Semantic Challenges for Volunteered Geographic Information

Andrea Ballatore
Center for Spatial Studies
University of California, Santa Barbara

RICH-VGI Workshop
AGILE 2015, Lisbon
#ThinkingSpatially
Talk outline

1. The “Semantic Gulf”
2. Semantics of OpenStreetMap
3. Measuring Conceptual Quality
The Semantic Gulf
Vagueness in geography

- **Where** is Lisbon?
- **Where** is the Tejo River?
- **What** is a city?
- **What** is a river?

- **Semantics** is important for any geo-data, but particularly for **VGI**
OpenStreetMap.org
The semantics in OSM

• Shared conceptualization
• Semi-structured **folksonomy** (tags)
• **Semantic ecosystem**: vector, wiki, mailing lists
• **Negotiation**: alternative | incommensurable conceptualizations
OSM Wiki website

(Forthcoming in IJGIS, Ballatore & Mooney 2015)
OSM Wiki website

<table>
<thead>
<tr>
<th>Relation</th>
<th>Top 1%</th>
<th>Top 10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revisions per contributor</td>
<td>39.1</td>
<td>81.5</td>
</tr>
<tr>
<td>Added words per contributor</td>
<td>48.6</td>
<td>87.9</td>
</tr>
<tr>
<td>Deleted words per contributor</td>
<td>51.0</td>
<td>90.3</td>
</tr>
<tr>
<td>Revisions per page</td>
<td>30.2</td>
<td>71.0</td>
</tr>
<tr>
<td>Added words per page</td>
<td>39.5</td>
<td>81.4</td>
</tr>
<tr>
<td>Deleted words per page</td>
<td>71.5</td>
<td>98.7</td>
</tr>
</tbody>
</table>

i.e., extreme contribution inequality!
OSM Wiki website

Most negotiated terms

<table>
<thead>
<tr>
<th>$N_p$</th>
<th>Key/tag page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>highway</td>
</tr>
<tr>
<td>2</td>
<td>boundary</td>
</tr>
<tr>
<td>3</td>
<td>building</td>
</tr>
<tr>
<td>4</td>
<td>barrier</td>
</tr>
<tr>
<td>5</td>
<td>access</td>
</tr>
<tr>
<td>6</td>
<td>traffic calming</td>
</tr>
<tr>
<td>7</td>
<td>amenity=place of worship</td>
</tr>
<tr>
<td>8</td>
<td>waterway=riverbank</td>
</tr>
<tr>
<td>9</td>
<td>service</td>
</tr>
<tr>
<td>10</td>
<td>wheelchair</td>
</tr>
<tr>
<td>11</td>
<td>natural=tree</td>
</tr>
<tr>
<td>12</td>
<td>shop</td>
</tr>
<tr>
<td>13</td>
<td>bridge</td>
</tr>
<tr>
<td>14</td>
<td>highway=turning circle</td>
</tr>
</tbody>
</table>
The dimensions of negotiation

Ontological negotiation

Mereo-topology

Entities and their attributes, relations, and parts
The dimensions of negotiation

Cultural and linguistic negotiation

Conceptualization in British English

Problems of equivalence

Tension between universal and local
The dimensions of negotiation

I don't give a flying monkey's for tag voting, automatic changebots, endless discussions, categories, or any of that crap, but prefer to get on and actually do stuff.
Poisonous people

‘disinfection’ of the project from ‘poisonous people’ who “drain, paralyse, slow, cause needless infighting and destroy the attention and focus of a community . . . [they] are wrecking the time, focus and goodwill of the majority of contributors, creating dissent out of nothing and even purposefully breaking our data”

(Coast, 2010)
EnRICHing: More explicit semantics

- Cooperation
- Exploration
- Retrieval
- Integration
Semantic technologies (for VGI)

- Semantic Web
- Linked (open) Data
- E.g., OSM Semantic Network
Conceptual quality

- **Information value** is not only in geometric and positional quality
- **Interpretation** and **intended meaning** of the data
- Operationalize conceptual quality to **measure the interpretability** of documentation, vocabularies, schemas, ontologies
Conceptual quality

Accuracy  Granularity  Completeness
Consistency  Compliance  Richness

(Forthcoming in COSIT, Ballatore & Zipf 2015)
Conceptual quality

**Accuracy**: Distance between conceptualization and domain knowledge. Degree of correctness in the classification of features.

**Granularity**: Level of thematic description present in the data, from abstract to specific concepts.

**Completeness**: Coverage in the conceptualization of the features of interest.
Conceptual quality

**Consistency**: Degree of homogeneity in the descriptions of geographic features.

**Compliance**: Degree of adherence to a given source (standard, schema, etc).

**Richness**: Amount and variety of dimensions that are included in the description of the real-world entity.
Compliance index with OSM Wiki in four regions
Thanks!

Andrea Ballatore

aballatore@spatial.ucsb.edu
@a_ballatore