Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 2020 Proceedings | Human-Computer Interaction (SIGHCI)

Aug 10th, 12:00 AM

Investigating Communicability Issues in the Open Data Manipulation Flow

Leonardo Pio Vasconcelos
Federal Fluminense University, lpiovasconcelos@gmail.com

Raissa Barcellos
Federal Fluminense University, rbarcellos@ic.uff.br

Jose Viterbo
Federal Fluminense University, viterbo@gmail.com

Flávia Bernardini
Federal Fluminense University, fcbernardini@id.uff.br

Luciana Cardoso de Castro Salgado
Fluminense Federal University (UFF), lsalgado2006@gmail.com

See next page for additional authors

Presenter Information
Leonardo Pio Vasconcelos, Raissa Barcellos, Jose Viterbo, Flávia Bernardini, Luciana Cardoso de Castro Salgado, and Daniela Trevisan

Investigating Communicability Issues in the Open Data Manipulation Flow

Completed Research

Leonardo Vasconcelos
Federal Fluminense University
lvasconcelos@ic.uff.br

José Viterbo
Federal Fluminense University
viterbo@ic.uff.br

Raissa Barcellos
Federal Fluminense University
raissabarcellos@id.uff.br

Flavia Bernardini
Federal Fluminense University
fcbernardini@ic.uff.br

Luciana Salgado
Federal Fluminense University
luciana@ic.uff.br

Daniela Trevisan
Federal Fluminense University
daniela@ic.uff.br

Abstract

The open data movement advocates that public data should be available in electronic format and accessible via the Internet. As a consequence, large volumes of data have been made available in open data portals. To tackle the complexity and possible social impacts resulting from the overwhelming production, collection, processing of data, the Human-Data Interaction providing mechanisms for citizens to interact with data. In this paper, we explore the flow of open data manipulation, aiming to find problems in the application of the whole flow, in practice, by the citizens. We used the Semiotic Inspection Method to find communication breakdowns in the data collection and data visualization interfaces. The results pointed to some communicability problems such as non-intuitive interfaces, lack of tutorials, excessive difficulties in accessing platforms, inconsistent data, and limited resources. These problems make it difficult for citizens across the flow of open data manipulation.

Keywords

Open data portals, human-data interaction, semiotic inspection method.

Introduction

Nowadays, citizens (from here on out, we will use the word "users" to refer to "citizens") live in an era in which the amount of collected data is abundant and diverse, resulting in a complex ecosystem (Hashem et al. 2015). This landscape raises a question: how to assist the users in the interaction with such large amounts of data (Lausch et al. 2014)? The open data movement (Dawes et al. 2016; Davies and Bawa 2012; Lee et al. 2015), a relatively new, expressive, and relevant emerging force, is part of this scenario of massive data availability. It advocates that public data should be available in electronic format and accessible via the Internet. This movement aims to make available local, regional, and national data so that the population can manipulate them directly, using software tools for tabulation, visualization, mapping, among others (Gurstein 2011). As a consequence, today, there are several open data portals, which provide access to large volumes of data in numbers that grow fast around the world. This openness is due to the increasing adherence to the concepts of public transparency (Barcellos et al. 2017).

However, the large amount of data available increases the users' difficulty in obtaining useful information. Defining the ideal datasets for solving a problem may, for instance, require many hours of research in different portals. Also, users will have to spend a considerable amount of time integrating different datasets or identifying relationships between them. This task could be even more costly for a user with no experience (Pinto et al. 2018). While the production, collection, processing, and use of data have taken on new dimensions in terms of complexity and possible social impacts, researchers have proposed different approaches to tackle such issues. In a relatively new area of research called Human-Data Interaction (HDI), Mortier et al. (2014) intend to place the human at the center of data flows, providing mechanisms for users to interact explicitly with data. The primary purpose of HDI is to analyze the procedures used in data collection, filtering, and processing, as well as to verify the impacts of data productions on the individuals and society in which they live.

Brazil, according to Ruediger et al. (2017), occupies the eighth position in the world and the first in Latin America in the Open Data Index (ODI) ranking. The ODI compares the ability to make data available to civil society, people, and the media among countries around the world. However, most problems are tied to the usability of the data. Brazil lacks development in two points. The first is the ability to produce an understanding of the data. The second is the facility to access the data. The HDI approach proposed by Mortier et al. (2014) may be useful to point out solutions to the problems in the flow of open data manipulation (i.e., the whole process of collecting, merging, cleaning and processing open data, creating data visualizations, analyzing and disseminating results).

In this work, we aim to explore the flow of open data manipulation and to identify where the user's most significant difficulty lies. Our object of study is a Brazilian open data portal. Our general research question is: "How the interface communicability may harm the user's cognitive performance when assimilating all the flow of manipulation of open data?". To access the communicability of mechanisms that disable human-data interaction, we adopt Semiotic Engineering, a theory that considers human-computer interaction as a particular case of metacommunication (Souza and Leitão 2009). Hence, we applied the Semiotic Engineering Inspection Method (SIM) for scientific purposes to find out potential communication breakdowns at human-data interaction time in the flow of open data manipulation. This article is organized into six sections, including this one. In the next section, we present the conceptual background for our study. Then we discuss the related work to our research. Next, we offer the methodology used in our study. In Section 5, we present our findings. Lastly, in Section 6, we conclude the paper, summarize our contributions, and share new research questions and future work items.

Theoretical Reference

Open Data

The movement of open data is the first step towards public transparency (Barcellos et al. 2017). In the Brazilian context, the law on access to information - Law nº 12,527 of November 18, 2011 - was sanctioned, which establishes that, among other aspects, data must be made available in a primary, the complete, authentic and update in all public agencies, mixed capital companies, among others. However, there is still a long way for data considered public to be made available to the population. Yet, among some requirements is greater access to the massive databases maintained by the administrative authorities themselves (Belkindas and Swanson 2014). The plea for greater openness and transparency of data has been perceived by governments around the world.

Nowadays, there are several open data portals, providing access to large volumes of data around the world. This is a consequence of a large amount of public information continuously generated, which can refer to finances, health, human development, among others. This openness is due to the growing adherence to the concepts of public transparency. The flow of open data manipulation can be observed in five steps. First, the data is collected from some open data portal. Second, a data junction is necessary in order to gather all the data inherent to the research question. Third, the data must be cleaned and treated to remove empty, null, and inconsistent fields. Fourth, the data is ready, and data visualizations need to be created in order to reveal aspects associated with data processing. And finally, conclusions and inferences about the context of the data must be made. Unconsciously incorporating the data context by merely looking at its graphing is still a challenge for many. An open question in the scientific community is to analyze where the most significant difficulty of the user lies in the flow of manipulation of open data. There is a need to verify, from the result of this analysis, how we can provide mechanisms for users to interact explicitly with data, given the qualities and defects of this flow, in practice.

Human-Data Interaction – HDI

HDI is a concept proposed by Mortier et al. (2014). HDI stands at the multidisciplinary intersection between behavioral economics, statistics, sociology, psychology, computer science, among others (Strey et al. 2018). The researchers present HDI as a set composed of three central themes, being: readability, agency, and negotiability. Readability focuses on making clear and easy-to-understand analytical algorithms and data for the people who will use them, both in their presentation and in processing issues. The agency focuses on providing autonomy for users to act within data systems. Its purpose is to allow the user to choose whether or not to inform, repair, infer, or perform any such task in the system. Negotiability focuses on the dynamic relationships built around the data and its processing. The incessant practice of accumulation of data and the remarkable importance of these data and inferences extracted from them leverages the need for the study of HDI (Barreto et al. 2018). Mortier et al. (2014) state that if the HDI is deployed correctly, it can provide mechanisms to manage human interaction with data and data processing. Designers/developers should take on the challenge of building systems that not only give the users intentional action, but also that the results of unintentional behavior are predictable by users, both as individuals and as groups.

The Semiotic Inspection Method

SIM is a qualitative inspection method to evaluate the communicability of interactive systems. It focuses on the meanings expressed by the interactive interface elements (Souza et al. 2010). The application of SIM does not require the user's testing with the system. Quite the contrary, it is an inspection method, that involves specialists, that advocates for users while investigating potential communicability problems in the interface and its consequences for human-computer interaction, in our case, to human-data interaction. SIM can be applied either by a group or by only one researcher (Souza and Leitão 2009). The method is composed of a preparation phase, followed by five evaluation steps.

In the preparation, the evaluator should select which portion of the system will be evaluated and define the user profile. Another critical point is to identify the interactive scenarios that will guide the inspection. The scenarios should also determine the primary user's tasks (Souza et al. 2010). In the evaluation stage, the evaluator examines the interface and classifies each interface element (i.e., signs) as metalinguistic, static, or dynamic. Metalinguistic signs are usually found throughout the interface, either in instructions, explanations, warnings, and error messages, with a focus on online help and user manuals (Souza and Leitão 2009). They usually explain other signs. Static signs are those that communicate their meaning regardless of cause and effect relationships and can be interpreted from instant screen pictures. Dynamic signs, in turn, are usually represented by animations, opening and closing dialogues, transitions between screens, or modifications to the elements of a screen.

In the first three steps, the main goal is to reconstruct the metacommunication of the designer for each category of signs (metalinguistic, static, and dynamic). In Step 1 (Inspection of metalinguistic signs), the evaluator explores the documentation and help system. In Step 2 (Inspection of static signs), the evaluator inspects the static signs of the interface. In Step 3 (Inspection of dynamic signs), the evaluator inspects the signs that emerge from the interaction. In Step 4, the evaluator contrasts and compares the metacommunication messages from steps 1, 2, and 3 and records possible problematic interpretations that may occur in user interaction time. Finally, in Step 5, the evaluator appreciates the quality of metacommunication. In this step, the evaluator produces a report containing the communicability problems encountered, which may frustrate or prevent the user from understanding the message intended by the designer, affecting his productivity. In this method, the inspector defends the interests of the target user.

Related Work

Human-Data Interaction in Data Visualization

In the context of HDI, in order to make information more comprehensible by users, data visualization techniques can be used. The related works sought tend to make explicit a general picture between HDI and data visualization. In the work of Hornung et al. (2015), the authors focus on understanding human-data interaction as a process of meaning, identifying the goal of designing the interaction efficiently, employing a semiotic perspective, and setting research challenges for the field. The authors point out that the production, collection, processing, and use of data need to be systematically investigated in the projection of a human-data interaction, with a greater focus on the social impact they cause. Understanding the interaction with data as a process of signs, researchers warn of an approach that goes beyond the challenges of representation and meaning-making, also considering pragmatic and social issues related to meaning in the context, intentions, negotiations, and effects of data usage.

Chang et al. (2018) discusses the objective of supporting users in performing data analysis, observing how users interact with this data. The authors highlighted two important issues that human-data interaction systems still face: (i) Users tend to rely on a very small subset of the analytical tools these systems provide making them not take advantage of the total capacity and resources of that category of systems. (ii) Easy access to tools can easily lead users to misuse, faulty workflows, and false discoveries. This problem is accentuated by the automation of the stages of the analysis process, since humans, with the naked eye alone, cannot effectively process and reason over a large amount of data.

Human-Data Interaction in Application Scenarios

Currently, the academic community has been directing its focus on the data generated and not of the physical artifacts. Thus Roberts et al. (2014) studied Human Data Interaction in a museum and the feedbacks sent by the visitors. Assuming that a high amount of complex data can be arduous to interpret, especially for beginners, they have organized a way of exploiting this data for visitors. The information was personalized so that visitors could explore a slice of this data and have control over it. The researchers chose people outside the museum to conduct the experiment. They were recruited into groups larger than or equal to two people. All selected were over nine years old. Participants were encouraged to test a display that showed the census data. Each visitor received a tablet with a preconfigured profile to answer census questions, view custom data, and address their queries with the searcher through an app. The visitors' responses were sent to the server on the display network. Through it, you could view personalized information. Display datasets were changed by users when they clapped twice. Participants were free to interact with the system at any time they feel necessary. In the end, they completed a survey that assessed their perceptions of control and their use of the exhibit. Responses were evaluated using open comments and Likert scales with five points. Five video cameras and a microphone were used to record the interactions. The results are presented in two ways. Children's perspectives were influenced when they interact with the census data maps displayed in the museum. Individuals had their view correlated with the interactive presentations of the data and the ways of thinking used when analyzing the data.

Widjojo et al. (2017) presented a study using HDI based on Virtual Reality. The Human-Data Interaction was used as an interface between a visual and human representation of the data. According to them, the virtual reality can offer immersion and thus aid the thought in the analysis of the data. They also presented four research questions that are still open. The first is the need to determine what is needed to facilitate the understanding of users in the specific data set. The second is the need to choose which attributes of Virtual Reality are impressive to assist the people's understanding. The third is the need to select the appropriate visualization techniques and visualization that can activate the essential attributes of virtual reality. And, to conclude, the need to offer methods of evaluation of HDI systems compared to the conventional desktop system.

Methodology

The methodology used in this work consists, in large part, in the application of the Scientific Semiotic Inspection Method (SIM) (Souza and Leitão 2009) with the focus on the evaluation of the communicability considering the concepts presented by the HDI perspective. We use, therefore, the predictive paradigm, making use of an interpretative and qualitative method (Lewis 2015). In this work, we performed two experiments, applying SIM both in an open data portal and in the Power BI data visualization tool - a complete visualization tool widely used nowadays. We chose these two items, due to the constant need to use them, in the entire flow of open data manipulation - the first used in the data collection and the second in the data fusion, data cleaning, data visualization, and data report. The inspections were carried out by three authors of this work, two junior-level inspectors, with the supervision of a senior-level assessor (specialist in Semiotic Engineering). We wanted to answer our general research questions through applying

SIM investigating: (a) the users' main difficulties in the flow of open data manipulation; and (b) the potential breakdowns in the communication process at interaction time.

The user profile corresponds to an inexperienced and novice user in an open data portal, and reporting difficulties in using data visualization tools. Hence, SIM was run considering the scenario: (1) *Gabriel participated in a workshop of open data portals to cities at his university. To establish his knowledge, he volunteered to participate in an activity in which he could choose to solve two proposed challenges.* (2) *He is already used to research on the internet to address his day-to-day needs. He also uses some computing tools, such as IDEs, compilers, to aid him in programming studies. However, despite having a background in computer science, Gabriel has never worked with open data.* (3) *To solve the challenge, Gabriel will first need to enter an open data portal and download two databases. Then he will need to open the bases in a data visualization tool. And, to conclude, it will be necessary to analyze with possible solutions based on the data obtained and analyzed.*

To validate our study with SIM, we ran a triangulation step to identify the consistency and congruence of the results. To this, we performed empirical research, using the same domain and activities described in the scenario above with ten students of an Information Systems undergrad program, at Federal Fluminense University in Brazil. Everyone has necessary computer skills, such as accessing websites, programming, and using some computational tools. Following, we present more details about the procedures, analysis, and results of the SIM application and triangulation step.

Procedures and Results

SIM Applied To The Open Data Portal

In this subsection, we summarize the steps performed in the application of the SIM in the open data portal of the Public Security Institute of Rio de Janeiro[1]. As a matter of scope, only the signs on the link sent to the students were analyzed. Figure 1 presents some examples of the identified signs. When comparing the three metamessages (with metalinguistic, static, and dynamic signs), starting with the understanding of who, for the designer, is its user, it was verified that it is an individual who uses the internet a lot and has excellent experience in using data sets. When evaluating from the designer's perspective, what the user wants or needs to do, it has been verified that the user needs to quickly find a set of data on topics or groups of interest (Strong 2019).

On the way in which ways the user prefers to do this, it was verified that the efficiency, and clarity to achieve the results was a constant in the metamessages. Clearly, it can be observed that the most outstanding signs are: search field and menu of groups and themes. Lists (according to dynamic signs) play a central role in the communication between the user and the application. They were determined by the intelligence of the system to meet the user's own contexts and demands. This results in another feature of the user: he prefers to conduct the search and explore the site rather than the system offering recommendations.

Regarding the perspective of the system that the designer suggested for the user, it was verified that the user's expertise in downloading data allows the site to be explored without any clarification, both with the features and the tools. No distinction is made in the search for a database or its documentation. As to how you can or should use the system, you can see that it is designed to be efficient and straightforward, so your communication is predominantly textual and, in some cases, via clicks on the "themes or groups" menu. On achieving a range of objectives, it was observed that the designer expects the search to be high-speed and focused, without the need to go through any intermediate process, which may distract the user and consequently delay their search. It is thus perceived that the metamessages were mostly coincident or complementary. Except for the static sign that contained a link informing the redirect to another place in the portal with more information about the base, but clicking it was broken - better access to the data is necessary. It is worth noting that metamessages are an interpretation of what the evaluator believes to have been the designer's intention in developing the application and making its interface choices. However, after completing the SIM analysis, it was observed that the profile of our user was different from the user profile proposed by the designer of the open data portal of the Public Security Institute of Rio de Janeiro.

[1] http://www.ispdados.rj.gov.br/EstSeguranca.html

Our user is inexperienced, and he is in his first contact with an open data portal. In his view, the system does not have an intuitive interface and is not offered a tutorial or some help from the system. In his view, there are several listed bases, but without any explanation about them. There are buttons with format names, but the user is unaware of all these formats - emphasizing the need to improve options for data format. When the subject clicks on the search field (S.2), the listed bases disappear, and a message is returned saying: "Zero results found" (M.2) - missing data. The "themes" and the "groups" are not divided by intuitive categories (S2), and neither is ordered in alphabetical order, there is apparent disorganization. There is a link on the portal informing you of the possibility of solving doubts (S.9). However, clicking on that same link is broken. The inexperienced user will feel lost and without help resources for a beginner.

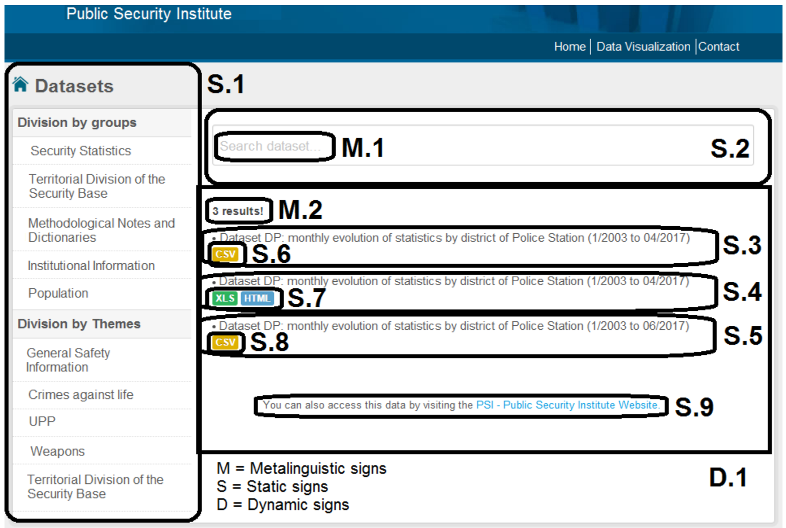

Figure 1. Signs In The Portal

The main communication problems are **Unclear interactive path in search:** Even though static signs describe titles (S.1 and D.1), formats (S.6, S.7, and S.8), search fields (S.2 and M.1), the portal fails because of the lack of clarity in how to direct your searches on the site. **Confusing Data:** Although the static "Data Sets" sign (S.1) presents the bases organized by "Groups" and "Themes," meaning that the data is held following a logical sequence, the bases are not ordered alphabetically nor is presented the criterion of the ordering of the subjects which confuses the user. Also, the item "institutional information" is presented in both groups. **Unclear formats for exporting data:** Although the static signs (S.6, S.7, and S.8) offer the export formats together with their description, meaning practicality and agility in downloading the data in the desired form, the formats do not appear in order. An explanation of each format is also not offered. **Confusing search field:** Although the static sign search field (S.2) means practicality and still be accompanied by a metalinguistic sign (M.1) explaining its functionality when the user clicks the field, the bases disappear from the screen, and a sign metalinguistic informs 0 results found. **Broken link:** The static sign "You can also access this data by going to the PSI's website (S.9)." It displays a redirect link, meaning that it will direct the user to another site. However, the link is broken, and when it is accessed, the error "404 - File or directory not found" is returned. **Little and misuse of metalinguistic signs:** Metalinguistic signs could help the user to understand all the functionalities offered by the portal, assisting the layman in to perform the tasks more efficiently.

SIM Applied To The Power BI Tool

In this subsection, we briefly summarize the steps that are performed in applying SIM in the Power BI tool version 2.69.5467.5201. Our inspection evaluated only the data manipulation and graphs construction due to our scope. Figure 2 presents some examples of the identified signs. When comparing the three metamessages (with metalinguistic, static, and dynamic signs), starting with the understanding of who for the designer, is its user, it was verified that it is an individual that uses Microsoft Power BI services and stays connected to the internet to ask recurring questions about the tool. He often requests several resources at the same time and is willing to update information in real-time.

When evaluating from the designer's perspective what the user wants or needs to do, it has been verified that the user wants to open a database, manipulate graphics, and make inferences in it quickly. Usability, clarity, and objectivity are considered fundamental by the designer to serve the user and assist him in his tasks. This is why menus assume a central role in the communication between the user and the application, as they are determined by the usability of the system to meet the users' contexts and demands. Thus it can be deduced that the user prefers to drive the program exploitation. Chart types and menu items (more emphasized by static signs) play an essential role both in the layout and in the use of the application. Although the tool is easy to use, graph information such as scale and color should be changed via tool features, which may present difficulty for beginners.

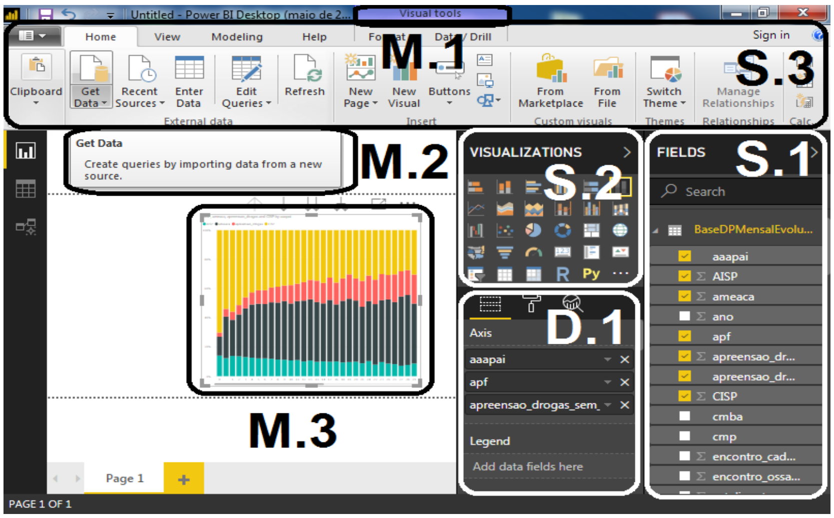

Figure 2. Signs In The Power BI Tool. M = Metalinguístic Signs, S = Static Signs, D = Dynamic Signs

About how the user prefers to do this was verified that because it is a desktop tool, the user can access it at any time on your computer. He wants to explore the system in general, minimizing tool interference, launching commands whenever he feels it is necessary. From the perspective of the system that the designer designed for the user, it was verified that the tool is not so intuitive, so an explanation is presented for each functionality. Also, there is a link that redirects to the main page of this tool, where it is possible to take courses for a better understanding. The tool needs to make explicit every feature and the possibility of use to expedite the user's decision making. About how you can or should use the system, it has been verified that the ability to connect via email makes the tool more robust to publish your results over the internet. However, if the user prefers, he can save his work on his local machine. Communication with the tool should be as accessible as possible. Each feature can be accessed most often by a click or by access to the two properties. Features can directly customize inferences to expedite user cravings. On achieving a range of objectives, it was observed that the designer expects the user to be able to manipulate the databases in an agile way. It is also likely that the user has the minimum of difficulty possible, so the features of this tool resemble those of the office product. The inferences are expected to be made in a clear and customizable manner. The user experience is expected to be productive, requiring as little reading as possible. In addition, the user is expected to have such an extensive immersion in the tool that no other distraction can delay him from reaching his goal.

After completing the SIM analysis, it was observed that our user profile was different from the user profile that SIM reveals as that intended by the Power BI tool designer. Our user is inexperienced and is in his first contact with an open data portal, reporting difficulties in using data visualization tools. In its view, the system has a beautiful interface, but not all functions are self-explanatory. Besides, not all explanations are clear. The main communication problems are potential: **Many tool options without good organization:** Although there is a logical division between the signs, separating the visual tool fields, types of visualizations (S.2 and S3), data fields (S.1) and filters, the user has many different options of functions (S3), which makes it difficult to perform tasks to change the color (D.1), data range and magnitudes, because the interface does not provide a proper organization of the available functions. **Difficulty in manipulating data and constructing views within the tool:** Despite the existence of static and dynamic signs that support the availability of several options of visualizations and functionalities, it is difficult to manipulate the data and construct the visualizations within the tool(S.1, S.2, S.3, and D.1), find relationships and draw conclusions about the data, it becomes a complex task for lay users. **Non-Intuitive Functions:** Although it resembles Microsoft Office products, not all features are intuitive (D.1). An effort is required to learn and record all available functions. **Lack of basic tutorials:** Despite the existence of metalinguistic signs (M.1 and M.2), the tool does not provide a basic, quick and efficient tutorial (S.3), teaching a layman the best way to handle the data, for example.

Triangulation

To have different perspectives of the same object, explored in different contexts, giving plausibility and consistency to the interpretation process, we performed the triangulation stage, a qualitative research procedure (Denzin 2003). To this end, we present an exogenous triangulation where participants - reported in the subsection Methodology - were encouraged to participate in two challenges, which involved the whole flow of open data manipulation. Participants were invited to visit an open data portal and download two datasets. After the data was collected, a junction, cleaning, and data processing were necessary. Students were also required to construct data visualizations in the Power BI tool BI (Parks 2014) and then perform various analyzes and inferences through the data visualizations. At the end of the challenges, the participants answered a questionnaire.

From the questionnaire, we extracted answers from two questions for the triangulation of the results. We carried out a discourse analysis (Mayring 2004) on the participants' responses. We derived categories from investigating what are the user's difficulties, considering the flow of data collection and manipulation in open data portals. The first question was: "What open data portals should provide to improve the analysis of the available data?". The categories extracted after qualitative analysis are: (1) Improve data access: unavailable, missing, and outdated data; (2) Improve data format options: formats not accessible to laypeople; (3) Data filtering: disorganized; scrambled data; (4) Better visualizations integrated: graphics integrated into datasets; (5) Increased disclosure: Most users do not know open data portals. The second question was: "What is the most challenging task in the entire flow of open data manipulation? (1) Using the Data Visualization Building Tool - Power BI. (2) Searching for data in portals. (3) Accessing specific formats. (4) Finding relationships and conclusions between data. (5) Finding the best way to display collected data.

We drew some conclusions after analyzing the results through the data collected with the application of the questionnaire. They are: (i) The most significant difficulty among the participants is to clean and process the data, after collecting them, in the portals. The participants report the problem in accessing the data, that is, a widespread presence of unavailable, missing, and outdated data. These factors make clear the importance of providing better data quality in the portals. Participants also complained that the data were not consistent. (ii) Although most participants reported ease of building views, it was evidenced that using Power BI was a problem. Observations such as: "complicated," "not very intuitive," "limited," "difficulty in learning the tool," "use time demand," "difficult handling for a layperson," "critical difficulty in using" and "relate more than one table increases complexity "have been documented. That is, the task of building the visualizations only becomes easier after spending time and understanding the tool. (iii) On the data collection, participants reported that the sites are "confusing and not friendly." However, they found it easy to collect the data requested in the challenges, since a link to each data set was made available in the questionnaire. (iv) Most of the participants reported difficulty in working with the file formats provided by the open data portals. (v) Some participants reported that further disclosure of open data portals is still required. (vi) Participants reported difficulty in observing relationships between data and providing conclusions about them, even after the task of constructing visualizations in graphic form. This context is aggravated when we see that some respondents put the task of "Analyzing and Reporting Data" as a challenging task. (vii) Some participants reported that it is necessary to deploy a better tool integrated with the portals for a more efficient visualization of data.

The communication breaks encountered in the SIM application converge with the difficulties presented by the user in the open data handling flow. We observed some critical aspects, within the context of HDI, that are visible from the results found in this work. Regarding readability, the participants showed difficulties during the flow, as they reported that there are inconsistent data, problems in data formats, and filtering of them. As for the agency, we conclude that, currently, in the analyzed portals, the user cannot be involved in collecting and storing the data, but only to use them. Regarding negotiability, we find that this term permeates all our research because our goal is to understand the difficulty of the user to obtain value from the data collection. Thus, all the inductions performed in this work, after the experiments, are factors that allow the continuous engagement of users so that they can perform data processing and obtain a value of them.

Conclusions

The main objective of this work was to find out communication breakdowns in the flow of open data manipulation by the users, aligning the conclusions obtained with the context of HDI. Looking to achieve our purpose, we applied SIM in an open data portal and a data visualization tool. To consolidate the obtained results using SIM, we conducted an empirical study with qualitative analysis through applying a questionnaire. We found that the results are converging and several factors that hinder the user throughout the flow of data manipulation, notorious problems, both in the open data portal analyzed, and in the data visualization tool. Non-intuitive interfaces, lack of tutorials, excessive difficulties in accessing platforms, inconsistent data, and limited resources are examples of problems that permeate the open data flow. In the context of HDI, we perceive elements contained in its three central themes and report points to be considered.

This paper brings one main contribution, the application of SIM in a yet unexplored context of HDI related to digital government research, showing that the implementation of the method was relevant in that context. Complementarily, it identifies a set of communicative problems used by designers that challenge the user's understanding. Besides, the description of aspects that must be worked out opens the venue for better interaction design to support users and open data. As a limitation, our study had a unique user profile. A more heterogeneous pattern could offer us more diverse feedback on the problems encountered. As future work, we intend to increase the sample of data comprising a heterogeneous group of participants. Then, report a set of suggestions to improve the communicability of the tools evaluated in our research. And finally, develop a platform that will facilitate the processing and interpretation of open data, given the difficulties encountered.

Acknowledgements

We would like to thank the precious comments of the anonymous referees that helped us to carefully and deeply improve our work. We also thank Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), connected to the Brazilian Ministry of Education, FAPERJ (Foundation of Support to the Research of State of Rio de Janeiro) and CNPq (Brazilian National Council for Research) for partially funding this work.

REFERENCES

Barcellos, R., Viterbo, J., Miranda, L., Bernardini, F., Maciel, C., and Trevisan, D. 2017. "Transparency in Practice: Using Visualization to Enhance the Interpretability of Open Data," in ACM International Conference Proceeding Series (Vol. Part F1282), New York, New York, USA: Association for Computing Machinery, June 7, pp. 139–148. (https://doi.org/10.1145/3085228.3085294).

Barreto, P., Salgado, L., and Viterbo, J. (n.d.). Assessing the Communicability of Human-Data Interaction Mechanisms in Transparency Enhancing Tools. (https://doi.org/10.15439/2018F174).

Belkindas, M. V., and Swanson, E. V. (n.d.). "International Support for Data Openness and Transparency." (http://opendatahandbook.org).

Chang, R., Fekete, J., Freire, J., Scheidegger, C., and Reports, D. (n.d.). "Connecting Visualization and Data Management Research," Dagstuhl Reports (7:11), pp. 46–58. (https://doi.org/10.4230/DagRep.7.11.46).

Davies, T., and Bawa, Z. A. 2012. "The Promises and Perils of Open Government Data (OGD)," The Journal of Community Informatics. (http://ci-journal.net/index.php/ciej/article/view/929/926, accessed April 25, 2020).

Dawes, S. S., Vidiasova, L., and Parkhimovich, O. 2016. "Planning and Designing Open Government Data Programs: An Ecosystem Approach," Government Information Quarterly (33:1), pp. 15–27. (https://doi.org/10.1016/j.giq.2016.01.003).

Denzin, N., and Lincoln, Y. 2003. "Collecting and Interpreting Qualitative Materials," p. 696. (http://books.google.pt/books/about/Collecting_and_Interpreting_Qualitative.html?id=CdCGek5KJ_QC&redir_esc=y, accessed April 25, 2020).

Gurstein, M. B. 2011. "Open Data: Empowering the Empowered or Effective Data Use for Everyone?," First Monday (16:2), First Monday. (https://doi.org/10.5210/fm.v16i2.3316).

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., and Ullah Khan, S. 2015. "The Rise of 'Big Data' on Cloud Computing: Review and Open Research Issues," Information Systems (47), pp. 98–115. (https://doi.org/10.1016/j.is.2014.07.006).

Hornung, H., Pereira, R., Cecilia, M., Baranauskas, C., Liu, K., and Baranauskas, C. 2015. Challenges for Human-Data Interaction-A Semiotic Perspective Publicação Realizada Durante o Doutorado View Project Exploring Principles for Affectibility in the Design of Mobile Applications View Project Challenges for Human-Data Interaction-A Semiotic Pe. (https://doi.org/10.1007/978-3-319-20901-2_4).

Lausch, A., Schmidt, A., and Tischendorf, L. 2014. Special Issue: "Remote Sensing of Inland Waters and Their Catchments" View Project Teaching and Learning in Remote Sensing View Project Data Mining and Linked Open Data-New Perspectives for Data Analysis in Environmental Research. (https://doi.org/10.1016/j.ecolmodel.2014.09.018).

Lee, M., Almirall, E., and Wareham, J. 2016. "Open Data and Civic Apps: First-Generation Failures, Second-Generation Improvements," Communications of the ACM (59:1), Association for Computing Machinery, pp. 82–89. (https://doi.org/10.1145/2756542).

Lewis, S. 2015. "Qualitative Inquiry and Research Design: Choosing Among Five Approaches," Health Promotion Practice (16:4), SAGE Publications Inc., pp. 473–475. (https://doi.org/10.1177/1524839915580941).

Mayring, P. (n.d.). "A Companion to Qualitative Research - Google Livros." (https://books.google.com.br/books?hl=pt-BR&lr=&id=6lwPkSo2XW8C&oi=fnd&pg=PP282&dq=Qualitative+content+analysis&ots=Zq-Z5nRoGl&sig=4tt2B6NdMw6F4hIobzKM2y7rYiY&redir_esc=y#v=onepage&q=Qualitative content analysis&f=false, accessed April 25, 2020).

Mortier, R., and Henderson, T. 2015. "Human-Data Interaction: The Human Face of the Data-Driven Society." (http://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:31995L0046).

Parks, M. (n.d.). "Microsoft Business Intelligence. POWER BI | Guide Books." (https://dl.acm.org/doi/book/10.5555/2721718, accessed April 25, 2020).

Pinto, H. D. S., Bernardini, F., and Viterbo, J. 2018. "How Cities Categorize Datasets in Their Open Data Portals: An Exploratory Analysis," in ACM International Conference Proceeding Series, Association for Computing Machinery, May 30. (https://doi.org/10.1145/3209281.3209377).

Roberts, J., Lyons, L., Cafaro, F., and Eydt, R. 2014. "Interpreting Data from within: Supporting Human-Data Interaction in Museum Exhibits through Perspective Taking," in ACM International Conference Proceeding Series, Association for Computing Machinery, pp. 7–16. (https://doi.org/10.1145/2593968.2593974).

Ruediger, M. A., Grassi, A., A., B., B., F., J. d. M., and Oliveira, W. (n.d.). "Open Data Index for Cities: Rio de Janeiro-RJ." (http://dapp.fgv.br/wp-content/uploads/2017/05/OpenDataIndexRJ.pdf, accessed April 25, 2020).

Souza, C. S., and Leitão, C. F. 2009. "Semiotic Engineering Methods for Scientific Research in HCI," Synthesis Lectures on Human-Centered Informatics (2:1), Morgan & Claypool Publishers LLC, pp. 1–122. (https://doi.org/10.2200/s00173ed1v01y200901hci002).

Souza, C. S., Leitão, C. F., Prates, R. O., Bim, S. A., and Silva, E. J. 2010. "Can Inspection Methods Generate Valid New Knowledge in HCI? The Case of Semiotic Inspection," International Journal of Human Computer Studies (68:1–2), pp. 22–40. (https://doi.org/10.1016/j.ijhcs.2009.08.006).

Strey, M. R., Pereira, R., and de Castro Salgado, L. C. 2018. "Human Data-Interaction: A Systematic Mapping," in ACM International Conference Proceeding Series, Association for Computing Machinery, October 22. (https://doi.org/10.1145/3274192.3274219).

Strong, S. 2019. "User Experience-Driven Innovation in Smart and Connected Worlds," AIS Transactions on Human-Computer Interaction (11:4), p. 215. (https://doi.org/10.17705/1thci.00121).

Widjojo, E. A., Chinthammit, W., and Engelke, U. 2017. "Virtual Reality-Based Human-Data Interaction," in 2017 International Symposium on Big Data Visual Analytics, BDVA 2017, Institute of Electrical and Electronics Engineers Inc., November 17. (https://doi.org/10.1109/BDVA.2017.8114627).