

City Data 3.0 - A generic initiative to promote and assess the reuse of geographic information in cities - Early steps

Fernando Benitez *

*Universitat Jaume I **

*Av Vicente Sos Baynat, S/N, Castellon - Spain **

Sergi Trilles

Universitat Jaume I

Av Vicente Sos Baynat, S/N, Castellon – Spain

Joaquín Huerta

Universitat Jaume I

Av Vicente Sos Baynat, S/N, Castellon – Spain

ABSTRACT

Nowadays, authorities around the world discuss on the way to engage their data users and involve to being part of open data initiatives in cities. Thereby, assess the re-use level of available information through those portals is a significant challenge to achieving the real potential of open data in cities. This article presents the early steps of an initiative called CityData 3,0 which is a generic strategy to improve and assess the re-use and discovery level of available open geographic data in cities. This initiative has been developed in three phases, to analyze, define and test a group of recommendations for both data consumers and data authorities. Using an on-line survey to know the current status of cities' open data portals and a set of participatory workshops in cities in Colombia (Bogotá, Medellín, Cali) and Spain (Valencia, Castellon). The project aim is to understand which barriers can affect the data selection process of data re-users and attend data producers authorities to assess the re-use level of open data initiatives in cities.

KEYWORDS

Open Data, Open Cities, Data Re-use, Open Government, Data Discoverability.

1. INTRODUCTION

For a long time, one of the main authorities' concern was to set up and populate open data portals with high data quality services (Attard et al. 2015), so that they were enriched by extensive data catalogs that follow most of the current standards.

However, the current challenge is not only downloading data or publishing web services; it is about data consumer requirements, their needs, engagement and about finding the way to involve them to be a part of the data selection process. What data selection process means, is the group of activities, that data users should face when they are looking for data.

The amount of data and comprehensive metadata could make challenging the data selection process (Janssen et al. 2012). Some official open data portals provided no more than simple data downloads and limited functionalities (Alexopoulos et al. 2013; Attard et al. 2015). Thereby, the availability of data has created a strain of barriers (Janssen et al. 2012) that data consumers must consider before using and transforming in new information; in many cases, they must download the data and after they can use it in their projects.

According to, The Belmont Forum's Open Data Survey (Schmidt et al. 2016) understand how to interpret and re-use data, how to access data and, varying degrees of quality in different datasets, and varying data formats, are the most significant aspects to take into account in open data portals.

Data consumers are demanding different and new ways to use and identify the relevant data for supporting applications to societal problems according to (Schmidt et al. 2016).

City Data 3.0 is a user-center initiative that can be beneficial for two city stakeholders, initially for authorities (Data producer or compilers) who need to assess and improve the re-use and impact level of available open data in cities, and data users who need to find the data for their projects.

Re-usability and discoverability are most frequent topics of discussion in events on open data and geographic standards. Several talks presented about data user's engagement mechanisms, tools to improve available information.

The main contributions of this work are: (1) Understand from perspective's data user the current barriers that avoid (developers and analysts) taking full advantage of available open data in cities, especially open geographic data. (2) Study the current status of cities open data portals. (3) Validate the responses gathered from on-line survey against data user suggestions in the hands-on activity. Once the analysis phase finished, the second aspect of this initiative is creating a generic strategy with indicators to assess the reuse level of available open data, which is a constant concern of the selected authorities.

2. METHOD DEFINED

This section describes the methodology employed, including three phases to cover the requirements and analysis steps, the definition of strategy and a final stage to testing. Whereby we started to develop the initial analysis stage for a better understanding of the real needs and barriers of data consumers might have once they are looking for open data in cities.

Phase No1. Contains two sub-phases, **Discovery** which was an on-line survey to assess the current status of open data portals in cities, it was public and share to everyone. The second sub-phase was a set of **participatory workshops** that allowed to interact with data users especially in places where we got more survey online responses, to prove what people say and what people do, not matter which role they have.

Phase No2. Which is the current stage, also contains two sub-phases, **Strategy design** and **Proof-of-concept** this stage is focused on the strategy and the indicators to assess the reuse level creating a guideline for selected cities.

Moreover, Phase No.3 **Test and Evaluate** is about testing the strategy and usefulness of the created guidelines with targeted cities stakeholders to have feedback.

3. INTINIAL RESULTS PHASE 1

The aim of this stage was studying the current barriers present among data consumer and cities' open data portals. For better understanding about how data consumers find available data in cities and which could be the barriers to re-use the official open data portals. The first step was the developing an on-line survey.

On-line Survey. We have used Google Forms to create a survey, and we shared to open data users such as, researchers, developers, analysts, managers and overall people who are actively involved with data especially open geographic data in cities.

To learn about the knowledge level of open data portals in cities, backgrounds of the data consumer, barriers to reusing open data, feedback from data user perspective and still rooms where open data portals can be improved. The survey was shared in several countries, but most of the responses were from Colombia and Spain.

Participatory Workshops. To contrast the perception' data users (on-line survey) and issues mentioned by data users, we created Open Data for Open Cities workshop which had two parts, finding the available open data and discussion related to re-use level in cities.

During the first part, we observed activities made by data users, and we attempt to identify how data consumers are looking for data. Then, the second part of this workshop was an open discussion over the reuse importance, impact of use available data in cities, which barriers they have to overcome and which issues they found in the first part of the workshop.

3.1 On-line survey results.

From 9 August 2016 to 16 March 2017, 195 responses were gathered from the on-line survey that was publicly shared. The survey had 22 questions some of them were non-mandatory, i.e. could be skipped by respondents. However, this article presents the most significant results. The survey had several sections as follows.

General information. Overall data gathered from 16 countries and several cities based on the information of 195 responses. Bogota, Cali, Medellin and Spanish cities were the most frequent chosen cities in this survey (Cities with 1 or 2 replies were clustered).

Your Work. For the section related to respondent's role, 25 % of roles are geographical data analysts, and the next categories are researchers (19%), project managers (18%) and developers with 17% (n = 195 responses). It means most of the respondents have a critical role to improve the re-use level; they are data consumers by definition, they have knowledge to create analysis processes, to develop applications, and to lead projects where the geographic data is considered.

City open data. In this section, the respondents were consulted about their knowledge over cities' open data portals and the reasons to use this geographic information. Using three categories, the respondents had to choose among "very important," "neutral" and "not important." Those who provided a response (n=195), only 37,4% consider "very important" use city open data for economic benefits. However, 95,38 % consider very importantly the accessibility of geographic information and 92,2% recognize data quality as important feature as we expected. From our point of view is relevant that 69,23% of respondents mentioned academic and research improvement as "very important" category when open data is reused.

Barriers and features. Overall, the most important barriers to use cities' open data portals were a lack of update published data, low integration of data sources, and low relevance of URL to access published data (see Figure 1). Over 90 % of respondents (n=194 valid responses) considered the time spent searching open data was likewise marked a major and moderate barrier. Barriers such as technology used for publishing data and understanding about how to re-use the data were least marked up as a major barrier.

In contrast to barriers, the respondents provided information related to the most used functionalities available in cities' open data portals (n=186 valid responses). Filters for advanced search (61,29%), Data categories (59,14%) and access data URL (54,30%) were marked up as every time. Nevertheless, feedback from other data users (25,27%), terms of re-use were less pronounced (22,58%).

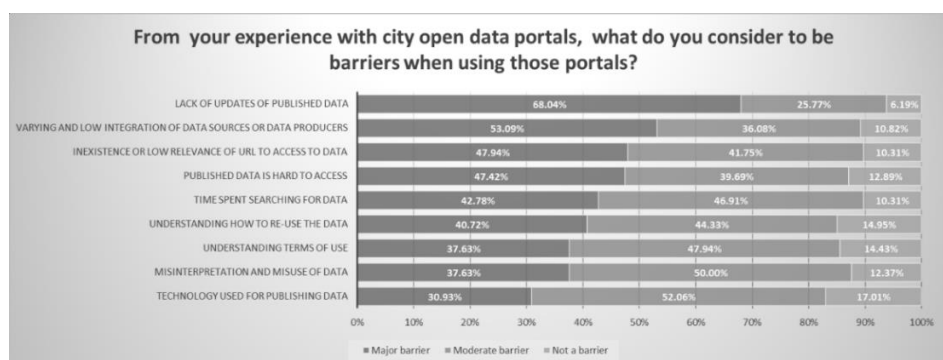


Figure 1. Barriers to use open data portals, n= 194 valid responses.

Searching for geographic data. Related to which features are considered when data users are choosing available open data in cities. With 184 valid responses, the creation/ publication date (77.17%) and data quality (73.91%) are main features marked up by respondents.

Although, the term of use and re-use which is currently a major discussion topic only considered by 35.87% of responses. Note that 65,22% of answers determined author as the main feature to choose available open data, even though the geometry (64.67%) of available data is highly considered. The survey also asked about output formats or web services more suitable where data consumers are choosing open data in the city.

Overall respondents (80.11%) still want to have access to shapefiles format and downloadable files (64.52%) in another way around of the new capabilities of open data portals with functionalities to create API to serve direct connection between data and users.

Surprisingly, one of the recent formats to access open data RDF only has 16.67%, and formats as JSON and REST services are seen least useful (44.62%, 30,11%). KML and OGC standards (WMS, WFS, WMTS) have 50% and 54.30 % of 186 valid responses

3.2 Open Data for Open Cities - participatory workshops results

The second sub-phase after on-line survey was a set of participatory workshops; those activities were prepared to collect information related to barriers and suggestions to compare the first data user's perception with the real actions when users are looking for data in specific challenge.

Taking into account only cities where we got more responses, Bogotá, Medellín, Cali (Colombia), Castellón and Valencia (Spain) those workshops had more than 140 attendants and cooperation of local authorities (IDECA¹ which is the Spatial Data Infrastructure of Bogota and Medellin², Cali³, and Valencia⁴ City Hall).

The aim was to discuss and evaluate suggestions to improve the re-use level of open geographic data. Using the open data categories, we request to data users find open data and explain to data authorities why they are facing issues to access and reuse open data in their city. Collecting suggestions, and observing how data users find open data, those participatory workshops were an interesting opportunity to engage more data consumers with the current cities' open data initiatives and enrich the open data reuse discussion.

During these hands-on activities, the initiative collected several different sorts of barriers that were not mentioned during the on-line survey for instance in Valencia the main concerns of data users was the integration and cooperation among the data producer departments, especially in mobility sector. In Bogota, re-users such as entrepreneurs do not have enough knowledge about available services, although IDECA has integrated city geographic data. Medellín is in the middle of open data initiative; however, data users expect to know more over the process about, with concerns like, how city open data project will be release? What technology will be used? Moreover, how data user communities are involved?

4. DISCUSSION

Currently, most of cities have their open data initiatives, through those projects accessibility, discoverability, and reusability of open data have been promoted. As Andrus Ansip, Vice-President for the Digital Single Market said "If we want our data economy to produce growth and jobs, data needs to be used. However, to be used, it also needs to be available and analyzed."⁵

Notwithstanding, cities have different ways to promote their open data. Through a local SDI in Bogota and Cali web services of mobility, environment, cadastral, security and more are accessible. In Medellín and Valencia, is the open data projects the way to release data from several categories.

These initiatives also include geographic data which are not the core of those initiatives. The way to display data can affect the re-use level and engagement level of new data consumers, likely the data communities in both situations are different.

Regarding barriers faced by data consumers, we found that terms to use and understanding about how to re-use the data were the options least marked in contrast with the argue by workshop attendants mentioned where terms of use are not clear in current cities' open data portals. Despite most of the open data portals have the terms of use available, probably the language used is not natural, these terms come from legal perspectives which are not a common skill and language for data re-users.

The on-line survey and workshop results also have some mutual points. Access to URL and advanced filters to search data were common requirements present in both methods. It means that data users expect to have the enough accessibility to find and re-use the available data in external applications. Data quality criterion was likewise an important fact mentioned by data consumers.

Criteria such as date of publication, how data was produced and which authority has created the data are important when users are looking for data. The authorities should put more relevance in those features to engage more users and most importantly, not lose data user confidence.

Another interesting finding of the survey was the persistence to have access to downloadable data. Re-users want to have easy and quick access to download the data and process it by themselves. The question is, that option promote the re-use level of available data, considering all barriers that data producers should overcome.

¹ <https://www.ideca.gov.co>

² <https://www.medellin.gov.co>

³ <http://idesc.cali.gov.co>

⁴ <https://www.valencia.es>

⁵ http://europa.eu/rapid/press-release_IP-17-5_en.htm

How authorities can assess the impact of available data if data consumers download data. This one is an important discussion point that this strategy wants to involve and present some recommendation for both data consumers and data producers as a future work.

However, there is a limitation regarding last point of discussion which is the sample group in the on-line survey, is likely that respondents that prefer local access and downloadable formats belong to the Geoscience community, where those formats and processes in local environments are common. Nevertheless, participatory workshops were critical comparison tool, promoted by the local administration and with more 140 participants with different roles, to contrast if this finding is correct or not. We observed that when data users mentioned the accessibility barrier was when they could not download the data, not for lack of web services and during the discussion the most requirement by data consumers was more efficient ways to download data.

5. CONCLUSION

Data consumers who need an easy tool to find the right way available open data around the city, filtering gaps of data, creation/publication date, geographic properties, the term of use without spending much time to understand how data should be utilized.

In this article, we present the preliminary results of phase 1 of this initiative. We identified the important barriers and features requirements of cities' open data portals based on 195 data re-users responses, and finally, we tested with other data users with different roles through a set of participatory workshops called *Open data for Open Cities* to discuss the economic and social benefits and real impact in targeted cities.

We plan to integrate those results and conclusions of an on-line survey and set of workshops to analyze the barriers to re-use available open data in the selected cities, through MCA (Multiple Correspondence Analysis) we will determine which barriers are most statistical significance for the current requirements of open data consumers.

This initiative attends authorities to have a generic strategy to engage and assess the reuse of the open data. For better understanding of cities' behavior, the next step is to integrate the project with other research themes such as a participatory framework to understand which information the authority should consider and focus on improving their upcoming decisions.

ACKNOWLEDGEMENT

The European Commission has funded this work through the GEO-C project (H2020-MSCA-ITN-2014, Grant Agreement number 642332, <http://www.geo-c.eu/>).

REFERENCES

- Alexopoulos, C., Spiliotopoulou, L. & Charalabidis, Y., 2013. Open data movement in Greece: A Case Study on Open Government Data Sources. *Proceedings of the 17th Panhellenic Conference on Informatics - PCI '13*, p.279. Available at: <http://dl.acm.org/citation.cfm?doid=2491845.2491876>.
- Attard, J. et al., 2015. A systematic review of open government data initiatives. *Government Information Quarterly*, 32(4), pp.399–418. Available at: <http://www.sciencedirect.com/science/article/pii/S0740624X1500091X> [Accessed December 14, 2015].
- Janssen, M., Charalabidis, Y. & Zuiderwijk, A., 2012. Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, 29(4), pp.258–268.
- Schmidt, B. et al., 2016. Open Data in Global Environmental Research: The Belmont Forum's Open Data Survey R. Guralnick, ed. *PLOS ONE*, 11(1), p.e0146695. Available at: <http://dx.plos.org/10.1371/journal.pone.0146695> [Accessed July 11, 2016].