WG4 - Map evaluation

Cidália Fonte

Faculty of Sciences and Technology / University of Coimbra
Institute for Systems Engineering and Computers at Coimbra / INESCC
Summary

- WG4 Review
  - Overview
- Proposals for the next step
- Interactions with IC1203
Review paper

- Land Cover Maps validation using VGI

  - Contents:
    - Approaches to address VGI quality
    - Methodologies to assess the quality of VGI
    - Types of VGI used for LCM validation
    - Uses of VGI for quality assessment of Land Cover Maps
    - Conclusions
Approaches to address VGI quality

- Categorization proposed by Goodchild and Li (2012) – for acquisition and compilation phase
  - Crowd-sourced
    - Rely on consensus and agreement. As the number of contributors increases it is more likely that the results have higher quality, since errors are more easily identified and corrected
  - Social
    - Based on using contributors with a reliable history of quality contributions to review the work of other contributors
  - Geographical
    - Identify rules that enable the use of connections between the several types of information for each location, to assess the possibility that a certain attribute is correct at a certain location
Approaches to address VGI quality

- Categorization proposed by Allahbakhsh et al. (2013) – run-time approaches
  - Expert review
  - Output agreement and majority consensus
  - Ground truth
  - Contributor evaluation
  - Real-time support
  - Workflow management (quality control on-the-fly)

- Also identify design-time approaches
  - Task preparation
  - Selection of workers
Approaches to address VGI quality

- **Result – oriented**
  - Guide the volunteers and establish procedures to produce VGI with higher quality

- **Evaluation – oriented**
  - Assess the quality of the obtained information
## Approaches to address VGI quality

We propose a categorization based on the level of automation and type of information required to perform the quality assessment.

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Increasing degree of reliability and costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Methods are fully automated and use available information in the application</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>Corrections or assessments made by the crowd - they do require human intervention, but it is made by the crowd, not by the system managers or other experts</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>Corrections or assessments made by selected volunteers, which are supposed to create more reliable information. This require the intervention of the system managers or other experts, in the choice of the reliable volunteers</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>Corrections or assessments made by experts</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Increasing degree of outcomes, reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Data which can be automatically registered and extracted, such as some types of metadata (e.g. number of contributions, geographic location, date of contribution)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Crowdsourced data</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Proprietary data which can be accessed freely and can be used as base data to assess some types of accuracy, such as positioning over satellite images (these may have several degrees of reliability)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Authoritative data, ground truth</td>
<td></td>
</tr>
</tbody>
</table>
Approaches to address VGI quality

- We propose a categorization based on the level of automation and type of information required to perform the quality assessment.

<table>
<thead>
<tr>
<th>Types of additional data used for quality assessment</th>
<th>Degree of automation and involvement of the professionals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Automated (A)</td>
</tr>
<tr>
<td>Available data / metadata (1)</td>
<td>A1</td>
</tr>
<tr>
<td>Crowdsourced data (2)</td>
<td>A2</td>
</tr>
<tr>
<td>Base proprietary data freely available (3)</td>
<td>A3</td>
</tr>
<tr>
<td>Authoritative data (4)</td>
<td>A4</td>
</tr>
</tbody>
</table>
Methodologies to assess the quality of VGI

- Credibility
- Positional quality
- Thematic quality
- Completeness
- Currency
- Logical consistency
Types of VGI used for LCM validation

- Photographs and descriptions collected by the Degrees of Confluence Project
- Photographs posted by volunteers at collaborative sites
  - Panoramio, Flickr and Geograph
- Land cover data collected by projects
  - GeoWiki Project and VIEW-IT
- Volunteer initiatives to map the world
  - OpenStreetMap
Review paper

- **Description of projects**
  - Degrees of Confluence Project ([http://confluence.org/](http://confluence.org/))
  - Flickr ([https://www.flickr.com/](https://www.flickr.com/))
  - Virtual Interpretation of Earth Web-Interface Tool (VIEW-IT)
  - OpenStreetMap ([http://www.openstreetmap.org/](http://www.openstreetmap.org/))
Review paper

- Uses of VGI for quality assessment of Land Cover Maps
  - Indicating the data used
  - Procedures applied to assess the quality of the VGI
  - Additional approaches to guarantee the quality of the accuracy assessment results
Uses of VGI for quality assessment of Land Cover Maps

- Using photos and descriptions
  - DCP
  - Panoramio and Flickr photos

- Using classifications made by the crowd
  - Geo-Wiki
  - VIEW-IT
Some conclusions

<table>
<thead>
<tr>
<th>Degree of automation and involvement of the professionals</th>
<th>Automated (A)</th>
<th>Made by the crowd (B)</th>
<th>Selected volunteers (C)</th>
<th>Experts (D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Available data / metadata (1)</td>
<td>A1</td>
<td>B1</td>
<td>C1</td>
<td>D1</td>
</tr>
<tr>
<td>Crowdsourced data (2)</td>
<td>A2</td>
<td>B2</td>
<td>C2</td>
<td>D2</td>
</tr>
<tr>
<td>Base proprietary data freely available (3)</td>
<td>A3</td>
<td>B3</td>
<td>C3</td>
<td>D3</td>
</tr>
<tr>
<td>Authoritative data (4)</td>
<td>A4</td>
<td>B4</td>
<td>C4</td>
<td>D4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Types of approaches used</th>
<th>Types of approaches not used</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Positional quality</td>
<td></td>
</tr>
<tr>
<td>B(2,3,4)</td>
<td>A(1,2,3,4)</td>
</tr>
<tr>
<td>C(2,3,4)</td>
<td>B(1)</td>
</tr>
<tr>
<td>D(1,3,4)</td>
<td>C(1)</td>
</tr>
<tr>
<td></td>
<td>D(2)</td>
</tr>
<tr>
<td>Thematic quality</td>
<td></td>
</tr>
<tr>
<td>B(2,3)</td>
<td>A(1,2,3,4)</td>
</tr>
<tr>
<td>C(2,3,4)</td>
<td>B(1,4)</td>
</tr>
<tr>
<td>D(1,3,4)</td>
<td>C(1)</td>
</tr>
<tr>
<td></td>
<td>D(2)</td>
</tr>
<tr>
<td>Completeness</td>
<td></td>
</tr>
<tr>
<td>D(3,4)</td>
<td>A(1,2,3,4)</td>
</tr>
<tr>
<td></td>
<td>B(1,2,3,4)</td>
</tr>
<tr>
<td></td>
<td>C(1,2,3,4)</td>
</tr>
<tr>
<td></td>
<td>D(1,2)</td>
</tr>
</tbody>
</table>
Some conclusions

- Only a few automated methods have been developed so far, mostly to assess contributor reliability.
- Develop more methods of types A1 and A2, including, for example, several sources of VGI, as suggested by Goodchild and Li (2012) (geographic approach).
- Develop approaches that do not perform expert checks (type D) to assess the information quality (enabling a preliminary assessment of fitness for use).
- Only few methods use additional data of type 1 (including metadata) – develop more?
Proposals for the next step

- We need to decide what to do next
  - Review on another focused topic?
  - Research on selected topics?
  - Need to identify who is interested in what…
What to keep in mind

Main tasks and deliverables of WG4 according to the MoU

- Identify the data needs for map evaluation
- Identify the sensitivity and tolerance of validation methods to different types of error and uncertainty in VGI
- Assess the suitability of current VGI efforts to map evaluation
- Identify good practices and propose how VGI could be enhanced to increase its applicability to map validation
Proposals for the next step

- Identify the **topics** that will be addressed
- Find a **coordinator** for the work in each topic
- Gather into relatively **small groups** to get the work done
- ...
Some suggestions for future work

- Provide **definitions for terms** related to data quality, such as:
  - Quality
  - Credibility
  - Reliability
  - Trust
  - Confidence
  - …

- Find the meaning given to these terms in the literature and come up with the definition that will be adopted within the COST action
Some suggestions for future work

- **Result-oriented approaches:**
  - Identify and structure procedures used in sites to improve the quality of data:
    - Type of instructions provided
    - Real-time support?
    - Discussions between contributors?
    - Selection of volunteers?
    - Some kind of expert control?
    - Procedures used to incorporate data change
    - …
Some suggestions for future work

- **Evaluation-oriented approaches:**
  - Develop methodologies to assess data quality
    - Automated methods?
    - Methods that do not use authoritative data?
    - Methods that use the geographical approach proposed by Goodchild and Li?
    - ...

<table>
<thead>
<tr>
<th></th>
<th>Types of approaches used</th>
<th>Types of approaches not used</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positional quality</strong></td>
<td>B(2,3,4) C(2,3,4) D(1,3,4)</td>
<td>A(1,2,3,4) B(1) C(1) D(2)</td>
</tr>
<tr>
<td><strong>Thematic quality</strong></td>
<td>B(2,3) C(2,3,4) D(1,3,4)</td>
<td>A(1,2,3,4) B(1,4) C(1) D(2)</td>
</tr>
<tr>
<td><strong>Completeness</strong></td>
<td>D(3,4)</td>
<td>A(1,2,3,4) B(1,2,3,4) C(1,2,3,4) D(1,2)</td>
</tr>
</tbody>
</table>
Interactions with IC1203

- WG2 main tasks and deliverables:
  - VGI producers and users; data quality and reliability

- Last April there was a joint meeting of both actions
  - WG4 of TD1202 and WG2 of IC1203
  - Only 4 members of TD1202 attended

- There are plans for more future joint meetings
Meeting with WG2 of IC1203

- A list of possible topics where collaboration seems possible for reviews and/or research:
  - Quality visualization
  - Use of OSM data to train/validate land cover maps
  - Quality assessment without authoritative datasets
  - Use of photos to train/validate land cover maps
  - Elements to assess VGI quality
  - Integration of data from multiple sources for mapping
  - Assessment of VGI fitness for use
  - Integration of VGI with sensors networks
  - Combination / integration of VGI and SDI
  - …
Feedback please…