Crowdsourcing with Statistics Canada:
a pilot project to explore new frontiers in data collection, mapping, and open data

Presented by
Alessandro Alasia
alessandro.alasia@canada.ca
Data Exploration and Integration Lab (DEIL)
Centre for Special Business Projects

Institute for Data Science, Carleton University
January 18, 2017
Evolution of data collection methods

- In-person interviews
- Mail-out of paper questionnaires
- Telephone interviews
- Electronic questionnaires
- Can we use open data platforms?
Why this pilot project?

- We know that crowdsourcing exists, is out there, and works
- What is needed is an in-depth understanding of how a national statistical office can mobilize contributors and existing technologies for the purpose of official statistics
Crowdsourced data: a working definition for this project

Information of **statistical value** that is **volunteered** on an **open data platform**, which is **freely accessible, editable, and reusable** by anybody
What data are we crowdsourcing?

- Georeferenced attributes of buildings (e.g., building footprint, address, name of business in the building, etc.)
- Open data platform: OpenStreetMap (OSM)
- Selected municipalities. For the pilot project: Ottawa/Gatineau
- Two years pilot project
www.openstreetmap.org

A free, editable map of the whole world that is being built by volunteers largely from scratch and released with an open-content license.

OSM is open data

Why OpenStreetMap? OSM Wiki
Why the world needs OpenStreetMap The Guardian
Tim Berners-Lee: The next web TED Talk 2009
# Project implementation

<table>
<thead>
<tr>
<th>Date Range</th>
<th>Task Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>May-August, 2016</td>
<td>• Background research /development (communication, outreach, IT)</td>
</tr>
<tr>
<td>September 15, 2016</td>
<td>• Launch of web page; start of communication campaign</td>
</tr>
<tr>
<td>October 17, 2016</td>
<td>• Deployment of customized OSM editor, social media campaign</td>
</tr>
<tr>
<td>October 2016 to February 2017</td>
<td>• Outreach and monitoring</td>
</tr>
<tr>
<td>March 2017</td>
<td>• Preliminary assessment of results; possible revisions of work plan</td>
</tr>
<tr>
<td>Year II</td>
<td>• Experiment with a mobile application</td>
</tr>
</tbody>
</table>
Governance and collaborations

Collaborative agreement: **Open North**, leading open data non-profit

- Statistics Canada
  - Working group (5 divisions)
  - Steering committee

- OSM community
  - Local OSM Meet-Up group
  - Canadian community @Talk-ca
  - OSM Board of Directors
  - OSM Institutions WG

- Educational institutions
- Municipalities
A customized version of an open source OSM editor (iD-Editor) to volunteer data to an open platform

(Link to source code upcoming)
Mapping your community through crowdsourcing

Welcome to crowdsourcing at Statistics Canada—a new pilot project
https://crowdid.osmcanada.com
Preliminary results

It is open data. You can monitor the project!
We have developed an open source dashboard (beta version)

(Link to source code upcoming)
Ottawa

Crowdsourcing - December 2016

29,758
Buildings

184
Users

2.39
Average number of tags

Buildings mapped

Buildings by type

Missing address fields in percentage

Number of tags
Gatineau

41,732 buildings
103 users
5.76 average number of tags

Buildings mapped

Buildings by type:
- residential
- yes
- house
- apartments
- static_caravan
- shed
- commercial
- retail
- garage
- terrace
- school

Missing address fields in percentage:
- Province
- Postal Code
- Street Number
- Street
- City

Number of tags:
- Dec
- Nov
- Oct
- Sep
- Aug
Ottawa - Gatineau

Crowdsourcing - December 2016

71,489 Buildings
247 Users
4.36 Average number of tags

Buildings mapped

Buildings by type
- residential
- house
- apartments
- commercial
- retail
- garage
- school
- terrace
- industrial
- church
- shed
- static_caravan
- detached
- civic
- roof

Missing address fields in percentage
- Province
- Postal Code
- Street Number
- City
- Street

Number of tags
- Dec
- Nov
- Oct
- Sep
- Aug

Statistics Canada • Statistique Canada
It is open data! You can monitor progress using OSM online tools

First 17 days of January 2017

```plaintext
/*
This has been generated by the overpass-turbo wizard.
The original search was:
"building and newer:"17 days"
*/
[out:json][timeout:25];
// gather results
(
  // query part for: "building and newer:"17 days"
  node["building"]{newer:"{$date:17 days}"}
  ([bbox]);
  way["building"]{newer:"{$date:17 days}"}
  ([bbox]);
  relation["building"]{newer:"{$date:17 days}"}
  ([bbox]);
); // print results
out body;
>;
out skel qt;
```
Lessons Learned

it is still too early to make an assessment of data quality issues, but we have learned something!
We are not alone....

- Over 1.7 mln buildings currently mapped on OSM across Canada

Rural: Dauphin (MB)
about 4,400 buildings
about 8,300 people

Urban: Regina (SK)
about 95,000 buildings
about 210,000 people
A new data ecosystem: open data platform

We are a new species in this ecosystem

We need to learn about and adapt to it
Expand collaborations and partnerships

Local(s): communities, organizations, institutions

Global: communities, organizations, institutions

OSM is a global open data platform
....starting by connecting at the local level

- Municipalities (and municipal open data)
- Local OSM groups and expert mappers
- Local for-profit and non-profit organizations
- Interest groups
- Educational institutions (high schools, colleges, universities)
Improve OSM data usability

- OSM data are great, but how easy is it to access, analyze and process these data?
- We are working on improving the processing of OSM data and enriching the tools to analyze them
- (OSM) data literacy is also a challenge
Foster the diversity of motivations to volunteer on an open data platform

- There are a variety of motivations
  - Fun, curiosity
  - Civic engagement
  - Passion for maps and cartography, open data
  - Professional interest or development (learning)

- Ideally: collaborate with a core group of volunteer mappers as well as engaging the general public
Recognizing the potential of open data and open source

- The fundamental motivations:
  - Contributing to a public good (open data)
  - Learning and sharing (open source code)
- Explore the potential of source code development on collaborative platforms
- Explore the potential of light IT infrastructures that could be accessible by any organization at low cost
Next step: mobile version

- Engage users of mobile/portable devices, with the goal of “making data entry easy”
Can we work together?

- Engage communities
  - Community mapping and capacity building
  - High school projects: “map your neighborhood”

- Develop (open source) tools for organizations
  - [New York City and OpenStreetMap Collaborating Through Open Data](https://github.com).
    Code available through [GitHub](https://github.com).
  - Crowd-id Editor
  - Project dashboard and monitoring tool (beta version)

- Use the data and develop data applications
  - Automated applications for analysis, data processing and management
  - Community-based applications, or applications for businesses
Acknowledgments

- Over 200 OSM contributors
- Openstreetmap – Ottawa Meetup group and core group of expert contributors
- Inputs and ideas form many OSM developers including Mapbox
- City of Ottawa and Gatineau
- Open North
- Working Group and Steering Committee @Statcan
Questions?