Mobility Open Data: Use Case for Curitiba and New York

Elis C. Nakonetchnei[1], Nádia P. Kozievitch[1], Cinzia Cappiello[2], Monica Vitali[2], Monika Akbar[3]

[1]Dep. de Informática, UTFPR, Curitiba, PR, Brazil

[2]Politecnico di Milano, Milan, Italy

[3]University of Texas at El Paso, El Paso, Texas, USA

elisan@alunos.utfpr.edu.br, nadiap@utfpr.edu.br,

{cinzia.cappiello,monica.vitali}@polimi.it, makbar@utep.edu

Abstract. *As cities are becoming green and smart, transportation systems are being revamped to adopt digital technologies. Open transportation data might include GIS maps for bus routes and bus stops. Such data provide a number of opportunities to analyze a segment of the current state of the transportation system of a city. In this paper, we address and discuss some of the urban mobility open data available and their characterization from the perspectives of two metropolitan cities: Curitiba (Brazil) and New York City (USA). Finally, we present challenges regarding their use.*

1. Introduction

The quality of life in a city greatly depends on the quality of its public transportation. The growing population of a city creates demand on existing infrastructure. Accommodating such demand calls for introducing new measures or changing existing infrastructure, both of which are time-consuming process and require in-depth analyses of the current state of public transportation as well as the current and future demand. One of the first steps of understanding urban transportation challenges is to understand the current state of public transportation. In this paper, we analyze the open data related to transportation for two metropolitan cities (Curitiba and New York) to gather cities open data for project EuBra-BigSea [1]. In particular, we address and discuss some of the urban mobility open data available and their characterization from the perspectives these two cities, along with challenges regarding their use.

2. Comparison

Curitiba encompasses 75 districts, and has developed and implemented mass transport corridors, densification of land-use along these corridors, and mobility solutions using Bus Rapid Transit (BRT) systems in the 1970s, where one main feature of the success of the system is its complex network of feeder lines [Duarte et al. 2016]. The city has also been participating in the open data initiative, through several government stakeholders, such as Instituto de Planejamento de Curitiba (IPPUC)[2] and the Municipality of Curitiba[3].

[1] Availableathttp://www.eubra-bigsea.eu/ - Last accessed on 11/03/2017.
[2] http://ippuc.org.br/Lastaccessedon9/03/2016
[3] http://www.curitiba.pr.gov.br/DADOSABERTOS/ – Last accessed on 09/03/2016.

General Transit Feed Specification (GTFS) data is available for buses [4], but still some differences are presented (such as missing lines, compared to the total number of bus lines available). Although Curitiba has a small network of trains, they are not used for citizen mobility (just one line, for example, is used for tourism). All the data is georreferenced and can also be downloaded at IPPUC. Curitiba does not have subways neither ferries. The datasets used in this paper come from IPPUC, the Municipality of Curitiba, along with data from Open Street Maps. In particular, the bus mobility system has the following characterization.

Bus Stops. The city has 9940 bus stops detected. Bus stops are divided among tube stations (officially 342), regular bus stops and terminals. The districts named Cidade Industrial de Curitiba (CIC) and Centro have the majority of regular bus stops, with a total of 1628 and 667 units in each one. Figure 1-D represents the bus stops in the city.

Bus Routes. The city has 482 bus routes distributed within 11 categories (*Metropolitano, Linha Direta*, among others). The categories Alimentador and Convencional have the majority of lines, with 265 and 65 units respectively. Figure 1-E represents the bus routes in the city. Different colours indicate different categories.

Bus Terminals. The city has 23 terminals (buses) and one terminal which also use trains. The oldest one is named Guadalupe, from January 1st, 1956. The districts CIC and Boqueirão concentrate their majority, with 3 and 2 units each one.

New York is composed of five boroughs: Manhattan, the Bronx, Queens, Brooklyn, and Staten Island.Metropolitan Transportation Authority (MTA) [5] runs most of the transit system of this city. There are more than 238 local routes, 62 express routes, and 7 Select Bus Service routes. GTFS data is available [6], integrating several mobility categories. The New York City Subway is the largest rapid transit system in the world by number of stations, with an average 469 stations in operation, and 25 train services. All the data is georreferenced and can also be downloaded. The datasets used in this paper come from the Department of City Planning from New York [7], the City University of New York [8], along with data from Open Street Maps. In particular, the bus mobility system has the following characterization.

Bus Routes. The city has 261 bus routes distributed within 3 categories (Express Service, Limited-Stops and Local Service). The category Local Service has the majority of lines. Figure 1-G presents the bus routes, where darker lines present intersections of routes at the same street.

Bus Stops. The city has 16,254 bus stops, and 200 bus stop names, within regular bus stops. Queens has the majority regular bus stops, with a total of 5,329. Figure 1-F presents the bus stops.

[4] `http://www.urbs.curitiba.pr.gov.br/faleconosco` – Last accessed on 02/12/2016.

[5] `http://web.mta.info/developers/developer-data-terms.html#data` – Last accessed on 09/03/2016

[6] `http://tracker.geops.ch/?z=13&s=1&x=-8226561.0174&y=4962941.5682&l=transport` – Last accessed on 02/12/2016.

[7] `http://www1.nyc.gov/site/planning/data-maps/open-data/districts-download-metadata.page` – Last accessed on 09/03/2016.

[8] `https://www.baruch.cuny.edu/confluence/display/geoportal/NYC+Mass+Transit+Spatial+Layers` – Last accessed on 09/03/2016.

Bus Terminals. The city has 3 terminals(The Port Authority Bus Terminal, George Washington Bridge Bus Station and Journal Square Transportation Center). No shapefiles were found for bus terminals.

Discussion. The distribution of the public transportation and related open data regarding bus is different along the two cities. Curitiba concentrates the bus traffic along the downtown area and some streets, but presents several bus terminals along the city. New York do not concentrate the bus lines, but present other types of transportation (such as subways and trains). Regarding data presentation, both cities have information which can be easily integrated (shapefiles along with other data types), but both also could be more helpful, for example, adding some metadata (such as date of the last update, or historical list of shapefiles) or an integrated portal, with all data sources from transportation within the municipality. Nevertheless, data from different sources (in Curitiba, for example) still present different information. Thus being necessary to integrate them to allowing a better handling of this data simultaneously.

3. Challenges and Data

In Brazil, the vehicle fleet in major cities grew more than the road structure [9]. Nevertheless, open data is still a new trend regarding data availability. Mobility challenges have already gained attention of the computer science society in Brazil [10]. In particular, these challenges can be grouped in the following areas: (i) pattern discovery, (ii) data statistics, (iii) data integration, (iv) location and tracking, (v) open and connected data, (vi) contextual information, among others. These challenges can have other subtypes such as [Kozievitch et al. 2016]: (i) **Different File Formats**: the different coordinate systems (UTM, Latitude and Longitude, etc.), and information along different sources keep the data integration as a challenge. In general, formats such as CSV, Excel, Json and Shapefile are preferred (some use APIs and KML). Not all of them provide metadata or the data visualization; (ii) **Different Reference Systems**: The data from IPPUC uses SAD69 (South American Datum from 1969), but the official standard in Brazil is SIRGAS2000 (Geocentric Reference System of the Americas). Global Positioning Systems, for example, uses WGS84 (World Geodetic System 1984). Note that in parallel, data such as street names and districts change over time. (iii) **Different File Structures / Precision / Accuracy within Data**: the official bus line data from IPPUC and URBS (listed in Figure 1-top), for example, is different. Data accuracy issue (Figure 1 - bottom left), presents the source with an atribute value which is different to its value in real world. (iv) **Open and Connected Data**: In Brazil , decrees for Open Data (decree 1135/2012, from law 12.527 [11]), competitions named Hackatons[12], as well as integration of partners for future

[9] http://www1.folha.uol.com.br/cotidiano/2014/08/1503030-frota-de-veiculos-cresce-mais-rapido-que-a-estrutura-viaria-no-pais.shtml – Last accessed on 09/03/2016.

[10] http://www.sbc.org.br/documentos-da-sbc/send/141-grandes-desafios/802-grandesdesafiosdacomputaonobrasil – Last accessed on 09/03/2016.

[11] http://multimidia.curitiba.pr.gov.br/2014/00147194.pdf – Last accessed on 09/03/2016.

[12] http://hackathon.curitiba.pr.gov.br/ – Last accessed on 09/03/2016.

projects[13] [14] are helping the data use by population. Within USA, several open data sites are available (such as Open Government [15]), but not all states are participating. (v) **Data Quality:** in order to better explore the data, metrics might be considered to verify non relevant data: errors, missing or outdated values can negatively affect decisions.

Figure 1. Different Precision/Accuracy and Values for bus data.

4. Conclusions

This paper presented an initial investigation and characterization in order to identify scenarios and implications from the urban mobility open data from the city of Curitiba and New York. Detected characteristics can be used for further analysis in order to optimize the transportation system, and contribute to standards to transportation data worldwide. This work will be extended to provide characterization in order to provide data mining and suggestions enhance mobility in both cities. **Acknowledgments**. Thanks to Municipality of Curitiba, IPPUC, NSF grant HRD-1242122 and EU-BR EUBra-BigSea project (MCTI/RNP 3rd Coordinated Call).

References

[Duarte et al. 2016] Duarte, F., Gadda, T., Luna, C. A. M., and Souza, F. T. (2016). What to expect from the future leaders of Bogotá and Curitiba in terms of public transport: Opinions and practices among university students. *Transp. R. Part F: Traffic Psychology and Behaviour*, 38:7 – 21.

[Kozievitch et al. 2016] Kozievitch, N. P., Gadda, T. M. C., Fonseca, K. V. O., Rosa, M. O., Gomes-Jr, L. C., and Akbar, M. (2016). Exploratory Analysis of Public Transportation Data in Curitiba. In *XXXVI CSBC*, pages 1656–1666. Sociedade Brasileira de Computação.

[13] http://www.curitiba.pr.gov.br/noticias/curitiba-e-holanda-vao-trabalhar-juntas-em-projetos-de-ciclomobilidade-para-a-cidade/37601 – Last accessed on 09/03/2016.

[14] http://multimidia.curitiba.pr.gov.br/2015/00166636.pdf – Last accessed on 09/03/2016.

[15] https://www.data.gov/open-gov/ Last accessed on 09/03/2016