Proceeding Paper

Analysis of the Types of Argentine Geospatial Public Open Data [†]

Roxana Martínez * [ID] **and Agustín Simón**

Facultad de Tecnología Informática, Centro de Altos Estudios en Tecnología Informática (CAETI),
Universidad Abierta Interamericana (UAI), C1270AAH Buenos Aires, Argentina;
miguelagustin.simon@alumnos.uai.edu.ar
* Correspondence: roxana.martinez@uai.edu.ar
† Presented at the II International Congress on the Technology and Innovation in Engineering and Computing,
Online, 21–25 November 2022.

Abstract: Massive data, public and in open formats, are essential to improving citizens' confidence in their countries. Open data generate value, as long as they can be standardized in terms of data quality, accessibility, and publication through user-friendly formats. This work consists of an analysis and study of the different types of open geospatial data that are available in the government website portals of the Argentine Republic. This analysis allowed us to garner the status of different geospatial datasets, understand the quality of their content, and detect the shortcomings of these types of datasets.

Keywords: open data; geospatial public open data; quality metrics; open government; datasets

1. Introduction

In the digital age, data is becoming the new "gold" and is part of countries' economic growth. Massive data, public and in open formats, are essential to improve the confidence of citizens in their countries. In addition, this favors citizens, facilitating access to information and improving the quality of public information services. It is essential to understand that open data can generate value as long as it can standardize in terms of data quality, accessibility, and publication through user-friendly and user-friendly formats.

The opening of government data faces multiple political, legal, and technical challenges, including issues such as the reliability of the published data, the protection of the privacy of individuals, and the quality of the content of datasets and the available data. Despite these challenges, Latin America is a region highly committed to the open data agenda in the context of Open Government. This work consists of an analysis and study of the types of open geospatial data that are available in the government website portals of the Argentine Republic. This analysis enables us to comprehend the status of the different geospatial datasets, understand the quality of their content, and detect the shortcomings of this type of format. Another contribution of this work is the presentation of a prototype that verifies some aspects corresponding to the measurement of geospatial quality metrics.

2. Open Government and Open Public Data

Open Government makes it possible to guarantee that the administration and operation of all the public services that the nation-state offers can supervise by the community.

The Open Data Program in Science and Technology of the Argentine Republic defines public data as any data that are generated in the governmental sphere or under its custody and that are not access restricted by any specific legislation. On the other hand, public data "is everything that can be freely accessed or consulted by any per-son or organization, although it is not necessarily digitized data" [1]. Martínez [2] indicates that this comprises the public data that are available in a digital medium under an open license and using

an open standard format. Additionally, and to belong to this category, the data must be complete, primary, up-to-date, processable by machines, and susceptible to treatment, and must not be discriminatory, proprietary, or subject to copyrights, patents, trademarks, or trade secrets regulation.

In [3], the authors clarify that open does not mean free but, rather, at a reasonable cost or proportional to its value. Reusable data means that the data must be available in a convenient form so that they can add to other datasets and can use by citizens or other public or private entities. On the other hand, redistributable means that this data must provide with licenses or terms of agreement that allow its use without commercial or other restrictions. Garriga [4] indicates that it is essential to have a standardized process that makes public data from the public administration available to the society at large in digital and standardized formats as well as in open ones.

In the context of open government, it is important to include the concepts of reuse and interoperability, so it is necessary to define a standardization protocol for the process of opening datasets and the content of those datasets. "This reuse of open data allows the development of new digital products and services, creating opportunities for social and economic development" [5].

3. Problem and Proposal

Within this context, inconveniences are generated, and some of these are:

- The datasets provided in the open data portals do not meet a standard.
- Although there are international principles and criteria for open data, there is no focus on the analysis of their content.
- There are problems that can mitigate beforehand in structural and format aspects (interoperability). Datasets are not always sufficient or easily readable.
- The importance of measuring the quality of what is available in order to favor an adequate analysis of the results.

This proposal is based on the study of guides and good practices of government open data publications [6] and guides prepared by the National Public Administration (APN) [7] for opening and processing the content of public datasets. In each of these, aspects of the quality of open public geospatial data are identified and analyzed.

4. Analysis of Results

Based on the analysis of the sample of fifty-two datasets from the following organizations, Buenos Aires Data [8], Bahía Blanca Data [9], Data.gob.ar [1], and the Argentine Ministry of Culture [10], the results shown in Figure 1 were found. It is shown that the type of open geolocation format with the largest amount is GeoJSON with twenty-one datasets, and then the CSV format is presented with more datasets, before, lastly, we see the format SHP with five datasets. This implies that one of the most-used formats in Argentine open data portals is GeoJSON.
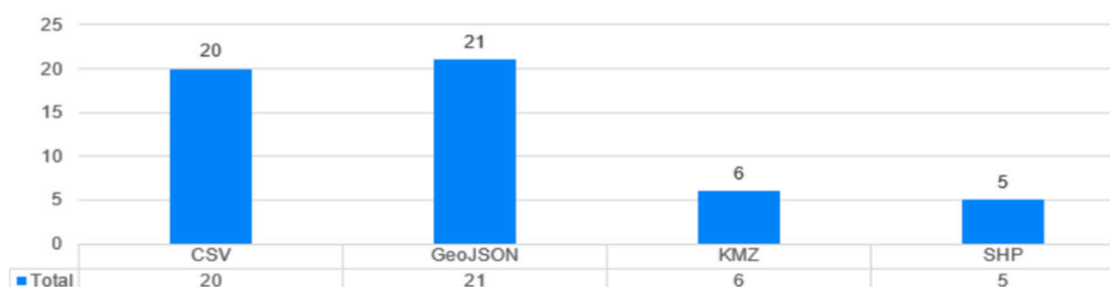

Figure 1. The number of datasets by types of formats available on the analyzed websites.

Based on the analysis, some of the advantages found for each type of geolocation open data format detected in the Argentine government portals are presented in Tables 1 and 2.

Table 1. Advantages of open geolocation data types, part I.

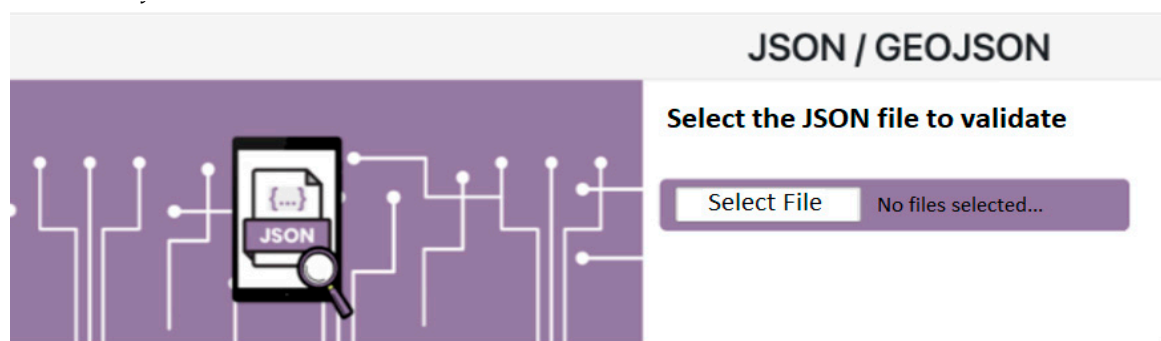| GeoJson	TopoJson	CSV
Self-descriptive and easy to understand, and its simplicity has allowed it to position itself as an alternative to XML. Fast anywhere. Browser. Easier to read than XML. High processing speed. Can be natively understood by JavaScript parsers.	Eliminate redundancies. Quantify Coordinates. A total of 80% reduction in volume relative to GeoJson. json file extension and geojson.	Easy to create. Readable. Easy to analyze.

Table 2. Advantages of open geolocation data types, part II.

SHP	KMZ/KML	WKT
Format supported by various applications (due to the fact that it is open), free map processing, and free code.	Allows user to easily insert points, polygons, etc. on the map with the Google Earth tool. User interfaces are simple and elegant. Editing functions, Calculation, and geoprocessing are becoming more advanced.	Supports several types of geometries, such as: Point, Linestring, Polygon, Multipoint, and Multipolygon.

5. Proposal of a Prototype

Based on the shortcomings detected and the types of open geospatial data, we worked on the development of our own prototype, which allows a basic validation of the structure of the JSON/GeoJSON geospatial data type of a dataset analyzed and extracted from a portal for public open data. The choice of this type of format was based on the identification of several datasets of this type.

The technical aspects of the prototype are: (a) programming a web application in Angular; (b) Development of an API with NodeJS to process and validate the data sets; (c) A database engine with MongoDB technology used. Figure 2 shows a validation splash screen in JSON format.

Figure 2. The initial screen of the developed prototype is shown.

The prototype proposed by the authors validates the predefined schemes. For example, in the "Geometry" property, it is validated that the data of the latitudes and longitudes between brackets ("[]") are displayed. If the dataset passes all of the internal validations scheduled for the prototype, the software will show the geolocation data of each of the validated dataset records on the map. In the event that the content does not adapt to any scheme, the system will return an error message to the user indicating the problem.

6. Conclusions and Future Work

This work analyzed the datasets focused on geolocation, using longitude and latitude coordinates, in addition, it used the analysis of data in geospatial formats, for example, files of the type: WKT (coordinate points), and SHP (geographic coordinate points), among others. As explained in the previous sections, this paper has presented the detection of flaws in government datasets made available for geolocation, and, as a result, we developed a small dataset content validation prototype. This work contributes to the analysis, verification, and understanding of the current state of the most relevant types of formats of the datasets generated by the government entities of Argentina according to the analyzed sample.

For future work, the survey and the study of new quality aspects in geospatial datasets could be expanded, and, in addition, work can continue on expanding new validations for the self-developed tool as well as incorporating geospatial machine learning techniques.

References

1. Datos.gob.ar. Datos Argentina. Available online: https://www.datos.gob.ar/ (accessed on 10 October 2022).
2. Martínez, R. Métricas de Calidad Para Validar Los Conjuntos de Datos Abiertos Públicos Gubernamentales. Doctoral Thesis, Universidad Nacional de La Plata, Buenos Aires, Argentina, 2022.
3. Hernández-Pérez, T. En la era de la web de los datos: Primero datos abiertos, después datos masivos. *Profesional de la Información* **2016**, *25*, 517–525.
4. Garriga-Portolà, M. ¿Datos Abiertos? sí, pero de forma sostenible. *El Profesional de la Información* **2011**, *20*, 298–303. [CrossRef]
5. Cadena-Vela, S. Marco de Referencia Para la Publicación de Datos Abiertos Comprensibles Basado en Estándares de Calidad. Doctoral Thesis, Universidad de Alicante, Alicante, Spain, 2019.
6. Datos.gob.ar. Guía Para la Publicación de Datos en Formatos Abiertos. Available online: https://datosgobar.github.io/paquete-apertura-datos/guia_abiertos/ (accessed on 10 October 2022).
7. Datos.gob.ar. Guía Para la Apertura de Datos en Organismos de la Administración Pública Nacional. Available online: https://datosgobar.github.io/paquete-apertura-datos/guia-apn/ (accessed on 10 October 2022).
8. BA Data. Buenos Aires Data. Available online: https://data.buenosaires.gob.ar/dataset/ (accessed on 10 October 2022).
9. Datos.bahia.gob.ar. Datos Bahía. Available online: https://datos.bahia.gob.ar/ (accessed on 10 October 2022).
10. Argentina.gob.ar. Ministerio de Cultura. Available online: https://www.argentina.gob.ar/cultura (accessed on 10 October 2022).