Data Quality Control:
Crowd-Sourcing Geospatial Information

Xizi Hotel, Hangzhou
24-26 May, 2012

Presented:  Prof. Dr. Woosug Cho, Senior Advisor, NGII, Republic of Korea
Prepared:   Prof. Dr. Ki-Joune Lee, Senior Advisor, NGII, Republic of Korea
            Mr. Jun-Sung Kim, Deputy Director, NGII, Republic of Korea
Definition of Crowd-Sourcing

Crowd-sourcing is a process that involves outsourcing tasks to a distributed group of people. This process can occur both online and offline, and the difference between crowdsourcing and ordinary outsourcing is that a task or problem is outsourced to an undefined public rather than a specific body, such as paid employees.

Shortly, a distributed problem-solving and production model.

The term "crowdsourcing" is a portmanteau of "crowd" and "outsourcing," coined by Jeff Howe in 2006.
Crowd-Sourcing Geospatial Information

VGI (Volunteered Geographic Information): the harnessing of tools to create, assemble, and disseminate geographic data provided voluntarily by individuals (Goodchild, 2007)

Examples on Crowd-Sourcing Geospatial Information:
1. OpenStreetMap (http://www.openstreetmap.org/)
http://www.youtube.com/watch?v=lpXk5W6b298&feature=channel&list=UL
2. Project Haiti: http://vimeo.com/9182869
http://www.ted.com/talks/lalitesh_katragadda_making_maps_to_fight_disaster_build_economies.html
Project Haiti
Google Map Maker: Myanmar Case
Example in UK – Google Map & OSM

http://tools.geofabrik.de/mc
Example in Korea – Google Map & OSM

http://tools.geofabrik.de/mc
PotLatch

JOSM (Java OSM editor)

Data Quality Control of Crowd-Sourcing Geospatial Data
Merkaartor
Vespucci

Data Quality Control of Crowd-Sourcing Geospatial Data
Mapen POI Collector

Data Quality Control of Crowd-Sourcing Geospatial Data
A Comparative Study of Crowd-Sourcing GI

OSM vs. OS datasets
- M. Haklay, Environment and Planning B: Vol.37, 2010

Length comparison between OSM and OS Meridian 2

<table>
<thead>
<tr>
<th>Cells</th>
<th>Area (km²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empty cells</td>
<td>16,300 (13.2)</td>
</tr>
<tr>
<td>Meridian 2 more detailed than OSM</td>
<td>75,977 (61.4)</td>
</tr>
<tr>
<td>OSM more detailed than Meridian 2</td>
<td>31,437 (25.4)</td>
</tr>
<tr>
<td>Total</td>
<td>123,714</td>
</tr>
</tbody>
</table>

Length comparison with attributes between OSM and OS Meridian 2

<table>
<thead>
<tr>
<th>Cells</th>
<th>Area (km²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empty cells</td>
<td>17,632 (14.3)</td>
</tr>
<tr>
<td>Meridian 2 more detailed than OSM</td>
<td>80,041 (64.7)</td>
</tr>
<tr>
<td>OSM more detailed than Meridian 2</td>
<td>26,041 (21.0)</td>
</tr>
<tr>
<td>Total</td>
<td>123,714</td>
</tr>
</tbody>
</table>
A Comparative Study of Crowd-Sourcing GI

OSM vs. Tele Atlas in Germany
- D. Zielstra and A. Zipf, Agile Conf., 2010

Total Length Differences of all Street Network Data between OpenStreetMap and Tele Atlas per Km²
Weakness of Crowd-Sourcing GI

NO Quality Control on GI and NO Metadata

- Data quality depends on each individual volunteer
- No metadata
- Only some lineage data (change-sets) is provided
Data Quality Control in ISO 191xx

- Data Quality (ISO 19157)
- Metadata (ISO 19115)
- Data Product Spec. (ISO 19131)

- Quality Principles (ISO 19113)
- Quality Evaluation (ISO 19114)
- Quality Measures (ISO 19138)
Data Quality Evaluation – ISO

For each QC item,
- scope,
- measure,
- Evaluation method,
Should be defined

Metadata (ISO 19115)

<table>
<thead>
<tr>
<th>Data specified by a data quality scope</th>
<th>Positional accuracy/absolute or external accuracy data quality result</th>
<th>Reported positional accuracy/absolute or external accuracy data quality result</th>
</tr>
</thead>
<tbody>
<tr>
<td>The dataset</td>
<td>1,35</td>
<td>1,35</td>
</tr>
<tr>
<td>Roads only</td>
<td>1,10</td>
<td>1,10</td>
</tr>
<tr>
<td>Streams only</td>
<td>1,35</td>
<td>Not reported</td>
</tr>
<tr>
<td>Railroads only</td>
<td>1,20</td>
<td>1,20</td>
</tr>
<tr>
<td>Pipelines only</td>
<td>1,60</td>
<td>1,80</td>
</tr>
</tbody>
</table>

Data Quality Control of Crowd-Sourcing Geospatial Data
Quality Evaluation Method – ISO 19114

- Quality evaluation methods
  - Quality evaluation method
    - Internal
    - External
  - Indirect evaluation method
- Evaluation from external data such as lineage information
  - Evaluation of geospatial data
  - Comparison with other data set
Data Quality Control for Crowd-Sourcing GI

Impossible to enforce data quality control process to volunteer map makers
  – Possibly by governmental or semi-governmental institutes

ISO Quality Control
  – As defined by ISO 19113, 19114, and 19138
  – Direct Evaluation
    • Comparison with other map data such as national maps
  – Indirect Evaluation, such as
    • Number of participants at given area,
    • Profile of each participant, and
    • Frequency of updates and last update date
Data Quality Control for Crowd-Sourcing GI

(-) May discourage volunteer map production
(-) Poor environment to implement quality control
(-) Limited use of data due to the lack of quality control and metadata
Data Quality Control for Crowd-Sourcing GI: How to break the vicious circle

(-) May discourage volunteer map production
(-) Poor environment to implement quality control

(+): Quality control by means of map production environments

GI Production by Volunteers

(+): Providing data quality information (metadata)

Quality Control

(-): Limited use of data due to the lack of quality control and metadata

Improving VGI environment of quality control

Monitoring quality
Data Quality Control for Crowd-Sourcing GI: Improving Map Production Environment
Data Quality Control for Crowd-Sourcing GI: Monitoring Data Quality

- Monitoring Tools
  - Quality Evaluation From Lineage Data
  - Quality Evaluation by Comparison

- Crowd-Sourcing GI
- Reference Map
- Metadata
Data Quality Control for Crowd-Sourcing GI: Proposed Flow

- Consumers of Crowd-Sourcing GI
- Monitoring Data Quality
- Quality Control from Map Production Tools
- Organization with Authority
- Metadata
- Volunteers of GI production
- Crowd-Sourcing GI
- Metadata

Data Quality Control of Crowd-Sourcing Geospatial Data
Summary

Crowd-Sourcing GI

- An Important Source of GI
- Very helpful for developing countries

Difficult to enforce rigorous quality control

Quality Control of Crowd-Sourcing GI

- Map Production Tool with QC functions
- Monitoring QC of Crowd-Sourcing GI