OSMatrix – Grid-based Analysis and Visualization of OpenStreetMap

Oliver Roick, Julian Hagenauer, Alexander Zipf

Chair of GIScience, Institute of Geography, Heidelberg University, Berliner Str. 48, 69120 Heidelberg

Abstract. Since data in OpenStreetMap is mostly surveyed by untrained private citizens rigorous quality control is required in order to produce a reliable data set. “Hidden” information, such as user activities, topicality of data, which is typically not visualized in maps might indicate areas with increased need for quality assessments. OSMatrix is a visual analytics approach that provides such information. The paper presents a short introduction to the OSMatrix application and some preliminary findings drawn from the resulting maps.

Keywords. Visual analysis, data quality, Volunteered Geographic Information

1. Introduction

Web 2.0 (O’Reilly 2005) technologies enable users to contribute to web-based data-stores, accessible through the World Wide Web. This change of perspective on how the World Wide Web is used resulted in a variety of projects, which are based on participatory collaboration (e.g. social networks such as Facebook or content communities like Flickr and Youtube). This development has also changed the way geographic information is acquired, managed, and distributed through the World Wide Web. Michael Goodchild coined the term Volunteered Geographic Information to subsume this development and to describe „the widespread engagement of large numbers of private citizens, often with little in the way of formal qualifications, in the creation of geographic information, a function that for centuries has been reserved to official agencies“ (Goodchild 2007).

The most prominent example for the phenomenon of Volunteered Geographic Information is OpenStreetMap (OSM), a project with the aim to

create a free world map solely through voluntary effort. OSM was founded in 2004 by Steve Coast and has since become a serious free alternative to data sets provided by commercial vendors and also gained the attention of the scientific community. There is a wide range of applications, which demonstrate the applicability of OSM as a data basis, e.g. the routing service OpenRouteService[1] (Neis & Zipf 2008), the virtual globe OSM3D (Neubauer et al. 2009) or disaster logistics applications (Neis et al. 2010). Furthermore, several studies on the quality of OSM data (Hakley 2010, Neis et al. 2010, Helbich et al. 2010, Ludwig et al. 2011) have shown that – at least in urban areas – the data quality is mostly comparable to commercial data sets.

However, since the data of OSM is mainly captured by untrained amateurs, it is helpful to learn about user behavior, frequency of updates, the topicality of the data, or the number of objects mapped for a certain feature type and to identify areas that stand out from their surroundings and thus may be subject for further investigations. This information provides interesting insights on the spatial diversity of OSM data, e.g. different activities and mapping habits of users in different countries or quality issues, and therefore is a valuable source to improve the overall data quality of OSM.

Because of the large size and complex structures of geospatial data sets, visual analytics have been proven as helpful methods to provide an overview and easily reveal patterns or anomalies within the data set (DiBiase 1990, Dykes et al. 2005). In the context of Volunteered Geographic Information visual exploration has already been applied successfully to provide insights on the given data set: Trame & Keßler (2011) created heat maps to identify objects in the OSM data set that are subject to a large number of edits. MacEachren et al. (2011) introduced a web-based geovisual analysis approach to allow for sense making of Twitter tweets in order to support crisis management agencies.

OSMatrix[2] is an attempt to supply the above-mentioned information using a visual analytics approach. Essentially, OSMatrix is a web-based application which portrays the spatial distribution of several key attributes, that where derived from the OSM dataset.

The paper is structured as follows: First an overview of data procession and the application itself is given. Second, some findings in OSM applying the OSMatrix and preliminary results are presented. The paper concludes with

---

[1] http://openrouteservice.org

[2] http://osmatrix.uni-hd.de

some final remarks and presents an overview of planed activities in order to improve the application.

2. Application Design and Implementation

In order to illustrate the spatial distribution of OSM data for Europe the area is divided into hexagonal cells using the open source library GeoTools. For each cell a given number of attributes is then calculated. The results are tagged with a timestamp and stored into a PostGIS database for further analysis and visualization.

The calculated attributes include information on user behavior (number of objects modified per user, number of contributing users), the topicality of data (version numbers and timestamps of the mapped objects), the relative completeness (number of features and number of attributes) or simply aggregated numbers of several feature types (area covered by buildings, farmland, commercial, residential and industrial areas) in OSM.

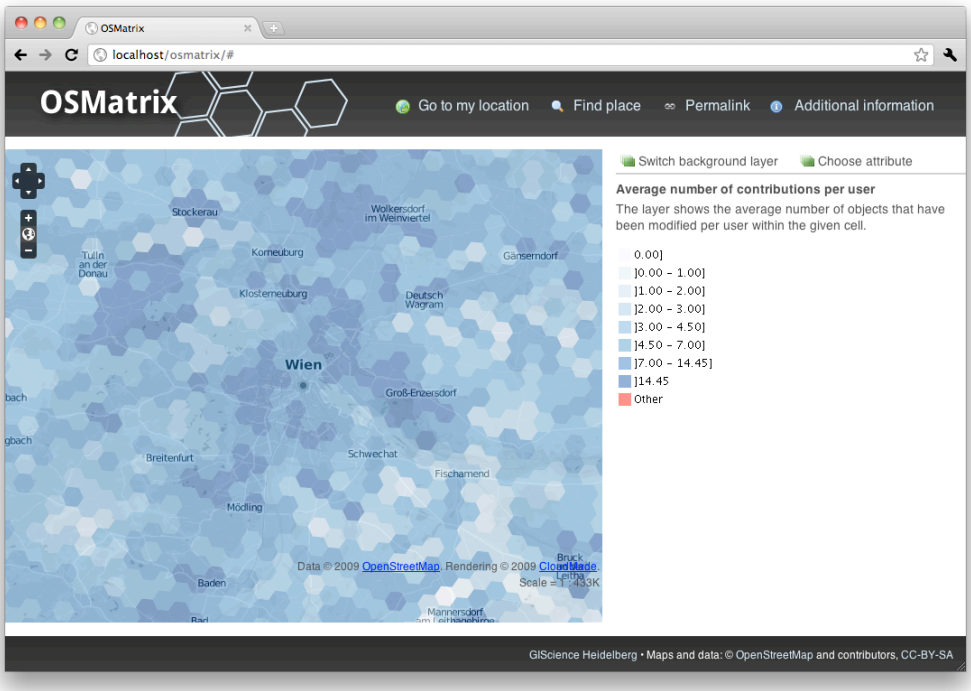

Figure 1: Screenshot of the web-based OSMatrix client.

In order to provide a geovisual analysis tool, a web-based application has been implemented. Therefore, the aforementioned PostGIS database is lin-

linked to a Geoserver based map service, which provides access to maps of the attributes through an OGC compliant Web Map Service interface. The client application is based on OpenLayers and allows for the map-based portrayal of the calculated attributes (Figure 1). Furthermore, additional information on the values for a given cell can be queried using the GetFeatureInfo tool.

3. Results

In the following section some preliminary insights that were derived from the visualization of the attributes mentioned in the previous section will be presented. The focus of this study concentrates on aspects of user activity and overall number of features and attributes. Single feature types have not been considered yet, except for buildings, because these are subject to increased mapping activities since Bing areal image were conferred right to use for mapping.

3.1. Germany

For Germany, the images of the number of features and the sum of all attributes within a cell are similar. For both, in the southern part the numbers are very high except in mountainous areas, whereas in the northern part those high numbers are restricted to urban areas. In rural parts the sum of attributes is mostly below 10 and the number of features is below 2.

Compared to other countries, the average version number is larger for Germany. Again, there are huge differences between urban and rural areas with greater values in urban areas. Same goes for the average number of contributions per user as well as the number of contributing users per cell: In urban areas the number of contributions is mostly above 7, in rural areas mostly below 3. The number of contributions is above 10 in cities and below 2 in rural areas.

In conclusion there is a very active community in Germany, which is reflected by the great number of features and attributes and the number of contributions and contributing users. Nevertheless, there are differences between urban and rural areas, with more active users in urban areas.

As the area covered by buildings implies, houses are mapped quite well in Germany's urban areas. However, the images are still somewhat patchy, indicating that buildings are still not mapped completely. In the countryside the numbers are mostly 0, so there are no buildings mapped at all.

3.2. The Netherlands

The Netherlands seem to be mapped almost completely. The sum of features and the number of attributes show the largest values compared to all other countries in Europe and the numbers are constant across the whole country.

The average version number is between 1 and 2 throughout the whole country, except for some smaller areas in the west, which is very low compared to other countries, which happen to have an active user community.

The average number of objects modified per user is constantly above 14.75 in the whole country and is thus amongst the highest in Europe.

Across the whole country the number of contributing users per cell is larger than in other European countries, but comparable to the numbers in Germany. But still, the numbers in rural areas are larger than in Germany.

Also, the area covered by buildings evolves constantly across the country. As expected, urban areas show larger numbers, but also rural areas show values between 30.98 and 76.29. Therefore, we conclude, that buildings are mapped almost completely throughout the country.

The large numbers of attributes and features throughout the whole country along with the low version numbers and large numbers of contributions per user may indicate that there has been a data import for the Netherlands. The large area covered by buildings reflects that as well.

3.3. France

The number of contributing users shows the expected situation for France: Numbers are significantly larger in urban areas than in rural areas. Hotspots are Paris, Marseille, Montpellier, Toulouse, Bordeaux and Lyon.

Contrary to the number of contributing users, the number of features mapped per cell and the sum of attributes behaves differently. One can find hot spots around the larger cities as well, but there are also areas with high numbers identifiable, where –based on the number of contributing users– lower values were expected. These areas include parts in western France around the cities of Tours, Poitiers, Nantes, Rennes and Brest as well as Perpignan and close to the Spanish border near Saint-Jean-de-Luz.

The aforementioned regions also show a larger area covered by buildings within a given cell. That's interesting since these areas do not match urban areas in France, where we expect a higher coverage of buildings. Furthermore, the boundaries of the patches match the boundaries of several administrative boundaries, which may indicate that the data originates from data imports.

The average version number also does not shape up as expected. In general the number is lower compared to Germany. But there are some patches, which stand out: Large average version numbers can be found around Amiens in north, between Paris and Reims (Figure 2), around Dijon in central France as well as around Bordeaux in the west.

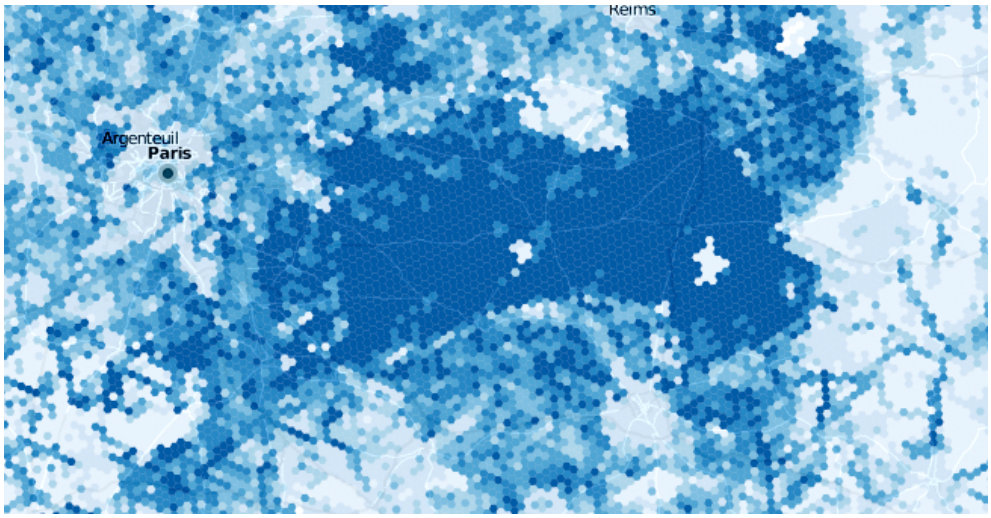

Figure 2: Increased version numbers in the area west of Paris.

In France, active user groups can be found mostly in and around large cities. Some areas might be subject to data imports, indicated by the number of features along with the number of contributing users.

3.4. Spain

The number of features and the sum of all attributes per cell show similar images in Spain. In cities, such as Madrid, Barcelona or Valencia the sum of all attributes is mostly above 2000 and the number of mapped features is at least above 100, which is amongst the largest values amongst Europe. In rural areas the values are mostly below 4.

The average version number shows some interesting patterns. Again, the urban areas around the larger cities show significantly larger values (between 2 and 4) than their surrounding rural parts (below 1). Furthermore, in rural areas linear patterns with values between 4 and 10 can be identified, which mostly matches OSM's tertiary highways (Figure 3).

In general, the number of contributing users and the number of objects modified per user and cell is notably lower than in Germany, the Netherlands and even France. Exceptions are Madrid, the urban areas at coastlines

and some cities in the northern part of Spain. In rural areas the values are mostly around 0.

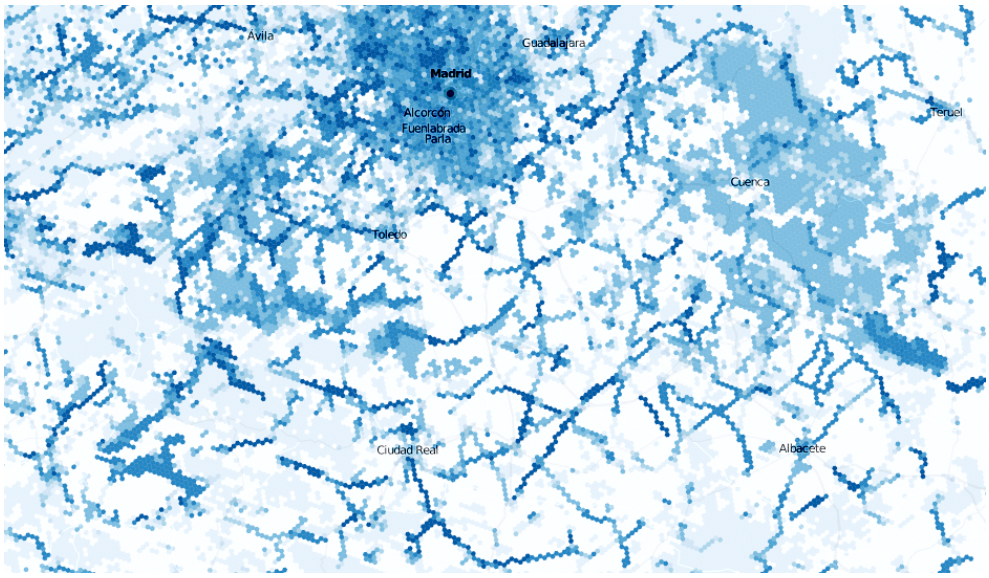

Figure 3: Linear patterns of increased average version numbers in rural Spain.

Also, the number of mapped buildings is lower than in countries we discussed before. Except for Madrid, Barcelona and cities at the northern coast the area covered by buildings is mostly 0. But also the areas of those patterns showing larger numbers do not completely match the areas of the corresponding cities. Thus, it can be expected that even urban areas buildings are largely unmapped. Therefore, it can be concluded, that mapping of buildings has started partly in larger cities but is still far from complete.

In conclusion, most user activities are present in urban areas. Exceptions are few highways in rural areas, which were subject to an increased number of edits. In rural parts of the country almost no activities were visible. This might be related to sparse population densities and a lack of features that can be mapped.

3.5. Portugal

In general, the situation of Portugal is very similar to Spain's. Cities have a more active user community and are thus mapped more completely. On the contrary in rural areas features besides the road network are mostly not mapped at all.

But areas with a significantly larger average number of attributes and a larger average version number can be identified. These patterns match ar-

eas of national parks or preservation areas (Figure 4). A reason for this might be data imports based on data sets provided by the management of the preservation areas.

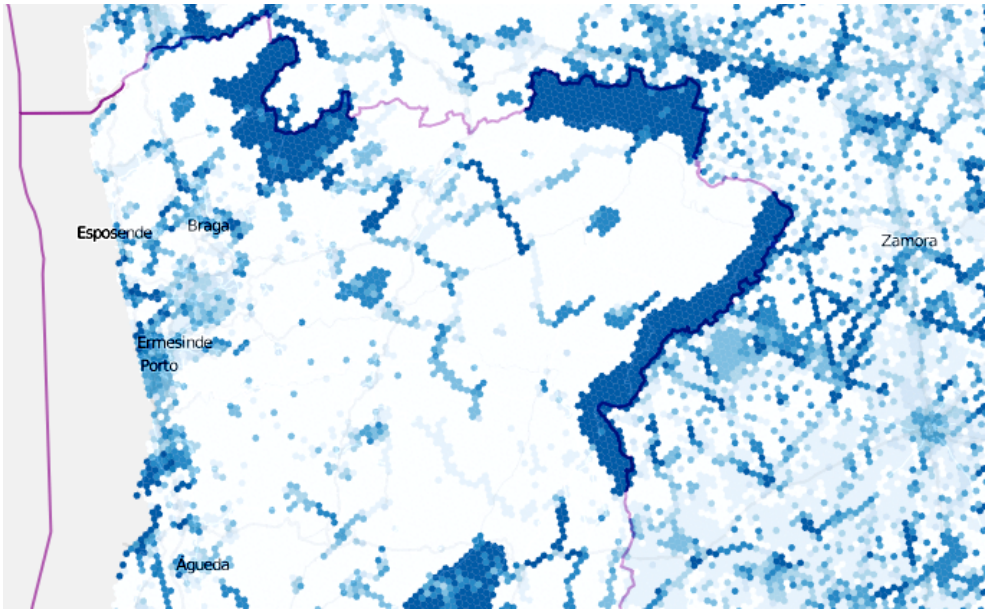

Figure 4: Natural preservation areas show increased numbers of average version number in Portugal.

3.6. Poland

Poland is in many terms different to the previously examined countries. Many attributes show significant rectangular patterns, which unlikely reflect natural activities of OSM users.

First of all, the number of mapped features per cell shows the expected increased values around urban areas such as Warsaw and Poznan. Additionally there are several rectangular patterns in the north and in the south (Figure 5) which may indicate regional data imports. The average number of contributions per user as well as the number of contributing users reflects these patterns as well.

The average number of attributes per object is generally comparable with the numbers in other countries. Northeast of Warsaw and in the south close to Kielce, there are rectangular patterns with values around 9 to 10 in places. In the western part of the country close to Gorzow Wielkopolski one can find an outstanding area with values larger than 10.

Again, identified rectangular patters might be subject to data imports. This assumption has been approved by comparison with the Mapnik rendering

of OSM data where the same patterns were also visible at natural features or the density of the street network. Thus, quality control and mapping activities are necessary in the surrounding areas.

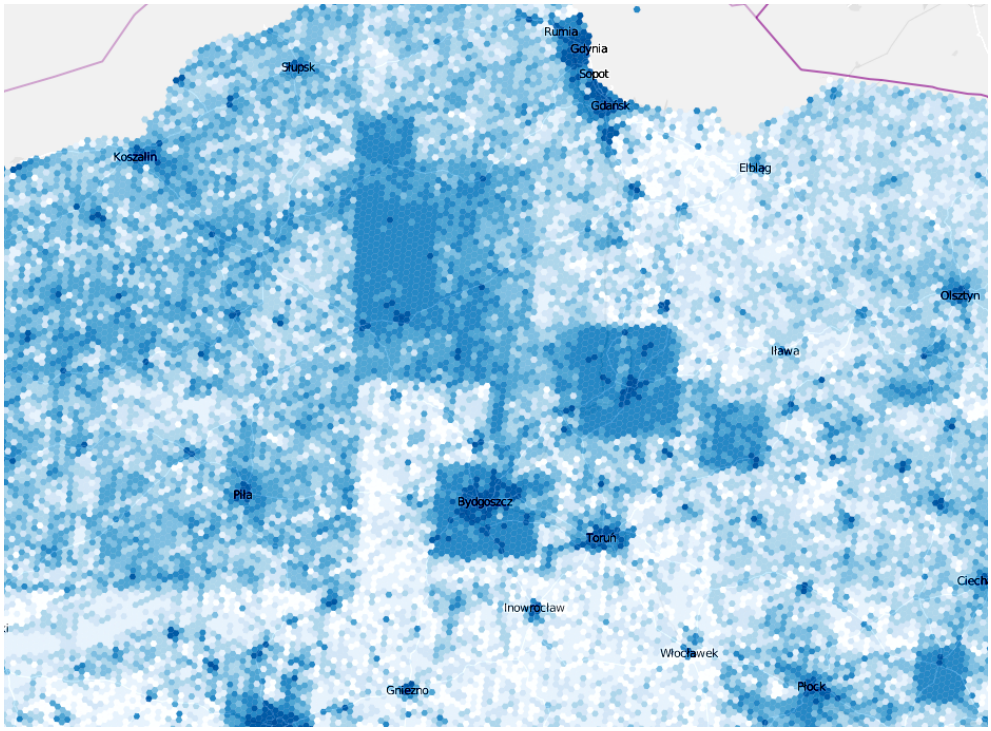

Figure 5: Rectangular patterns of the number of features mapped per cell in Poland.

4. Conclusion

The analysis of the results of the OSMatrix in the previous section revealed, that there are huge differences in terms of user activity and numbers of mapped features between selected countries across Europe.

In Germany there is a very active community and thus a lot of features got mapped, whereas in the Netherlands the relative completes of the map can be explained with data imports covering the whole country.

For countries of southern Europe it was shown that most activities are restricted to urban areas. This may be related to lower population densities or lesser objects available to be mapped.

In Poland we identified several areas where the values revealed distinct different patterns. These areas might be subject to data imports and thus

we assume that the surrounding areas are still not close to be mapped completely.

5. Future Work

A first goal in future is to improve the overall computational performance of the application. Because of the large number of features that is displayed simultaneously both client and server are temporarily overloaded especially at lower zoom levels. Solutions might be the use of a tile cache service or the use of a grid coverage based data structure at large scales instead of a feature-based structure.

Furthermore, we also aim at providing information on the temporal evolution of the attributes. Therefore, the attributes will be calculated every two months. The results may be portrayed in a single map for each timestamp or as histograms showing the evolution for a given cell. The client's functionality has to be extended accordingly.

Concerning the interpretation of the maps it may be reasonable to relate the found insights to the population density or compare the findings to those of other data sets, for instance those of commercial vendors. Furthermore, it is necessary to investigate how single feature types are mapped in order to find out what gets mapped where and why.

Further, it is important to research, how size and shape of the cells influences the results of the map: Larger or even rectangular cells may alter the results significantly and thus may be taken into account by future implementations of OSMatrix.

References

Dykes, J., MacEchren, A.M. & Kraak, M.-J. (2005): Exploring geovisualization. Elsevier.

Goodchild, M. (2007): Citizens as sensors: the world of volunteered geography. GeoJournal, 69, 211-221.

Haklay, M. (2010): How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. Environment and Planning B: Planning and Design, 37, 682 - 703.

Helbich, M., Amelunxen, C., Neis, P. & Zipf, A. (2010): Investigations on Locational Accuracy of Volunteered Geographic Information Using OpenStreetMap Data. GIScience 2010 Workshop. Zurich, Switzerland.

Ludwig, I., Voss, A. & Krause-Traudes, M. (2011): A Comparison of the Street Networks of Navteq and OSM in Germany. In: Geertman, S., Reinhardt, W. & Toppen, F. (Eds.): Advancing Geoinformation Science for a Changing World. Springer.

MacEachren, A. M., Robinson, A. C., Jaiswal, A., Pezanowski, S., Savelyev, A., Blanford, J. & Mitra, P. (2011): Geo-Twitter Analytics: Applications in Crisis Management. 25th International Cartographic Conference. Paris, France.

Neis, P. & Zipf, A. (2008): OpenRouteService.org is three times „Open“: Combining Open-Source, OpenLS and OpenStreetMaps. GIS Research UK (GISRUK 08). Manchester.

Neis, P., Singler, P. & Zipf, A. (2010): Collaborative mapping and Emergency Routing for Disaster Logistics - Case studies from the Haiti earthquake and the UN portal for Afrika. Geospatial Crossroads @ GI_Forum '10. Proceedings of the Geoinformatics Forum Salzburg.

Neis, P., Zielstra, D. Zipf, A. & Strunck, A. (2010): Empirische Untersuchungen zur Datenqualität von OpenStreetMap - Erfahrungen aus zwei Jahren Betrieb mehrerer OSM-Online-Dienste. AGIT 2010. Symposium für Angewandte Geoinformatik. Salzburg. Austria.

Neubauer, S., Over, M., Schilling, A. & Zipf, A. (2009): Virtual Cities 2.0: Generating web-based 3D city models and landscapes based on free and user generated data (OpenStreetMap). GeoViz 2009. Contribution of Geovisualization to the concept of the Digital City. Workshop. Hamburg. Germany.

O'Reilly, T. (2005): What Is Web 2.0? Design Patterns and Business Models for the Next Generation of Software. URL: http://oreilly.com/web2/archive/what-is-web-20.html. Accessed on: May 12th, 2011.

Trame, J. & Keßler, C. (2011): Exploring the Lineage of Volunteered Geographic Information with Heat Maps. GeoViz, Hamburg, Germany.