Towards a framework for

Intrinsic Data Quality Analysis of
Volunteered Geographic Information

Alexander Zipf,
GIScience Research Group
Heidelberg University
zipf@uni-heidelberg.de
Smart Mobility
OpenRouteService.org

VGI for Humanitarian Aid

Big Spatial Data Analytics
Dimensions of Spatial Data Quality

- Lineage
- Completeness
- Logical Consistency
- Positional Accuracy
- Attribute Accuracy
- etc...

OSMatrix.uni-hd.de

uni-heidelberg.de/gis
VGI Data Quality


iOSManalyser: Intrinsic OSM Analysis

iOSManalyser: Intrinsic Quality Indicators


uni-heidelberg.de/gis
<table>
<thead>
<tr>
<th>ID</th>
<th>Method</th>
<th>Description</th>
<th>Primary Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>Geographic context</td>
<td>Investigating the area surrounding the location of CGI to determine its geographic features and employ them to assess the quality of CGI.</td>
<td>Senaratne et al. (2013); Zielstra &amp; Hochmair (2013)</td>
</tr>
<tr>
<td>M2</td>
<td>Redundancy of volunteers’ contribution</td>
<td>Requesting several volunteers to provide information about the same geographic feature to find out if or not there is a convergence of the information produced by different volunteers.</td>
<td>Comber et al. (2013); Foody (2014); See et al. (2013)</td>
</tr>
<tr>
<td>M3</td>
<td>Scoring volunteered contribution</td>
<td>Asking volunteers to rate every piece of CGI that is contributed by other volunteers.</td>
<td>Lertnattee et al. (2010)</td>
</tr>
<tr>
<td>M4</td>
<td>Expert assessment</td>
<td>Submitting CGI to experts who are responsible for checking the information content and correcting it if necessary.</td>
<td>Foody et al. (2013); Karam &amp; Melchiori (2013)</td>
</tr>
<tr>
<td>M5</td>
<td>Automatic location checking</td>
<td>Estimating the quality of CGI by the distance between geocoded coordinates, obtained from multiple geocoding services, and the location (i.e., an address) provided by the volunteer.</td>
<td>Cui (2013)</td>
</tr>
<tr>
<td>M6</td>
<td>Spatiotemporal clustering</td>
<td>Creating spatiotemporal clusters of CGI elements using prior information about a phenomenon of interest and, later, evaluating the significance of the resulting clusters for a specific purpose.</td>
<td>Longueville et al. (2010)</td>
</tr>
<tr>
<td>M7</td>
<td>Volunteer’s profile; reputation</td>
<td>Analyzing volunteer’s profile or reputation and using it to estimate the quality of CGI.</td>
<td>Bishr &amp; Janowicz (2010); Bishr &amp; Kuhn (2013); Bodnar et al. (2014)</td>
</tr>
<tr>
<td>M8</td>
<td>Error detection /correct by crowd</td>
<td>Several volunteers acting as gatekeepers and, thus, correcting errors introduced by other volunteers.</td>
<td>Haklay et al. (2010)</td>
</tr>
<tr>
<td>M9</td>
<td>Extracting/learning of characteristics</td>
<td>Extracting characteristics from each type of geographic feature, learning the information implicit in them and, later, using the information to estimate the quality of CGI.</td>
<td>Ali &amp; Schmid (2014); Jilani &amp; Corcoran (2014); Mohammadi &amp; Malek (2015)</td>
</tr>
<tr>
<td>M10</td>
<td>Ranking/filtering by linguistic terms</td>
<td>Evaluate CGI items based on different criteria that are expressed linguistically, rank them in degrees of criteria satisfaction and, later, filter them based on the constraints of the application domain.</td>
<td>Bordogna et al. (2014)</td>
</tr>
<tr>
<td>M11</td>
<td>Historical data analysis</td>
<td>Deriving (intrinscic) indicators from the history of the data and, later, using them to make statements regarding the quality of CGI.</td>
<td>Keller &amp; de Groot (2013)</td>
</tr>
</tbody>
</table>
Conceptual Quality Dimension for VGI

- **Accuracy**
  
  Indicators: E.g. Number of features with multiple classifications; number of contributors.

  *Distance between conceptualization & domain knowledge. Degree of correctness in the classification of features in classes*

- **Granularity**
  
  E.g. Depth of classes in the class hierarchy

  *Level of thematic description, from abstract to specific concepts*

- **Completeness**
  
  E.g. Number of classes; number of attributes

  *Coverage in the conceptualization of the features of interest.*

- **Consistency**
  
  E.g. Number of features in a class with the same attributes; ratio between consistent features or attributes to all others...

  *Degree of homogeneity in the descriptions of geographic features*

- **Compliance**
  
  E.g. Ratio between the number of classes and attributes dened in an external source S and the total number of classes & attributes.

  *Degree of adherence of an attribute, a feature, or a set of features to a given source S*

- **Richness**
  
  E.g. Number of attributes describing a feature.

  *Amount and variety of dimensions that are included in the description of the real-world entity.*

---

Ballatore, A. & A. Zipf (2015):
A Conceptual Quality Framework for Volunteered Geographic Information.
COSIT 2015, Santa Fe.
Investigating the OSM folksonomy

Development of the OSM Wiki


osm-vis.geog.uni-heidelberg.de

uni-heidelberg.de/gis
ohsome.org
OSM-HISTORY PLATFORM

ohsome API
→ REST web service
→ interactive analyses and visualizations in browser

uni-heidelberg.de/gis
Example

https://api.ohsome.org/v0.9/elements/count/groupBy/boundary?bbox=Kathmandu:85.2,27.6,85.45,27.81&Pokhara:83.9142,28.1693,84.0775,28.2687&types=way&time=2015-01-01/2017-01-01/P1M&keys=building&values=residential&showMetadata=true
ohsome dashboard (preview)

[Image: ohsome dashboard preview]

[Image: HEiGIT - Heidelberg Institute for Geoinformation Technology]

[Text: uni-heidelberg.de/gis]

[Text: ohsome.org/apps/dashboard]
ohsome examples

Evolution of OSM buildings in Africa

Context: HOT coordinates OSM building mapping tasks within their Malaria Elimination Campaign.

Processing using Big Data Frameworks (Apache Ignite, Spark etc.)
Routing Nepal Earthquake 2015

- Completeness:
  - street lengths aggregated for street categories
  - percentage of streets which include street names

- Topological correctness:
  - percentage of intersecting OSM ways that do not share a common node
  - percentage of OSM endpoints close (distance 1 m), but not connected to an OSM way

- User activity: Number of users that edited OSM streets

uni-heidelberg.de/gis
Routing Nepal Earthquake 2015

Auer, Eckle, Fendrich, Griesbaum, Kowatsch, Marx, Raifer, Schott, Troilo, Zipf (2018):
Towards Using the Potential of OpenStreetMap History for Disaster Activation Monitoring.

uni-heidelberg.de/gis
Evolution Topological Correctness

uni-heidelberg.de/gis
Want to try it yourself?

- [github.com/GIScience/oshdb](https://github.com/GIScience/oshdb)
- [api.ohsome.org/v0.9/swagger-ui.html](https://api.ohsome.org/v0.9/swagger-ui.html)
- [ohsome.org/apps/dashboard](https://ohsome.org/apps/dashboard)

Get in touch...
[zipf@uni-heidelberg.de](mailto:zipf@uni-heidelberg.de)

uni-heidelberg.de/gis
Conclusion & Outlook

• The *ohsome* platform enables analyses of data evolution & quality, the detection of contribution patterns etc.
• Future versions will facilitate monitoring OSM data in near-real time on a global scale and bring extended functionalities
• Your ideas welcome!
• Further work on *enhancing* OSM

uni-heidelberg.de/gis
Thank you very much!

Questions?

Alexander Zipf  
GIScience Research Group  
Heidelberg University  
zipf@uni-heidelberg.de

uni-heidelberg.de/gis
OSM Research Overview


OpenStreetMap in GIScience: experiences, research, applications.