

A framework for linking open environmental data

Juan Santiago Preisegger¹ , Ariel Pasini , Patricia Pesado 

Computer Science Research Institute LIDI (III-LIDI)*

Facultad de Informática – Universidad Nacional de La Plata 50 y 120 La Plata Buenos Aires

* Partner Center of the Scientific Research Agency of the Province of Buenos Aires (CIC)

¹ Fellow UNLP

{jspreisegger, apasini, ppesado}@lidi.info.unlp.edu.ar

Abstract. The concept of open government has promoted several initiatives in order to progress in their implementation in different governmental agencies. Generally, implantation is carried out, in the first place, based on transparency and the opening of data, giving citizens relevant information about their community. Within that information, data about the rational use of natural resources and care of the environment can be found. In this context, and taking into account the ongoing analysis, it was discovered that, apart from the positive measures taken for the progress of the implantation, the main existing flaw is the possibility to link and relate the existing data in different data sources. The linking of data will allow for a closer analysis that can generate a wider context for information. This paper is based on a PhD thesis in process whose aim is to generate a framework that allows the linking and relationship of information related to the environment published in different open government portals.

Keywords: Open government, Open data, Data linking, Related open data, Environment.

1 Introduction

The demands of society towards their rulers and government entities, such as transparency, inclusion in decision-making processes and collaboration, are, nowadays, enormous. Therefore, a new way of government that includes citizens, allowing them to help in public politics and to participate in decision-making processes of the government, emerges [1]. This method of governing is called open government and it is defined as "...a technological and institutional platform that turns governmental data into open data to let citizens use, protect and collaborate with public decision processes, accountability and improvement of public services" [2]. It is based on three main and well-defined principles: Transparency, Collaboration and Participation [3] [4]. In papers related to this area, it could be observed that there exist standards in the dataset publication formats [5] [6], however, the main flaw is how data is loaded in an orderly manner to improve the linking and relationship to others.

Due to changes that affect the world, a significant level of awareness about the rational use of resources and the environment care was generated in the world society. The governmental agencies are not indifferent to this matter, and great progress is observed regarding data published.

As part of the PhD thesis, it was analyzed the difficulties when linking open environmental data published by different organizations; particularly, the work is focused on energy, water and air data.

Section 2 presents the state of the art on issues of open government and open data. In the third section, the main existing problem in the linking of these data is described. In section 4, the framework for linking open data, that is expected to be carried out to conclude with the PhD thesis, is presented. Finally, in section 5, conclusions and the expected results are described.

2 Open data

Currently, citizens present enormous demands towards their rulers and government entities. Among these demands, transparency and the efficient management of public assets, inclusion in decision-making processes and collaboration with different areas of society, are included. According to these requirements, and with the help of new technologies, a new way of government that includes the citizen was generated; this way allows them to generate contributions to public politics and to participate in the decision-making processes [7].

Through the implantation of open government, there was an opening of data from governmental agencies. It strengthens two activities that are part of the performance of societies: critical thinking and decision-making process, now based on more information [8]. However, the main difficulty that citizens find is that such information is spread in different data sources, such as portals and catalogues, which makes it impossible to link and relate these data in order to obtain a global vision of the topic.

3 Linking open data

Through the boom generated by this new paradigm within the governmental agencies of the world, several applications and tools were generated from different sectors to make the search, integration and data processing automatic or to better them. One of these tools is Google Dataset Search, a search engine that is specialized in finding sets of data stored in the web via keywords, as long as they use schema.org dataset tags or equivalent structures represented in Data Catalog Vocabulary (DCAT) format [9].

Various searches were carried out to obtain datasets related with energy, water and air topics published worldwide, under these schemes. Additionally, datasets published by *UN Water* [10], *U.S. Energy Information Administration* [11] and *European Statics* [12] were analyzed. They contained relevant information but, as they didn't fulfill the scheme, they weren't reached by Google Dataset Search. The searches were carried out with the following words: *Drinking water quality*, *Energy generation*, *Air quality*. From

these searches, a group of dataset of each topic was selected and the shapes, the available formats and the structures of each of them were analyzed.

Regarding air and water quality, it was possible to raise awareness of certain standard to analyze the existent magnitudes in different characteristics. It could be observed that almost the same tests on different samples were carried out to analyze different characteristics and determine if they are within the healthy margins for the human use. Nevertheless, there exist differences in the structure of dataset, such as, variation from columns to rows or the division of a field into many, which complicate the linking between different datasets of such topics.

In the case of energy area datasets, the difference between published data from different governmental agencies is even greater. It was observed that: some countries simply publish the annual percentage generated and the type of energy obtained in which different published generation stations are based on; other countries simply publish the percentages of their energy generation sources, without discerning between the stations they have; and other countries publish geo-referenced data of the location of each station, the total capacities, the types of stations and even their owners. It is possible to observe that there is a potential for the interrelationship of data in this field, with the consequent standardization and the selection of certain fields in common among different dataset.

4 Framework for linking open data

As part of the PhD thesis, it is expected to carry out the state of the art study in relation to usual problems that are detected in the publication of environmental data in different portals of open government. It is expected to identify a group of existing patterns, similitudes and differences in the different technologies used, the data publication formats and their structure, which prevents interrelation and, due to the study aforementioned, it is expected to generate a framework for linking open data.

5 Conclusion

This work suggests analyzing different data sources that governmental agencies provide and generate a framework for open data, particularly in water, air and energy data. The framework of this study allows for the selection of data available in different portals in order to change them and combine them, according to the aim the parties concerned wish to analyze.

Acknowledgments. Project co-funded by the Erasmus+ Programme of the European Union. Grant no: 598273-EPP-1-2018-1-AT-EPPKA2-CBHE-JP.

6 References

- [1] S. A. Chun, S. Shulman, R. Sandoval, and E. Hovy, "Government 2.0: Making

- connections between citizens, data and government,” *Inf. Polity*, vol. 15, no. 1–2, pp. 1–9, 2010.
- [2] J. R. Gil-García and J. I. Criado, *Las Tecnologías de Información y Comunicación en las Administraciones Públicas Contemporáneas*. 2017.
- [3] C. Calderón and S. Lorenzo, *Open Government. Gobierno Abierto*. 2010.
- [4] A. Naser and A. Ramirez, “Plan de gobierno abierto. Una hoja de ruta para los Gobiernos de la Región,” *CEPAL - Manuales*, vol. 81, p. 80, 2017.
- [5] A. Pasini, J. S. Preisegger, and P. Pesado, “Modelos de evaluación de gobiernos abiertos , aplicado a los municipios de la provincia de Buenos Aires,” *XXIV Congr. Argentino Ciencias la Comput.*, vol. XXIV, pp. 0–10, 2018.
- [6] A. Pasini, J. S. Preisegger, and P. Pesado, “Open Government Assessment Models Applied to Province’s Capital Cities in Argentina and Municipalities in the Province of Buenos Aires,” in *Communications in Computer and Information Science*, 2019, vol. 995, pp. 355–366.
- [7] A. Naser, Á. Ramírez-Alujas, and D. R. Editores, *Desde el gobierno abierto al Estado abierto en America Latina y el Caribe: Planificación para el Desarrollo*. 2017.
- [8] J. Pastor Verdú, “Conceptos y fenómenos fundamentales de nuestro tiempo,” *Unam*, p. 10, 2012.
- [9] N. Noy, M. Burgess, and D. Brickley, “Google dataset search: Building a search engine for datasets in an open web ecosystem,” in *The Web Conference 2019 - Proceedings of the World Wide Web Conference, WWW 2019*, 2019, pp. 1365–1375.
- [10] <https://www.unwater.org/>
- [11] <https://www.eia.gov/>
- [12] <https://ec.europa.eu/eurostat/>