Community and Copyright: The Building of an OpenGIS Data Repository in Buffalo, NY

Monica Stephens[1]

[1]Department of Geography, University at Buffalo (SUNY), 105 Wilkeson Quadrangle, Buffalo, NY 14261, United States
Email: mgstephe@buffalo.edu

Abstract

This project examines the challenges and opportunities that arose in the first phase of building an open Geospatial data portal for Buffalo, NY. Community groups, academics, and non-profits articulated a desire to have an exchange of datasets and information that was accessible and stored in one place. However, the City of Buffalo, NY was unable and unwilling to share local data or provide a top-down initiative for collecting, storing, and disseminating geospatial information. Other locales like Detroit and Pittsburgh have relied on non-governmental actors to take the lead establishing data repositories with open data policies that were later adopted by the public sector. Through a Civic Engagement and Public Partnership grant with the United Way of Buffalo and Erie County, we built a prototype of an open GIS-data portal.

1. Introduction

Open data are often thought of as a foundation on which to develop new goods, services, and applications that could significantly improve quality of life and community development. Open data are freely available to everyone to use without the barriers conventionally imposed by copyrights, patents, or other forms of information control. Open data provides an opportunity for communities to improve transparency and democratic control of government, augment the efficiency and effectiveness of public services, and generate new knowledge and insight into processes of social, political, economic and environmental change.

Although open data systems have been strongly championed by city governments in places like New York City, Chicago, and San Francisco, other locales like Detroit (http://portal.datadrivendetroit.org) and Pittsburgh (http://www.wprdc.org/background) have relied on non-governmental actors to take the lead in establishing data repositories with open data policies subsequently following from the public sector. In Buffalo, New York, where local government has not adopted public policy related to open data, we partnered with community organizations to build a repository to collect and disseminate geospatial information.

1.1 Open Data in Buffalo

In Buffalo, NY local government has resisted sharing or disseminating datasets collected by the city or county. Community groups and non-profits provided motivation to share data by re-creating updated versions of community data. For example, the 'ReTree the District' project used smartphones and volunteers to crowdsource a house-by-house tree census of one neighbourhood in the city (Krolikowski 2015). The dataset they created was of higher quality and integrity than the data held by the city forester and was later adopted by the city.

Another community group, frustrated that the city budget was released only as single-page, non-machine readable PDFs (City of Buffalo 2015), scraped the budget website and posted on GitHub a version accessible for analysis, the city later submitted the budget to OpenBooks. The United Way of Buffalo and Erie County (UWBEC) recognized the need for data and technical applications available to the non-profit community, so in 2014 they

partnered with AT&T and organized a civic hackathon. This hackathon offered prize money to the applications that best addressed a number of municipal issues (housing blight, transportation, etc…) in Western New York. An initial goal in designing an open GIS repository was to disseminate this data as well as other data collected by community groups to the public.

1.2 Open Data and Community Control

Rhetoric surrounding open data implies that government and data empowered citizen can work together to usher in a new phase of information democracy. However, creating maps or geospatial analysis of local data is a process that has historically been imbued with power (Crampton 2005). Johnson (2014) criticized open data movements for marginalizing groups that are unable to contribute to the data by excluding them from resulting data sets.

Cities such as Buffalo have options to harness the power of motivated citizens through open data. Sieber and Johnson (2015) examined government open data systems and recommend using open data as a conduit for a larger "agenda of citizen inclusion and participation in decision-making." Johnson (2016) described different ways citizens can provide feedback and update governmental datasets at different levels of risk to the government. The city of Buffalo does not share geospatial data which precludes citizens from conducting independent analysis or participating in urban decision making.

2. Methods

To provide a service to the community, we decided to build a community data platform that would allow non-profits, academics, and community groups to post and access geospatial data without a human gatekeeper. Many open data systems (e.g. NYC Opendata, Detroit Open Data) rely on Socrata (https://www.socrata.com) which focuses on hosting open data, but not organizing or retrieving geospatial information. Additionally Socrata charges municipalities approximately $5,000/month for data hosting services. These limitations created the need to build a new platform that was geospatial and could build linkages within the data.

Our community data platform was implemented as a web application using C#, .Net, with MySQL for a database management system. Data is stored in a cloud storage system (cloudatcost), which also provides API access. Currently, the data can only be uploaded as .csv, .json, .xlsx (although we plan to expand this to shapefiles in the future). To add new data to the repository, any user can visit the portal and upload vector data. Figure 1 shows the current start-up screen that accepts uploaded files.

Figure 1. Data-submission page for community data portal

All aspects of the platform center around the master data repository that uploads, retrieves and presents data. The core concept for the repository is the use of a master ontology. This is a list of data attributes with a well defined name and meaning. Appropriate fields within a source data set are mapped to these master fields allowing users to establish linkages among previously disparate data sets. These linkages can subsequently be used to build interpretations and presentations of the data. Figure 2 identifies the major components and subcomponents of the system architecture.
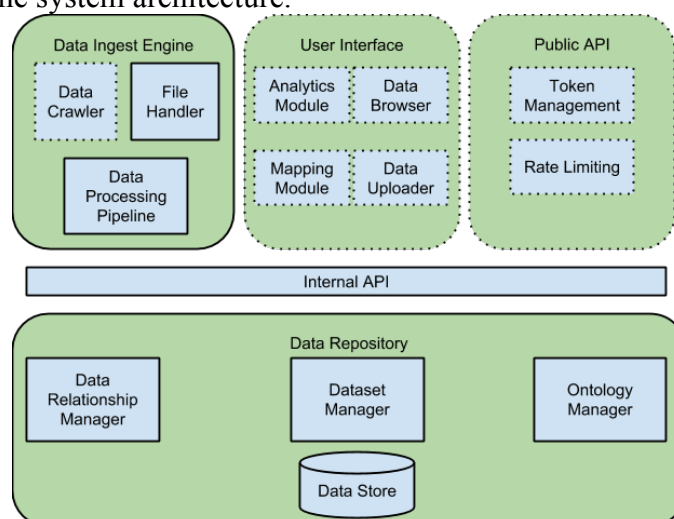

Figure 2. Major components of system architecture (dotted lines are phase 1)

In other words, if the data uploaded has an attribute similar to data that already exists in the data storage (such as a common GeoID or FIPS code from census data) then the new data is added to the existing data table. If the data is of a new type or similar data does not exist, a new data set will be created for that data in the storage. The data file is parsed based on the information provided by the uploader such as number of header rows. Phase two of this project will address the data retrieval and visualization aspects.

3. Results and Conclusion

Phase one of this project has introduced many challenges and some successes. Challenges have included training an interdisciplinary team, bridging the divide between the university and the community, and addressing cloud storage, security, copyright and server issues. Successes included building connections between the business community, non-profits, academics, and motivated community members.

Building an interdisciplinary team between Computer Science and GIS led to many disciplinary debates and conversations over ontology within systems architecture. The University at Buffalo Library was hesitant to host data without knowing the origins and was concerned about potential copyright issues if somebody uploaded copyrighted material. University at Buffalo Information Technology (UBIT) could not host this data in case somebody outside the university uploaded a malicious file. However the UB Library similarly struggles to obtain geospatial information about the local community and has referred patrons towards this geospatial repository. We hope that creating this repository will provide a forum to allow citizen scientists to compose their own analyses and see linkages within the geospatial information that describes their own community.

Acknowledgements

References

City of Buffalo, 2015, Adopted Budget. ed. Ms office. Buffalo, NY.

Crampton JW, Krygier, John, 2005, An introduction to critical cartography. *ACME: an International E-journal for Critical Geographies,* 4(1)**:** 11-33.

Johnson JA, 2014, From open data to information justice. *Ethics and Information Technology,* 16(4)**:** 263-274.

Johnson PA, 2016, Models of direct editing of government spatial data: challenges and constraints to the acceptance of contributed data. *Cartography and Geographic Information Science,* 1-11.

Krolikowski A, Cotton, D, 2015, Putting the sharing economy to work. In *TEDxBuffalo.*

Sieber RE and PA Johnson, 2015, Civic open data at a crossroads: Dominant models and current challenges. *Government Information Quarterly,* 32(3)**:** 308-315.