Pauliceia 2.0: A Computational Platform for Collaborative Historical Research

Karine R. Ferreira[1*], Luis Ferla[2*], Gilberto R. de Queiroz[1], Nandamudi L. Vijaykumar[1], Carlos A. Noronha[1], Rodrigo M. Mariano[1], Yasmin Wassef[1], Denis Taveira[1], Ivan B. Dardi[1], Gabriel Sansigolo[1], Orlando Guarnieri[2], Daniela L. Musa [2], Thomas Rogers[3], Jeffrey Lesser[3], Michael Page[3], Andrew G. Britt[3], Fernando Atique[2], Janaina Y. Santos[4], Diego S. Morais[4], Cristiane R. Miyasaka[2], Cintia R. de Almeida[2], Luanna G. M. do Nascimento[2], Jaíne A. Diniz [2] and Monaliza C. dos Santos[2]

[1]INPE - National Institute for Space Research, Sao Jose dos Campos – SP – Brazil
[2]UNIFESP - Sao Jose dos Campos and Guarulhos – SP – Brazil
[3]Emory University (The Halle Institute for Global Learning, The Emory Center for Digital Scholarship, and The Department of History) – Atlanta – United States
[4]Arquivo do Estado de Sao Paulo – Sao Paulo – SP – Brazil

*{karine.ferreira@inpe.br, ferla@unifesp.br}

Abstract. *Digital humanities research promotes the intersection between digital technologies and humanities, emphasizing free knowledge sharing and collaborative work. Based on the digital humanities features, this paper presents the architecture of a computational platform for collaborative historical research that is being designed and developed in an ongoing project called Pauliceia 2.0. This platform uses volunteered geographical information (VGI) and crowdsourcing concepts to produce historical geographic data and to allow historians to share historical data sets resulting from their researches.*

1. Introduction

Digital Humanities (DH) research takes place at the intersection of digital technologies and humanities, producing and using applications and models that make possible new kinds of teaching and research both in humanities and in computer science [Terras 2012]. The digital humanities community is interdisciplinary, linking together the humanistic and computational approaches. This community includes people with different disciplines and methodological approaches that come together around values such as openness and collaboration [Spiro 2012].

DH have drawn increasing institutional support and intellectual interest among scholars working on historical research in universities throughout the world. Historians working within digital humanities promote the use of Geographical Information Systems (GIS), among other tools, to understand historical data. DeBats and Gregory [DeBats and Gregory 2011] argue that GIS has directly contributed to the advancement of knowledge in history and that the principal topic so far developed within this field is urban history.

Free knowledge sharing and collaborative work have indeed become core features of digital humanities [Spiro 2012]. The role of the world network of computers, in particular web 2.0, has boosted these aspects of the field. It places value not only on the broad dissemination of studies and investigations, but also on opportunities for collaboration and putting into practice those theoretical values. Nowadays historians can benefit from a wide variety of technological options to disseminate their research widely and to participate in collaborative investigation across boundaries of time and space.

In literature, there are a variety of terms used to represent the general subject of collaborative work and citizen-derived geographical information, such as volunteered geographical information (VGI), science 2.0, crowdsourcing and collaborative mapping. See et al.[See et al. 2016] present a good review of these terms, providing some basic definitions and highlighting key issues in the current state of this subject. The authors categorize these terms according to three main aspects: (1) information or process that can be used to generate it; (2) active or passive contributions; and (3) spatial or non-spatial user-generated information.

The term VGI was first defined by Goodchild [Goodchild 2007] as “the harnessing of tools to create, assemble, and disseminate geographic data provided voluntarily by individuals”. According to [Estellés-Arolas and González-Ladrón-de Guevara 2012], crowdsourcing is “a type of participative online activity in which an individual, an institution, a non-profit organization or company, proposes to a group of individuals of varying knowledge, heterogeneity and number, via a flexible open call, the voluntary undertaking of a task. The undertaking of the task, of variable complexity and modularity, and in which the crowd should participate bringing their work, money, knowledge and/or experience, always entails mutual benefit.”.

In GIScience, many efforts have been made to propose general frameworks and protocols that can be followed by VGI projects. Davis Jr. et al. [Davis Jr et al. 2013] propose a general framework for VGI applications, based on coordinating both web-based tools and mobile applications. Mooney et. al. [Mooney et al. 2016] propose a protocol for the collection of vector data in VGI projects. Besides providing a standard for data collection, the protocol also guides users to contribute to improve the overall data quality of the project, which positively impacts their motivation to keep on providing information. In the recent years, an increasing number of projects have used crowdsourcing and VGI concepts to produce historical geographic data. Examples of crowdsourcing projects to collect historical geographic information are shown in next section.

This paper presents the architecture of a computational platform for collaborative historical research that is being designed and developed in an ongoing project called Pauliceia 2.0. This project has two main objectives. The first is to collect, select and digitize historical data of São Paulo city from 1870 to 1940. During this period the city went through a dramatic process of urbanization, almost unique in terms of contemporary history. This transformation was taken as a challenge by several historians to investigate a range of issues within this period. The second goal is to design and build a computational platform that allows humanities researchers to explore, integrate and publish urban historical data sets. This platform will appeal to historians to not only explore historical data sets provided by the project, but also to contribute by including and sharing their own knowledge and data sets.

2. Related work

In this section, we present some projects that have similar features to Pauliceia 2.0 and highlight differences between the computational platform proposed in this paper and the ones provided by such projects.

OpenStreetMap (OSM) is the most well-known general platform that implements VGI successfully. It allows users to edit and work with free geographical data, following an open content license [OpenStreetMap 2017]. There are many applications that are built on top of the OSM database. Two examples of OSM applications that focus on historical data sets are HistOSM[1] and OpenHistoricalMap[2]. HistOSM is a web application to visually explore historic objects stored in the OpenStreetMap database, such as monuments, churches and castles. OpenHistoricalMap is an effort to use the OSM infrastructure as a foundation for creating a universal, detailed, and historical map of the world.

The Atlanta Explorer project creates historical geodatabases, geocoders and 3D models of Atlanta city for post Civil War to 1940 [Page et al. 2013]. ATLMaps[3] web portal allows users to explore historical maps, Atlanta Explorer geodatabases, and students generated content of Atlanta city about different subjects, such as historical events, sites and land use. The project members argue that it presents a broad potential for using crowdsourced information about particular sites and structures. But for now, the project portal does not allow citizen-derived geographical information.

The New York Public Library promotes a crowdsourcing project to create polygonal representation of building footprints and attributes from insurance atlases from 1853 to 1930 of New York city. This project provides a web-based application called Building Inspector[4] that allows citizens to extract, correct and analyze data from historical maps. The volunteered information is used in training computers to recognize building shapes and other data on digitized insurance atlases. Budig et al. [Budig et al. 2016] propose a consensus polygon algorithm to extract a single polygon to represent each building from all polygons provided voluntarily by citizens in this project.

The Digital Harlem website[5] is based on legal records, newspapers, archives and published sources, to provide information on everyday life in New York City's Harlem neighborhood in the years 1915-1930. The website enables looking for events and places and creating interactive web maps. The Digital Harlem historical database was created by the project members, without using crowdsourcing and VGI concepts.

The British Library has a vast collection of maps. It developed a project to employ crowdsourcing to georeference most of the old maps, using an online georeferencer tool[6]. Such tools enable overlaying historic maps with modern ones from which one may compare the past with the present and georeference these historical maps. The Library originally turned to crowdsourcing in 2011 and since then five releases of maps have been made public, with extremely successful results. Participants georeferenced 8,000 maps and after undergoing a check for accuracy they were duly approved. They also developed

[1]http://histosm.org
[2]http://www.openhistoricalmap.org/
[3]https://atlmaps.org/
[4]http://buildinginspector.nypl.org/
[5]http://digitalharlem.org
[6]http://www.georeferencer.com

a portal Old Maps Online[7] with a geographic search interface to identify and view historic maps from a variety of available collections.

Perret et al. [Perret et al. 2015] describe a project that creates roads and streets of France from the 18th century by digitization of historical maps using collaborative methodology. However, it is not clear in the paper whether the collaboration is from trained operators or from the general public and specialist in history. Cura et al. [Cura et al. 2017] present another project from France that deals with collaborative geocoding in History. The authors propose a solution that is open source, open data and extensible for geocoding based on the building of gazetteers that have geohistorical objects collected from historical topographical maps. The case study was Paris in the 19th-20th centuries. The results can be visualized over modern or historical maps and even verified and/or edited in a collaborative manner. They store several instances of the same space at different moments in history that can be pictured as a snapshot of a given instant. The system enables collaborative editing, but the user profiles of those who collaborate and post content into the system are not clear.

ImagineRio[8] is another initiative from an American University that provides a platform to understand the evolution, both social and urban, of Rio de Janeiro, Brazil, looking into the entire history of the city. Several views from the perspective of artists, historical maps and architectural plans, in space and time have been organized. It is an open-access digital library. It is possible to relate elements within a web environment in which a streaming of data (vector, spatial and raster) is conducted. Such data can also be inspected, toggled, visualized and naturally queried. It is quite valuable for architects, urbanists, and scholars to consult or view some particular spatiotemporal aspects of the history of the city. An interesting aspect of this project is the availability of a mobile app to enable interested parties and tourists to explore the city.

Pauliceia 2.0 project has many similarities with these projects and has been influenced by most of them, following a strong trend towards urban history and its relationship with space. Several projects described in this section use crowdsourcing and VGI concepts to vectorize specific features from historical maps, such as building footprints and streets, to georeference historical maps as well as to geocode historical places. The main difference between the similar projects and Pauliceia 2.0 is that we are proposing a computational platform that allows historians to share historical geographic data sets resulting from their researches on São Paulo city. Using crowdsourcing and VGI concepts, the Pauliceia 2.0 platform will allow citizens to vectorize streets and buildings from historical maps as well as researchers to share their historical data sets, providing a proper environment for collaborative work. Pauliceia 2.0 is a platform for digital humanities that adheres to the field's main features of free knowledge sharing and collaborative work.

3. Platform architecture

The Pauliceia 2.0 platform is open source, web-based and service-oriented. Its architecture is shown in Figure 1. It is being implemented using the GIS library TerraLib and the web geoportal framework TerraBrasilis developed by INPE [Câmara et al. 2008]. Service-oriented architectures are suitable for data and functionality exchanging across

[7]http://www.oldmapsonline.org/

[8]http://hrc.rice.edu/imagineRio/home

systems, promoting a better integration and interoperability among technologies. The Pauliceia 2.0 spatiotemporal vector data is stored in a PostGIS database system and raster data in Geotiff files.

The architecture has two groups of web services. The first group is composed of geographical web services defined by the Open Geospatial Consortium (OGC): Web Map Service (WMS) for map images, Web Feature Service (WFS) for vector data, Web Coverage Service (WCS) for coverage data, and Catalogue Service Web (CSW) for metadata of spatiotemporal data, services and related objects [Open Geospatial Consortium 2017]. OGC has played a crucial role in geospatial data interoperability by proposing web services standards for visualizing, disseminating and processing geospatial data.

The dissemination of the Pauliceia database through OGC web services is important for interoperability, integration with other applications and data sharing. The Brazilian National Infrastructure for Spatial Data (INDE) specification is based on OGC web services. The purpose of INDE is to catalog, integrate and accommodate the existing geospatial data produced and maintained by agencies of the Brazilian Government so that they are easily located, explored and accessed for a wide variety of uses through the internet. The Pauliceia 2.0 historical data sets will be disseminated according to INDE specification.

The second group is composed of three web services that are being designed and implemented to augment the functionalities of the OGC standard services, attending to the specific demands of the Pauliceia 2.0 project. One of the services is aimed to handle volunteered historic geographical information (VGI). The second one deals with spatiotemporal geocoding and the third one with new means to visualize spatiotemporal data.

3.1. Web service for historical VGI

In the past decade, VGI has become an interesting area of research due to its challenges and advantages. Even though it is possible to reach high standards of data quality with VGI projects, comparable to those collected by National Mapping Agencies (NMAs) and Commercial Surveying Companies (CSC) ([Ludwig et al. 2011], [Graser et al. 2014], [Ciepłuch et al. 2010]), the lack of a rigorous protocol is often a major source of errors and an obstacle to the wider dissemination of VGI initiatives [Mooney et al. 2016].

To ensure the collection of quality data and the reuse of VGI for applications beyond the ones originally intended, it is necessary to establish a protocol that balances the need for meticulous data collection strategies and the motivation for contributors to follow its guidelines. Given the importance of creating a VGI protocol, Pauliceia 2.0 project will create its own protocol based on the guidelines proposed by Mooney et. al. [Mooney et al. 2016], that include recommendations and best practices for VGI projects. In this section, some topics stipulated by these guidelines, such as data types, data collection methods, metadata, quality control mechanisms and feedback to the community are detailed in the Pauliceia 2.0 context.

In the Pauliceia 2.0 project, we intend to use VGI and crowdsourcing to vectorize streets and buildings from historical maps as well as to collect and share historical data sets resulting from researches. All these data sets should be restricted to the urbanization period of the São Paulo city from 1870 to 1940, which is the historical scope of the Pauliceia 2.0 project.

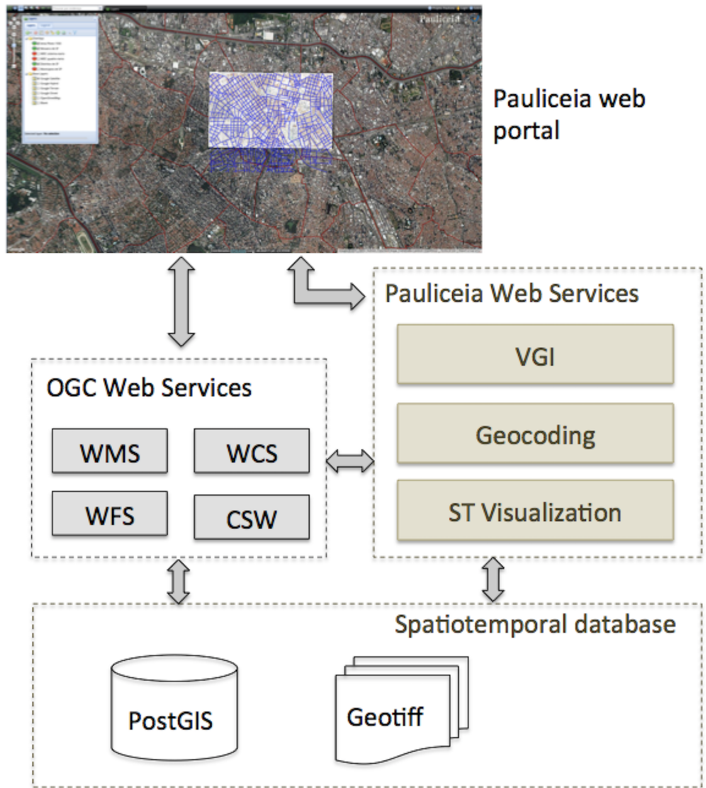

Figure 1. Pauliceia platform architecture

The vectorization of streets and buildings from historical maps will be done manually by volunteers through the Pauliceia 2.0 web portal. This data consists of lines and polygons representing streets and buildings, respectively. In this case, the data set gathered by volunteers will have a set of geometries, polygons or lines, to represent the same object, buildings or streets. To extract the most accurate geometry to represent a single object from this data set, we will employ methods that compute a single geometry that represents the majority opinion, as proposed by Budig et al. [Budig et al. 2016].

Using the Pauliceia 2.0 platform, historians can share historical geographic data resulting from their researches. Such data sets can vary among different themes, for example, crimes that occurred in São Paulo city in 1930 and industrial sites from the period 1870-1930. The platform will accept geographical vector data with different types of attributes, including links to photos, videos and documents that must be stored in other platforms such as YouTube and Dropbox. Besides the historical data sets, historians should validate the metadata extracted automatically by the platform from such data sets as well as inform missing metadata.

To promote and motivate volunteers to vectorize streets and buildings as well as historians to share their historical data sets, we intend to organize events oriented to this purpose. Such events will contain tutorials about the platform and how to contribute, following the same idea of the events called mapathons promoted by Google Maps and OpenStreetMap. These events can be organized in universities with historians and their students to promote the mass contribution of vector data in the Pauliceia 2.0 platform.

To organize the contribution process, the Pauliceia 2.0 platform implements two main concepts, user and project. Everyone can visualize and access the project data sets through the web portal, but only registered users can add contributions and edit them. Users can register in the platform using social logins from Facebook and Google, after accepting the project Use Policy terms. These terms specify that the data provided by the platform is public and the platform is not responsible for any issues that may arise from users providing copyrighted data.

A project is a group of data sets related to a specific theme. Each project has an administration user who created it and a group of users, called collaborators, that are able to edit its data sets. Administration users can add other users as collaborators in the project. If a user wishes to be part of a project as collaborator, it is necessary to make a formal request to the project administration. Users can only edit data sets in projects where they are collaborators. Users are also able to submit reviews and comments about the project data sets through the platform.

In the platform, users can contribute with a single feature or a group of features. For single feature, users will choose a location on the historical map, either clicking on the map or by typing an address, and will inform attributes about this location. For a bulk of features, users must provide the vector data sets following a specific format defined in the Pauliceia 2.0 guidelines. Before inserting into the Pauliceia 2.0 database, such data sets must be approved by the project administrators.

With respect to quality control, users are expected to self-assess the data, checking for coherence, adequate quality and correctness of the attributes, before submitting it to the platform. Once the data is submitted to a project, its collaborators can edit it. Taking into consideration the fact that the target audience of this platform is people with prior knowledge of the field (historians and students), it is expected that they use their own knowledge to point out errors or inconsistencies in the project database. This also works as a mechanism of quality control maintained by the Pauliceia 2.0 community.

A collaborative project progresses as more users participate in it. Therefore, it is important to improve the user experience as a means of encouraging more contributors to join the platform. The user will be encouraged to provide feedback about his experience with the platform, commenting about the positive aspects and what needs improvement, giving opinions, making observations and suggesting changes. This feedback can be provided via mailing lists or social media, and will be used as an important base for improving the platform.

3.2. Geocoding web service for historical data

As described in previous section, a crucial feature of the Pauliceia 2.0 platform is to provide functionalities that allow historians to share historical geographic data resulting from their researches. In this case, most historical data sets have textual addresses to indicate spatial locations in the past. Thus, it is necessary to provide a geocoding algorithm able to transform historical textual addresses into geographical coordinates.

Geocoding is the process of transforming textual data into geographical information. Obtaining coordinates from textual addresses is one of the most important geocoding methods [Martins et al. 2012]. Address geocoding has to deal with challenges related to variations in textual addresses, such as abbreviations and missing parts. Martins et. al [Martins et al. 2012] propose a geocoding method for urban addresses whose output includes a geographic certainty indicator, which informs the expected quality of the results.

In the literature, there are many proposals of efficient geocoders for current addresses, but they do not deal with historical data. A geocoder for historical information must operate on spatiotemporal data sets, that is, spatial entities whose geometries and attributes vary over time. The challenges of creating an address geocoding system for historical data are mainly related to the variation of names, geometries and numerations of streets and buildings over time. In the Pauliceia 2.0 database, every spatial entity, such as a street segment and a place with an address, has an associated period that indicates when it is valid. Thus, the geocoding method for this database has to take into account all valid periods associated to spatial entities.

Cura et al [Cura et al. 2017] argue that historical geocoding requires dedicated approaches and tools due to three reasons. The first is that existing geocoding services do not consider the temporal aspect of the data sets they rely on. They implicitly work on a valid time that is the present. The second reason is that traditional geocoders are based on a complete and strict hierarchy, such as city, street, and house number, which is verifiable. Historical data, however, are full of uncertainties and are not directly verifiable. One has to check possibly incomplete and conflicting available historical sources and, very often, make assumptions or hypotheses. The third is that historical sources available to construct a geocoding database are sparse (both spatially and temporally), heterogeneous, and complex. Based on these reasons, Cura et al [Cura et al. 2017] propose an open source solution for geocoding that is based on gazetteers of geohistorical objects extracted from historical maps.

The geocoding web service that is being designed and implemented in the Pauliceia 2.0 project has to consider all these particularities of historical data sets. Using this service, historians can geocode a single address or as a set of addresses via CSV files. Each address has to contain its street name, number and year. The service computes geographical coordinates associated with the addresses using the historical places and street segments stored in the project database. Besides the geographical coordinates, the service returns a certainty degree associated with each coordinate. This degree indicates how confident a geographical coordinate is, based on the number of available historical entities that were used in the geocoding process.

To populate the project database, we designed and implemented a web portal for historical address edition [9] shown in Figure 2. This portal provides functionalities to insert, delete, edit and search historical addresses. Through this portal, project members are collecting and inserting historical addresses of São Paulo city from 1870 to 1940, such as houses, buildings, churches and squares, into the project database. Each address has a street name, a location number, a period when it is valid and a geographical coordinate that is informed by clicking on the historical map in the portal. The stored addresses can be viewed in an intuitive and simultaneous way by several users registered in the system.

The geocoding service relies on the historical addresses and street segments stored

[9] http://www.pauliceia.dpi.inpe.br/edit

in the project database. So, the construction of a good quality database is crucial to the success of the geocoding process. The greater the number of addresses identified, the greater the accuracy of the resulting base. The project members are using different types of historical sources, such as legislative documents, newspapers, license plate books and advertisement leaflets, to collect these addresses.

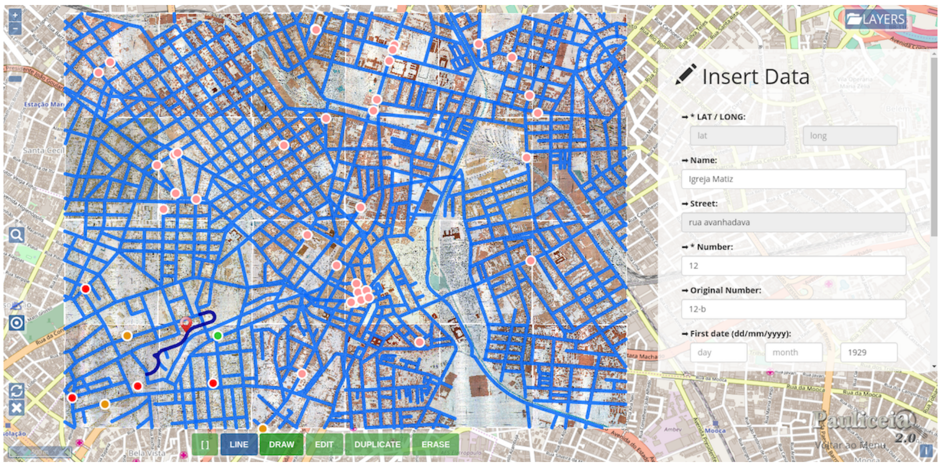

Figure 2. Web portal for historical address edition

3.3. Spatiotemporal visualization

When dealing with spatiotemporal data, visualization is a key aspect. It is an important and essential step in order to provide insights for understanding not only behavior but also to enable decision makers to opt for suitable actions [Mazumdar and Kauppinen 2014]. Visualization enables a clear and effective communication of the presented data.

With respect to Pauliceia 2.0, spatiotemporal visualization is a natural aspect to be envisioned or considered. We intend to create a service that provides techniques for spatiotemporal visualization. Time is an important factor in historical data sets and should be considered as a filter to display the same location over different time periods.

Therefore, it seems more than natural to analyze spatiotemporal data, such as the one being dealt with in Pauliceia 2.0, to understand certain relationships and patterns among the events and locations duly stored. It is important that the developed service provides means to visualize location and time in an integrated manner [Shrestha et al. 2013].

4. Preliminary results and final remarks

This paper presents the architecture of a computational platform that contains crucial modules to build an environment for collaborative historical research. At the moment, the Pauliceia 2.0 database contains the following data sets: (1) two mosaics of georeferenced historical maps of São Paulo city, one of 1930 and another of 1924, stored as geotiff files; (2) streets extracted by project members from such maps of 1930 and 1920 stored in the PostGIS database as vector data; (3) historical addresses, such as churches, houses and buildings, that were collected by project members through the web portal shown in Figure 2 and stored in the PostGIS database as vector data.

Figure 3 shows the historical map and streets (blue line segments) of São Paulo city, both from 1930. To compare the historical map of 1930 with the current map of São Paulo city, Figure 4 presents the streets of 1930 (blue line segments) over a present map from OpenStreetMap. Figures 3 and 4 show the same region of Sao Paulo city that includes the "Mercado Municipal" and "Palácio das Indústrias" (where the "Museu de Ciências Catavento" is currently located). Comparing the two figures, we can observe many differences, including the "Viaduto Diário Popular" that exists today but not in 1930. It was created in 1969.

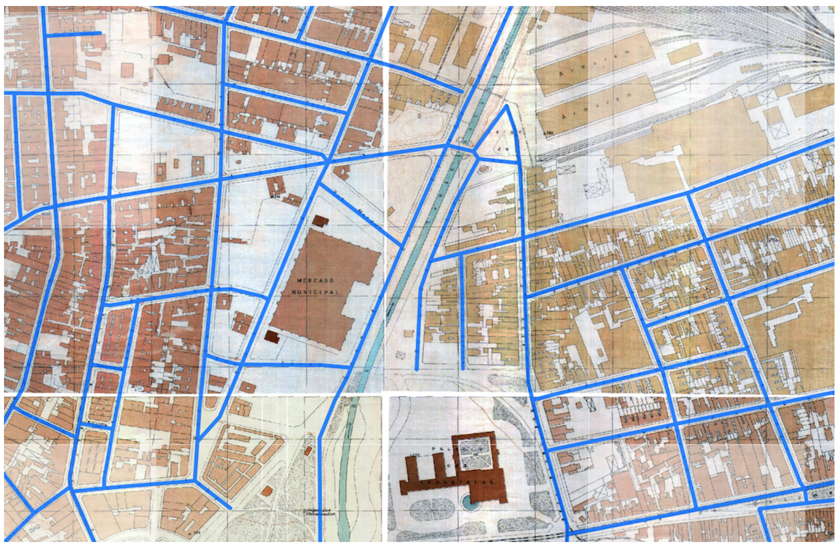

Figure 3. Sao Paulo city - historical map and streets (blue line segments) of 1930

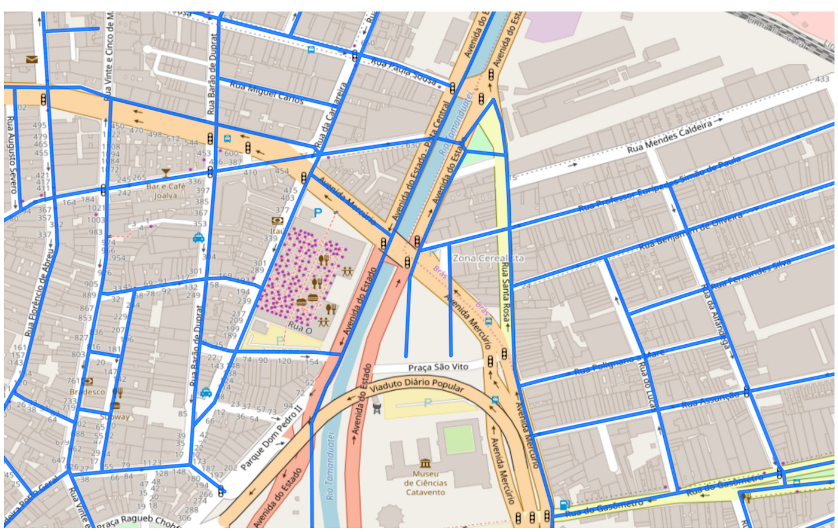

Figure 4. Sao Paulo city - current map from OpenStreetMap and streets (blue line segments) of 1930

Some OGC web services are available for this Pauliceia 2.0 database through the link http://www.pauliceia.dpi.inpe.br/geoserver and all source code developed is available in the github link www.github.com/Pauliceia. The historical data sets of the Pauliceia 2.0 project are mainly created by the following processes:

1. **Digitization, mosaic creation and georeferencing of historical maps.** This process produces raster data as geotiff files.
2. **Vectorization of streets and buildings from historical maps.** This process produces vector data to represent the old streets and buildings, based on the historical maps generated in item (1). At the moment, the project members are vectorizing the streets from such maps manually. After building the platform, we will use VGI and crowdsourcing for the vectorization of streets and buildings from historical maps, as described in Section 3.1. We also intend to evaluate the use of automatic methods in this process.
3. **Gathering and georeferencing of historical addresses.** The project members are collecting and georeferencing historical addresses through the web portal shown in Figure 2. These addresses are crucial for the geocoding processes, as described in Section 3.2. We are still evaluating the use of VGI and crowdsourcing in this process.
4. **Sharing of historical data sets resulting from researches.** After building the entire platform, historians are allowed to geocode and share historical data sets resulting from their researches, as described in Section 3.1. This process is based on VGI and crowdsourcing concepts.

5. Acknowledgment

The funding of the Pauliceia project is provided by the eScience Program of FAPESP (Grant 2016/04846-0). We thank FAPESP for granting students scholarships: #2017/03852-9, #2017/11637-0, #2017/11625-2 and #2017/11674-3.

References

Budig, B., van Dijk, T. C., Feitsch, F., and Arteaga, M. G. (2016). Polygon consensus: smart crowdsourcing for extracting building footprints from historical maps. In *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, page 66. ACM.

Câmara, G., Vinhas, L., Ferreira, K. R., Queiroz, G. R. D., Souza, R. C. M. D., Monteiro, A. M. V., Carvalho, M. T. D., Casanova, M. A., and Freitas, U. M. D. (2008). Terralib: An open source gis library for large-scale environmental and socio-economic applications. *Open source approaches in spatial data handling*, pages 247–270.

Ciepłuch, B., Jacob, R., Mooney, P., and Winstanley, A. C. (2010). Comparison of the accuracy of openstreetmap for ireland with google maps and bing maps. In *Proceedings of the Ninth International Symposium on Spatial Accuracy Assessment in Natural Resuorces and Enviromental Sciences 20-23rd July 2010*, page 337. University of Leicester.

Cura, R., Dumenieu, B., Perret, J., and Gribaudi, M. (2017). Historical collaborative geocoding. *Working paper submitted to Humanities. arXiv preprint arXiv:1703.07138*.

Davis Jr, C. A., de Souza Vellozo, H., and Pinheiro, M. B. (2013). A framework for web and mobile volunteered geographic information applications. In *Proceedings of XIV Brazilian Symposium on Geoinformatics (GeoInfo 2013)*, pages 147–157.

DeBats, D. A. and Gregory, I. N. (2011). Introduction to historical gis and the study of urban history. *Social Science History*, 35(4):455–463.

Estellés-Arolas, E. and González-Ladrón-de Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information science*, 38(2):189–200.

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4):211–221.

Graser, A., Straub, M., and Dragaschnig, M. (2014). Towards an open source analysis toolbox for street network comparison: Indicators, tools and results of a comparison of osm and the official austrian reference graph. *Transactions in GIS*, 18(4):510–526.

Ludwig, I., Voss, A., and Krause-Traudes, M. (2011). A comparison of the street networks of navteq and osm in germany. In *Advancing geoinformation science for a changing world*, pages 65–84. Springer.

Martins, D., Davis Jr, C. A., and Fonseca, F. T. (2012). Geocodificação de endereços urbanos com indicação de qualidade. *Proceedings of XIII Brazilian Symposium on Geoinformatics (GeoInfo 2012)*, pages 36–41.

Mazumdar, S. and Kauppinen, T. (2014). Visualizing and animating large-scale spatiotemporal data with elbar explorer. In *Proceedings of the 2014 International Conference on Posters & Demonstrations Track-Volume 1272*, pages 161–164. CEUR-WS. org.

Mooney, P., Minghini, M., Laakso, M., Antoniou, V., Olteanu-Raimond, A.-M., and Skopeliti, A. (2016). Towards a protocol for the collection of vgi vector data. *ISPRS International Journal of Geo-Information*, 5(11):217.

Open Geospatial Consortium (2017). Ogc® standards and supporting documents. `http://www.opengeospatial.org/standards/`. Accessed on 2017-09-20.

OpenStreetMap (2017). About openstreetmap. `http://wiki.openstreetmap.org/wiki/About_OpenStreetMap`. Accessed on 01/09/2017.

Page, M. C., Durante, K., and Gue, R. (2013). Modeling the history of the city. *Journal of Map & Geography Libraries*, 9(1-2):128–139.

Perret, J., Gribaudi, M., and Barthelemy, M. (2015). Roads and cities of 18th century france. *Scientific Data*, 2(150048):1–13.

See, L., Mooney, P., Foody, G., Bastin, L., Comber, A., Estima, J., Fritz, S., Kerle, N., Jiang, B., Laakso, M., et al. (2016). Crowdsourcing, citizen science or volunteered geographic information? the current state of crowdsourced geographic information. *ISPRS International Journal of Geo-Information*, 5(5):55.

Shrestha, A., Zhu, Y., Miller, B., and Zhao, Y. (2013). Storygraph: Telling stories from spatio-temporal data. In *International Symposium on Visual Computing*, pages 693–702. Springer.

Spiro, L. (2012). "this is why we fight": Defining the values of the digital humanities. *Debates in the digital humanities*, page 16.

Terras, M. (2012). Quantifying digital humanities. *Melissa Terra's Blog*.