Empirical Evaluation of Data Visualizations by Non-Expert Users

Elena Ornig, Jolon Faichney, Bela Stantic
School of Information and Communication Technology
Griffith University
Gold Coast, Australia
email: elena.ornig@griffithuni.edu.au
email: {j.faichney, b.stantic}@griffith.edu.au

***Abstract* — With the increased release of Open Government Data (OGD), several problems hinder the breakthrough of the Open Data agenda into the mainstream. One of these problems is the slow acceptance of OGD by non-expert end-users. They do not have the technical skills and prefer a *human-readable* format compared to the experts who demand *machine-readable* data. Recently, some OGD portals added interactive visualizations to ease the use of OGD by non-expert users. However, the question of human-usability or what makes it easier for non-experts to interact with OGD visualizations, remains open. With the aim to answer this question, we report results on the evaluation of OGD visualizations from the field experiment conducted with non-expert users. We discuss results and insights to inform designers and OGD providers.**

Keywords - data visualization; empirical evaluation; open government data; non-expert users.

I. INTRODUCTION

This paper is an extended version of [1]. In this paper we provide more detail on the conducted field experiment; explain how the Visualization Evaluation Design Constructor (VEDC) was applied and provide more detail about experiment results and observations, including informal comparisons with Data USA portal and an extension to the previous discussion.

The current number of released datasets is now over 18 million [2]. According to dataportal.org, there are 520 registered government portals [3]. On the International Open Government Dataset Search, there are 192 catalogs in 24 languages, representing 43 countries [4]. These numbers represent a growing supply of OGD for users. However, there are many barriers preventing OGD breakthrough into the mainstream [5]. One possible barrier, which we investigate, is the limited usability of open data.

A study on barriers to Open Data Agenda, found that the initial focus has been on the supply side, followed by a focus on data discovery and integration; only recently, with the increased development of data applications, the concern has been raised on the demand side [5]. Additionally, the desired format of the data is one that is machine-readable. The motivation is based on the principle of completeness so that the community has access to raw information from datasets [6].

A downside to this motivation is that it is only usable by a small percentage of the community, those with technical computer skills, such as computer programmers and data analysts, i.e. the experts in data creation, modification, and manipulation. The focus on machine-readability has limited the human-usability of open data. The users with a lack of technical skills, particularly *common citizens,* will find using OGD in the form of reports, visualizations and applications more usable [7]. On the other hand, *informed citizens* can view visualizations and analytical results [8].

Several studies addressed the question of OGD demand, its consumption and the lack of user's technical skills. Shadbolt et al. [8] listed several lessons that can form part of a roadmap to move away from raw government data to a Linked-data Web (LDW) that can be regularly consumed by citizens. Ding et al. [9] developed the Semantic Web-based Tetherless World Constellation (TWC) Linked Open Government Data (LOGD) portal to support LOGD production and consumption. They concluded that LOGD must provide service to a diverse set of stakeholders, including *average citizens*. The MIT Media Lab created the free software portal DataViva, as an information visualization engine, to make open government data more comprehensible for the *average user* [10].

Furthermore, the MIT Media Lab, in partnership with Deloitte and Datawheel, released "the most comprehensive website and visualization engine," Data USA, to make it easier to use OGD for people without technical skills [11]. Graves and Hendler [7] proposed use of visualizations to deal with a lack of technical expertise and developed a prototype tool to simplify the creation of visualization based on Open Data for non-expert users.

Though, none of these studies had provided a formal evaluation of human usability of OGD, they acknowledged the need for visualization [8], its potential to lower demand on technical expertise [7][9][11] and the need to evaluate how citizens can participate, and what makes it easier for them to consume OGD [7]. We characterize a *common citizen*, an *average user* or a *user without technical skills* as a non-expert user.

This paper is organized as follows. First, we investigate what stops non-expert users utilize OGD. Secondly, we evaluate what limits usability for non-expert citizens in using existing OGD visualizations incorporated into portals [1]. In Section II we provide an overview of the identified problems and challenges in visualization and its evaluation.

Next, motivated to better understand the complexity of visualization evaluation, we conceptualized the Evaluation Method Mapping (EMM) approach. This approach and how it was applied, is described in Section III. In order to access possible evaluation methods and techniques, we developed the Visualization Evaluation Design Constructor (VEDC). The VEDC is a framework that consists of four, essential for any evaluation, elements: general goals, evaluation methods, theoretical implications and practical aspects. The VEDC was applied to construct a task-based field experiment. The use of VEDC is also explained in Section III.

A formal experiment was conducted to evaluate the usability of three different visualizations: 1) TreeMap, which represents data in percentage terms; 2) Map, which represents the spatial distribution of variables, and 3) Stacked, which represents growth of a variable over a period of time. The results and findings are shown in Section IV. Furthermore, we discuss significance and implications of the results in Section V. Finally, Section VI contains our conclusions and offers directions for future work.

II. Background

A. Multi-disciplinary fusion of Visualization

Visualization is an effective technique for the communication of data, due to our natural ability to understand patterns. Ware [12] provides a scientific explanation:

"The human visual system is a pattern seeker of enormous power and subtlety. The eye and the visual cortex of the brain form a massively parallel processor that provides the highest bandwidth channel into human cognitive centers."

Ware [12] views the role of visualization in cognitive systems as small but crucial and expanding. Though, Ware highlights several capabilities of information visualization, including its ability to help humans comprehend large amounts of data; he takes a view that all people have the same visual system, which can only perceive presented data in a particular way. Thus, he argues, if we can understand how we perceive data, we can build better visual displays. Shneiderman [13], from the field of Human Computer Interaction (HCI) views information visualization as a subfield. There is no strict formula for a successful interface but only a few basic approaches. The computer is seen as a 'tool' to extend the user's body, in order to create experience where the user is in control, confident and focused on their goal. This *optimal experience* is achieved through a balance when the interface is simple, not confusing, but at the same time—not boring.

Two decades ago, Butler, Almond, Bergeron, Brodlie, and Haber [14], in their discussion of the general understanding of visualization, asked if visualization is a general process or "a collection of unique, unrelated techniques?" They queried if the scope of the visualization reference model should include related domains: visual perception, computer-human interface and computer graphics? Who should use it—providers, developers or users? Would they use it to learn techniques, to evaluate systems, to design systems or to define standards? Since then, several visualization reference models and taxonomies of visualization techniques were developed, including: a data-oriented taxonomy by Card and Mackinlay [15] and a type-by-task taxonomy by Shneiderman [17]. Also, Khan and Khan [16] have published a collection of all visualization techniques, giving each a brief introduction to guide young researchers through their work in visualization.

A decade ago, Lengler and Eppler [18] overviewed the discipline of visualization studies and found it a highly unstructured domain of research in the context of applicable visualization methods. To provide assistance for researchers and practitioners, a user-centered periodic table of 100 visualization methods was created as a prototypical example based on Shneiderman's Visual Information-Seeking Mantra. In their table of visualization methods, they highlighted the fact that there is not necessarily one appropriate method but rather a few different methods that could be applied for a particular requirement. By using this table, a designer could see which methods are providing overview, overview *and* details on demand, and which methods are good at providing additional details.

They also categorized visualization methods according to cognitive processes: convergent and divergent thinking. For example, an area chart, which is a type of data visualization method and a data map, which is another type of information visualization method, can both be used to *overview* an entire collection of items (Shneiderman's design principle [17]). The treemap, an information visualization method, can be used for simultaneous *overview* and *detail* (Shneiderman's design principle [17]). These three methods of visualization are applicable to the cognitive process of convergent thinking. Lengler and Eppler [18] used several selection criteria before a specific method was included in the table: a method must be fully documented, must be put into practice in real-life, must illustrate complex issues, must be applicable by non-experts and previously evaluated.

These criteria reflect the underlining multi-disciplinary fusion of visualization in general, and the information visualization field, which originated from low level perception and statistics, and in modern times includes [19]: "color theory, visual cognition, visual grammars, interaction theory, visual analytics, and information theory." This inherited multi-disciplinary fusion causes a challenge for scientists to define a unified theory of visualization.

Traditionally, a general theory can be formulated through the process of eliminating or unifying competing and complementary theories, from determined domains [20]. In regards to data and information visualization, some possible theories were discussed by a group of scientists from Brown University in the US [20]. Demiralp [20] identified a need for specific and restricted theoretic models that would provide explicit methods for effective visualizations. He concludes that the question of how to measure and construct effective visualizations, in general, is an unsolved problem. Laidlaw [20] observed a controversy in identifying what defines a theory of visualization. Wijk [20] stated that the discipline of visualization is a technology and not a science.

In order to understand what works and does not work, there is a need to develop methods and techniques and a need for cross-cutting insights as a guide in searching for new visualization solutions. Ware [20] argued that the reason why visualization works is in its transformation of data, which creates visual patterns. These patterns, due to natural human perception skills, then help to solve problems. Since theory is based on generalized experimental results, then, in "the case of data visualization in large part this has to be the theory of perception" [20]. Thus, applied perception and distributed cognitive algorithms are all that is needed for the theory of visualization design. In addition, the panelists argued [20] that evaluating visualization with user studies is insufficient and inefficient when an inductive approach is used.

B. Evaluation challenges

However, the advantage of evaluating visualization has many different values for other researchers. Plaisant [21] sees evaluation value, particularly by *potential adopters* or *new users*, in the discoveries of the same data through new perspectives i.e. in answers to questions "you didn't know you had" and even possible changes in work practices. She argues that controlled experiments and usability studies help to recognize the tool's potential and limitations. Lam et al. [22] define evaluation as a complex science which aids in the detailed understanding of a tool or system and their supportive processes. This includes "exploratory data analysis and reasoning, communication through visualization or collaborative data analysis." They specified evaluation as an assessment of the visualizations themselves and contributed a new, scenario-based approach for the information visualization research community [22]. Carpendale [23] emphasized the importance of empirical research and called for more convincing evaluations to encourage wider adoption of information visualization tools.

Despite the difference in opinion on the benefit in evaluation of visualization and the existing lack of a unified theory, in terms of design principles, significant and well-established work has been done in the fields of data and information visualization and HCI.

Shneiderman [17], the inventor of treemap visualization, developed a type-by-task taxonomy to guide designers of advanced graphical user interfaces: *overview first* ("Gain an overview of the entire collection"); *zoom* ("Zoom in on items of interest") *and filter* ("filter out uninteresting items")*; then details-on-demand ("Select an item or group and get details when needed")*. He defined these as basic principles, commonly known as the Visual Information Seeking Mantra. He used this mantra as a starting point to propose a type-by-task taxonomy (TTT) of information visualization, adding new tasks: *relate*, *history*, and *extract*. These seven tasks represent a high level of abstraction based on the user's problems, to be solved in seven data types: "1-, 2-, 3-dimensional data, temporal and multi-dimensional data, and tree and network data" for controlled exploration by users [17]. Shneiderman [14] also laid the philosophical foundation for designers to make systems comprehensible, the interfaces predictable and controllable, and the features understandable for the tasks. The design must amplify user's capabilities and make users feel like masters who can accomplish their tasks with pride.

To achieve this, the theory of visualization needs methodologies to integrate its rules into visualization software [19]. The designers of visualizations need a reminder that serving a human need is the purpose of technology [24]. Information visualization needs new evaluative methodologies for usability studies, with a learning-centered perspective [25]. The evaluators need improvement of usability testing. This will help to conduct more rigorous empirical research, where the methodology fits a proposed research question, a given situation and a research goal [23]. Plaisant [21] recommends evaluations where tools are matched with users, tasks, and real problems. She describes recorded observations of users as "the basis for refinement or redesigns, leading to better implementations, guidelines for designers and the refinement of theories."

Additionally, there are still ten major unsolved information visualization problems [25]. They are usability; understanding of elementary perceptual-cognitive tasks; prior knowledge in operating devices and domain knowledge to interpret content; education and training through accessible tutorials for the general public to promote awareness of the potential and problems of information visualization; quality metrics to enhance advances in evaluation and selection of visualizations; the enduring scalability problem; understanding interaction of insights and aesthetics; necessity to distinguish visualization processes with built-in trend identification mechanisms and without; algorithms resolving conflicting evidence; and the challenge of knowledge domain visualization (KDViz) [25]. These unsolved problems [25] add complexity to information visualization in general and its evaluation.

Finally, there is an important element in the process of evaluation—the human factor [16]. Since the evaluation of visualization is directly related to human-computer interaction and interaction with an interface to complete tasks, finding an appropriate sample of participants can be challenging [23]. In Graves and Handler's [7] paper which evaluated tools and visualization techniques for OGD visualization, the majority of users had some technical or domain expertise. In the papers that investigated multiple cases of evaluations, concern was raised on the overreliance on students [22][23]. The reasons are varied. In some cases, the expertise of the participants is necessary [23] and in some, it is simply difficult to find the intended users, have a large enough sample and conduct an effective empirical evaluation. The most challenging part is to relate [23] "a new set of results to previous research and to existing theory." In our case, we could not find any related formal evaluation of OGD interactive visualizations by non-expert users, nor could we confidently use one general theory.

However, to evaluate OGD visualizations, we identified our intended users. We found and adopted two simple arguments made by Barrence [10] and Hammer [27]: "There's not a lot of value for data without the right visualization," and "Open data has little value if people can't use it."

III. Methodology

Our overall approach was based on systematic investigation of what was clearly understood and what was not in the evaluation of visualizations. However, through our literature research, we realized that all problems surrounding OGD can be divided into two types: inherent and accumulated.

A. *The Evaluation Method Mapping approach*

The inherent types are interoperability, scalability, accessibility, integrity, reusability, integration, visualization, production, quality, and interaction. These are not new problems for researchers and some of these problems can be defined as general problems. These problems have already been investigated and their evaluation methods can be easily found through literature research.

The accumulated problems or the new problems are transparency, social barriers, cultural barriers, participation, technical barriers, legislative barriers, regulation, supply and demand, economic impact, cost of release, maintenance cost, management and resource allocation, which occurred recently with OGD release or are directly related to OGD. However, when both inherent and accumulated problems are broken down into specific issues, the similarities of these issues can be matched. Then, through the matching of issues, the methods of evaluation can be found much easier by viewing directly related sources. This is how we arrived at the idea to conceptualize the EMM approach and created the first version of a manually compiled repository (See Appendix A).

For example, Martin [5] investigated: implementation barriers and barriers to use in relation to the open data agenda. One of the found issues/barriers [5] is “limited interoperability between government ICT systems.” If we look at the inherent general problem in our repository as shown in the Table I, we find a list of investigated specific issues, including “system interoperability.”

Table I. Partial excerpt from Appendix A repository.

Accumulated Issues	**Inherent problems with specific issues**	**Known evaluation methods**	**Source with links**
Limited interoperability between government ICT systems	Interoperability (general problem) Specific issues: System interoperability	System Interoperability Framework	Authors and links to the source

These listed issues are directly related to the next column and its known evaluation methods, including practiced and proposed methods, models, frameworks, measures, metrics, and evaluation criteria. These methods are connected to the subsequent column, which provides a source of information and a link. If a researcher decides to proceed with evaluation, they will find actual links as shown in Appendix A (not shown in Table I for brevity). There, they can find the examples of methods and examples of how to collect and analyse data.

Initially, we used this repository to find methods for our evaluation of interactive visualizations. In the inherent problems of visualization is a list of 18 different known and investigated issues. It is easy to see the listed methods in the next column: high-dimensional data visualization analysis, practice of evaluating visualization, evaluation methods, user interface evaluation, simple visual prototypes and task sets based on a visual taxonomy, heuristic evaluation and an evaluation of several quality predictors for model simplification. For our empirical evaluation of interactive visualization techniques, we chose a field experiment, which is usually conducted in a realistic setting and allows an experimenter to have some degree of observation [23]. The provided sources of information revealed several challenges for information visualization empirical research: difficulty in finding the right focus, asking the right questions and working out sufficient and precise procedures for data collection.

B. *The Visualization Evaluation Design Constructor*

Further investigation uncovered that evaluation of information visualization is closely related to HCI evaluations, when tasks are based on interaction with an interface: overview, zoom, filter and getting necessary details [17]. Furthermore, it relates to the usability of a system, interface or device. Thus, the challenge is to understand results clearly in order to identify where the problem is: in the application, in a specific technique [23] or in the design of device.

To overcome these challenges, we obtained inspiration from Lam’s et al. [22] suggestion to reflect on goals and questions prior to a decision of applying specific methods. As a result, we used meta-data analysis to generalize research questions into more generic groups These groups were further classified into general research goals based on their strategic orientation: problem-oriented, theory-oriented, product-oriented, process-oriented and user-oriented. On a higher, conceptual level, their complex interrelation allowed us to classify them based on their key strategic focus. Furthermore, by analyzing and generalizing theoretical implications [23] and practical aspects [22] of evaluation, we defined four essential elements common to all related fields. These are general goals, evaluation methods, theoretical implications, and practical aspects. Based on these elements we developed the VEDC framework as shown in Figure 1.

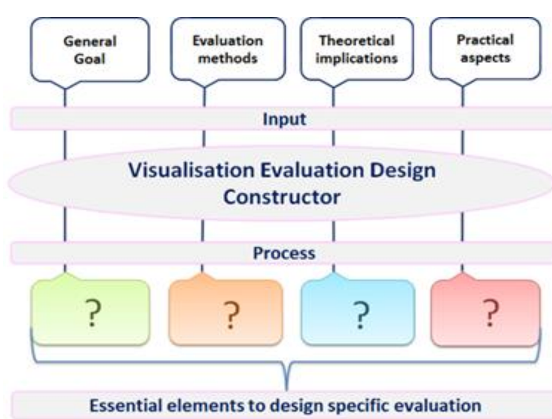

Figure 1. Visualization Evaluation Design Constructor (VEDC) framework.

The VEDC emphasizes the interconnectedness and interrelation of the essential elements for evaluation of any visualization. We argue that for any evaluation there are at least four essential elements: a general goal with a particular strategic focus, one evaluation method or combination of different methods, an underlying theory or a set of theories and subsequently, some practical aspects.

Our overall approach to this study is a combination of qualitative and quantitative approaches which complement one another [23] in order to find potential usability issues and inform designers [22]. The general goal was strategically focused on the user (user-oriented). Based on the general goal, we overviewed the literature related to the users' requirements, needs, wants, desires and user interaction with systems, devices, applications, interfaces and visualizations, and their evaluations [7][17][21][23][24].

This helped us to choose our method – a field experiment to obtain empirical evidence with an emphasis on realism [16][17][22][23]. The perception of data visualization [22], the choice of design [28], the Visual Information-Seeking Mantra principles of information visualization application design [17], and the usability test for use of data visualization tool [22] provided theoretical background. Subsequently, we could not avoid the consideration of practical aspects such as: procedures and techniques [21][22][23]; sample size [22][23][26]; data collection and analysis [21][22][23]; research ethics; and observer-experimenter-evaluator effects [21][22][23].

The VEDC was created to overview the Literature on the existing evaluations and their analysis. We view the VEDC framework as an advanced method for organizing literature related to evaluation research. The VEDC is currently limited but can be used as a guide (See Appendix B, C, D and E) with the existing four repositories of collected information and related sources.

Each repository has a set of the most common goals with related (most) common research questions (See Appendix B), a set of related evaluation methods with related possible theoretical implications (See Appendix C), a set of theoretical implications with direct relations to the theories (See Appendix D) and a set of practical aspects that could have implications on the research (See Appendix E).

The first repository has four strategically-oriented goals. Each goal has generically grouped problem question(s) directly related to the common research questions. These generic questions lead to the directly related existing sources of information. The information in the sources includes solutions and recommendations, helping to clarify research questions. Once the research question is clarified, the next step is to look for evaluation methods.

The second repository represents existing evaluation methods: a perception based evaluation, empirical studies, quantitative and qualitative evaluations, etc. Each method has information on possible known theoretical implications and specifically related descriptions of the existing methods. These are: a controlled user study, scenarios for understanding data analysis, quantitative experimental research, etc. The provided descriptions indicate what can be found in the existing sources of information.

The third repository classifies theoretical implications and existing theories under interrelated fields. They are: data visualization, information visualization, human-computer interaction, cognitive psychology, computer graphics, etc. The descriptions of theoretical implications lead directly to the existing sources of information.

The last repository defines practical implications in evaluations under general titles. These are: procedures and techniques, evaluators, participant's sample sizes, data, observer (or experimenter or evaluator) effect, tools for data collection and research ethics. Each general title has more specific descriptions. Each description leads to the existing source of information.

As an example, researchers can find how to compare heuristic evaluation and cognitive walkthrough, etc. The steps from one repository to another are not essential which makes the VEDC more flexible in use.

C. *Web-based field experiment*

Our overall method is based on a set of task-based experiments and observations.

To evaluate the usability of open data interactive visualization techniques, we performed a web-based field experiment using the DataViva [10] as a tool for interaction with visualizations. DataViva is a web portal for Brazil's open data developed in partnership with the MIT Media Lab [10]. Since starting this investigation, the MIT Media Lab has also launched the Data USA open data portal, which contains updated visualizations [11].

We evaluated Data USA informally in an attempt to compare our findings. The field experiment focused on three visualization techniques provided by DataViva: TreeMap, Map (data map), and Stacked (area chart). All three belong to the category of descriptive applications. Figures 2, 3 and 4 showing examples of these visualizations.

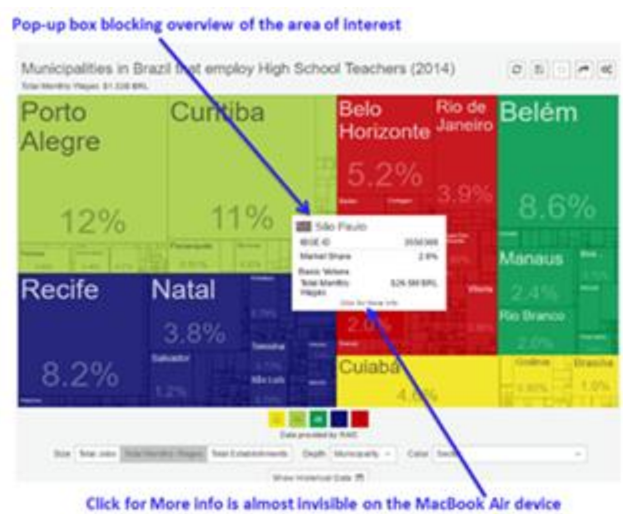

Figure 2. DataViva TreeMap visualization.

We engaged our users at 7 different locations around Gold Coast city, Australia, in public places where Wi-Fi access was freely available. To conduct the experiment, we used a MacBook Air laptop for internet access; DataViva website created specifically for OGD of Brazil to evaluate its visualizations; and software Debut as a tool for video and audio data collection. The software Debut allowed to capture audio and video recordings for every single task conducted by our participants in parallel with actual observations. Our goal was to test at least 10 participants as this is a suitable number according to Faulkner [26]. He showed that for usability testing, 10 users are sufficient enough, to find 80% of the problems.

To balance the control between observer and the users and to balance the trade-offs between generalization, precision, and realism [23], the experiment was broken down into two stages: a preliminary stage and a controlled-testing stage.

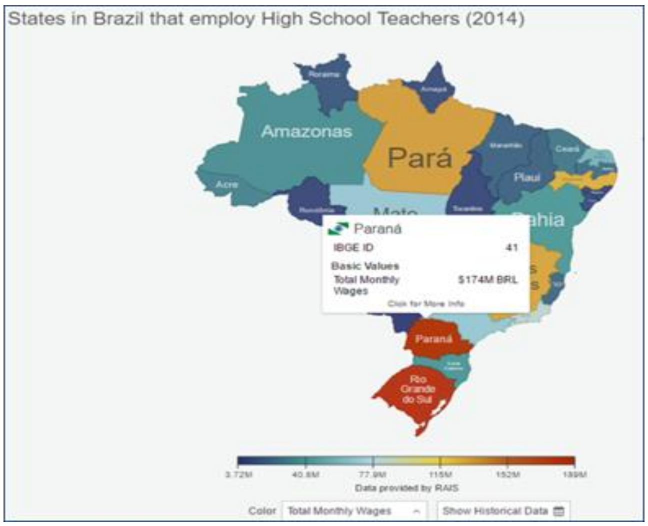

Figure 3. DataViva Map visualization.

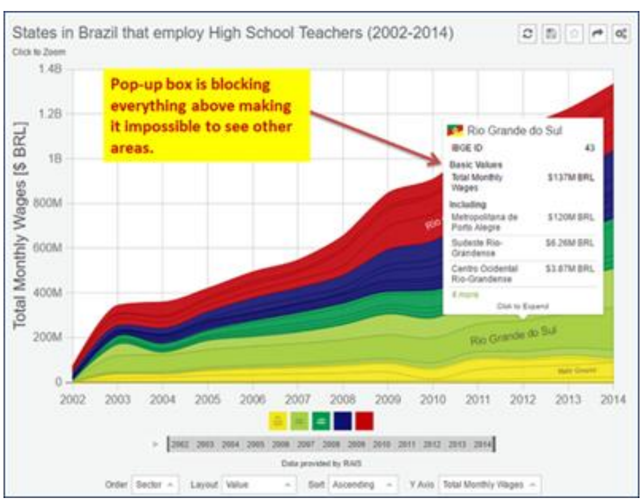

Figure 4. DataViva Stacked visualization.

The preliminary stage included presenting the participant with an information sheet about the study. This was followed by conversational questioning, to find out what stops non-expert users from using OGD. This stage was concluded with the formal signing of the consent form. The controlled-testing stage included 5 minutes of device and interface familiarization. This was followed by performance tasks designed as a motivational scenario based on an envisaged real situation. Tasks were designed to solve real problems with real data, in a real setting. This was designed in such a way, that the participants would always interact with a new interface, with every new task.

The user's interaction was captured with screen recording software and audio that were later analyzed to calculate completion time. We used an unenforced think-aloud protocol [23] to support the identification of possible usability issues. Specifically, for visualizations, the users were given 3 tasks to complete, each using a different visualization technique and a different task for that visualization.

The controlled tasks were designed on data about high-school teachers in Sao Paulo. The flow of all tasks mirrored Shneiderman's [17] visual mantra: overview first, zoom and filter, then detail-on-demand. Task 0 was designed to find a specific area to evaluate navigation through the DataViva web portal. The participants needed to start with the *Home page*, find *Occupations*, then find the *High School Teachers* page. On this page, they needed to find the *Preview* area for *Wages and Jobs* and then click on the *Municipality* under *WAGES BY* title to open a drop-down list of visualizations: TreeMap, Map and Stacked. It did not have a predicted completion time but it was measured later via video recordings. This task had three different possible paths, leading to the same information. Each path was mapped by the number of clicks: first – four, second – five and last – six, averaging at five clicks.

Tasks 1, 2 and 3 were designed to search for specific information in order to give a correct answer. The answers for Tasks 1, 2 and 3 were located in the listed visualizations. Task 1 required finding the total amount of jobs, where hierarchical data was graphically represented by TreeMap visualization, based on 2014 data. The correct answer was "7.08 k." The predicted time was 30 sec.

Task 2 required users to find the nominal wage growth, visually represented by Map (similar to choropleth/thematic or data map) visualization and based on 2014 data. The correct answer was 11%. The predicted time was 30 sec. Task 3 required users to find total monthly wages, visually represented by Stacked (similar to area chart/stacked area graph) visualization. It was based on the volume of an aggregated summary of 2012 data. The predicted time was 20 sec. The predicted time for Task 1, 2 and 3 included average download times.

This was followed by rating based on user's preferences to quantify user's subjective opinion for overall assessment of each single visualization interface. The participants' subjective judgments were turned into numbers, with the use of a rating scale: first choice = 1, second choice = 2 and the last choice = 3. Finally, the participants were asked a single open-ended question: "Why do you prefer this particular visualisation compare to others?

In addition, the experiment observer was provided with the designed templates to make observational notes of the participant behavior and to confirm the accuracy of the information found by participants in all tasks. These notes (qualitative data) were analysed and compared to the tasks' measured results (quantitative data) in order to obtain insights into the process of evaluation and the participant's interaction with visualizations.

IV. RESULTS

Our experiment sample was based on 12 users, selected randomly, at seven pre-defined locations with free access to the Internet via Wi-Fi, to achieve realism of a pre-defined scenario for a realistic setting, with realistic tasks and real users. The target number was 10. First, we knew [23] that with a realistic setting it would be difficult "to get a large enough participant sample."

Secondly, we were familiar with reported successes of usability tests to evaluate a data visualization tool with eight [22] or ten [26] participants. Thirdly, we were not generalizing our findings to make statistically significant statements. Also, the practical part of research was conducted by a novice investigator taking on the role of experimenter and observer [23]. However, we do understand that with only 12 participants there is a high rick of bias.

The participants average age was 54 years. As shown in Figure 5, 33% had a university degree, 42% had a college education and 25% were educated at TAFE (a technical training institution). 80% of the participants were female.

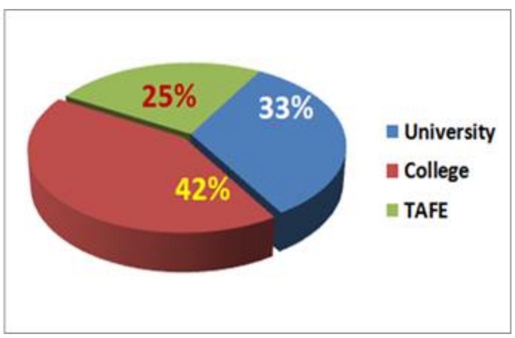

Figure 5. Distribution of participants occupations.

Their professional occupations were very diverse: an international shipping company accountant, a business consultant, the CFO of a mid-sized engineering company, a fashion designer (single operator), special needs teacher, administration clerk from a small company, administrator of small reselling company, retired real estate consultant, kitchen equipment installer, private college administrator, a retired construction worker and a retired nurse.

A. Results from preliminary stage

The time spent per participant to complete the tasks took on average 11 minutes, excluding 5 minutes given to participants to familiarize with the DataViva interface and the time spent to answer the open-ended question. More than 80 hours were spent on the preliminary stage by the novice investigator on approaching random people and conversing in order to select and sign up participants. This means that the time to find one suitable participant took considerable time.

At the preliminary stage, we approached participants with conversational questioning to find out what stops them from using OGD. The presented results reflect an analysis of the answers from the 12 selected participants. 83.2% of participants answered that they had never heard of OGD; did not know OGD existed; or what it meant. However, after their interaction with open data, 66.6% had expressed an interest to know more.

The average completion time for Task 0 (navigating from the home page to the visualization) was two minutes and ten seconds, and on average took 7 clicks. Only two participants were familiar with how to operate the laptop. 66% of participants failed to remember that one click is sufficient to select an item and 75% forgot to scroll with two fingers. Some users blamed their double-clicking habit on primarily using a mouse instead of a touch pad. It was our assumption that the participants were familiar with the Mac look-and-feel, but the majority were not. This wrong assumption might severely have influenced the results of our study.

However, the size of the Mac screen compared to often bigger sized displays of personal home computers is likely to had affected the visibility of the title *Explore our database*, which can be seen only when scrolled down. It was observed that more than 80% of participants did not use the *Get started* button, located in the middle of the screen or the *Search* option, located on the top bar of the *Home page*. As shown in Figure 6, only a few participants commented that they could not see it clearly or that the background image was too busy.

Figure 6. DataViva Home page.

B. Results from Controlled-testing stage

No user errors were recorded through Tasks 1, 2 and 3. The average time to complete each task as shown in Table II, was calculated and linked to ratings.

The Map visualization was the quickest, followed by Stacked, and then TreeMap. Participants were asked to rate the visualizations in order of preference. Figure 7 shows the results of the preferences rating. The participants then were asked open question: "Why do you prefer this particular visualisation compare to others?" Their comments were recorded, later analysed and compared with our observations.

Table II. Correlation between time performance and rating.

Visualizations	Average time per participant	Rating
Map	1 min	First
Stacked	1 min 13 sec	Second
TreeMap	1 min 19 sec	Last

The Map visualization was rated as the first choice, it was also the most frequent second choice; not one participant rated Map as their last choice. The ratings of TreeMap and Stacked were very similar. Stacked having one extra rating for second place and one less for the last. As a result, the order of preference for the participants was Map, Stacked, and TreeMap, as shown in Table II, which correlates with the time it took to complete each task.

Participants also provided reasons why they gave visualizations the particular rating. The Map visualization was chosen because it was perceived as a familiar shape, that of a geographic map, and easy to use. The Stacked visualization had contradictory perceptions. Some perceived it as easy to understand and clear. Others found it confusing and reported that it "didn't make sense." Participants that rated the TreeMap first, found it easy to find information. Those that rated it second stated that it was "not clear." Those that rated it last said it was confusing, busy, and more difficult to find information.

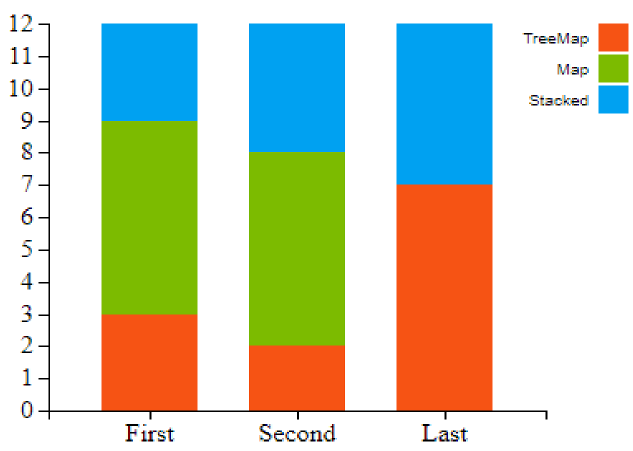

Figure 7. Rated preferences for each visualization type.

Through analysis of observational notes and recorded comments, the participants revealed their perceptions on shapes, sizes, color contrast, and features in the visual presentation of data. "Easy to find info," "I like TreeMap screen" and "The TreeMap was the easiest one" - were the most favorable comments for TreeMap visualization. The majority of participants complained: "Not clear enough" (in regards to color contrast), "… busy, visually it is busy" (too many areas), "… small to find and navigate" (in regards to headings), "confusing" (in regards to low contrast between headings and colored areas), "Boarders between small squares unclear," "Hard to read headings," and "more difficult" (to find information).

The recorded comments: "The Map is similar to the world map…" or "…map is easy to see" and similar comments, clearly demonstrated favorable preferences for Map visualization due to its shape familiarity.

With Stacked visualization, the perceptions were polarized: from "clear" to "confused." This was due to the inability of some of the participants to read the graph.

The popup box interference was the major reason for slowing down task completion in Tasks 1, 2 and 3. The recorded comments and observational notes confirmed this as a major issue for all three visualizations.

The results of the controlled-testing stage, which measured our participants' performance with three visualizations are conclusive even if they are not statistically significant to make generalizations. Furthermore, the performance results, shown in Table II, which are supported by their ratings of preferences shown in Figure 7. However, the participants' comments and our own observations of what had affected their performance gave us a more insightful picture revealing usability issues.

One might assume that these results reflect familiarity as a single contributor to the performance, supported by ratings. Though the familiar shape was perceived more favorably and more likely contributed to better performance, it was not the only factor.

The average downloading time for each visualization was calculated into the predicted performance time. The TreeMap and Stacked downloading time was between 7-8 seconds, more than twice longer than Map (about 3 seconds). Taking into consideration that the majority of participants were already getting frustrated with navigation, the slow downloading of TreeMap and Staked increased their negative perception.

The downloading time is a usability issue and more likely contributed to the negative perception of TreeMap and Stacked visualisations. This means that the familiarity of the Map shape was not necessarily the only factor for user's perception. Furthermore, shape and size of the display, color and color contrast, the size of text and the use of features are usability issues. They are meant to enhance usability of visualizations and not frustrate or confuse. However, according to participants' comments, the TreeMap and Stacked were perceived as confusing, cluttering (visually busy) and unclear. These contributors to usability are matters of display layout and the effectiveness of style.

The uncontrollable popup boxes, incorporated into each visualization, were a major usability issue. Users did not feel in control of this feature which appeared unexpectedly on the mouse rollover in each evaluated visualization. The details should have been given when they are needed. That is why they referred to as *details-on-demand* [17] appearing when one clicks on the selected item and not during the *overview*.

C. Data USA Comparison

MIT Media Lab also produced the Data USA portal. We overviewed the portal to see if it could be compared with our findings. From our perspective, the *Home page* of Data USA, as seen in Figure 8, is much clearer on where to start searching.

Figure 8. Data USA Home page.

In contrast, the DataViva *Home page* has the *Search* option in the header bar, the *Get started* button in the middle of the left side of the page, the *Explore our database* option and the additional several icons are located at the bottom of the page, as shown in Figure 6. This confused the participants, as there were too many options for the same outcome. With regards to visualizations, we found that the TreeMap on both portals looked almost identical, however geographical maps appear differently. We did not find stacked charts, only line charts.

V. Discussion

The goal of this investigation was strategically focused on users who had no technical skills in data creation, modification, and manipulation and no knowledge in data domain. Also, we realized that the majority of the participants were not familiar with the Mac look-and-feel and unaware that OGD existed. In addition, each participant had different cognitive limitations, age, gender and level of education.

These factors, including differences in external noise and lighting in various cafes and factors that we might not yet be aware of, had some effect on the participants' performance and perception. What then is the point to report the results from a field experiment which had a questionable number of sample size, diminishing its statistical significance?

The point is to learn and to inform about potential usability issues of mainstream applications for general users such as OGD portals with incorporated interactive visualisations.

First, for the concern that has been raised on the demand side of utilizing OGD by non-expert users. Only two participants had previously heard of open data. However, the majority of participants demonstrated their interest to know more about OGD. After completing their tasks, they asked what OGD represents, where to find existing portals, and how to use OGD for their benefit.

This indicates that if citizens were more aware of OGD it might increase their interest to utilize OGD potentially contributing to the increase of its demand [5]. Though this is not a statistically meaningful conclusion, it is a possible indicator on the issue of awareness that could be further investigated by OGD suppliers and developers.

Secondly, TreeMap is a very common visualization tool, often used in data journalism, however, we found that participants had the most trouble with it, both in terms of taking the longest time to complete the task, and also in response to the open question. This can be explained by non-expert users' unfamiliarity with TreeMap visualization compared with Map and Stacked.

Additionally, this visualization represents a significant amount of information in one space, increasing demand on the end-users to find specific information. The demand to find specific information, under constraint, could be a second explanation for difficulties experienced. If participants were asked to explore data at their own pace and interest, their opinion and overall experience with TreeMap visualization could have had a different outcome.

Furthermore, if we take into consideration that when TreeMap was first prototyped 27 years ago, it required training for effective use [21]. The current version, deployed in DataViva, was used by people without technical skills, for the first time. The 100% correct answers, found by participants in 1 min and 19 secs on average, without any preliminary demonstration gives us a different perspective.

The interactive choropleth map was first prototyped 24 years ago. At the time, novice users reported difficulties in even starting to use it, perceiving it as too complicated [21]. The modern version, the Map, deployed by DataViva, was perceived by our participants as the easiest to use. Though only one person used zooming, and none of the participants noticed a slider.

The stacked chart is a kind of area chart, which was first published in 1786 [28]. However, the average completion time for the task was more than three times over the predicted time. Several participants did not know how to read a chart. The majority had a substantial level of education and according to their professional occupations, one could assume they would understand how to navigate through a chart. Further analysis revealed that those who understood a chart, completed their task faster, compared to those who did not.

Also, there was confusion with the differences in the area sizes. The participants did not understand why some areas were too narrow, compared with others. None of them acknowledged the slider, but later, one participant, after completion of their last task asked what it was. When the experimenter explained, the participant commented that she had no idea that such a feature exists.

However, for TreeMap and Stacked, the slow time of downloading; the small size of visualization in contrast to the size of display; the difficulty of some users to see a contrast between neighboring areas; and the low contrast between headings and colored areas indicate that these are usability issues that could be improved by designers. These are well-known usability issues in conveying information.

The most significant usability problem with all three visualizations was a feature known as the tooltip plugin or more commonly known, as a popup box. With all three visualizations, the popup box was blocking the overview. Taking into consideration the extended principles for designers of data visualizations: *overview first*, *zoom* and *filter*; then *details-on-demand* [17], we demonstrated, as shown in Figure 2, 3 and 4, that this feature was blocking overview with details even before they were demanded by the users.

The problem with the feature is that it appears on a mouse rollover and cannot be controlled by the users [13]. Thus, this very useful feature is poorly implemented. As the user is navigating to interact with the visualization, the popup box occludes the area they want to interact with. We have provided possible solutions to the popup box issue for each of the visualizations, shown in Figures 9, 10 and 11. The solution is generally to display the popup box to the side or it should only appear [17] when the area of interest is clicked on. Overall, the usability issues with the DataViva interface might appear to be insignificant to designers, but it had a negative effect on the non-expert end-users. Also, our own experience in conducting this field experiment proved how difficult it is to find intended users, chose the right sample and conduct an effective empirical evaluation [23].

In summary, we assumed that if we could find and describe potential usability issues we could inform designers and help them to understand what can be improved to make interactive visualization more user friendly and easier to use.

Other usability issues were not new and are avoidable by designers if they would follow basic approaches for successful interfaces [13] and well-established design principles [17] for interactive visualizations.

VI. CONCLUSION AND FUTURE WORK

The OGD movement is maturing with large quantities of data being released by governments around the world. The embracing of the open data agenda has not necessarily translated into uptake by OGD consumers. We propose that this is because of the focus on machine-readability rather than human-usability. Recent efforts are focusing on providing interactive visualizations of OGD to make it easier for non-expert users to get engaged with OGD.

In this paper, we evaluated three visualizations from one OGD portal, to identify strengths and weaknesses of visualization techniques, specifically for non-expert users, which currently has not been investigated in literature. Even though our participants were unfamiliar with OGD, after a short introduction they were able to answer the problems set before them, under 2 minutes on average. This demonstrates the advantage visualizations have over technical and raw data. This serves as a strong argument for OGD portals to provide visualizations to increase end-user uptake by non-expert users.

Comparing three different methods of OGD visualization, the clear preference was for Map visualization which represents data on a geographical map. The basis for Map being the greatest preference, both qualitatively and quantitatively, is more likely due to its shape familiarity to the non-expert user. Concrete concepts are quicker to grasp than abstract concepts. However, we cannot dismiss other usability factors that contributed to the performance of participants and their rating based on their perception.

The TreeMap and Stacked visualizations represented data more abstractly, which requires a greater conceptual leap for non-expert users to make. However, other usability issues did not help the ease of use. Therefore, to encourage end-user uptake of OGD, visualizations should be selected that are concrete and familiar to end-users, such as Map visualizations. The more abstract visualizations containing large amounts of information in one space, need to be simplified further. It is in line with the basic purpose of visually representing data, that insight must be represented as easily as possible [16].

Note that visualizations such as TreeMap have been designed to address many usability and visualization factors, however, we have found that for non-expert users, concreteness and familiarity are important factors. However, with resolved usability issues with TreeMap and Stacked, the overall perception by non-expert users could be much more positive.

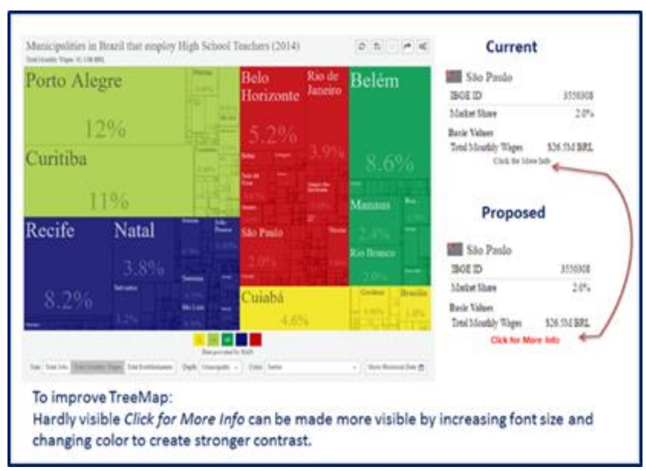

Figure 9. Non-occluding popup box for TreeMap visualization.

Our study also identified an issue with popups, where a simple and useful feature, when poorly implemented, can grossly impact the effectiveness of a visualization. This reinforces the need not just for visualizations, but for end-user testing to verify the effectiveness of the visualizations' features.

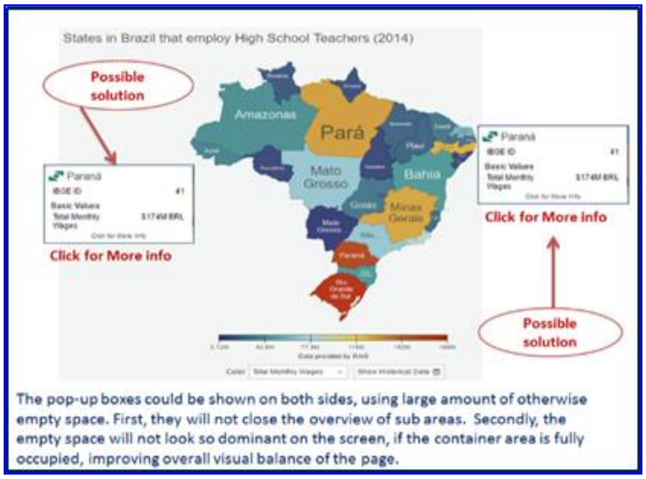

Figure 10. Non-occluding popup box for Map visualization.

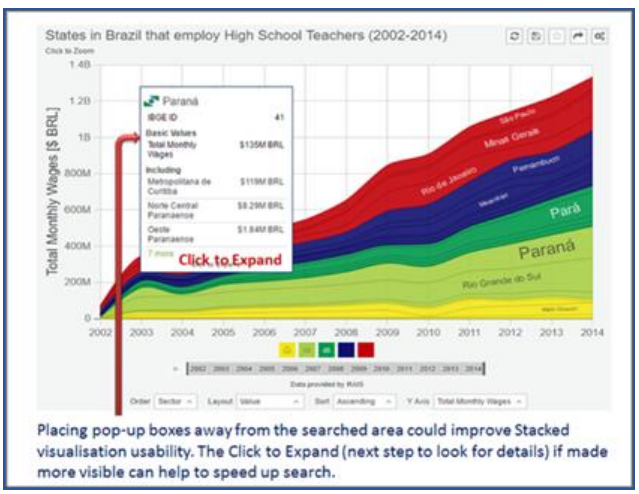

Figure 11. Non-occluding popup box for Stacked visualization.

Our study supports the argument for the need of an optimized visualization that is easy to use, with a comprehensible information visualization language. These usability issues should be tackled with the consideration of several fields, including the human factor [16]. What we observed is that users are primarily concerned with ease and simplicity of use, which supports the argument that usability is all about ease of use [13]. This also supports the argument that the value of data is in the right visualization [10] and if people cannot use open data, then it does not hold value for them [27].

Subsequently, if non-expert end-users will not use it, then the uptake of OGD would remain limited.

ACKNOWLEDGMENT

We would like to acknowledge all participants who greatly contributed their time and effort to support this project.

REFERENCES

[1] E. Ornig, J. Faichney and B. Stantic, "Empirical Evaluation of Open Government Data Visualisations," The Third International Conference on Big Data, Small Data, Linked Data and Open Data (ALLDATA 2017), Apr. 2017.

[2] data.world, "data.world Launches to Make the World's Data Easier to Find, Use, and Share." 11 July, 2016, retrieved: March, 2017. [Online]. Available: https://globenewswire.com/news-release/2016/07/11/855045/0/en/data-world-Launches-to-Make-the-World-s-Data-Easier-to-Find-Use-and-Share.html

[3] Data Portals, "A Comprehensive List of Open Data Portals from Around the World," retrieved: March, 2017. [Online]. Available: http://dataportals.org/

[4] Linking Open Government Data, "IOGDS Analytics," retrieved: March, 2017. [Online]. Available: https://logd.tw.rpi.edu/iogds_analytics_2

[5] C. Martin, "Barriers to the Open Government Data Agenda: Taking a Multi‐Level Perspective," Policy & Internet, 6 (3), pp 217-240, 2014.

[6] Sunlight Foundation, "Ten principles for opening up government," August 11, 2010. Retrieved: July, 2017. [Online]. Available: http://sunlightfoundation.com/policy/documents/ten-opendata- principles/

[7] A. Graves and J. Hendler, (2014). "A study on the use of visualizations for Open Government Data," Information Polity: The International Journal Of Government & Democracy In The Information Age, 19(1/2), pp.73-91. doi:10.3233/IP-140333

[8] N. Shadbolt, K. O'Hara, T. Berners-Lee, N. Gibbins, H. Glaser and W. Hall, (2012). "Linked open government data: Lessons from data. gov. uk," IEEE Intelligent Systems, 27(3), pp. 16-24.

[9] L. Ding, T. Lebo, J, S. Erickson, D. DiFranzo, G. T. Williams, X. Li, J. Michaelis, A. Graves, J. G. Zheng, Z. Shangguan, J. Flores, D. L. Mcguinness and J. A. Hendler, "TWC LOGD: A Portal for Linked Open Government Data Ecosystems," Journal of Web Semantics, 9 (3), pp. 325-333, 2011.

[10] S. Ferro, "New MIT Media Lab Tool Lets Anyone Visualize Unwieldy Government Data," CO.DESIGN. [Online]. Available: https://www.fastcodesign.com/3022701/new-mit-media-lab-tool-lets-anyone-visualize-unwieldy-government-data

[11] S. Lohr, "Website Seeks to Make Government Data Easier to Sift Through," *The new York Times*, Technology, Apr., 2016. [Online] Available: https://www.nytimes.com/2016/04/05/technology/datausa-government-data.html?mcubz=0

[12] C. Ware, "Information visualization: Perception for design," Elsevier. Third edition, 2013.

[13] B. Shneiderman and B. B. Bederson, "The Craft of Information Visualization: Readings and Reflections," Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2003. ISBN:1558609156

[14] D. M. Butler, C. James, R. Almond, D. Bergeron, K. W. Brodlie and R. B. Haber, 1993. "Visualization reference models," In Proceedings of the 4th conference on Visualization '93 (VIS '93), Dan Bergeron and Greg Nielson

(Eds.). IEEE Computer Society, Washington, DC, USA, pp. 337-342.

[15] S. K. Card, J. D. Mackinlay. "The Structure of the Information Visualization Design Space," Proceedings of IEEE Symposium on Information Visualization (InfoVis '97), Phoenix, Arizona, 92-99 Color Plate 125, 1997.

[16] M. Khan and S. S. Khan, "Data and Information Visualization Methods, and Interactive Mechanisms: A Survey," International Journal of computer Applications, vol. 34, no. 1, pp.1-12, November. 2011.

[17] B. Shneiderman, "The Eye Have It: A Task by Data Type Taxonomy for Information Visualizations," In Proceedings of 1996 IEEE Symposium on Visual Languages, Boulder, CO, USA. DOI: 10.1109/VL.1996.545307.

[18] R. Lengler and M. J. Eppler. 2007. "Towards a periodic table of visualization methods of management," In Proceedings of the IASTED International Conference on Graphics and Visualization in Engineering (GVE '07), ACTA Press, Anaheim, CA, USA, pp. 83-88.

[19] C. Ziemkiewicz, P. Kinnaird, R. Kosara, J. Mackinlay, B. Rogowitz, and J. S. Yi. 2010. "Visualization Theory: Putting the Pieces Together," 2014-10-13. [Online] Available: https://pdfs.semanticscholar.org/42f3/dc15a3d5577b42143be1b8d7eb91e16d9d82.pdf

[20] C. Demiralp, D. H. Laidlaw, J. J. Van Wijk, and C. Ware, "Theories of Visualization—Are There Any?" Brown University. Panel discussion. Sep, 2016.[Online]. Available: http://hci.stanford.edu/~cagatay/projects/vismodel/TheoriesOfVisualization-Vis11.pdf

[21] C. Plaisant, "The challenge of information visualization evaluation," In Proceedings of the working conference on Advanced visual interfaces (pp. 109-116). ACM, May, 2004.

[22] H. Lam, E. Bertini, P. Isenberg, C. Plaisant and S. Carpendale, "Empirical Studies in Information Visualization: Seven Scenarios," in IEEE Transactions on Visualization and Computer Graphics, vol. 18, no. 9, pp. 1520-1536, Sept. DOI=2012.doi: 10.1109/TVCG.2011.279, 2012.

[23] S. Carpendale, "Evaluating information visualizations," Information Visualization: Human-Centered Issues and Perspectives, vol. 4950, pp. 19-45, 2008

[24] B. Shneiderman, "A Grander Goal: A Thousand-fold Increase in Human Capabilities," Educom Review, 32, 6, 410. HCIL-97-23, Nov-Dec 1997. [Online]. Available from: http://hcil2.cs.umd.edu/trs/97-23/97-23.html

[25] C. Chen, "Top 10 Unsolved Information Visualization Problems," Ed. Theresa-Marie Rhyne. IEEE Computer Graphics and Applications, Volume: 25, Issue: 4, pp. 12-16. 11 July, 2005

[26] L. Faulkner, "Beyond the five-user assumption: Benefits of increased sample sizes in usability testing," Behavior Research Methods, Instruments and Computers, Volume: 35, Issue: 3, pp. 379-383, 2003

[27] C. Hammer, "Open Data Has Little Value If People Can't Use It." Harvard Business School Review, 29 Mar., 2013. Retrieved 20 Aug. 2016. [Online] Available: https://hbr.org/2013/03/open-data-has-little-value-if

[28] E. Tufte, "The Visual Display of Quantitative Information," Cheshire, Connecticut, pp. 13. 1983.

Appendix A - Evaluation Method Mapping (partial presentation).

Accumulated Issues	**Inherent problems with specific issues**	**Known evaluation methods, measures, metrics, models, evaluation criteria, etc.**	**Source with links**
Interoperability in an open data ecosystem	Interoperability (general problem)	Evaluation models Existing interoperability evaluation models, the similarities and differences in their philosophy and implementation, assessment process of the system.	Rezaei, R., Chiew, T. K., Lee, S. P., & Aliee, Z. S. (2014). Interoperability evaluation models: A systematic review. Computers in Industry, 65(1), 1. doi: 10.1016/j.compind.2013.09.001; http://www.sciencedirect.com/science/article/pii/S0166361513001887
Limited Interoperability between government ICT systems	Specific issues: Data exchange; Information exchange; Service exchange; System interoperability; Application interoperability; Infrastructure interoperability; Knowledge exchange; Network interoperability; Technical interoperability; Operational interoperability.	System Interoperability Framework e-Business inspired eHealth interoperability framework from an overall system perspective.	Craig E. Kuziemsky and Jens H. Weber-Jahnke, "An eBusiness-based Framework for eHealth Interoperability," Journal of Emerging Technologies in Web Intelligence, Vol. 1, No. 2, pp. 129-136, November 2009. doi:10.4304/jetwi.1.2.129-136 http://www.jetwi.us/uploadfile/2014/1226/20141226054221610.
Open standards of interoperability for open data end-users		Technical Interoperability Maturity Model The Information Systems; Interoperability Maturity Model (ISIMM); Interoperability of hardware, software, data, communication and physical interoperability for Government, including measures.	Staden, S. V., & Mbale, J. (2012). The information systems interoperability maturity model (ISIMM): Towards standardizing technical interoperability and assessment within government. International Journal of Information Engineering and Electronic Business, 4(5), 36-41. http://www.mecs-press.org/ijieeb/ijieeb-v4-n5/IJIEEB-V4-N5-5.pdf
		Conceptual Evaluation and Selection Framework The parts and forms of the evaluation framework; Interoperability levels; Activities of the evaluation process.	Mykkänen, J. A., & Tuomainen, M. P. (2008). An evaluation and selection framework for interoperability standards. Information and Software Technology, 50(3), 176-197. doi: 10.1016/j.infsof.2006.12.001 http://www.sciencedirect.com/science/article/pii/S0950584906001960
Identifying scalable solutions across government	Scalability (general problem) Specific issues: Operating system; Database server; Application server; Hardware performance;	Design and Evaluation Efficient parallel hash algorithms for processing large-scale data, including a theoretical analysis of different hashing; Frameworks and testing scalability.	Cheng, L., Kotoulas, S., Ward, T. E., & Theodoropoulos, G. (2014). Design and evaluation of parallel hashing over large-scale data. Paper presented at the 1-10. doi:10.1109/HiPC.2014.7116909
			Liu, H. H., & Books24x7, I. (2009). Software performance and scalability: A quantitative approach (1st ed.). Hoboken, N.J: John Wiley & Sons.
Government yet to improve technical accessibility of Open Data	Accessibility (general problem) Specific issues: Data accessibility; Website accessibility.	Potential improvements of UX research The products, dimensions of experience, and methodologies across a systematically selected sample of 51 publications from 2005-2009, reporting a total of 66 empirical studies.	Bargas-Avila, J.A., Hornbæk, K., 2011. Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI '11. ACM, New York, NY, USA. pp. 2689–2698. http://dl.acm.org/citation.cfm?id=1979336

Appendix B – VEDC: General goal based on strategic focus.

General goal	Generically grouped	Commonly encountered	Source of information https://www.researchgate.net/publication/310843052_Evaluation_of_Open_Government_Data_Visualisations
Problem oriented	What is the problem and how to solve it?	To identify a problem (potential or open)	Graves and Hendler [22], Chen [25], Bederson et al [28], P Plaisant [29], Teyseyre at al [30], Khan [31], Trochim [32], Shneiderman [33], Eliane [34], Ware [36], Heer [38], Carpendale [42], Dong [41], Ellis et al [43], Çağatay [45], Andrews [47], Faulkner [49], Hilbert et al [50].
		To identify its cause (issues and limitations)	Plaisant [29], Lam et al [23], Bederson et al [28], Teyseyre at al [30], Khan [31], Trochim [32], Shneiderman [33], Ware [36], Heer [38], Carpendale [42], Ellis et al [43], Goebel [46], Faulkner [49], Hilbert et al [50]
		To find solution	Lengler et al [26], Bederson et al [28], Teyseyre at al [30], Carpendale [42], Ellis et al [43], Çağatay [45], Andrews [47], Hilbert et al [50].
		To provide recommendations	Lam et al [23], Shneiderman [27], Carpendale [42], Andrews [47], Faulkner [49].
Theory oriented	How to test hypothesis?	To prove it	Lam et al [23], Graves and Hendler [22], Zhu [24], Bederson et al [28], Ware [36], Heer [38].
		To generate	Shneiderman [33], Eliane [34], Carpendale [42], Çağatay [45], Zhai [48].
Product oriented	How to develop new application? How to develop new system? How to develop new feature?	To evaluate design	Shneiderman [27], Lam et al [23], Zhu [24], Bederson et al [28], Plaisant [29], Eliane [43].
		To evaluate prototype	Graves and Hendler [22], Lam et al [23], Bederson et al [28], Plaisant [29], Teyseyre at al [30], Khan [31], Ware [36], Ellis et al [43], Hilbert et al [50], Shneiderman [51].
		To evaluate product	Lam et al [23], Bederson et al [28], Plaisant [29], Teyseyre at al [30], Trochim [32], Shneiderman [33], Ware [36], Carpendale [42], Ellis et al [43], Andrews [47], Zhai [48].
Process oriented	How to evaluate process?	Workflow process	Lam et al [23], Eliane [34], Carpendale [42]
		Data exploration	Lam et al [23], Shneiderman [27], Bederson et al [28], Plaisant [29], Teyseyre at al [30].
		Data analytics	Lam et al [23], Plaisant [29], Teyseyre at al [30], Khan [31], Shneiderman [33], Eliane [34].
		Descriptive process	Lam et al [23], Bederson et al [28], Plaisant [29], Teyseyre at al [30], Khan [31], Ware [36].
		Displaying data	Lam et al [23], Shneiderman [27], Bederson et al [28], Plaisant [29], Teyseyre at al [30],
User oriented	How to satisfy their requirements? How to satisfy their needs? How to satisfy their wants? Good interaction	Evaluate requirements	Bederson et al [28], Teyseyre at al [30], Khan [31], Çağatay [45], Hilbert et al [50].
		Evaluate needs	Bederson et al [28], Plaisant [29], Teyseyre at al [30], Khan [31], Trochim [32].
		Evaluate wants	Bederson et al [28], Shneiderman [33], Heer [38], Rogowitz [44], Hilbert et al [50].
		Evaluate interaction	Lam et al [23], Bederson et al [28], Plaisant [29], Teyseyre at al [30], Khan [31], Eliane [34].

Appendix C – VEDC: Evaluation Methods Repository (partial presentation).

Evaluation methods	Description	Source with links
A Perception-Based Evaluation Possible theoretical implication: Data Visualization; Computer Graphics; Perception theory.	A controlled user study to test against the following hypotheses: Projection performance is task-dependent; Certain projections perform better on certain types of tasks; Projection performance depends on the nature of the data; Subjects prefer projections with good segregation capability.	Etemadpour, Ronak, et al. "Perception-based evaluation of projection methods for multidimensional data visualization." IEEE transactions on visualization and computer graphics 21.1 (2015): 81-94. http://ieeexplore.ieee.org/abstract/document/6832613/
An evaluation for categorical data Possible theoretical implications: Information visualization; Data Visualization.	A task based performance evaluation: A method designed for categorical data; Approaches in the context of two basic data analysis tasks; Efficiency of the quantification approach.	Fernstad, Sara Johansson, and Jimmy Johansson. "A task based performance evaluation of visualization approaches for categorical data analysis." Information Visualisation (IV), 2011 15th International Conference on. IEEE, 2011. http://ieeexplore.ieee.org/abstract/document/6004026/
Nine common evaluation methods Possible theoretical implications: Information Visualization; Data Visualization; Visual cognition; Human-Computer Interaction (HCI).	Methods & Types of Evaluation: Classification of types to perform evaluation; Evaluation of information visualization techniques;	K. Andrews. Evaluation comes in many guises. In CHI workshop on BEyond time and errors: novel evaluation methods for Information Visualization (BELIV), pages 7–8, 2008. http://www.dis.uniroma1.it/beliv08/pospap/andrews.pdf
Multi-dimensional In-depth Long-term Case Studies (MILCs) Possible theoretical implications: Information Visualization; HCI; Visual cognition; Information theory; Visual perception; Color theory.	Strategies for evaluation: Evaluation methods; Multi-dimensional In-depth Long-term Case studies guidelines.	Shneiderman, Ben, and Catherine Plaisant. "Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies." Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization. ACM, 2006. http://dl.acm.org/citation.cfm?id=1168158
User studies Possible theoretical implications: Data Visualization; Information visualization; Computer Graphics; Visual cognition; Perceptual psychology.	Why, how and when: Building experiments that include human participants.	Kosara, R., Healey, C. G., Interrante, V., Laidlaw, D. H., & Ware, C. (2003). Thoughts on user studies: Why, how, and when. IEEE Computer Graphics and Applications, 23(4), 20-25. https://pdfs.semanticscholar.org/d39b/f307f99188ff66404d2cda78590f3b24127c.pdf
Task and Insight methods Possible theoretical implications: Information Visualization; Data Visualization; Computer Graphics; Bioinformatics.	Empirical evaluation methods: Tasks benchmarking; Comparison of information visualization studies; The insight method's ability to confirm results of the task method.	North, Chris, Purvi Saraiya, and Karen Duca. "A comparison of benchmark task and insight evaluation methods for information visualization." Information Visualization 10.3 (2011): 162-181. http://journals.sagepub.com/doi/abs/10.1177/1473871611415989

Appendix D – VEDC: Theoretical Implications Repository (partial presentation).

Title	Theoretical implications to consider	Source with links
Data Visualization	Theory and practice in the design of data graphics; Graphical presentation of statistics.	Tufte, Edward R., and Glenn M. Schmieg. "The visual display of quantitative information." American Journal of Physics 53.11 (1985): 1117-1118. http://aapt.scitation.org/doi/abs/10.1119/1.14057
	How to model summarizes data; Three strategies for visualizing statistical models (includes case studies).	Wickham, Hadley, Dianne Cook, and Heike Hofmann. "Visualizing statistical models: removing the blindfold." Statistical Analysis and Data Mining: The ASA Data Science Journal 8.4 (2015): 203-225. http://onlinelibrary.wiley.com/doi/10.1002/sam.11271/full
	A new theoretical framework for data visualization approaches; Iterative stochastic matrix approximation for data visualization (includes a set of experiments).	Labiod, Lazhar, and Mohamed Nadif. "A unified framework for data visualization and coclustering." IEEE transactions on neural networks and learning systems 26.9 (2015): 2194-2199. http://ieeexplore.ieee.org/abstract/document/6945382/
Information Visualization	Information Visualization – theories and understanding	Ben Shneiderman and Benjamin B. Bederson. 2003. The Craft of Information Visualization: Readings and Reflections. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. Chapter 8., pp.349 - 351 http://dl.acm.org/citation.cfm?id=961853
	Theories of Visualization	Demiralp, Çağatay, David Laidlaw H., Jarke Van Wijk J., and Colin Ware. "Theories of Visualization—Are There Any?" Theories of Visualization—Are There Any? (n.d.) Brown University. Panel discussion. Web. 02 Sept. 2016. http://hci.stanford.edu/~cagatay/projects/vismodel/TheoriesOfVisualization-Vis11.pdf
	Different approaches to the theoretical foundations of Information Visualization: data-centric predictive theory, information theory, and scientific modeling.	Purchase, H. C., Andrienko, N., Jankun-Kelly, T. J., & Ward, M. Theoretical Foundations of Information Visualization. http://geoanalytics.net/and/papers/springer08a.pdf
Human-Computer Interaction	Design Theories, Methods and Tools	Kurosu, Masaaki, ed. Human-Computer Interaction Theories, Methods, and Tools: 16th International Conference, HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings. Vol. 8510. Springer, 2014. https://books.google.com.au/books?hl=en&lr=&id=D1m7BQAAQBAJ&oi=fnd&pg=PR6v
Information Theory	An information-theoretic framework for visualization	Chen, Min, and Heike Jänicke. "An information-theoretic framework for visualization." IEEE Transactions on Visualization and Computer Graphics. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.372.5270
Color Theory	A theory of color combination (theoretical model): Dimensions of the color combinatorics model; Similarities between color percepts & examples of various order rhythms regarding colors. (visual appearance, texture, basic elements).	Hård, A., & Sivik, L. (2001). A theory of colors in combination? A descriptive model related to the NCS color-order system. Color Research & Application, 26(1), 4-28. doi:10.1002/1520-6378(200102)26:1<4: AID-COL3>3.0.CO;2-T http://onlinelibrary.wiley.com/doi/10.1002/1520-6378(200102)26:1%3C4::AID-COL3%3E3.0.CO;2-T/abstract http://www.lacambrecouleur.be/pdf/A_Theory_of_Colors_in_Combination.pdf

Appendix E – VEDC: Practical Aspects Repository (partial presentation).

Title	Description	Source with links
Procedures & Techniques	Heuristic evaluation compared to user testing	Doubleday, Ann, et al. "A comparison of usability techniques for evaluating design." Proceedings of the 2nd conference on Designing interactive systems: processes, practices, methods, and techniques. ACM, 1997. http://dl.acm.org/citation.cfm?id=263583
	What influences users' preferences?	De Angeli, A., Sutcliffe, A., & Hartmann, J. (2006). Interaction, usability and aesthetics: What influences users' preferences? 2006 271-280. doi:10.1145/1142405.1142446 http://dl.acm.org/citation.cfm?id=1142446
Sample Size	Why you only need to test with 5 users?	Nielsen, J. (2000, March). Why you only need to test with 5 users: Alertbox. Retrieved 3 Sept, 2016 from https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/
	Why and when five test users aren't enough?	Woolrych, Alan, and Gilbert Cockton. "Why and when five test users aren't enough." Proceedings of IHM-HCI 2001 conference. Vol. 2. Eds.) (Cépaduès Editions, Toulouse, FR, 2001), 2001. https://www.researchgate.net/publication/200553185_Why_and_when_five_test_users_aren%27t_enough
	How many users do you need to test?	Lewis, James R. "Sample sizes for usability tests: mostly math, not magic." interactions 13.6 (2006): 29-33. http://dl.acm.org/citation.cfm?id=1167973
Evaluators	Non-expert's performance compares to experts.	Laidlaw, David H., et al. "Comparing 2D vector field visualization methods: A user study." IEEE Transactions on Visualization and Computer Graphics 11.1 (2005): 59-70.
	Domain experts	Brewer, Isaac, et al. "Collaborative geographic visualization: Enabling shared understanding of environmental processes." Information Visualization, 2000. InfoVis 2000. IEEE Symposium on. IEEE, 2000. http://ieeexplore.ieee.org/abstract/document/885102/
	Usability experts	Tory, Melanie, and Torsten Moller. "Evaluating visualizations: do expert reviews work?" IEEE computer graphics and applications 25.5 (2005): 8-11. https://pdfs.semanticscholar.org/21dc/f2a158b05f24b48a3624fee46e8cea6d53c6.pdf
Tools for Data Collection	Testing tools	Teixeira, Carlos, Bernardo Santos, and Ana Respicio. "Usability testing tools for web graphical interfaces." Informatica Dec. 2013: 435. Expanded Academic ASAP. Web. 3 Sept. 2016. http://www.informatica.si/index.php/informatica/article/viewFile/473/477
	Eye-tracking methods	Tarasewich, P. and Fillion, S. Discount Eye Tracking: The Enhanced Restricted Focus Viewer. Proc. AMCIS (2004). http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1961&context=amcis2004
Observer Evaluator Effect	User experience and evaluator intervention	Held JE, Biers DW. Software usability testing: Do evaluator intervention and task structure make any difference? In Proceedings of the Human Factors and Ergonomics Society Annual Meeting 1992 Oct (Vol. 36, No. 16, pp. 1215-1219). Sage CA: Los Angeles, CA: SAGE Publications. http://journals.sagepub.com/doi/abs/10.1177/154193129203601607
	Observer accuracy in usability testing	Adriane M. Donkers, Jo W. Tombaugh, and Richard F. Dillon. 1992. Observer accuracy in usability testing: the effects of obviousness and prior knowledge of usability problems. In Posters and Short Talks of the 1992 SIGCHI Conference on Human Factors in Computing Systems (CHI '92). ACM, New York, NY, USA, 127-128. http://dx.doi.org/10.1145/1125021.1125116