Visualization tools for Open Government Data

ABSTRACT

In recent years many government organizations have implemented Open Government Data (OGD) policies to make their data publicly available. This data usually covers a broad set of domains, from financial to ecological information. While these initiatives often report anecdotal success regarding improved efficiency and governmental savings, the potential applications of OGD remain a largely uncharted territory. In this paper, we claim that there is an important portion of the population who could benefit from the use of OGD, but who cannot do so because they cannot perform the essential operations needed to collect, process, merge, and make sense of the data. The reasons behind these problems are multiple, the most critical one being a fundamental lack of expertise and technical knowledge. We propose the use of visualizations as a way to alleviate this situation. Visualizations provide a simple mechanism to understand and communicate large amounts of data. We also show evidence that there is a need for exploratory mechanisms to navigate the data and metadata in these visualizations. Finally, we provide a discussion on a set of features that tools should have in order to facilitate the creation of visualizations by users.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces

General Terms

Visualization, Tools, Open Data, collaboration, sharing

Keywords

Open Government Data, Visualization, Data Exploration, Smart government

1. MOTIVATION

The steady increase in Open Government Data (OGD) initiatives has created both opportunities and challenges for a wide range of stakeholders including government employees, journalists, researchers, scientists and engineers. Currently more than one million datasets have been made available by governments around the world[27], at national, regional and local levels. These datasets cover all activities in which governments are involved, namely: political boundaries, transportation networks, education performance, health related data, budgets and financial reports.

The release of all this data by government organizations has opened new possibilities in terms of increased transparency, improved reuse of government information, the possibility of more efficient government services and enhanced citizen participation[2], among others. While it is not clear if all of these promises have been fulfilled[3], there are many success stories that indicate OGD is a beneficial initiative to society. From mapping financial transactions[20] to garbage collection notification systems[28], citizens with different interests and expertise are using and consuming OGD for different purposes.

To mitigate these problems, we propose using visualizations as a medium to consume, share and interact with data. The use of diverse styles of visualization has proven useful for understanding large quantities of data in multiple fields, ranging from military to economics to basic science. The problem with existing visualization tools and techniques is that they treat visualizations as finished artifacts. Except in rare situations, current tools do not empower users to explore how data was used, from where it was obtained and how it was displayed. Such tools do not help users create derivative visualizations from existing ones, forcing users who want only a slightly modified version to create it from scratch. We claim that the use of visualizations can greatly enrich the use of Open Government Data if these constraints are overcome.

This paper is organized as follows. Section 2 describes what a visualization is, the current problems for using them with OGD and why most of the tools currently available for creating visualizations are not suitable to solve the problem in the case of OGD. Section 3 shows common tools used to create visualizations based on data. Section 4 describes briefly the stakeholders related to the consumption of OGD via visualizations and their characterization in the Open Government Data Ecosystem. Section 5 shows data related to

a survey focused on evaluating the interest in the use of visualizations by stakeholders. We develop a discussion on a set of requirements that visualization tools for OGD should support in Section 6. Sections 7 describes future work focused on providing such tools and Section 8 provides the conclusions of this paper.

2. VISUALIZATIONS AS A MEDIUM FOR COMMUNICATION

For centuries humans have used visualizations as a way to consume large amounts of data[10], since they provide a mechanism to comprehend, analyze and understand such data in an easy way[11]. For example, by visualizing data it is possible to discover trends and outliers that otherwise would be hard to detect, specially in cases where the observer does not know exactly what she is looking for. It is for these reasons that the use of visualizations is widely spread in finance, science and other data-intensive areas. In this sense, creating visualizations based on Open Government Data can help people to have a better understanding of it. These visualizations can also help people to answer questions such as: *What lessons can be learned from this dataset? What trends can be discovered? How can this data be used to improve government's efficiency?* These and other questions are critical to researchers, data journalists and government employees who use data as a raw material to generate reports, articles and support public policy proposals.

Despite the fact that there are several tools to create visualizations, few deal with Open Data on the Web: Most tools consider whether the data sources are local (Excel or SPSS files for example) or accessible in an internal network (e.g., relational or NoSQL databases). In some cases, this implies a certain level of confidence about the authenticity and immutability of the data. In others, there is an implicit access to the data steward, who might work in the same organization as the person using this data. In general these tools do not acknowledge that data may come from multiple sources, some of which may not be completely reliable. These sources might also change the datasets —thus creating completely new versions or simply removing them from their repositories. Usually these tools do not provide built-in capabilities to maintain the data accountable (e.g., providing the exact URL and date when this particular dataset was obtained).

One of the biggest barriers to creating visualizations is the need of technical knowledge from stakeholders. The open nature of OGD and loose use of standards implies in many cases the need of processing data before and during the process of creating a visualization. For example, *Data.gov* provides many geographical datasets in Shapefile[9] format. In order to create a web-friendly map using *Google Maps*, it is necessary to convert the data to KML[19]. While there are some tools that allow people to perform such a transformation, this action requires a set of extra steps and a minimum knowledge of geographical formats that not all stakeholders may have.

Another problem with existing tools is that they conceive visualizations as finished, immutable resources. Each visualization on the Web can be seen as a read-only, immutable artifact that can be shared "as is" or cannot be shared at all. The problem with this is that it is not possible to reuse or repurpose these visualizations, forcing users to create new versions from scratch or trust the original visualization without any possibility to validate or verify it. For example, in scientific papers, researchers use visualizations such as graphics, tables and pictures to present and support evidence of their claims[11]. Most of the time these visualizations represents data and processes that are not available to the reader. This implies that the audience is forced to trust the author's claims *bona fide*. This is specially difficult to do when the claims made are exceptional[30] or controversial[5]. It is also possible that the claims supported by a visualization are not controversial or exceptional, but users still want to know more about it. An example of such types of problems can be found in the report done by the Bureau of Labor Statistics about unemployment data in the Albany-Schenectady-Troy area[18], as seen in Figure 1. The graph presented is not directly related to any table in the document. In fact, it is not clear where to obtain the raw data, how to change the visualization to show absolute numbers (instead of percentages) or how to compare unemployment rates for Albany with other areas.

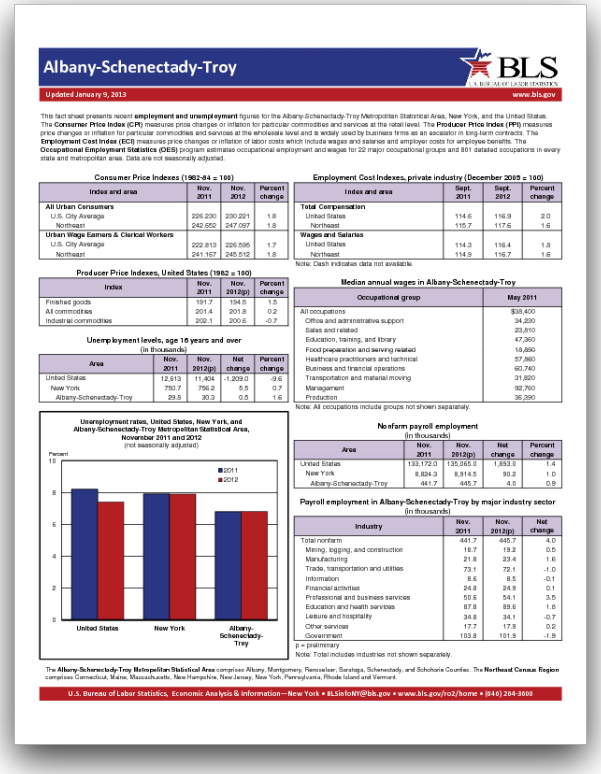

Figure 1: Report of the Bureau of Labor Statistics on unemployment in the Albany Metropolitan area.

In general, a user will ask questions regarding the visualization that cannot be answered by current technology, such as *Was the data used adequate?*, *Is the processing of the data correct and sound?* or *Is this visualization the most suitable for delivering the intended message?*. Following this line of thought, another problem is that it is not possible for users to create new versions of a visualization, limiting the way people can reuse the data and show their own way of visual-

izing this data. This is particularly dramatic to non experts who do not have the knowledge to manage data.

3. RELATED WORK

There is a large number of tools that allow users to create visualizations from data. Reviewing each one of them would be out of the scope of this paper. However, we will present some of the more relevant ones, grouped by the level of complexity or knowledge required by a person to use them. Office Suites are general purpose applications that allow users to create charts from spreadsheets. Business Intelligence tools provide powerful mechanisms to explore data via visualizations and data analysis, including a set of techniques for data drilling. Tools for analysis are generally more powerful and they are used by specialists in a certain domain (finance, chemistry, etc). These people may have a deeper knowledge of their domain but not necessarily of the internal mechanics of the tools. Finally, different libraries and APIs provide a great amount of flexibility in creating different types of visual representations. Using such libraries usually requires some knowledge about programming from the user, since these libraries offer little or no guidance at all.

3.1 Office suites

Currently, there is a large number of libraries and tools available to create visualizations. Office suites, like Microsoft Excel and Google Spreadsheets, allow users to make use of tabular data to create different types of charts. Users select a set of cells and follow a wizard-like workflow to create a specific type of chart (Pie, Bar, Line, etc). The major advantage of these tools is the ease of use, allowing people without significant technical knowledge of computer programming to create charts. On the other hand, given the step-by-step approach used by these tools, they are rather limited and they can only be tweaked on minor aesthetic aspects[29]. These tools are widely used by people without a background in programming or computer science. Another limitation of these tools is that users need the data locally (on their own computer) available in a spreadsheet format[1].

3.2 Business Intelligence software

Many applications focused on Business Intelligence (BI) and Data analytics allow users to create visualizations based on data. In many cases, this is a critical feature for users to explore multidimensional data that otherwise would be too complex for decision makers to understand. An example of such tools is Tableau[16], a software suite that allow users to generate visualizations (mostly charts, but also maps) with data from different sources, including CSV files and relational databases, to name a few. Tableau also offers powerful features for data exploration, such as sorting, filtering, data drilling, grouping and data pivoting. In general Tableau and other similar software offer great flexibility in creating different types of charts. In the same way as office suites, business intelligence software assume the data is correct and immutable. While many visualizations can be created without a deep technical knowledge, there are many cases where the user may be limited by complex data sources. For example, in many cases the data used may be distributed in multiple tables in a relational database. In order to obtain that data, it is necessary to execute a SQL[13] query, something that the user may not know how to do.

[1]In our understanding, it is possible to load remote data using some scripting programming language, however this is beyond the knowledge of common users.

3.3 Analysis and specialized tools

Many tools focused on statistical and numerical computing offer features that help users to create charts based on data. Some of the more well known ones are Matlab[17], Mathematica[31] and the R programming language[12], though many others offer similar capabilities. The audience of these tools are usually people in the fields of science, engineering and technology or people with a strong background in algebra and statistics. These tools usually provide capabilities for generating charts and graphics based on data, such as the one in Table 1. These tools provide powerful mechanisms to use data from a varied number of sources and formats, but they also require some technical expertise for people to use them.

```
maxtemp <- c(10, 15, 16, 14, 20)
mintemp <- c(3, 5, 0, 5, 12)
plot(maxtemp,type="o",col="blue",ylim=c(0,20))
lines(mintemp,type="o",pch=22,lty=2,col="red")
title(main="Temp",col.main="red",font.main=4)
```

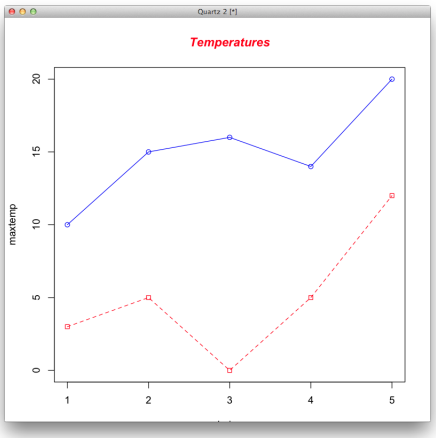

Table 1: Code in the R programming language (upper section) used to generate a chart (lower section)

3.4 Visualization libraries and APIs

Technical-savvy users can make use of more powerful (but also more complex) tools and libraries. These libraries provide great flexibility and an elegant syntax at the cost of a steep learning curve. The tools usually assume knowledge of at least one programming language. In the case of web-based tools, this usually implies a certain experience with JavaScript plus HTML[23] and Cascading Style Sheets(CSS)[15]. Among the most popular libraries are ProcessingJS[25], Protovis[6], D3 (Data-Driven Documents)[7] and Raphaël[1] .

A well-known library is ProcessingJS[25], a JavaScript port of the C-based library Processing[24]. ProcessingJS provides a series of low-level primitives (points, lines, ellipses) that

are applied in the `canvas` element available in HTML5. ProcessingJS uses code written in the Processing programming language (similar to C++) to generate 2D and 3D visualizations.

Protovis provides primitives to basic visual elements (lines, rectangles) called *marks*, which are associated with data elements. Thus, data can be represented as visualizations by composing multiple marks that represent a set of data (e.g., a bar chart is a group of rectangle marks, and each one is associated with a datum).

D3 (Data Driven Documents) is a framework for visualizations and data manipulation in web documents derived from Protovis. D3 allows JavaScript developers to create aesthetically appealing visualizations using Scalable Vector Graphics(SVG)[22]. One of the main advantages of D3 is the ease of binding data with visual elements in the document. This feature gives expert developers great flexibility in creating a wide range of visualizations. This library provides mechanisms for easy consumption of data from JSON and CSV documents, as well as useful auxiliary functions such as color ranges.

Raphaël is another library that makes use of the SVG component in HTML. It also provides primitives to create a series of SVG elements, add events and so on. Raphaël's features allow skilled developers to create compelling visualizations using JavaScript. gRaphaël is a library based on Raphaël focused on creating web-based charts. While Raphaël provides a complete framework for creating any type of visualization, gRaphaël makes it easier for the users who only need to create a chart, such as column or pie charts.

4. OGD STAKEHOLDERS

People related to the Open Government Data Ecosystem (OGDE) have a wide range of interests, abilities and knowledge relative to data management. It is possible to find certain analogies between OGDE and a ecological ecosystem. There are a set of participants that provide a specific good (data) that is later consumed by others. Also, these consumers may create new subproducts that may be used by other participants. Another important feature of this environment is the diversity of interests, goals and method of the participants: Some organizations release data simply to comply mandatory regulations, while others may have a more proactive approach. On the other hand, developers are interested in using the available data for creating applications, while businesses may seek ways to make profit out of it. Finally, common citizens may utilize this data via reports, visualizations or applications. Maybe one of the more important characteristics of OGDE is the asynchronous, uncoordinated way participants relate to each other and with the data; there is no central control or messaging center that coordinates interests, purposes and efforts of the participants. This differs greatly from other data-focused environment, like enterprise data repositories, where all the aspects of it (from data formats to goals of the participants) are controlled by a central organization with a set of well-defined objectives. Similar to a natural ecosystem, the boundaries of an OGDE are hard to define and are blurry at best. For some participants it may involve one data repository and a couple of collaborators, for others it will imply data available in dozens of data portals, the communication with others via Social Media with a group of collaborators.

4.1 User profiles in OGDE

Based on a series of 12 interviews with different people related or interested about OGD, it was possible to characterize a set of profiles that identify people involved in an OGDE. In some cases some persons may fit in two or more of these profiles, since they represent a role rather than a group of specific people.

4.1.1 Government Data Provider

This profile identifies a group of people that work in the government and are focused on making government data available. In many cases, Government Data Providers are in charge of projects related to Open Data in their own organization. In other cases they are focused on improving interoperability inside the government.

4.1.2 Government Data Consumer

Government Data Consumer represents a group of people that consume OGD *from* the government. Their work implies the use of government data, whether to perform *ex-post* analysis (e.g., analytics, annual reports), or generate proposals for future public policies. There are two distinctive aspects of Government Data Consumers. First, in general, they use OGD for specific purposes related to their work, so the subproducts (e.g., reports) created by them are defined by the tasks they need to accomplish. Second, they have better access to data stewards, whether in their own organization or others in the government. This helps Government Data Consumers solve problems when data is not available or when the data representation is unclear.

4.1.3 Researcher/Journalist

Researchers and Journalists refer to people out of the government that work in newspapers or Non Governmental Organizations (NGOs) and that are interested in using OGD for their work. For example, many NGOs focused on public policies related to education or health, use government data to perform studies, write reports and lobbying to promote their agendas.

4.1.4 Civil Programmer

The term Civil Programmer refers to a group of people with a high level of technical expertise in areas such as programming languages, databases, and other information technologies. These people work with OGD usually in their spare time, creating applications, mashups or adapting the data to their needs. In contrast to Government Data Consumers, Civil Programmers have a broader set of objectives when it comes to use OGD. In general, Civil Programmers are motivated by personal interest: some may want to look for potential cases of corruption in the government (e.g., a chart showing an analysis of government budgets), while others simply want to offer a service that society may find useful (e.g., a mobile application displaying the location of farmers markets).

4.1.5 Common Citizen

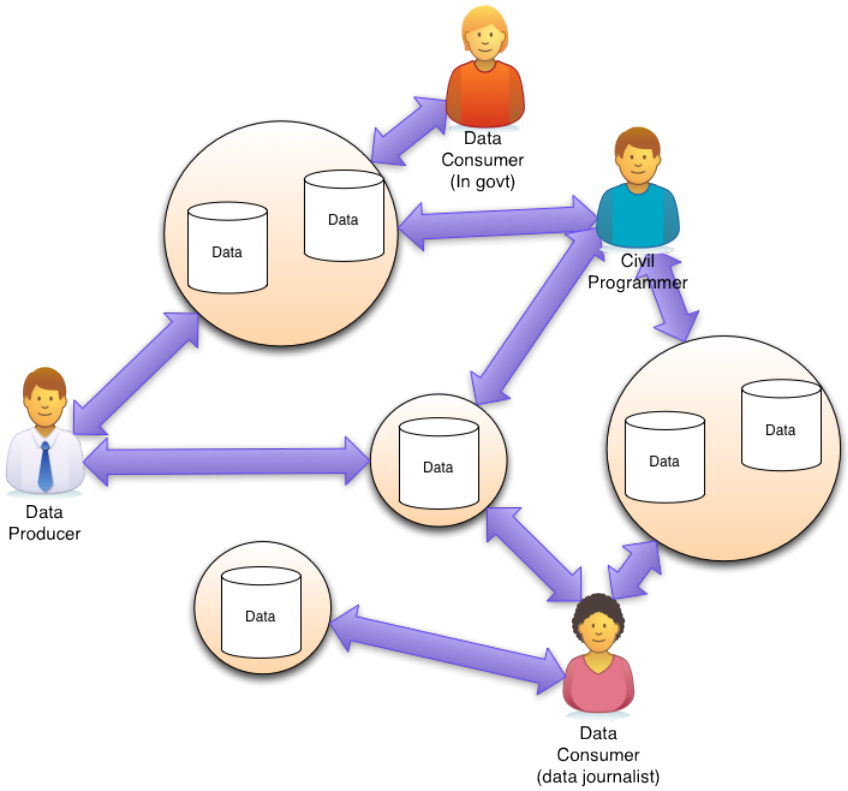

Figure 2: Diagram showing an Open Government Data Ecosystem. Several users use one or more datasets (not necessarily from the same source) in a distributed, asynchronous way.

Finally, what we refer to as Common Citizen is the large majority of the population that eventually may consume data via an application, a visualization or a report, but it is not directly involved with OGD.

4.2 Profiles considered for this work

We claim that a certain group of people want to create and use visualizations based on OGD but they cannot, since they do not have the technical knowledge to do so. Out of the five profiles shown above, two were considered not relevant for this work.

First, Civil Programmers define a group of technically savvy people with knowledge in data management, databases and programming languages. This characteristic makes them unlikely to have technical problems in creating data visualizations based on OGD. Thus, it is reasonable to consider that Civil Programmers are of no concern in this scenario. Second, Common Citizens describe a group of people that are not directly interested in OGD, although they may consume it via applications or visualizations. There is evidence on a large scale[21] as well as at a small scale[8] showing that social participation is mostly done by a reduced group of people who are really interested in a certain theme. Thus, it would not be reasonable to try to engage Common Citizens in the creation and reuse of visualizations based on OGD.

Thus, only people that can be classified as Government Data Provider, Government Data Consumer or Researcher/Journalist will be considered as relevant for the purpose of this study.

5. SURVEY ON USE OF VISUALIZATIONS

Finding people that fit the defined profiles that were also willing to be interviewed was a time-consuming activity. Most of these people did not have much time for the interviews due their work load. In fact, several potential people to be interviewed did not have time available at all for the next several months. For this reason, we created an online questionnaire that could be easily answered by many people, without the need to coordinate face to face interviews. The survey can be answered in 5 minutes or less. It starts by asking a few demographic questions and later it focuses on what type of data people use and their main limitations. The full survey can be found at `http://0n.cl/3`.

Although this questionnaire was not specific about OGD, it was focused on the use of data in general, since we did not wanted to restrict potential useful answers and feedback from other domains.

5.1 Survey results

We obtained 257 answers from people, most of them working in academia, government, private sector and NGOs. 85% of them were between the ages of 26 and 55 years old, and almost 75% of them had a masters or a doctoral degree. Over 80% of the people who answered the questionnaire had a background related to science or technology. The majority of them (95%) declared using data (in any form) in their

work.

One important question asked was how they defined themselves in terms of their technical skills and expertise with information technologies. As described in Section 4.2, Our focus was on people that are not expert programmers, but still have minimal knowledge on how to use a computer, and ideally how to use some basic data management tools, such as Microsoft Excel. 42% of the people who answered the questionnaire claimed to *know how to surf the Web, check emails, twitter, etc* (2%), *know how to use Excel, Word and other office software* (13%), *know some basic programming language* (25%) or *Other* (2%). We defined the people that fall into one of those categories as our group of interest. The following results are based only on this group of interest.

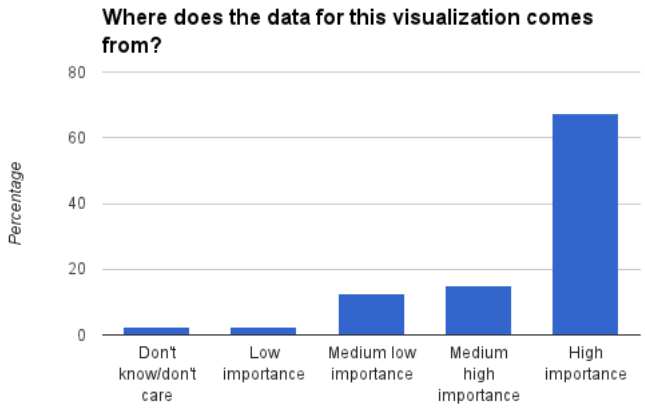

Figure 3: Number of votes assigned by non-experts expressing the importance of the question: *Where does the data for this visualization comes from?*

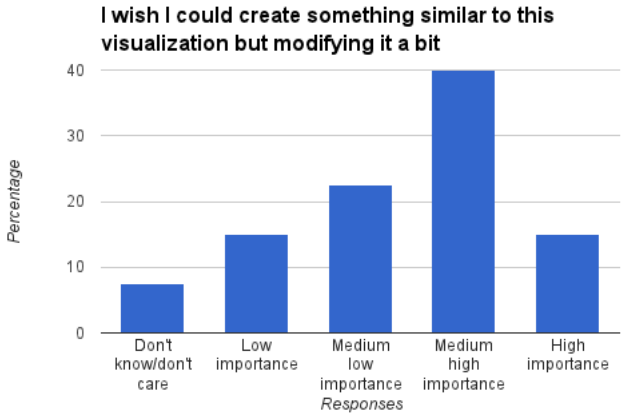

Figure 4: Number of votes assigned by non-experts expressing the importance of the statement: *I want to create something similar to this visualization but modifying it a bit*.

One of the questions asked was how do they value the question *Where does the data for this visualization comes from?*. In Figure 3, it is clear that the ability to find the original source from which a visualization was created is considered of high importance. In contrast, the idea of being able to "tweak" a visualization (based on the value of the phrase *I want to create something similar to this visualization but modifying it a bit*) was not considered as critical, as seen in Figure 4. One possible explanation for this is that this option may be considered desirable but not critical, given that a high number of answers still consider it of "Medium high" importance. This may be justified by the notion that analyzing or exploring a visualization is a simpler task than creating or tweaking a new one [14]. Also, usually exploring or analyzing a visualization is a required step to tweak it. In that sense, many people may find the ability to explore a visualization more critical than the ability to modify one.

Analogously, Figure 5 shows the results of the importance of knowing how a visualization was processed (e.g., which algorithms, filters and heuristics were applied to the data). A large portion of the respondents considered this question of "high importance" or "medium-high importance." Figure 6 shows the importance given to the phrase "I have other data and I want to create a visualization similar to this one with it." Similar to the results from Figure 4, the majority of the respondents indicated that this is of "medium-high importance," followed by "high importance." A similar explanation, as in the case of Figure 4, is that this option implies a higher effort than finding how the data was processed or where the data comes from. Thus, it is not surprising to find that people consider it more critical to discover how the visualization was produced rather than being able to create a similar one based on their own data.

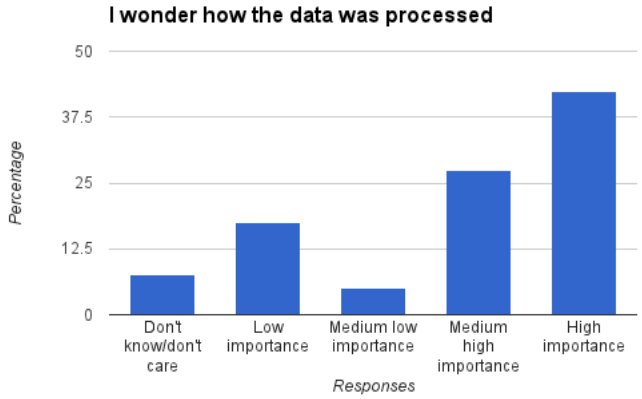

Figure 5: Number of votes assigned by non-experts expressing the importance of the question: *I wonder how the data was processed?*.

6. DISCUSSION

From the survey described in Section 5, we can elaborate a set of requirements or recommendations that will help stakeholders in the use of OGD via the creation and use of visualizations. We list a set of requirements that OGD visualization tools should consider.

6.1 Facilitate visualization creation

One of the most critical requirements for these tools is to simplify the process of creating a new visualization based on OGD. For example, one of the final comments in the survey said, "I want a tool where I can create a visualization in, say, no more than 6 clicks." Thus, it is recommended

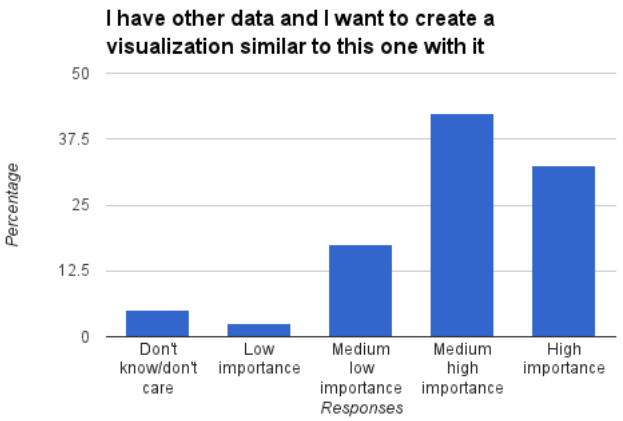

Figure 6: Number of votes assigned by non-experts expressing the importance of the phrase: *I have other data and I want to create a visualization similar to this one with it*

that tools should provide a preliminary visualization (even if it is a scaffold) after a few interactions from the user. This preliminary visualization can be later debugged based on further input from the user.

6.2 Exploratory mechanisms

Figures 3 and 5 show a significant interest in further exploration of data provenance of a visualization by stakeholders. Thus, tools focused on OGD visualizations, should implement a set of mechanisms to explore the data and metadata behind a visualization. These mechanisms should provide the current location of the data that is visually represented, the original location of the data where it was taken from (if they differ), the time of creation of the visualization, along with other pieces of information.

6.3 Ease of communication with data providers

Many stakeholders may not be clear about the meaning of the data released (e.g., *what does column* `FY1998` *mean?*) or what dimensions the data covers (e.g, temporal, spatial). Also, it is possible in many cases the the data provided may not be well documented or the conventions used are not clear enough for others to use it. In this sense, a tool for creating visualizations based on OGD should facilitate the communication between data consumers (Government Data Consumers or Researchers and Journalists) and Government Data Providers. In this way it would be possible to clarify questions that consumers have in respect to the data during the process of creating a visualization. Ideally, the feedback and clarifications from the Government Data Providers should be made available for others, to avoid or minimize repeated questions. Another useful characteristic would be if these tools could provide mechanisms to contact the person or organization responsible for the integrity, correctness and consistency of the data. In many cases the visual representation of data permit the discovery of errors in the data. A further step in the data lifecycle would be to contact the Government Data Providers to report these errors. In this way it is possible to close a loop in the data lifecycle, benefiting from the community by improving data quality based on such reports.

6.4 Viralization and sharing

In many cases, the process of creating a visualization is followed by the subsequent action of sharing it. It is necessary to simplify the access to different sharing methods that stakeholders often use to share information. For example in many communities, the use of Social Media, such as Facebook or Twitter, is a valid and useful communication channel. Thus, it is necessary to provide simplified mechanisms to share the newly created visualizations via these media.

6.5 Reuse of visualizations

In some cases it would be desirable for stakeholders to create revised versions of an existing visualization created by a third party, as can be seen from Figures 4 and 6. Thus, tools that allow the deconstruction, modification and creation of a new version of a visualization could be useful for the stakeholders. This requires that the tool has to record and encode metadata on how the visualization was built, including which sources were used and how they were processed. This metadata needs to be retrieved to provide the schematics of how the visualization was built, so the stakeholders can edit such schematics to create a modified version of the visualization.

7. PROTOTYPE AND FUTURE WORK

We have developed a demo that features most of the capabilities described in Section 6, such as providing provenance metadata about where the data came from, when the visualization was created and what parameters were used for it. Along with that, this prototype provides a copy of the original copy of the datasets used, including its MD5[26] hash signature. The reason behind this is to provide accountability, so when a dataset used in a visualization is changed in the original source (e.g., the dataset was updated), it is possible to acknowledge the difference between the current version of the data and the data used to create the visualization. By providing an offline copy of the data and its MD5 hash, it is possible to at least to confirm that the data has been changed (although MD5 does not indicate *what* changed, that is left for further investigation). It is important to note that all the data in the visualization is published both in human as well as machine-readable formats, by using semantic technologies and applying the principles of Linked Data[4]. In this way, it is also possible for humans to consume a visualization, but for machine and automatic agents to obtain all the necessary data to recreate such visualizations. Our demo provides a set A screenshot of our prototype can be seen in Figure 7.

There are several paths to consider as future work. First, there are still some features we want to add in our prototype, such as including aggregated functions. This will provide users more options to create meaningful visualizations based on raw data. We also want to perform a user study to test this prototype with real stakeholders in order to confirm our hypothesis and fix potential usability problems and software bugs. We specified that our intended audience was people with an interest and certain expertise in OGD. However, the question of how we can empower Common Citizens, as described in Section , remains an open question. Thus, further research is needed to evaluate how citizens can participate in the OGD ecosystem and how we can make it easier for them to consume OGD.

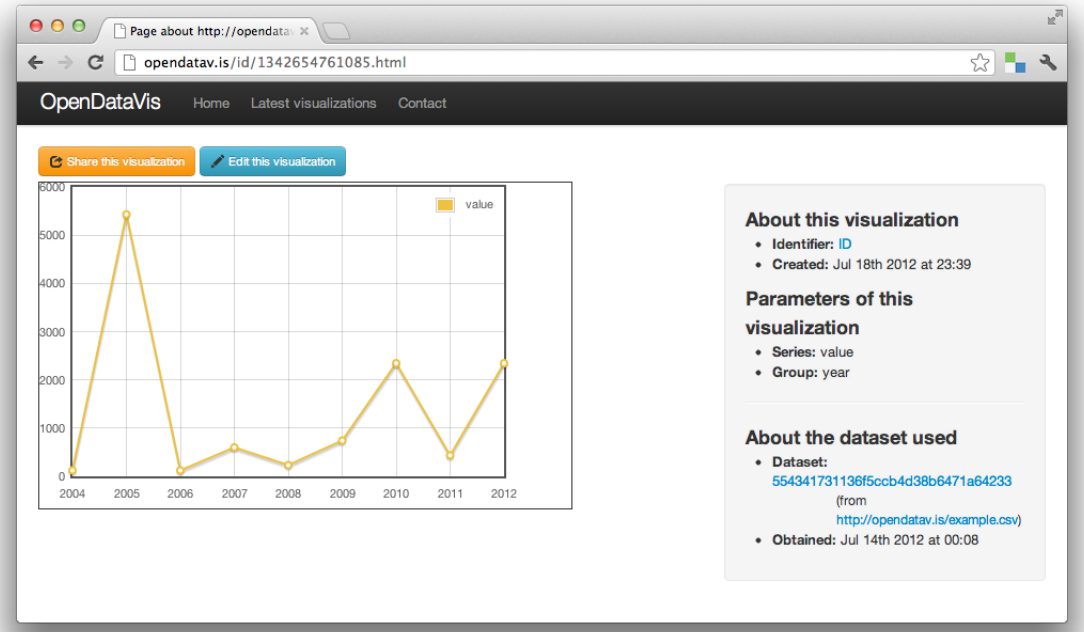

Figure 7: Prototype of a tool to create visualizations based on OGD. One of the distinctive features is the ability to provide metadata

8. CONCLUSIONS

In this paper we have shown that there is a gap between current Open Data initiatives and an important part of the stakeholders of the Open Government Data Ecosystem. We showed that this gap prevents or limits stakeholders in making extensive use of publicly available data, due to lack of technical knowledge. We showed that a possible way to facilitate data consumption is via the use of visualizations, but in order to do that providing better tools to the stakeholders is needed. We showed evidence that there is a real interest by stakeholders to create, reuse and explore such visualizations and that, in many cases, the existing tools are not enough. We also discussed a list of requirements that these new tools would need to cover to support the needs of the stakeholders, whether for creating, exploring or reusing visualizations. Finally, we showed the progress in a tool that support most of the requirements discussed in Section 6, along with a few others, such as machine-readable metadata and data accountability.

9. REFERENCES

[1] D. Baranovskiy. Raphael–javascript library `http://raphaeljs.com/`, 2010.

[2] D. Bennet and A. Harvey. Publishing open government data (w3c working draft 8 september 2009). *World Wide Web Consortium. http://www. w3. org/TR/2009/WD-gov-data-20090908*, 2009.

[3] J. C. Bertot, P. T. Jaeger, and J. M. Grimes. Using icts to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. *Government Information Quarterly*, 27(3):264 – 271, 2010.

[4] C. Bizer, T. Heath, and T. Berners-Lee. Linked data-the story so far. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 5(3):1–22, 2009.

[5] J. Borowski. Bush-era tax cuts reduce deficit and increase employment. `http://www.freedomworks.org/blog/jborowski/bush-era-tax-cuts-reduce-deficit-and-increase-empl`.

[6] M. Bostock and J. Heer. Protovis: A graphical toolkit for visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 15(6):1121–1128, 2009.

[7] M. Bostock, V. Ogievetsky, and J. Heer. D^3 data-driven documents. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2301–2309, 2011.

[8] D. DiFranzo and A. Graves. A farm in every window: A study into the incentives for participation in the windowfarm virtual community. 2011.

[9] ESRI, E. Shapefile technical description. esri. *INC, http://www.esri.com*, 1998.

[10] S. Few. *Data Visualization for Human Perception*. The Interaction-Design.org Foundation, Aarhus, Denmark, 2010.

[11] P. Fox and J. Hendler. Changing the equation on scientific data visualization. *Science*, 331(6018):705, 2011.

[12] R. Ihaka and R. Gentleman. R: A language for data analysis and graphics. *Journal of computational and graphical statistics*, 5(3):299–314, 1996.

[13] ISO. ISO/IEC 9075-1:2008: Information technology. Database languages. SQL, Part 1: Framework (SQL/Framework), 2008.

[14] S. Kalyuga and J. Sweller. Rapid dynamic assessment of expertise to improve the efficiency of adaptive e-learning. *Educational Technology Research and Development*, 53(3):83–93, 2005.

[15] H. Lie and B. Bos. *Cascading style sheets*. Addison-Wesley, 1997.

[16] J. Mackinlay, P. Hanrahan, and C. Stolte. Show me: Automatic presentation for visual analysis. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1137–1144, 2007.

[17] MATLAB. *version 7.10.0 (R2010a)*. The MathWorks Inc., Natick, Massachusetts, 2010.

[18] B. of Labor Statistics. Selected bls economic indicators albany-schenectady-troy.

[19] Open Geospatial Consortium. Keyhole markup language specification. `http://www.opengeospatial.org/standards/kml/`.

[20] Open Knowledge Foundation. Open Spending. `http://openspending.org/`, 2011.

[21] J. Preece and B. Shneiderman. The reader-to-leader framework: Motivating technology-mediated social participation. *AIS Transactions on Human-Computer Interaction*, 1(1):13–32, 2009.

[22] A. Quint. Scalable vector graphics. *Multimedia, IEEE*, 10(3):99–102, 2003.

[23] D. Raggett, A. Le Hors, I. Jacobs, et al. Html 4.01 specification. *W3C recommendation*, 24, 1999.

[24] C. Reas and B. Fry. *Processing: a programming handbook for visual designers and artists*. Mit Press, 2007.

[25] J. Resig, B. Fry, and C. Reas. Processing. js. *Aufgerufen am*, 20:2010, 2010.

[26] R. Rivest. The md5 message-digest algorithm. 1992.

[27] E. Rozell, J. Zheng, Y. Shi, T. Lebo, J. Erickson, J. Shinavier, G. T. Williams, and A. Graves. IOGDS: International Open Government Dataset Search. `http://logd.tw.rpi.edu/demo/international_dataset_catalog_search`.

[28] Tim Bray. Recollect.net. `http://recollect.net/`, 2012.

[29] L. Wilkinson. *The grammar of graphics*. Springer, 2005.

[30] F. Wolfe-Simon, J. Blum, T. Kulp, G. Gordon, S. Hoeft, J. Pett-Ridge, J. Stolz, S. Webb, P. Weber, P. Davies, et al. A bacterium that can grow by using arsenic instead of phosphorus. *Science*, 332(6034):1163, 2011.

[31] S. Wolfram. *The MATHEMATICA® Book, Version 4*. Cambridge university press, 1999.