Developing an Open Source WebGIS Framework for Census Data Visualization

Meng Li[1], Xiaojing Lu[1], Weiping Zeng[2], Jason Disano[2], Joe Garcea[3], Lori Wilkinson[4], Stan Yu[5], Dazawray Landrie-Parker[5], Scott Bell[1]

[1]Department of Geography and Planning, [2]Social Sciences Research Laboratories (SSRL), ssrl@usask.ca,
[3]Department of Political Studies, [5]International Centre for Northern Governance and Development,
University of Saskatchewan, Canada
[4]Department of Sociology, University of Manitoba, Canada

Abstract

Census data is extensively used within research. WebGIS offers great potential for census data dissemination and analysis through its rich and flexible spatial visualization options. However, few current WebGIS frameworks are designed to specifically address the unique requirement of census data mapping, mainly suffering from limited visualization functionalities and restricted developing environment. In this paper, we developed an open source WebGIS framework prototype to address the challenges with its rich, dynamic, and interactive visualization and analysis options. The prototype was developed using a typical census dataset and was applied in two projects: Immigration Research West and Northern Saskatchewan Communities. Results show that our open-sourced framework offers design and implementation flexibility, and is robust and effective in performance.

Background and Relevance

Census data are used extensively in academic, government, and industrial research. Data is often disseminated in tabular form, representing different geographic units (e.g., province, census division, census subdivision). However, the inherent geographic nature of census data is not apparent in a spreadsheet because of a lack of spatial visualization. Maps are indispensable for summarizing large amounts of raw data concisely into usable information, encouraging analyses ranging from a simple comparison between geographic units to sophisticated spatial functions like hotspot identification. Since its conception census mapping has been used in diverse research areas, some examples include environmental health, nutrition research, retail location analysis, and political and urban studies (Kohli et al., 1997; Wilson, Denalane, & Martinez, 1997; Benoit & Clarke, 1997; Lupton & Mather, 1997; Wong, 1996). In the early stages of census data visualization, the work was undertaken in a desktop environment (Thoen, 1995; Zhang & Griffith, 1997). More recently, WebGIS has become a more prominent means of visualizing data. WebGIS is a technology used to display and analyze spatial data on the ubiquitous Internet. It offers online users a convenient way to access spatial information without the cost of a challenging desktop GIS environment (Goodchild, 2008). WebGIS is the ideal platform for disseminating census data through spatial visualization and promoting knowledge discovery with little restrictions on time, location and cost, thus improving the decision-making process.

WebGIS has problematic features; the most apparent problem is its performance. Since

a WebGIS typically targets the web browser as its client, computing power on the client-side is significantly limited compared with a desktop GIS environment, resulting in unsatisfactory performance in graphic rendering and other computing-intensive operations. Unpredictable network environment (e.g. mobile cellular network) further complicates the situation. To promote efficiency and robustness of the WebGIS, long-term efforts have been made (ESRI, 2017; MapBox, 2017a; OGC, 2017) regarding various implementation standards as well as specifications since the 1990s. Today the landscape of WebGIS is still evolving with a mixed collection of available frameworks from a simple ad hoc mashup exercise (Hunt, 2013) to a comprehensive framework (ESRI, 2016). The United States Census Bureau's "Fact Finder" (United States Census Bureau, 2017) is a prime example of publishing large volumes of census data online. It is easily observed that mapping is not the core function of the product as minimum visualization options are implemented despite using ESRI's comprehensive framework. To remedy this situation, "Census Data Mapper" (United States Census Bureau, 2015) was released for dedicated web mapping built using ESRI's framework. Its limitations include (1) proprietary and outdated Adobe Flash platform; (2) lacking support for visualizing data with multiple years and multiple scales; (3) limited visualization options are implemented, especially map types. Canada has also released a similar platform to the Census Data Mapper.

Statistics Canada's "data visualization" product is built with ESRI's framework (Statistics Canada, 2016). This tool shows the Canadian government's effort in integrating census data with WebGIS. It includes a mix of simple ad hoc WebGIS applications for a few selected census variables. Other commercial, as well as open-source, WebGIS solutions exist (Carto, 2016; GeoNode Contributors, 2016; Government of Canada, 2014, 2016a, 2016b; MountainMath Software, 2016; Toronto Star, 2013) that can be used for census data visualization and analysis. However, these systems are not designed to systematically address the specific challenges of the census data commonly identified in the census visualization endeavor: (1) effective visualization of dataset with multiple years, scales and variables; (2) rich selection of visualization parameters to interrogate a dataset from various perspectives; (3) efficient rendering of many geographic objects especially at the finest census scale; (4) necessary tools to facilitate data comparison, reporting and other analysis; (5) well-structured references designed for various audiences to facilitate dataset navigation, promote data discovery; (6) cost-effective and transparent framework that prevents vendor lock-ins and promotes flexibility.

This study intends to develop a WebGIS framework with open source tools to effectively and economically disseminate census data through spatial visualization. Results have demonstrated that the proposed framework delivered quality mapping content from the solid back-end infrastructure to diverse front-end browsers. This framework is configurable, flexible, solid, and high performing, and has been applied to two WebGIS applications: the Immigration Research West (IRW) a project which compiles and analyzes immigration data for several Canadian provinces, and the Northern Saskatchewan Communities project, the International Centre for Northern Governance and Development (ICNGD) project which compiles and analyzes census data of several

Aboriginal communities in Saskatchewan.

Methods and Data

Census Data

The Immigration dataset for Western Canada was used to develop the prototype of the proposed WebGIS framework. Subsequently, the Northern Saskatchewan Aboriginal Communities dataset was used to test the WebGIS framework. Both datasets were typical census datasets found in census data mapping. The first census dataset received was in various formats (e.g. spreadsheet, report, website), at various census scales. Data across scales were later reconciled mainly through data aggregation from census subdivision level to other census scales including census metropolitan, census division, provincial and national levels (Figure 1). Most census data was from 2012 to 2014, while some (e.g. age and gender) had a longer time span from 2001 to 2014. The Northern Saskatchewan Aboriginal dataset lacked explicit scales. The heterogeneous data sources and data structure requires effective data organization and database design, especially considering the multispatial and multi-temporal challenges in the framework design.

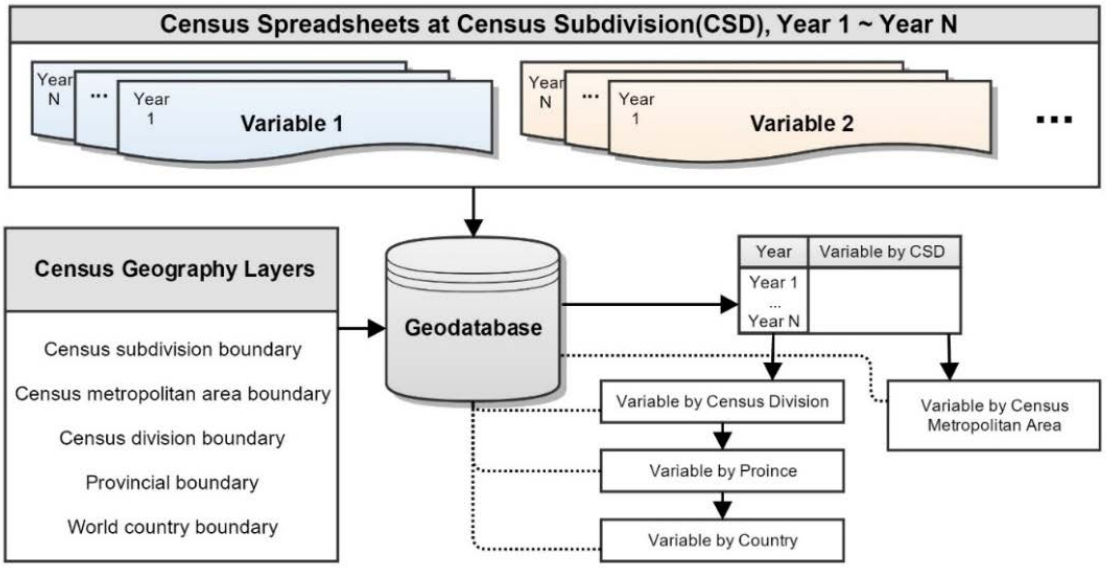

Figure 1. Dataset used to develop the WebGIS prototype.

We deployed a spatial-enabled geodatabase to store spatial layers and census tables separately. The spatial layers were census geography boundaries acquired from Statistics Canada at various census levels. Since the census boundary changed over time, we managed to download spatial layers by year and join to the corresponding census tables using a common ID. Census tables in a spreadsheet format were restructured into a simple and consistent schema (geography-pivoted with multiyear records) to facilitate efficient table management. This not only makes future projects easier to update data but also makes implementing the database maintenance module (e.g., troubleshooting functionalities) possible.

Framework design

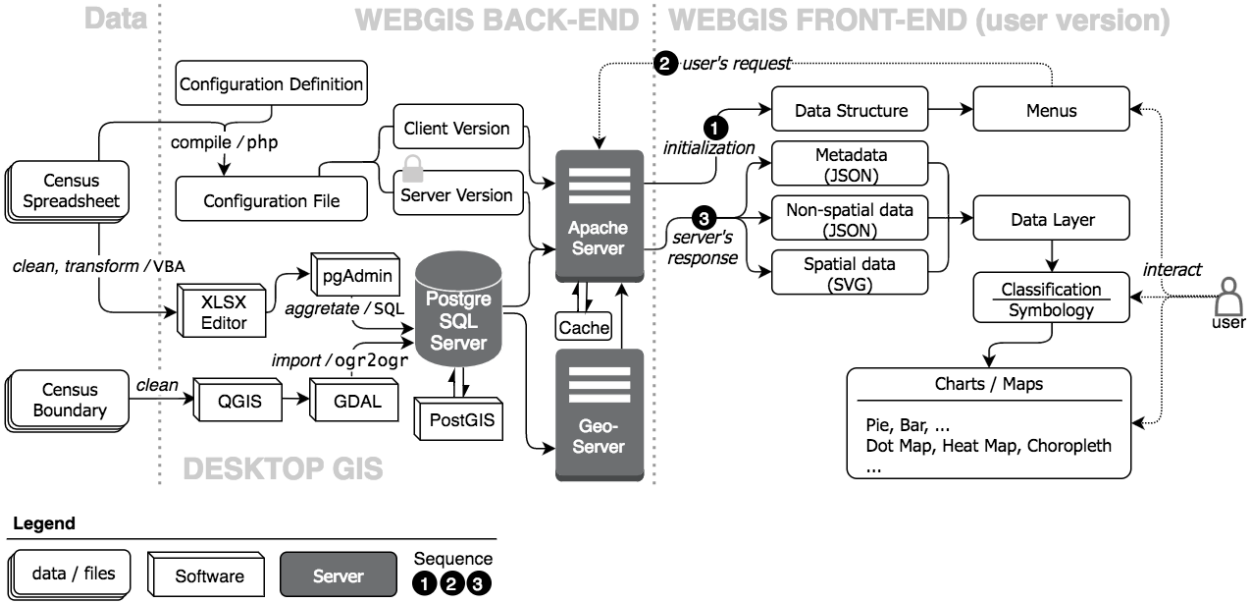

Figure 2. Overall design of the WebGIS framework.

Figure 2 shows the WebGIS framework. Three components make up the framework and cover the project's entire lifecycle from source and input (Data) to processing and modeling (Desktop GIS and WebGIS backend), to client-side presentation (WebGIS frontend). In this project, we focus more on the design of the WebGIS component. The WebGIS consists of a back-end for publishing GIS/non-spatial data and server-side analysis, as well as a front-end for consuming and presenting published services in a user-friendly graphic user interface (GUI) as shown below:

WebGIS back-end: For a WebGIS, a back-end GIS server is used for publishing GIS data and functionalities as web services consumed by the WebGIS's front-end web application. After we had prepared the database, additional WebGIS-related metadata was registered in the configuration file. "Configuration" in Figure 2 defines all WebGIS-related data structure and presentation rules. Specifically, configuration for each data layer included (1) table schema used to administer the published layer and to apply table operations requested by the front-end WebGIS, (2) the logical data structure for the front-end WebGIS to better portray the inherently nested structure of census variables, and (3) predefined permissible rendering options and specific level styling rules for the front-end presentation. These were compiled at runtime producing two configuration files required by the WebGIS –a server version with complete metadata (e.g., database design) used by the back-end WebGIS (protected by the authorization module) and a client version with limited data that were only necessary for front-end WebGIS consumption; all sensitive content was filtered out in the client version. The configuration files were designed using the JavaScript Object Notation (JSON) format to benefit from its conciseness and readability.

Although we see large amounts of WebGIS applications use ESRI's ArcGIS Server as the back-end, there are capable open-source GIS server alternatives (such as Open Source Geospatial Foundation, 2014, 2016; Carto, 2016; GeoNode Contributors, 2016), as well as open-source GIS libraries (Open Source Geospatial Foundation, 2017) that can equip a regular web server with GIS functionalities. This project used GeoServer to provide WMS services for census boundaries. The basemap layers were provided by OpenStreetMap, Google, Mapbox, and ESRI.

WebGIS front-end: The front-end consists of a graphical user interface (GUI) that allows users to visualize and interact with census maps and charts. The front-end was built with the open-source mapping application program interfaces (APIs) Leaflet (Agafonkin, 2015), heatmap.js (Patrik, 2016), and other JavaScript framework libraries (Bootstrap, jQuery). The WebGIS was specifically designed to overcome other frameworks' limitation on census data visualization and analysis. It is more focused on providing the end user with as many visualization options as possible by generating rich, dynamic and interactive charts on the map canvas at the front-end, which is critical for providing information needed to support the discussions and decisions of Immigration, Refugees, and Citizenship Canada (IRCC) officials and related stakeholders.

Figure 2 demonstrations the activities: the user requesting a data layer, the server responding with necessary components, and finally the selected layer being shown as a user-defined chart or map. When the WebGIS front-end is initialized, a configuration file will be created and the project data structure based on the configuration will be created and presented as menus for the user interaction. When the user requests a layer through their interaction with the GUI, the server will return the target layer's necessary components (spatial, non-spatial, metadata). From there it will then be assembled (technically, manipulation on the document object model) and rendered properly at the front-end in the user's browser using HTML5, JavaScript, and CSS.

Results

Immigration Research West project

The IRW project (available at http://gistest.usask.ca/irw) profiles recent immigration data at various census scales in four western provinces (British Colombia, Alberta, Saskatchewan, and Manitoba) for the development of resettlement and integration policies and programs. The immigration data in census subdivision units were provided by IRCC and some provincial government agencies through collaboration with professors Lori Wilkinson at the University of Manitoba and Joe Garcea at the University of Saskatchewan. The user interface is shown in Figure 3. Currently, the mapping system has 51 census variables from 8 categories, sourced by 123 tables in the database, and can produce 2,649 online maps. Some highlights of features within the online mapping system include:

- It is aware of census data's temporal and scale dimension (legend control in Figure 4). For census variable with more than 2 years of data, playback animation is supported. This facilitates changes detection. Moreover, the WebGIS has a scale translator to show data variables at the scale suitable for the current map view. This

allows the system to implement data generalization strategy which ensures high performance while retaining desirable data aggregation levels.

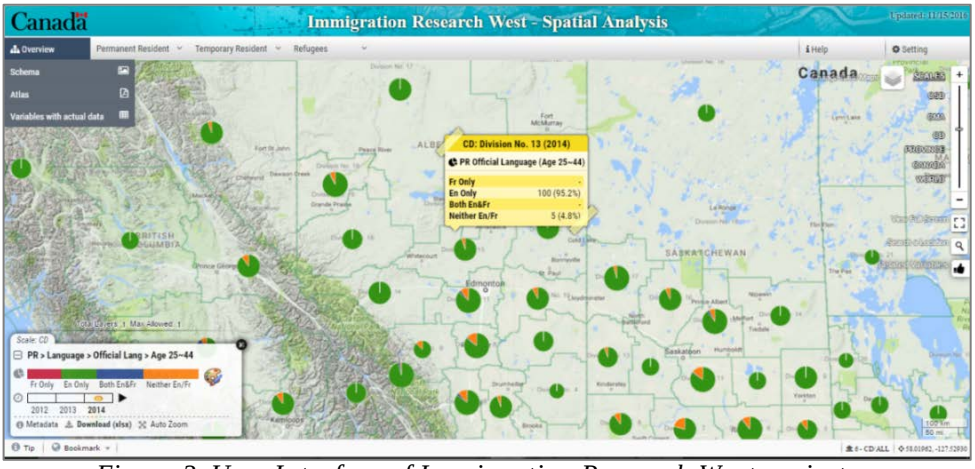

Figure 3. User Interface of Immigration Research West project.

- It supports a mixed overlay of 5 map types (dot map, heat map, bar chart, pie chart, and choropleth) to meet different needs of data interpretation. The open source framework has a programming interface to support new customized map types. Bar chart can effectively portray the long-term trend of a desired variable (Figure 4 bottom left). Whereas the pie chart is better at visualizing proportion of the interested group within any chosen variable (Figure 4 bottom right). The legend control changes accordingly for each map type, and allows rendering dot map and heat map for each year to facilitate comparison and find hotspots.

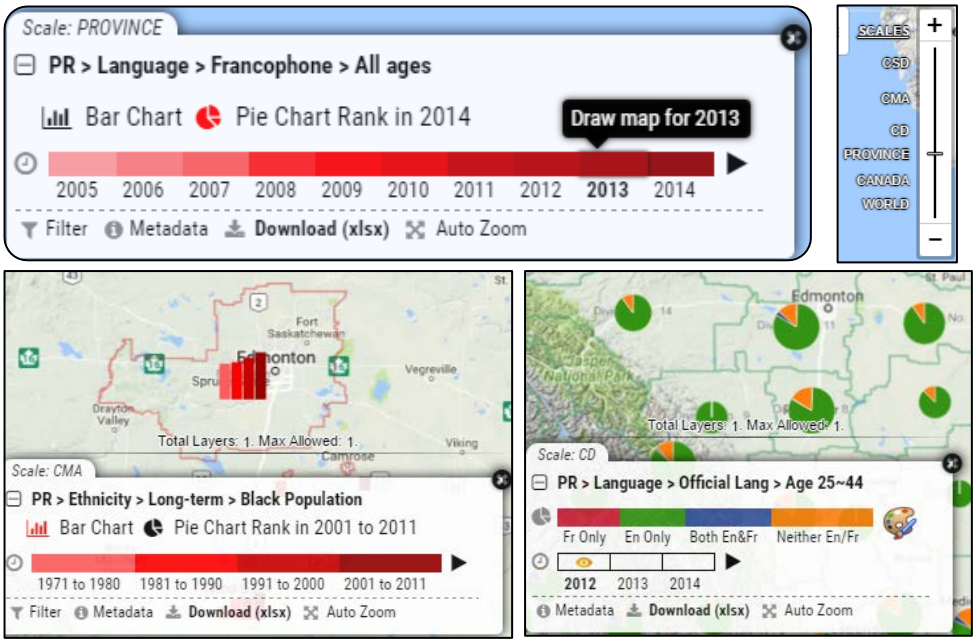

Figure 4. Legend control (top left), scale control (top right), bar chart example (bottom left) and pie chart example (bottom right).

- One challenge in publishing comprehensive census datasets is the accessibility for end users to find desired variables. This will become more obvious when the census database grows over time. In case of IRW, the following measures are taken to address this issue. A virtualized menu layer is constructed on top of the raw data variables to better portray the data's logical structure. This allows end users to access datasets through hierarchical menus. In addition, a data catalog infographic (Figure 5) was designed as a high-level overview of the census database. Each census variable in the infographic is hyperlinked to its corresponding thematic map.

Figure 5. Project schema for quick mapping navigation.
Full resolution image is available at http://gis.usask.ca/irw/schema.html.

- IRW is a thin-client WebGIS with more focus on spatial visualization than analytical of data. Nevertheless, relevant basic analysis tools were implemented, including filtering data dynamically using spatial and non-spatial expressions, as well as downloading the original census data table for power users to perform analysis in their desktop environment (Figure 6). More sophisticated analysis functions are being implemented using GeoServer.

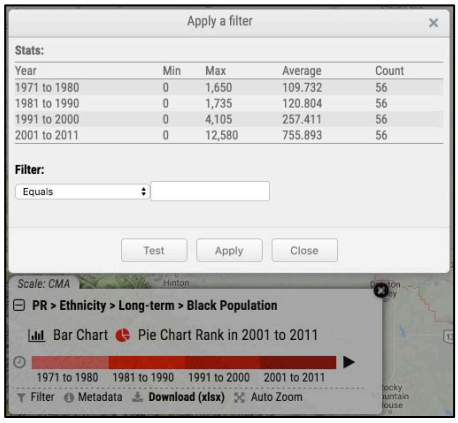

Figure 6. Applying filter to select black permanent residents based on its population.

Northern Saskatchewan Communities project

The Northern Saskatchewan Communities project (user interface as Figure 7, website available at http://gistest.usask.ca/north3/) consists of census data for 60 selected Aboriginal communities. The data was provided by the International Centre for Northern Governance and Development (ICNGD) at the University of Saskatchewan. Instead of menus as used in the Immigration Research West project, this project uses a tree-like structure (Figure 7) to present its entire data catalog (Figure 8). Except for the difference in its front-end GUI, the Northern Saskatchewan Communities project has the same underlying framework used in the IRW project. Besides all the features mentioned in IRW project, a new presentation module is added to facilitate census data comparison. This "Card Deck" module functions like a bookmark manager by allowing end users to collect interested census chart elements for later comparative studies across multiple years, locations or variables (Figure 9). This module is especially friendly for the touch-enabled mobile devices. Furthermore, it helps end users to highlight interesting information on the map.

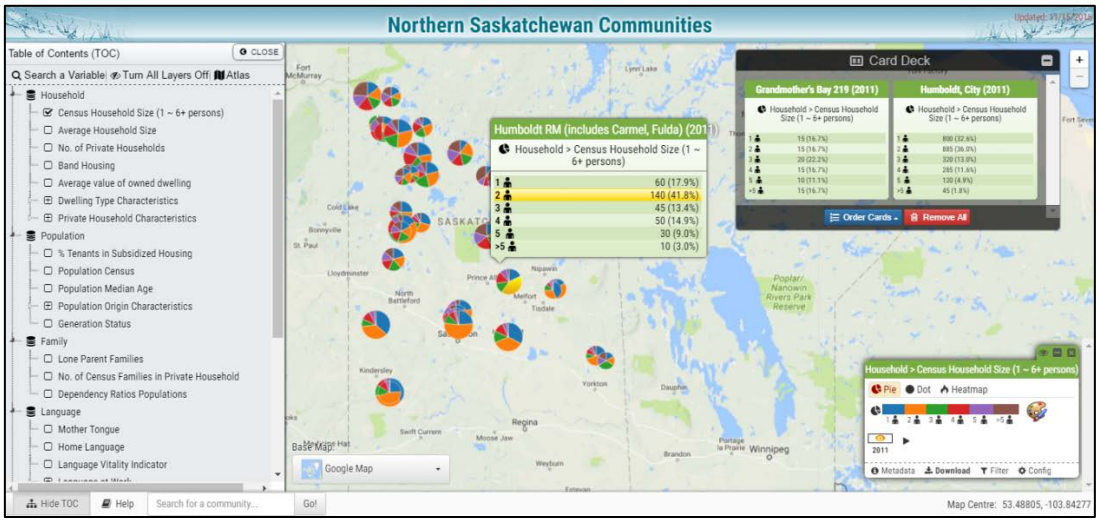

Figure 7. User interface of Northern Saskatchewan Communities project.

Category	Sub-category/Table (Key Properties)	
1 household	Average Household Size	
	Number of Private Households	
	Band Housing	
	Dwelling Type Characteristics	Dwelling Counts by property types
		Occupied Dwelling Tenure
	Household Value	Average Value of Owned Dwelling
	Private Household Characteristics	Dwellings by the period of construction (total)
		Dwellings by the period of construction (2006-2011)
		Number of dwellings by condition
		Major repairs needed
		Number of dwellings by housing suitability
		Not suitable percentage
	Census Household Size (1, 2, 3, 4, 5, 6+ persons)	
2 Population	Percentage of Tenants in Subsidized Housing	
	Population Census Data	
	Median Age of Population	
	Origin Characteristics	Aboriginal/Non-Aboriginal Population
		First Nation/Metis
		Registered/Treaty Indian Status
	Generation Status in Canada Population 15+ 1st, 2nd, 3rd Generation	
3 Family	Number of Lone Parent Families	Total
		By gender
	Number of Census Families in Private Household	
	Dependency Ratio Data (0 – 65+)	
4 Language	Mother Tongue	
	Home Language	
	Language Vitality Indicator	
	Language at Work	
5 Workforce	Census Labour Statistics (Employed/Unemployed)	
	Census Labour Statistics in 2010 (full time/part time/Unemployed)	
	Labour Force by 12 Occupation Sectors	
	Census Employment	By 12 Sectors (initial letters A~F)
		By 13 Sectors (initial letters M~P)
6 Education	Education levels (6 levels with total number)	
7 Income	Census Income Data 2010 (by Market, Gov't Transfer, Other Money)	

Figure 8. Project schema for effective mapping navigation.

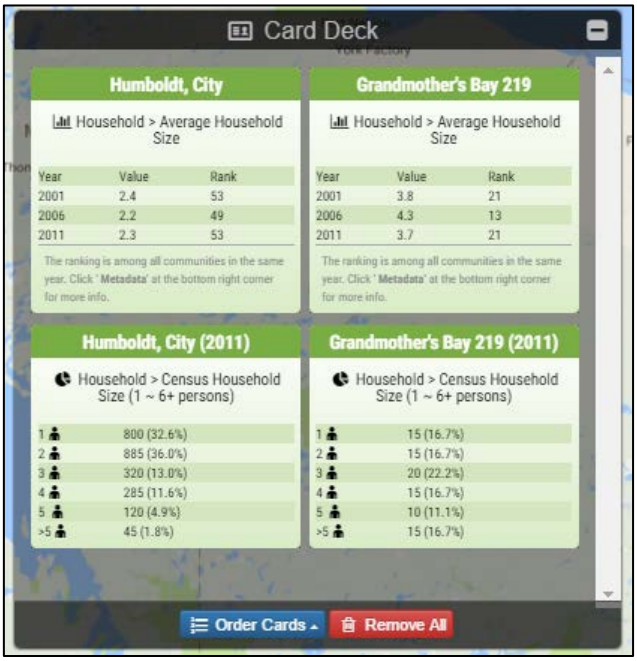

Figure 9. Card deck for collecting interested data for further investigation.

Conclusion and Future Work

WebGIS is an effective tool for census data dissemination through its rich visualization options and analytical functions. However, few WebGIS frameworks can be tailored to the unique requirement for census data visualization and analysis. This paper proposes a WebGIS framework designed to overcome the common limitations of available WebGIS alternatives with its robust, integrated, open, and extensible architecture. The proposed WebGIS framework renders spreadsheet data into rich, dynamic, and interactive charts and maps that are optimized for census data visualization. The prototype is open-sourced, portable, flexible to design and implement, and robust in performance. Application of the framework on two census datasets demonstrated improved performance from this specific design and implementation. The initial release has made an impact on several refugee-related conferences and workshops (e.g. Social Sciences Research Laboratories, 2016a,b). Stakeholders have shown a keen interest in the online mapping system and have given feedback for an upcoming phase of this mapping project.

The current prototype can be further improved from a few perspectives. Changes in census boundaries across different census years can be reconciled as described by Walford (2013). Vector tiles can be used to improve the WebGIS' performance with binary data compression, producing high resolution and custom rendering with considerably smaller file sizes (MapBox, 2017b). Sophisticated GIS analysis such as spatial queries and analyses can be implemented by taking advantage of GeoServer's spatial processing and analysis functions (Boundless Spatial Inc., 2016). Furthermore,

mapping data at a census scale that overlaps other scales requires better design. The impending next stage of the project will try to address the aspects mentioned above.

Acknowledgements

The authors are thankful to IRCC for access to the Landed Immigrant Data File information, IRCC-western region, for funding the project and to the several NGOs and provincial and federal government staff who gave feedback on the mapping project. The authors are also thankful to ICNGD and SSRL for funding this project.

References

Agafonkin, V. (2015). Leaflet. Retrieved November 1, 2016, from http://leafletjs.com/

Benoit, D., & Clarke, G. P. (1997). Assessing GIS for retail location planning. *Journal of Retailing and Consumer Services*, *4*(4), 239–258.

Boundless Spatial Inc. (2016). Introduction to Spatial Processing. Retrieved January 4, 2017, from http://suite.opengeo.org/docs/latest/processing/intro.html

Carto. (2016). CartoDB. Retrieved November 1, 2016, from https://github.com/CartoDB/cartodb

ESRI. (2016). ArcGIS API for JavaScript. Retrieved January 4, 2017, from https://developers.arcgis.com/javascript/

ESRI. (2017). Open Standards. Retrieved January 4, 2017, from http://www.esri.com/software/open/open-standards

GeoNode Contributors. (2016). GeoNode. Retrieved November 1, 2016, from http://geonode.org/

Goodchild, M. F. (2008). What does Google Earth mean for the social sciences? In M. Dodge, M. McDerby, & M. Turner (Eds.), *Geographic Visualization: Concepts, Tools and Applications* (pp. 11–23). Oxford: John Wiley & Sons Ltd.

Government of Canada. (2014). GeoCanViz. Retrieved January 6, 2017, from https://github.com/GeoCanViz/GeoCanViz

Government of Canada. (2016a). Map of destination communities and service provider organizations.

Government of Canada. (2016b). Web Experience Toolkit. Retrieved November 1, 2016, from http://wet-boew.github.io/wet-boew/index-en.html

Hunt, P. (2013). Census mapping mashup. *Cartographic Perspectives*, *76*(76), 67–75.

Kohli, S., Sahlen, K., Lofman, O., Sivertun, A., Foldevi, M., Trell, E., & Wigertz, O. (1997). Individuals living in areas with high background radon: A GIS method to identify populations at risk. *Computer Methods and Programs in Biomedicine*, *53*(2), 105–112.

Lupton, M., & Mather, C. (1997). "The anti-politics machine": GIS and the reconstruction of the Johannesburg local state. *Political Geography*, *16*(7), 565–580.

MapBox. (2017a). Open Source > Standards. Retrieved January 4, 2017, from https://www.mapbox.com/about/open/

MapBox. (2017b). Vector Tiles. Retrieved January 4, 2017, from https://www.mapbox.com/vector-tiles/

MountainMath Software. (2016). Census Mapper. Retrieved January 4, 2017, from https://censusmapper.ca/

OGC. (2017). OGC Standards. Retrieved January 4, 2017, from http://www.opengeospatial.org/docs/is

Open Source Geospatial Foundation. (2014). GeoServer. Retrieved November 1, 2016, from

http://geoserver.org/
Open Source Geospatial Foundation. (2016). MapServer.
Open Source Geospatial Foundation. (2017). GDAL. Retrieved January 4, 2017, from http://www.gdal.org/
Patrik, W. (2016). heatmap.js: Dynamic Heatmaps for the Web. Retrieved November 1, 2016, from https://www.patrick-wied.at/static/heatmapjs/
Social Sciences Research Laboratories. (2016a). Immigration Research West: Spatial Analysis Demonstration. In *Alberta Settlement and Integration Consultation*. Calgary, AB: Alberta Association of Immigrant Serving Agencies. Retrieved from http://aaisa.ca/wp-content/uploads/2016/07/Joe-Garcia.pdf
Social Sciences Research Laboratories. (2016b). Immigration Research West: Spatial Analysis Demonstration. In *CIC & SaskSector Symposium*. Saskatoon, SK.
Statistics Canada. (2016). Data products. Retrieved October 31, 2016, from http://www12.statcan.gc.ca/census-recensement/2011/dp-pd/index-eng.cfm
Thoen, B. (1995). Internet resources for the geosciences, with an emphasis on GIS and mapping. *Computers and Geosciences*, *21*(6), 779–786.
Toronto Star. (2013). Immigration map explainer. Retrieved October 31, 2016, from https://www.thestar.com/news/2013/05/06/immigration_map_explainer.html
United States Census Bureau. (2015). Census Data Mapper. Retrieved January 4, 2017, from https://www.census.gov/geo/maps-data/maps/datamapper.html
United States Census Bureau. (2017). American Fact Finder. Retrieved January 4, 2017, from https://factfinder.census.gov
Walford, N. (2013). Development and design of a web-based interface to address geographical incompatibility in spatial units. *Environment and Planning A*, *45*(7), 1713–1733.
Wilson, J. F., Denalane, A. N., & Martinez, M. (1997). Application of Geographic Information Systems (GIS) for Nutrition Programs. *Journal of the American Dietetic Association*, *97*(9), A110.
Wong, D. (1996). Enhancing segregation studies using GIS. *Computers, Environment and Urban Systems*, *20*(2), 99–109.
Zhang, A., & Griffith, G. (1997). Developing user-friendly spatial staitical analysis modules for GIS: An example using ArcView. *Computers, Environment, and Urban Systems*, *21*(1), 5–29.