti, M. (2011). A proposal to promote Open Data from and for the schools. Retrieved January 23, 2015, from <http://mfioretti.com/2011/10/warsaw-open-data-and-education/>
- Freire, P. (2005). *Pedagogy of the Oppressed* (30th ed.). New York: Continuum.
doi:10.1007/BF02195128
- Gohn, M. da G. (2011). Movimentos sociais na contemporaneidade. *Revista Brasileira de Educação*, 16(47). Retrieved from <http://www.scielo.br/pdf/rbedu/v16n47/v16n47a05.pdf>
- Gurstein, M. B. (2011). Open data: Empowering the empowered or effective data use for everyone? *First Monday*, 16(2), 1–7.
- Huijboom, N., & Broek, T. Van Den. (2011). Open data: an international comparison of strategies. *European Journal of ePractice*, 1(12), 1–13.
- Michener, G., Moncau, L. F. M., & Velasco, R. (2014). *Estado Brasileiro e Transparência - Avaliando a aplicação da Lei de Acesso à Informação*. Rio de Janeiro: Fundação Getúlio Vargas.
- Morin, A. (2004). *Pesquisa-Ação Integral e Sistêmica: Uma Antropopedagogia Renovada* (1st ed.). Rio de Janeiro: DP&A.
- Mumford, E. (1987). Sociotechnical Systems Design: Evolving Theory and Practice. In G. BJERKNES, P. EHN, M. KYNG, & K. NYGAARD (Eds.), *Computers and democracy: a Scandinavian challenge* (pp. 59–76). Michigan: Avebury.
- School of Data. (2014). School of Data Report 2014. Retrieved January 19, 2015, from <http://2014.schoolofdata.org/>

- Thiollent, M. (2009). *Pesquisa-ação nas organizações* (2nd ed.). São Paulo: Atlas.
- Tygel, A. F., & Alvear, C. A. S. de. (2011). The Development of an Information System for the Solidarity Economy Movement. *Journal Of Community Informatics*, 7(1-2), 1–12.
- Ubaldi, B. (2013). *Towards Empirical Analysis of Open Government Data Initiatives*. *OECD Working Papers on Public Governance* (Vol. NO.22).
doi:10.1787/5k46bj4f03s7-en
- Veljković, N., Bogdanović-Dinić, S., & Stoimenov, L. (2014). Benchmarking open government: An open data perspective. *Government Information Quarterly*, 31(1), 278–290. doi:10.1016/j.giq.2013.10.011
- Zuiderwijk, A., & Janssen, M. (2014). Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly*, 31(1), 17–29. doi:10.1016/j.giq.2013.04.003
- Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., & Alibaks, R. S. (2012). Socio-technical Impediments of Open Data. *Electronic Journal of E-Government*, 10(2), 156–172.

Appendix

Table 6: Motivations, Impediments and Improvements indicated in answers to Question 4.

| Question 4: Why have you attended to the course? Why do you think open data is important? | | | |
|---|---|--|--|
| # | Motivations | Impediments | Improvements |
| 4.1 | Work with data and link different information to create arguments | There is a mismatch between amount of data released and the capacity of social movements to analyse it | Make investments in education for open data use |
| 4.2 | Be able to work with data driven journalism | There are many barriers to access information | Promote publicity about existence of data |
| 4.3 | Use data to denounce injustices | Open Data is unknown for most social movements | Improve knowledge about how to search for data |
| 4.4 | Data can give basis to stimulate new claims | There is no full transparency in government actions | Enable access to information, without discrimination |
| 4.5 | Translate data into information for readers | Most of the people have little informatics ability | |
| 4.6 | Produce data in juridical research | | |
| 4.7 | Open data can stimulate | | |

| | | | |
|------|--|--|--|
| | analysis | | |
| 4.8 | Open data can stimulate new data | | |
| 4.9 | Validate/legitimate arguments in communication with data | | |
| 4.10 | Use data to understand the capitalist society | | |
| 4.11 | Understand the resistances against oppression with data | | |
| 4.12 | Fight corruption using spending data | | |
| 4.13 | Make better use of information, a central point in class conflicts | | |
| 4.14 | Unveil data manipulation | | |

Table 7: Impediments pointed in answers to Question 8.

| <i>Question 8: What is the main impediment perceived by using data?</i> | |
|---|---|
| # | Impediments |
| 8.1 | The lack of knowledge about data production process makes interpretation difficult |
| 8.2 | It is hard to understand data connection and linking possibilities |
| 8.3 | Finding data in the web is hard /Open data portals are complicated |
| 8.4 | Access to data outside the web is hard / FoIA application is complicated |
| 8.5 | Data organization is confused |
| 8.6 | Data formats does not help its use |
| 8.7 | The state presents data through different platforms, which increase the need for training |
| 8.8 | The need of specific software tools makes data usage harder |
| 8.9 | Some important data is concealed |
| 8.10 | Most data is outdated |
| 8.11 | The querying interfaces present too much information |
| 8.12 | Access to raw data is hard |
| 8.13 | Government agencies do not follow common data standards |
| 8.14 | Data interpretation is difficult, especially on specific subjects |
| 8.15 | Linking data from different sources is difficult without appropriate tools and metadata |

Table 8: Improvements indicated in answers to Question 9.

| <i>Question 9: How do you imagine that the use of data could be improved?</i> | |
|---|---|
| # | Improvements |
| 9.1 | Provide user-friendly interfaces |
| 9.2 | Provide education on statistics/mathematics |
| 9.3 | Standardize open government data |

| | |
|------|--|
| 9.4 | Provide user-friendly language (avoid technical terms) |
| 9.5 | Provide wider training possibilities |
| 9.6 | Promote more advertising of open government data initiatives |
| 9.7 | Promote more advertising of social movements open data initiatives |
| 9.8 | Foster more research on open data and social movements |
| 9.9 | Improve data search engines |
| 9.10 | Increase the offer of open data sources |
| 9.11 | Avoid the need of intermediaries for data interpretation |
| 9.12 | Improve open data portals |