VGI and formal data

EEO / AGI (Scotland) seminar

David Fairbairn
Newcastle University
School of Civil Engineering & Geosciences
VGI and formal data

- Investigating the possible interaction between informal and formal data
  - Informal Volunteer Geographic Information
  - Formal official spatial data
- Examples of VGI data collection
  - Flickr
  - Panoramio
  - commercial satellite imagery
  - OpenStreetMap
Photographs of the South Side of Edinburgh

137 black-and-white photographs taken in 1929 taken by Alfred Henry Rushbrook on behalf of the City of Edinburgh Improvement Trust. Show tenements and shops around St Leonards before buildings were pulled down for slum clearance. Half-platePROCESS SILVER
Using satellite imagery to update spatial databases
OpenStreetMap is a free editable map of the whole world. It is made by people like you.

OpenStreetMap allows you to view, edit and use geographical data in a collaborative way from anywhere on Earth.

OpenStreetMap's hosting is kindly supported by the UCL VR Centre, Imperial College London.
OpenStreetMap is a free editable map of the whole world. It is made by people like you. OpenStreetMap allows you to view, edit and use geographical data in a collaborative way from anywhere on Earth. OpenStreetMap's hosting is kindly supported by the UCL VR Centre, Imperial College London.
OSM data

Manually select a different area
Hide areas
Display object list

Institue of Geography (New Surgical Hospital Building)

- amenity: university
- building: yes
- name: Institute of Geography (New Surgical Hospital Building)

History for Institute of Geography (New Surgical Hospital Building)

- Edited by jamesks at 2010-09-29T10:51:58Z
- Edited by jamesks at 2010-09-28T14:26:45Z
OSM data
Applications for data interaction

• Photosynth
• SatNav
  • TomTom
• Hazard and disaster mapping
  • Sri Lanka
  • Haiti
• SDI
  • INSPIRE
• ....
Photosynth pointcloud
Sri Lanka 2005

Sri Lanka South East Coast
East of Palatupana

Before and after disaster map

SPOT 2 - 08 February 2002

SPOT 4 - 13 January 2005

Disaster type: Tsunami
Disaster date: 26 December 2004

Data sources:
Reference image: SPOT 2 natural colour (10 m)
08 February 2002
© CNES 2002 : distribution Spot Image
Crisis image: SPOT 4 natural colour (10 m)
13 January 2005
© CNES 2005 : distribution Spot Image

Datum: WGS 84
Projection: UTM 44
Scale: 1/25 000 for A3 prints
Map created 13 January 2005 by SERTIT;
© SERTIT 2005
sертit@sертit.u-strasbg.fr
http://sertit.u-strasbg.fr/

Probably highly affected areas
Possibly affected roads
• 2007/2 EU Directive does not mention citizen input

• BUT can existing structures and short-term finances meet statutory needs?

• Whether / how to use ‘citizen-sourced’ data
Applications for data interaction

- Photosynth
- SatNav
  - TomTom
- Hazard and disaster mapping
  - Sri Lanka
  - Haiti
- SDI
  - INSPIRE

- … more mapping examples later
Mapping anew or combining with existing sources?

- Can VGI initiate new projects?
- VGI contributing to existing projects
  - updating
  - enhancing / extending
  - integrating
- Investigating integration
Interaction => integration?

- Need to standardise?
- Parameters to consider
  - geometry
  - semantics
  - currency
  - completeness
  - collection method
  - nature of collecting organisation
- Role of metadata
Experimental approaches

- Girres and Touya (2010) France – semantics (and other parameters)
- Haklay (2010) UK – completeness
Positional Accuracy of OpenStreetMap Data in Comparison to Tele Atlas in Five Major Cities in Germany

- Berlin
- Hamburg
- Munich
- Cologne
- Frankfurt

Data within 10 m buffer in %
Experimental approaches


- Girres and Touya (2010) France – semantics (and other parameters)

- Haklay (2010) UK – completeness
<table>
<thead>
<tr>
<th>Layer</th>
<th>Geometry</th>
<th>Objects</th>
<th>Field</th>
<th>Informed field</th>
<th>Uninformed field</th>
<th>Ratio %</th>
</tr>
</thead>
<tbody>
<tr>
<td>France_poi</td>
<td>Point</td>
<td>111,440</td>
<td>« NAME »</td>
<td>111,440</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>France_highway</td>
<td>Polyline</td>
<td>886,680</td>
<td>« TYPE »</td>
<td>756,655</td>
<td>130,025</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>« NAME »</td>
<td>382,896</td>
<td>503,784</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>« ONEWAY »</td>
<td>143,274</td>
<td>743,406</td>
<td>16</td>
</tr>
<tr>
<td>France_coastline</td>
<td>Polyline</td>
<td>1,200</td>
<td>« NAME »</td>
<td>47</td>
<td>1,153</td>
<td>4</td>
</tr>
<tr>
<td>France_administrative</td>
<td>Polyline</td>
<td>51,413</td>
<td>« ADMIN_LEVEL »</td>
<td>51,352</td>
<td>61</td>
<td>99.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>« NAME »</td>
<td>838</td>
<td>50,575</td>
<td>2</td>
</tr>
<tr>
<td>France_water</td>
<td>Polygon</td>
<td>12,507</td>
<td>« NATURAL »</td>
<td>12,507</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>« NAME »</td>
<td>2,041</td>
<td>10,466</td>
<td>16</td>
</tr>
<tr>
<td>France_natural</td>
<td>Polygon</td>
<td>24,756</td>
<td>« TYPE »</td>
<td>24,756</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>« NAME »</td>
<td>4,505</td>
<td>24,333</td>
<td>18</td>
</tr>
</tbody>
</table>
Experimental approaches

• Zielstra and Zipf (2010) Germany – accuracy

• Girres and Touya (2010) France – semantics (and other parameters)

• Haklay (2010) UK – completeness
Newcastle experiments on geometric comparability

- Large-scale, not small scale data
- Comparison for data integration purposes: OpenStreetMap / OS
- Concentration on determining factors:
  - type of feature / nature of landscape
  - data collection method
  - expertise of data collector
Study areas

- Cramlington, Northumberland
  - Urban area with hard features

- Clara Vale, Northumberland
  - Rural area with soft features

- Creation of a reference dataset
  - Comparison of ‘official data’ (Ordnance Survey) and ‘VGI data’ (OpenStreetMap)
Cramlington (Google Street View)
Clara Vale (Google Street View)
Geometrical comparison

- National Standard for Spatial Data Accuracy (NSSDA)
  - Tested points $\geq 20$ points
  - A minimum of 20% of the sample points are located in each quadrant
  - Distance between points $\geq \frac{d}{10}$

$$RMSE = \sqrt{\frac{1}{n} \left( \sum_{i=1}^{n} \delta x_i^2 + \sum_{i=1}^{n} \delta y_i^2 \right)}$$
### Geometrical comparison

#### Comparisons of RMSE and NSSDA accuracy of compared datasets

<table>
<thead>
<tr>
<th></th>
<th>RMSE (m)</th>
<th>NSSDA accuracy (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Urban area, UK</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FS/OS</td>
<td>0.492</td>
<td>0.846</td>
</tr>
<tr>
<td>FS/OSM</td>
<td>5.429</td>
<td>9.143</td>
</tr>
<tr>
<td>OS/OSM</td>
<td>5.331</td>
<td>8.989</td>
</tr>
<tr>
<td><strong>Rural area, UK</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FS/OS</td>
<td>1.843</td>
<td>3.190</td>
</tr>
<tr>
<td>FS/OSM</td>
<td>11.650</td>
<td>20.162</td>
</tr>
<tr>
<td>OS/OSM</td>
<td>10.887</td>
<td>18.832</td>
</tr>
</tbody>
</table>
FS / OS data
RMSE = 0.478 m

FS / OSM data
RMSE = 11.032 m
Linear displacement

An example of overlap information of roads in an urban area, UK.
Areal comparison

- Comparative measures of polygon characteristics

<table>
<thead>
<tr>
<th></th>
<th>FS</th>
<th>FM</th>
<th>OSM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Urban area, UK</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7-Moments</td>
<td>1.85108238646506</td>
<td>1.89001507034833</td>
<td>1.56908203370849</td>
</tr>
<tr>
<td></td>
<td>0.676600780054538</td>
<td>0.729962945130622</td>
<td>0.335373471713145</td>
</tr>
<tr>
<td></td>
<td>0.233448302839737</td>
<td>0.267667161267928</td>
<td>0.0605591551248832</td>
</tr>
<tr>
<td></td>
<td>0.750316226280423</td>
<td>0.822267015494153</td>
<td>0.33424273805489</td>
</tr>
<tr>
<td></td>
<td>0.153616129162584</td>
<td>0.189278623466423</td>
<td>0.0218145701863227</td>
</tr>
<tr>
<td></td>
<td>0.43296222290291</td>
<td>0.493429527929644</td>
<td>0.131961318307499</td>
</tr>
<tr>
<td></td>
<td>1.8312107781601e-08</td>
<td>7.4366209540819e-08</td>
<td>7.10871985189128e-07</td>
</tr>
<tr>
<td><strong>Rural area, UK</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7-Moments</td>
<td>2.0495184479386</td>
<td>75443.9928203039</td>
<td>48225499296.0036</td>
</tr>
<tr>
<td></td>
<td>58999866.0268901</td>
<td>1073970995.77469</td>
<td>9.07041885436089e+020</td>
</tr>
<tr>
<td></td>
<td>230717651378.386</td>
<td>1937014198097.6</td>
<td>2.09525131641157e+031</td>
</tr>
<tr>
<td></td>
<td>230747917719.997</td>
<td>19370660096412.9</td>
<td>2.09525131644362e+031</td>
</tr>
<tr>
<td></td>
<td>2.41806602897163e+022</td>
<td>1.61072173320842e+026</td>
<td>3.4451456295601e+062</td>
</tr>
<tr>
<td></td>
<td>1.02767825569229e+015</td>
<td>3.80120101545707e+017</td>
<td>5.18970613279153e+041</td>
</tr>
<tr>
<td></td>
<td>-1834908876800</td>
<td>-8238135246848</td>
<td>3.71382011785614e+027</td>
</tr>
</tbody>
</table>
• Regular moments (Hu, 1965)
  • Area moment invariant, invariant to translation, rotation and scaling
    \[ m_{pq} = \int\int_{N} x^p y^q f(x, y) \, dx \, dy \]

• Improved moment invariants (Chen, 1993)
  • Set of invariant functions evaluated on the shape boundary only
    \[ m_{pq} = \int_{C} x^p y^q \, dl \]

\[ \square \text{FS, } \circ \text{OS and } \triangle \text{ OSM} \]
Euclidean distance among 7 moment invariants of FS, OS and OSM - Cramlington / UK (B1)
Semantic comparison

- In effect, comparing specified attributes
- Equivalent to exercises to ‘combine legends’
- Use of ontologies
  - assessment of semantic similarity of official and VGI categories
- establishing synsets and hierarchies
- semantic connections established
- testing of relationships using dictionaries
Table 1. Results of feature-based approach in Cramlington (urban).

<table>
<thead>
<tr>
<th>Semantic similarity score</th>
<th>Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.00 \leq S &lt; 0.25$</td>
<td>54</td>
</tr>
<tr>
<td>$0.25 \leq S &lt; 0.50$</td>
<td>19</td>
</tr>
<tr>
<td>$0.50 \leq S &lt; 0.75$</td>
<td>27</td>
</tr>
<tr>
<td>$0.75 \leq S \leq 1.00$</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 2. Results of feature-based approach in Clara Vale (rural).

<table>
<thead>
<tr>
<th>Semantic similarity score</th>
<th>Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.00 \leq S &lt; 0.25$</td>
<td>58</td>
</tr>
<tr>
<td>$0.25 \leq S &lt; 0.50$</td>
<td>20</td>
</tr>
<tr>
<td>$0.50 \leq S &lt; 0.75$</td>
<td>11</td>
</tr>
<tr>
<td>$0.75 \leq S \leq 1.00$</td>
<td>11</td>
</tr>
</tbody>
</table>
Schema node relationships

- A1 -> B1: One-To-One
- A2 -> B2: One-To-Many
- A3 -> B3: Many-To-One
- A4 -> B4: Many-To-One
- A5 -> no: Target Lacks Data
- no -> B5: Source Lacks Data
### Results of schemas relationships in Cramlington

<table>
<thead>
<tr>
<th>Node relations</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single correspondences (One class in Source-To-One class in Target)</td>
<td>0%</td>
</tr>
<tr>
<td>Single correspondences (One class in Source-To-Many classes in Target)</td>
<td>44%</td>
</tr>
<tr>
<td>Single correspondences (Many classes in Source-To-One class in Target)</td>
<td>53%</td>
</tr>
<tr>
<td>Missing correspondence (Source Lacks Data)</td>
<td>47%</td>
</tr>
<tr>
<td>Missing correspondence (Target Lacks Data)</td>
<td>41%</td>
</tr>
</tbody>
</table>

### Results of schemas relationships in Clara Vale

<table>
<thead>
<tr>
<th>Node relations</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single correspondences (One class in Source-To-One class in Target)</td>
<td>0%</td>
</tr>
<tr>
<td>Single correspondences (One class in Source-To-Many classes in Target)</td>
<td>43%</td>
</tr>
<tr>
<td>Single correspondences (Many classes in Source-To-One class in Target)</td>
<td>57%</td>
</tr>
<tr>
<td>Missing correspondence (Source Lacks Data)</td>
<td>43%</td>
</tr>
<tr>
<td>Missing correspondence (Target Lacks Data)</td>
<td>39%</td>
</tr>
</tbody>
</table>
Similarity measures

- Semantic similarity
- Structural similarity
- Data-type similarity
Influences on the accuracy of OSM data

• Factorial analysis
  • three factors considered
    • number of contributors (“Linus’ Law”)
      • identifying individuals responsible
    • nature of the area
      • hard or soft features
  • method of data collection
    • field or office data sources?

• Data source was the major influence
Empirical testing in Baghdad

Comparisons of RMSE and NSSDA accuracy of compared datasets

<table>
<thead>
<tr>
<th></th>
<th>RMSE(m)</th>
<th>NSSDA accuracy(m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baghdad, Iraq</td>
<td>FS/GDS</td>
<td>1.246</td>
</tr>
<tr>
<td></td>
<td>FS/OSM</td>
<td>5.903</td>
</tr>
<tr>
<td></td>
<td>GDS/OSM</td>
<td>5.806</td>
</tr>
</tbody>
</table>
Identifying the magnitude and direction of differences between GDS-OSM data in Baghdad, Iraq
Semantic similarity tests in Baghdad
Applications for data interaction

- Photosynth
- SatNav
  - TomTom
- Hazard and disaster mapping
  - Sri Lanka
  - Haiti
- SDI
  - INSPIRE

- ... more mapping examples later
Further applications for interaction

... 

• Environmental monitoring
  • Natural England

• Mapping of informal settlements
  • example of Kibera, Nairobi

• Engineering and citizen-based development
  • water / sanitation engineering
OpenStreetMap, Kibera, Kenya
Earthquake calls blamed on jet sonic boom

Public reports of an earthquake felt in south east Scotland and north east England have been blamed on the sonic boom from a military jet.

The British Geological Survey (BGS) was called by people as far apart as Eyemouth in the Borders and just north of Sunderland on Thursday afternoon.
Summary

• Need for caution

• Approach to crowdsourcing
  • active data
  • passive crowdsourcing
  • variability of data collection / reality
  • high-quality / low-quality databases?

• Is integration possible?