The OpenStreetMap folksonomy and its evolution

Franz-Benjamin Mocnik, Alexander Zipf & Martin Raifer

To cite this article: Franz-Benjamin Mocnik, Alexander Zipf & Martin Raifer (2017) The OpenStreetMap folksonomy and its evolution, Geo-spatial Information Science, 20:3, 219-230

To link to this article: http://dx.doi.org/10.1080/10095020.2017.1368193

© 2017 Wuhan University. Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 18 Sep 2017.
The OpenStreetMap folksonomy and its evolution

Franz-Benjamin Mocnik, Alexander Zipf and Martin Raifer

Institute of Geography, Heidelberg University, Heidelberg, Germany

ABSTRACT
The comprehension of folksonomies is of high importance when making sense of Volunteered Geographic Information (VGI), in particular in the case of OpenStreetMap (OSM). So far, only little research has been conducted to understand the role and the evolution of folksonomies in VGI and OSM, which is despite the fact that without a comprehension of the folksonomies the thematic dimension of data can hardly be used. This article examines the history of the OSM folksonomy, with the aim to predict its future evolution. In particular, we explore how the documentation of the OSM folksonomy relates to its actual use in the data, and we investigate the historical and future scope and granularity of the folksonomy. Finally, a visualization technique is proposed to examine the folksonomy in more detail.

1. Introduction
Geographical information is often regarded as exposing spatial, temporal, and thematic aspects. Goodchild (2007) has, for example, coined the term geo-atom for data explicitly exposing spatial, temporal, and thematic dimensions (Goodchild 2007). Such a view on geographic information also applies to many examples of Volunteered Geographic Information (VGI), which expose these dimensions. Specifications for spatial and temporal aspects exist, for example, for a location represented by a pair of coordinates in a given coordinate system, or for a point in time represented in Coordinated Universal Time (UTC) and formatted according to the ISO 8601 (ISO 2004). Thematic aspects are though harder to be formalized in general due to their more manifold and often more complex nature, and taxonomies or ontologies have to be established for each data-set in order to translate between the formal symbols of the data and their meanings. As VGI is often created and improved in a community-driven process, the data as well as its taxonomy is heterogeneous and reflects the needs and views of the community members. The taxonomy is thus, in many cases, never entirely formally written down, and some classes of the taxonomy are used by many contributors while others are only adopted by single ones. In case of such a community-driven creation process that is not centrally steered nor coordinated, taxonomies are often called folksonomies to reflect the decisiveness of the community, the heterogeneity and the resulting rather weak formalization.

OpenStreetMap (OSM) can be regarded as being one if not the most characteristic example of VGI. With the aim to produce maps and to offer environmental data for other purposes, the OSM project targets at representing the environment. Each feature is represented by an element, either a point feature, called a node, having a location; a polyline, called a way, composed by several nodes; or a relation between other elements. These OSM elements are thematically characterized by tags. Each of these tags consists of a key and a value, often written as "key"="value". In principle, contributors can use such tags freely without any specification that would restrict possible keys or values. Accordingly, many different tags are used (more than 89 millions as of June 2017, Taginfo 2017), and their meanings are not necessarily communicated to other contributors or users. The most important tags are documented in a wiki. The documentation is, however, incomplete because a folksonomy is, by definition, open to changes by every contributor, and conflicting versions exist due to translations into different languages.

The thematic information represented in the data are, in case of OSM, reflected by the folksonomy, a fact which can be used to predict the future development of the data when analysing the folksonomy. Which scope of the data can be expected in the future? How fine-grained will the representation be? Can different phases of the evolution be identified? etc. Despite the obvious relevance of these questions, only little research about the OSM folksonomy, and even about folksonomies in VGI in general, has been conducted. This article approaches the general understanding of the evolution of the OSM folksonomy as a whole by statistically examining the properties of the folksonomy. A more detailed comprehension of single tags remains for further examination; we though provide a visualization technique to
tackle this issue. In particular, we address the following research questions (RQ) in this article:

RQ1: Acknowledging that there is no formal requirement to document the folksonomy, how does the folksonomy used in the OSM data-set relate to its documentation in the OSM wiki? This question is of particular interest because the documentation of the folksonomy is easy to analyse, while the analysis of the folksonomy used inside the OSM data-set would require extensive computation power and a more sophisticated statistical examination. We address this question by comparing when key-value pairs were first used in the data, and when they were first documented. (Section 3)

RQ2: More and more key-value pairs are introduced over time. How did the OSM folksonomy change in the past, and how will it evolve in the future? In particular, we aim at showing that only a limited number of keys and values will be introduced if current trends continue, and we estimate this number of keys and values. This approach renders possible an understanding of how the scope of OSM folksonomy may evolve, and how fine-grained the representation will become. (Section 4)

RQ3: The scope of the folksonomy can be expected to increase over time, and the folksonomy can be expected to become more fine-grained and to be increasingly documented. Can we identify several phases in the evolution of the OSM folksonomy? RQ1 and RQ2 aim at understanding the changing scope, granularity, and documentation in more detail. We address the third research question by comparing the results of the preceding research questions. (Section 5)

RQ4: The OSM folksonomy is complex and subject to regular modifications. Many decisions to modify the folksonomy, or its documentation, result from the need for new values, or even from planning processes. These factors can be understood by manually retracing when new values were introduced, or when values were deprecated. How can we visualize the OSM folksonomy in order to understand its evolution at the level of individual keys and values? The authors are not aware of any already available visualization of the history of the OSM folksonomy. We propose a new visualization technique, which is able to address this research question. (Section 6)

2. Related work

VGI, and OSM data in particular, has been examined in many studies, and a number of tools exist to browse the data, including its folksonomy. A commonly used method is to display parts of the OSM data-set in an interactive map, which provides additional information on request. There exist, for example, a number of software tools to view OSM data (www.openstreetmap.org, mobile viewers, etc.), to use data for further investigations (geographic information systems, in particular Quantum GIS, ArcGIS, etc.), and software tools to edit OSM data (iD, Potlach 2, JOSM, Maps.me, Vespucci, etc.). These tools concentrate on the examination of current OSM data, often including thematic information, but historic data are mostly excluded. It is, however, the temporal dimension which enables the examination of the evolution of OSM data.

Several software tools examine and visualize the creation process of only a small part of the OSM data-set. The application show-me-the-way (www.github.com/osmlab/show-me-the-way), for example, visualizes the recent changes of OSM data with only a short delay. While this application provides an understanding of how boundaries of elements are mapped, it does not provide holistic insights about the entire creation process. The history of an OSM element can be examined by the application osm-deep-history (www.github.com/osmlab/osm-deep-history); a collection of changes submitted as a “changeset” can be examined using the Augmented OSM Change Viewer (overpass-api.de/achavi); and the tool Who did it? (zverik.osm.rambler.ru/whodidit/) provides information about local changes. Similar tools exist or did exist.

Information about the folksonomy, in particular, about the tags used to thematically describe OSM elements, has been collected and aggregated by several websites such as Taginfo (taginfo.openstreetmap.org) and Tagfinder (tagfinder.herokuapp.com). These websites summarize information provided by the OSM wiki, whereby this information is further enhanced by considering statistics about the usage of tags in the OSM database, and by information about how other projects use these tags. The tool OSM Tag History (taghistory.raifer.tech) visualizes the usage of a tag in the OSM database by a line chart. The website OSMstats (osmstats.neis-one.org) examines even other statistical data about the OSM data-set and the users, and visualizes the data by line charts. The geospatial distribution of elements tagged as buildings or roads can be examined by OpenStreetMap Analytics (osm-analytics.org). The website OSMatrix provides tools to, among others, statistically analyse the use of tags (Roick et al. 2012, 2011). A detailed statistical analysis of OSM users has been provided by Mooney and Corcoran (2012b).

The evolution of OSM has been studied widely by tracing how metric properties and the topology of the represented street network evolve (Neis et al. 2012; Corcoran and Mooney 2013). Arsanjani et al. (2015) even
simulated the potential evolution of OSM through a cellular automata model. In a study of the road network in Beijing, Zhao et al. (2015) demonstrated how the mapping behaviour advances OSM data, in particular, how the evolution of the road network is shaped by exploration and densification activities. These papers examine, however, only to a minor extent the folksonomy but rather focus on spatial and temporal features of the data. Other studies relate OSM data, at least to some degree, to the folksonomy, but do not examine the history. Zielstra et al. (2013) have, for example, assessed the effect of data imports, differentiating between different tags that stand for different categories of roads.

The quality of OSM data has been discussed in respect to the folksonomy. Barron et al. (2014) have, for example, discussed the quality of OSM data in terms of different factors, one of which is the number of tags of an element. A more thorough discussion on the conceptual quality of OSM has been provided by Ballatore and Zipf (2015). They discuss different dimensions of conceptual quality, including the accuracy, the granularity, the completeness, the consistency, the compliance, and the richness of the data, by considering the folksonomy documented in the OSM wiki and the taxonomies provided in different editors. The discussion does, however, not consider the evolution of the folksonomy to a greater extent. The tagging practices related to OSM have been examined in greater detail by Davidovic et al. (2016). The study examines, in particular, how well features have been tagged by different users. Mooney and Corcoran (2012a) have examined how the tags associated to an element change over time, and how the lack of control mechanisms can affect data quality. Finally, Aliakbarian and Weibel (2016) have shown how to make use of the OSM folksonomy when generalizing information for maps.

Folksonomies have been examined in many studies (Trant 2009). Shen and Wu (2005) describe folksonomies as complex networks, whereby the latter might reflect the evolution of the former: the discussed laws of complex networks have been shown to often be the result of a temporal process. The dynamic aspects, resulting from the collaboration of many contributors, have been discussed by Golder and Huberman (2005). The general evolution of folksonomies has been discussed by Gendarmi and Lanubile (2006), with the aim to provide methods to apply community-driven evolution to ontologies.

3. Documentation of the folksonomy

The OSM folksonomy is created by the use of tags in the data, but a documentation in the OSM wiki is available to foster a common view on which tags are meaningful. As a folksonomy, the collection of tags is neither planned nor controlled by a central instance. It is rather the result of, at least in parts, independent decisions by individual contributors. These contributors, however, need to agree on common keys and values if their data shall be usable on a larger scale – how could the data otherwise be interpreted when, for example, creating a map? Such agreements are discussed in the community, using mailing lists or personal discussions, and they are subsequently often documented in the OSM wiki. As a result, the folksonomy can, at least in parts, be examined by analysing its documentation. This section examines how good the documentation of the folksonomy is, and what can accordingly be followed about the taxonomy by an analysis of its documentation.

The first use of a tag in the data and the first documentation of the tag are compared in Figure 1. While the date of the first documentation in the wiki is very clear, it is not clear when a tag shall be considered as being used in the data. There are more than 89 millions distinct tags being used as of June 2017 (Taginfo 2017), and a single use of a tag may thus not be considered as relevant. As can be seen in Figure 1(a), tags are, with only minor exceptions, used in the data before being documented. This behaviour reflects that the folksonomy is created by its use, rather than by a centrally coordinated process with a strong formalization. Before 2011, some tags were documented upon their first use. Corresponding contributors were thus most likely aware that they are the first to use certain tags in the data, and hence recognized the necessity to document these tags. The vast majority of tags documented after 2013 have, however, been used before their documentation. The documentation can thus be regarded as a representation of the folksonomy that was defined in the data, and not vice versa.

The majority of the relevant and frequently used tags are documented in the wiki, despite the fact that their documentation is, with minor exceptions, created after the first use of the tags in the data. Most tags have, for example, been documented before having reached 10% of its current use in the data (Figure 1(d)). The same effect can even be seen in case of the 100th use (Figure 1(b)) or 1% of its current use (Figure 1(c)). These figures do not depict those tags that are only used in the data but never have been documented. In fact, the general examination of tags like "name"="New York City" inside the documentation would make little sense, because they are only used for one or few specific features, in the above example, for the City of New York. While the documentation of the tags appears mostly after their first use, the first documentation and the first use in the data are only weakly correlated (tags are distributed in the lower triangle in Figure 1(a)). There is, however, a linear correlation between the first documentation and the time at which they become relevant (tags distributed around the diagonal in Figure 1(d)).
Majordocumentationeffortshavebeenmadeinthe earlyyears,withafocusonfrequently-usedtags.Taggene bylargerdisksinFigure1(a)—tagsthathave beenusedextensivelyinthedata-set—appear significantlymoreoftenbefore2009,regardingtoboth theiruseandtheirdocumentation.Whilethiscould beduetothighchancetoadopttheseprograms—they basicallywerewrittenandsourcedfromsomepointaone canassume that the most essential concepts were introducedinearlytimes,andmanyofthesescenarios canalsobeexpectedtobeverefrigerately used.Thisdemonstrates that themostfrequently-used tagswereusedanddocumentedveryearly.Notonlyfrequently-usedtagsbutalsoleinfrequently-usedones wereextensivelydocumentedbefore2010.Incaseof disksbeinghorizontallyalignedinFigure1,several tagsweredocumentedatthesametime,mostlikely inacoordinatedway,eventhoughthetagshavebeen usedinthedatatendifferentpointsintime.Such coordinatedeffortscanbeobservedbetween2008and 2015.

How can we determine how the completeness of the documentation of the tags has been changing over time? As has been discussed earlier, there is no sense in considering all tags, because many values are only used once or a few times in the data, as in the above example of the City of New York. Instead, only relevant tags should be considered, and most of them seem to be documented in the OSM wiki, according to our previous findings. This is why we consider as a statistical population only the currently documented tags $\tau$ that have been used more than 1000 times in the data. At a given point in time $t$, only a subset $\tau_t \subset \tau$ of these tags have been used in the data. The completeness of the documentation at a point in time $t$ is, in the scope of this paper, defined as the percentage of tags in $\tau_t$ that were documented at time $t$. While this definition necessarily implies that the documentation is complete in current times, it can reveal about how the completeness evolved over time (Figure 2).

After a period of ongoing documentation, the documentaion of the tags had reached a high level of com-
pleteness. Figure 2(a) shows that the completeness of the documentation is increasing over time, which also could be an artefact of considering only currently documented tags \( \tau \). The larger increase in the early years compared to later ones indicates, however, that the increase is not only such an artefact. This impression is amplified when considering, instead of \( \tau_t \), the set of tags \( \tau'_t \) that have reached 10% of its current use in the data at time \( t \) : a rapid increase of the completeness happened between 2008 and 2010, and the completeness in later years was always around 90% or above (Figure 2(d)). Before 2008, the documentation was very incomplete (Figure 2(b) and (c)).

The results of this section have demonstrated that there exists a close relationship between the folksonomy and its documentation, which answers RQ1. Tags are usually first documented after being introduced in the data, justifying the collection of tags to be called a folksonomy due to their, in large parts, uncoordinated use in the data. Most tags are, however, documented as soon as they have become relevant due to their frequent use in the data, making the documentation suitable for studying the folksonomy. It can even be hypothesized that the documentation and the adoption of tags in OSM editors have an impact on the use of the tags in the data.

4. Evolution of the folksonomy

The OSM folksonomy is evolving over time – it is extended and modified by the use of new tags during the contribution of data. As we have seen in the previous section, we can analyse relevant parts of the folksonomy by its documentation. This section tackles RQ2 by analysing how the documentation of the folksonomy has been changing over time.

The number of keys and tags is growing over time. Figure 3 depicts the number of keys and tags, that is, key-value pairs, that have been documented in the OSM wiki.

Figure 2. Completeness of the documentation of the tags in the OSM wiki. (a) First use of the tag in the data. (b) 100th use of the tag in the data. (c) 1% of the current use of the tag in the data. (d) 10% of the current use of the tag in the data. The depicted completeness refers to how many tags of the currently documented tags have or have not, at a given point in time, been documented, despite having already been introduced in the data (first/100th/etc. use). The documentation is necessarily 100% complete at the current date, because only tags that are documented in the OSM wiki and that are used at least 1000 times in the data are considered in the population of the statistics. Tags with value “*” are excluded. Data from the OSM wiki © OpenStreetMap contributors (cf. http://wiki.openstreetmap.org/wiki/Wiki_content_license).
wiki. After a period of slower documentation (Figure 3(a) and (b)), the number of documented keys and tags is increasingly growing since 2008. This growth follows an exponential law with negative exponents in both cases, approaching a constant value in future times. Assuming that the current behaviour continues, there will be about 194 keys documented in the limit case, and 98% of this number will statistically be reached in the third quarter of 2017. After late 2017, the number of documented keys will, by and large, stagnate, which means that the number of new keys will counterbalance the number of removed keys if current trends continue. As can be seen in Figure 3(a), deviations from this statistical trend may occur. The number of tags will approach about 1213. According to the current prognosis, 98% of these will be reached in the fourth quarter of 2031, if the current trend is not subject to future changes.

The evolution of the keys and tags can be interpreted in terms of scope and granularity of the folksonomy. Keys are used to represent different themes. As values only occur in certain combinations with these keys, the values can be regarded as being subordinate to the keys. The scope of the folksonomy is accordingly only determined by the keys, while the values determine the granularity. Each value represents a sub-concept of a key. The more values are used for a certain key, the more fine-grained the subconcepts are. As the documentation contains the relevant keys and values, we are able to measure the relevant scope and the relevant granularity respectively.

The folksonomy of OSM will have reached its maximal scope by late 2017, according to the above findings, while the granularity of the folksonomy is still becoming finer after this date. The granularity can be examined in more detail by analysing the average number of values per key (Figure 3(c)). This average number varies in early years, but follows a linear trend after 2010. This linear growth shows that the folksonomy becomes increasingly fine-grained. When the number of documented keys and values stagnates, the linear growth of the granularity will have to be stopped: the number of values per key will also stagnate.

Most keys have only one value, such as "amenity"="yes" or "width"="number". In the first case, the concept of a tunnel is not very fine-grained, and the value "yes" just indicates that the feature is, in fact, a tunnel. In the second case, the width is provided and represented as a number. Both examples are very typical, as can be seen in Figure 4(a): most keys have only one value, even when excluding the value "". Keys with many values were created in the first years (Figure 4(b)). The fact that these keys have currently many documented values is not an effect of the long time since their creation. Instead, the number of values of these keys has also been growing much quicker than for other keys. The keys "shop" and "amenity" have 140 and 106 values, respectively.

<table>
<thead>
<tr>
<th>Phase</th>
<th>Years (prognosis after 2016)</th>
<th>Documentation</th>
<th>Scope</th>
<th>Granularity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase I</td>
<td>–2007</td>
<td>Very little</td>
<td>Growing</td>
<td>Refining</td>
</tr>
<tr>
<td>Phase II</td>
<td>2008–2009</td>
<td>Growing</td>
<td>Growing</td>
<td>Refining</td>
</tr>
<tr>
<td>Phase III</td>
<td>2010–2017</td>
<td>Almost complete</td>
<td>Growing</td>
<td>Refining</td>
</tr>
<tr>
<td>Phase IV</td>
<td>2018–2031</td>
<td>Almost complete</td>
<td>Stable</td>
<td>Refining</td>
</tr>
<tr>
<td>Phase V</td>
<td>2032–</td>
<td>Almost complete</td>
<td>Stable</td>
<td>Stable</td>
</tr>
</tbody>
</table>

It comes not unexpected that these keys have most documented values, because thematic information in the geographic domain is often about places, and shops and amenities are very important types of places.

The history of the folksonomy did follow simple laws in the last years, which enables us to extrapolate its future development, as has been discussed in this section. This answers RQ2. In particular, we have argued that both the scope and the granularity will become stable over time, and we have derived the number of keys and values to expect in the limit case, if current trends continue.

5. Phases in the evolution of the folksonomy

The two preceding sections have shown how the folksonomy and its documentation change over time. These changes are very different in earlier and in later years, and different trends can be identified. In this section, we aim at identifying different phases in the evolution of the OSM folksonomy by considering in combination the different factors that already have been discussed. These considerations answer RQ3. An overview can be found in Table 1. It should be noted that the year specifications of the phases IV and V are predictions; they presume that current trends continue without change and can only be seen as a prognosis.

Phase I: Foundation Phase (~2007): In the early years of OSM the folksonomy emerges. There exists only very little documentation, and only very little can be followed by examining the documentation.

Phase II: Documentation Phase (2008–2009): The second phase is characterized by a growing documentation, until most relevant keys and values are documented. During this phase, the documentation reflects the folksonomy only in parts, and it can only be conjectured that the number of keys and values is growing.

Phase III: Phase of Growing Scope and Refining Granularity (2010–2017): The documentation is close to completion in this stage, and relevant parts of the folksonomy can accordingly be examined by its documentation in this and subsequent phases. The number of keys...
and values increases, indicating the scope of the folksonomy to be growing and the folksonomy to become more fine-grained. This growth of the number of keys follows an exponential law with negative exponent; the average number of values per key, a linear law.
Phase IV: Phase of Refining Granularity (2018–2031):
The number of keys has become stable, and the scope of the folksonomy is, accordingly, not growing any longer. The number of tags, that is, of key-value pairs, still grows. The folksonomy becomes accordingly more fine-grained, but it is not clear how exactly the granularity is changing. Conjectured that the number of tags follows an exponential growth with negative exponent also in this phase, the phase will last until the end of 2031.

Phase V: Phase of Stability (2032–): This phase is characterized by a non-changing number of documented keys and values. For each new key and value, there will, statistically, be an old key or value respectively be removed from the documentation. The relevant scope and granularity of the folksonomy are expected to not grow any longer, albeit they may still evolve.

At the time of publication, the evolution of the OSM folksonomy is at the end of phase III. Phases I to III are, accordingly, the result of the analysis of the previous evolution of the folksonomy and its documentation. Subsequent phases are, however, the extrapolation of current trends, and they are thus subject to unexpected influences. Will the folksonomy unexpectedly become even more fine-grained when the mapping of the environment, according to the existing folksonomy, reaches global completeness, while, at the same time, adhering to high quality standards? Will the aims of OSM change and the scope accordingly broaden in the future? While
the expected temporal boundaries of phases IV and V may not be proven to be true, the predicted evolution does not come unexpected. The increase in the number of keys has already significantly slowed down, whereas the number of values is still increasing. It comes not unexpected either that the number of values will stagnate, because the granularity can practically not refine forever.

6. Visualizing the evolution of the folksonomy

The OSM folksonomy is subject to constant change. The preceding sections only examine changes in the number of keys and values, but not all changes are reflected by these numbers. In particular, these numbers do not change when old keys and values are replaced by new ones in phase V. These changes can, accordingly, not be analysed with the methods discussed in the preceding sections. This section aims at finding visualization techniques to explore these changes of the folksonomy and thus provides answers to RQ4.

In Figures 5 and 6, the history of the documentation of the OSM folksonomy is visualized as a network. The nodes in the inner circle refer to the keys, and the nodes outside this circle refer to the values related to these keys. Both the keys and values are linked by lines in case they can occur in combination. The documented keys and values, and the combinations in which they occur, are changing over time. Accordingly, also their depiction varies over time. In the interactive visualization, which can be found online as part of the OSMvis-Project,² the point in time can be chosen by a time

---

² The OSMvis-Project is an open-source project for visualizing and analyzing OpenStreetMap data.
related values were introduced in early 2010, but the concept related to the key was already much more fine-grained at that time (Figure 5). The concept related to the key "sports" and related values were introduced in early 2010, but the concept of buildings has become more fine-grained first in late 2012 (Figure 6). These examples demonstrate that the folksonomy, or at least its documentation, has, in parts, developed in a possibly unexpected way. This is despite that the number of keys and values developed in a predictable way.

This section does not aim at discussing the keys and values in detail, but rather demonstrate that a detailed comprehension can be gained by the proposed visualization technique. Such knowledge is, in fact, useful for the understanding of how the concepts for OSM data are evolving and how they affect data quality. In fact, several metrics of how good an ontology might be in respect to different aspects and applications have been developed (Burton-Jones et al. 2005; Fernández et al. 2009). As long as the environment, that is, the subject to describe, and the purposes for which OSM data are used do not drastically change over time, the ontology – in our case the folksonomy – can be expected to be stable over time or to improve uniformly. The folksonomy can also be expected to be of uniform granularity, if there are no particular reasons to model certain aspects of the environment with different granularities. If there happen unexpected changes of the scope or the granularity over time, a detailed understanding of which keys and values were removed or introduced may provide insights about data quality.

7. Conclusions and future work

This article treats the OSM folksonomy and how it evolves over time. We have found evidence that tags usually are first used in the OSM data and then are documented, aligning well with the collection of tags being regarded as a folksonomy that is created and evolves in a community-driven process. Despite the documentation being created at a later point in time, it contains most of the relevant tags, and the documentation of the relevant tags seems to be close to completion since a longer time. It has been shown that the evolution of the folksonomy has followed, at least in recent years, simple laws, which provide insights into the future evolution of the folksonomy. We have, in particular, identified five phases in this evolution, including an increasing scope of the folksonomy (almost 200 keys expected in late 2017; end of phase III) and a refining granularity (more than 1200 values expected in late 2031; end of phase IV), assuming that the evolution of the folksonomy can be extrapolated from its history. In phase V, the scope and the granularity will both be stable if current trends continue. Finally, we have introduced a visualization technique to explore the folksonomy at the level of single keys and values.

We have, in this article, examined the folksonomy as a whole. While some aspects of the evolution of the folksonomy can be comprehended by such an examination, the motivation behind single changes can only be comprehended by a more detailed analysis. Which keys become more important or more fine-grained over time? Are values systematically renamed? Which new topics are reflected by the folksonomy? The visualization presented in Section 6 provides a possible approach to systematically examine the folksonomy in detail, but further research is needed to obtain a better understanding of the predominant patterns. Even supplementary or alternative visualizations might be developed to stress different aspects of the folksonomy.

This article examined the English documentation of the OSM folksonomy only, despite the folksonomy being documented in different languages in the OSM wiki. These language versions differ in their content and length, and the comparison of these versions might reveal more information about the creation process of the documentation, as well as about its completeness. Future research might address the evolution of the folksonomy and data quality issues by examining these differences between the different language versions in detail.

The OSM folksonomy is subject to change, because the environment and, even more important, also the purpose of the data are changing. The folksonomy not only adapts to these changes but also improves and reflects the zeitgeist. In consequence, data may refer to an outdated tag, that is, a tag which has been renamed or replaced in the documentation in the meantime, or to a
tag that has got another meaning. Such inconsistencies can hardly be avoided and can even be seen as a characteristic of VGI. How do changes of the documentation and changes in the data relate? What influence does the vocabulary that has been adopted by OSM editing software have? Future research may shed light on such interactions, in particular by viewing such interaction as a community-driven process.

We have discussed several phases of the evolution in Section 5. These phases incorporate the evolution of the folksonomy as well as its documentation. Twitter hashtags, tags in Flickr, and folksonomies in similar data collections share many properties with the OSM folksonomy. In how far do the phases of the evolution of the OSM folksonomy also apply to other folksonomies and to examples of social tagging? Which of the observations are specific to the evolution of the OSM folksonomy, and why are they?

Notes

1. In the scope of this paper, OSM wiki refers to the English language version of the wiki run by OSM, and the documentation of the folksonomy refers to the keys and values documented on http://wiki.openstreetmap.org/wiki/Map_Features as well as on linked pages.

Funding

This work has been partially supported by the Deutsche Forschungsgemeinschaft (DFG) project A framework for measuring the fitness for purpose of OpenStreetMap data based on intrinsic quality indicators [grant number FA 1189/3-1].

Notes on contributors

Franz-Benjamin Mocnik is a postdoctoral researcher at Heidelberg University. His main interests are structures and laws in geographical information science, often with a focus on data quality and Volunteered Geographic Information.

Alexander Zipf is a professor at Heidelberg University. He works on the analysis of Volunteered Geographic Information with a strong focus on data quality, as well as in crowdsourcing and citizens as sensors.

Martin Raifer is a researcher at Heidelberg University. He is working on innovative technology related to OpenStreetMap and open geodata in general, as well as on spatial data analysis and visualization.

ORCID

Franz-Benjamin Mocnik  http://orcid.org/0000-0002-1759-6336
Alexander Zipf  http://orcid.org/0000-0003-4916-9838

References


