Mocnik *et al. Open Geospatial Data, Software and Standards* (2018) 3:7
https://doi.org/10.1186/s40965-018-0047-6

Open Geospatial Data,
Software and Standards

SOFTWARE Open Access

Open source data mining infrastructure for exploring and analysing OpenStreetMap

Franz-Benjamin Mocnik*, Amin Mobasheri and Alexander Zipf

Abstract

OpenStreetMap and other Volunteered Geographic Information datasets have been explored in the last years, with the aim of understanding how their meaning is rendered, of assessing their quality, and of understanding the community-driven process that creates and maintains the data. Research mostly focuses either on the data themselves while ignoring the social processes behind, or solely discusses the community-driven process without making sense of the data at a larger scale. A holistic understanding that takes these and other aspects into account is, however, seldom gained. This article describes a server infrastructure to collect and process data about different aspects of OpenStreetMap. The resulting data are offered publicly in a common container format, which fosters the simultaneous examination of different aspects with the aim of gaining a more holistic view and facilitates the results' reproducibility. As an example of such uses, we discuss the project OSMvis. This project offers a number of visualizations, which use the datasets produced by the server infrastructure to explore and visually analyse different aspects of OpenStreetMap. While the server infrastructure can serve as a blueprint for similar endeavours, the created datasets are of interest themselves too.

Keywords: Infrastructure, Data repository, Volunteered geographic information (VGI), OpenStreetMap (OSM), Information visualization, Visual analysis, Data quality

Introduction and background

The way geographical information is collected and used, and also the characteristics of the data themselves, has changed in the last years and decades. Geographical information is not longer the domain of experts, but citizens, often without formal qualification, voluntarily collect information about the environment they are living or working in, or are familiar with for some other reason. Citizens also derive geographical information from aerial images and other sources. Such geographical information that is collected in a cummunity-driven process is called *Volunteered Geographic Information* (VGI) [1, 2].

One of the most prominent examples of VGI is OpenStreetMap (OSM) [3] – a community-driven effort to collect geographical information about the environment: streets and buildings; villages, cities, administrative divisions, and country boundaries; land use classification; natural and physical land features; amenities; leisure, tourism, sport sites, and shops; and many more information. The data derived and maintained by the OSM project can be used for different purposes [4]. While the data are often used to create street maps (www.openstreetmap.org), also other services have been created based on or using OSM data: topographic maps (www.opentopomap.org), thematic maps for cyclists (www.opencyclemap.org), for sailors (www.openseamap.org), and for users of ski pistes (www.opensnowmap.org); a (reverse) geocoder (nominatim.openstreetmap.org); a digital globe (marble.kde.org); routing services (www.project-osrm.org, www.openrouteservice.org); etc. OSM data are used by citizens, companies, and organizations for producing and viewing maps, for planning tasks, for humanitarian aid, and crisis management.

Data quality is a major issue for OSM and VGI in general, because a holistic and deep understanding of the data creating process is needed [5–7]. The characteristics of VGI to be collected and maintained by a community implies that the advancement of the dataset and its underlaying folksonomy, that is, the semantics of the tags used in OSM, are not controlled by a single person or

*Correspondence: mocnik@uni-heidelberg.de
Institute of Geography, Heidelberg University, Im Neuenheimer Feld 348, 69120 Heidelberg, Germany

authority but rather the result of a community-driven process [8–10]. The tools and services used to produce and process the data are neither. The data are rather edited and the folksonomy and tools created on a voluntary basis by individuals who coordinate their activities, leading to strong heterogeneity among the users and their mapping behaviour [11, 12]; among the choice of what to map; among the chosen representation; and among the sources used to map, for example, aerial images, local knowledge, and GPS tracks. This heterogeneity implies that some regions are mapped more complete than others [11, 12]; concepts are used differently [13]; the data are not always up to date [12]; and locations are described with differing precision [12].

The comprehension of the complex, community-driven process that leads to VGI datasets is necessary to understand the data and their quality, because it is foremost this process which distinguishes VGI data from other data and potentially causes heterogeneity. Information visualization investigates visual representations of complex information to reinforce human cognition and, in turn, make complex processes more tangible. This article demonstrates how techniques from information visualization can be used to gain a deeper understanding of the community-driven process of creating and modifying OSM data.

There are several studies and tools that inspect, analyse and/or employ VGI and OSM data in particular. In the case of OSM there exist several systems to display and investigate OSM data, for example, the OSM web platform, GIS systems such as *QGIS* (open source), and *ArcGIS*. Several other software also exist for editing OSM data, with *iD*, *Potlach 2*, *JOSM*, *Maps.me*, *Vespucci* being the most common tools for this purpose. A capability that is lacking in before-mentioned software is the possibility to analyse and display information about the creation process of data.

For this specific purpose, several other software tools exist that can only examine and visualize the creation process of a small part of the database. An example of such services is *show-me-the-way* (www.github.com/osmlab/show-me-the-way), which provides users the possibility to visualize the recent changes of OSM data (with short delay). Further examples include *osm-deep-history* (www.github.com/osmlab/osm-deep-history), which allows to analyze the history of OSM; *Augmented OSM Change Viewer* (overpass-api.de/achavi), which allows to analyze a collection of changes submitted to OSM in the form of a 'changeset'; and the tool *Who did it?* (zverik.osm.rambler.ru/whodidit/), which provides information about local changes. In addition, there exist websites that collect, aggregate, and analyse information about the tags used to describe OSM objects. The most famous websites are: *Taginfo* (taginfo.openstreetmap.org), *Tagfinder* (tagfinder.herokuapp.com), *OSM Tag History* (taghistory.raifer.tech), and *OSMstats* (osmstats.neis-one.org). Furthermore, a first statistical analysis of OSM users has been provided by Mooney and Corcoran [9, 14].

The quality of OSM data has been examined in respect to many purposes and many areas using different methods [7]. Trame et al. [15], for example, examined the lineage of OSM objects and visualized the result as a cartographic heat map. Roick et al. [16, 17] have visualized several statistical properties of OSM data by a cartographic heat map, with the aim to understand the quality of the data. An overview of how the quality of OSM data compares to the one of other datasets has been provided by Haklay [18] for the first time and repeated in several other studies and use case scenarios. In terms of geographic areas, the quality of OSM has, among others, been examined for Germany [19], France [20], Brazil [21] as well as for Iran [22]. OSM data have been widely used in various application domains including disaster management [23], routing and navigation [24, 25], and urban demographic estimation [26].

In the context of the community-driven process, data quality has widely been discussed, among others, by Keßler et al. [27] in respect to trust; by Arsanjani et al. [28] in respect to the quality of single contributors; by Rehrl et al. [29] in respect to contribution patterns; by Rehrl et al. [30] in respect to the motivations to contribute; and by Mooney et al. [9] in respect to the community consisting of individual contributors. Hashemi et al. [31] have assessed the logical consistency; Ballatore et al. [6], the conceptual quality; and Vandecasteele et al. [32] and Mooney et al. [13] have discussed the influence of the tagging process on data quality. Some intrinsic quality measures have been discussed by Mooney et al. [33] and Gröchenig et al. [11]. A general overview over methods and indicators to assess data quality of OSM has been provided by Barron et al. [34] and Senaratne et al. [7].

Despite the extensive literature body on the examination of OSM data and OSM mapping efforts as well as on data quality in respect to VGI and OSM in particular, technologies to support these aims have been discussed rarely. This is despite the fact that technology is an important tool for conducting research. This article aims at developing and improving the view on infrastructure to investigate OSM data and OSM mapping efforts in a scientific context. In particular, we present new infrastructure that

(1) fosters the analysis and the comprehension of different aspects of OSM,
(2) efficiently supports the analysis without overusing existing server infrastructure, and
(3) makes the analysis of the data reproducible.

In this article, we examine infrastructure to collect and mine data from and about the OSM project in order to explore and analyse them holistically, but the results can easily be transferred to other VGI projects. First, we discuss the OSM dataset and numerous additional data sources, among them the documentation of the folksonomy in the OSM wiki and statistical information about the OSM dataset ("Data sources" section). This discussion is followed by a general overview over the infrastructure that aims at achieving the above requirements 1 to 3 ("Conceptual overview" section). The infrastructure consists, in particular, of software to retrieve, merge, filter, and aggregate data from the data sources, which are then stored in a repository ("Technologies and resulting datasets" section). The resulting datasets of this repository can be used to examine several aspects of original data related to the OSM project in detail. As an example of such software, visualizations that use original data from the repository are discussed. These visualizations demonstrate, in particular, how the mapping behaviour and the folksonomy of OSM can be examined ("Results and discussion of use cases" section).

Implementation

A holistic understanding of the OSM data and its creation process can only be gained by examining several datasets. In the following sections, we discuss some important datasets by referring to their characteristics and their potential use, and we discuss how they can be mined and combined into new datasets.

Data sources

Several data sources are needed to get an overview over the community-driven data collection and maintenance process, which explains the heterogeneous nature of the data and eventually renders the datas' meaning. This section discusses the most important data sources related to OSM: the OSM database, which contains the collected data about the environment themselves; the OSM changesets, which contain information about which data of the database were edited, by whom, and under which circumstances; the OSM wiki, which can be seen as a documentation of the folksonomy of the OSM database and a documentation of the community activity; and some additional sources.

The OSM database. It is the aim of OSM to collect information about the environment, which can be represented in a map. This information is saved in the OSM database, which is accordingly a major component of OSM. The database is currently hosted at the Imperial College in London and technically maintained by the OpenStreetMap Foundation. Despite this technical custodianship, the data is created by a community-driven process. This process of data collection has a strong influence on how the environment is formally represented in the data, on which data are stored into the database, and thus on the quality of OSM data – data contained in the OSM database. While this community-driven process is not directly reflected by the data stored in the OSM database, information about this process can indirectly be concluded by understanding local differences between the data of different areas, by tracing the heterogeneity of the dataset, and by comparing the data with other information, for example, aerial images.

The OSM database is mostly accessed indirectly, for example, when using tiles to view a map, or when using a routing service or a geocoder based on OSM data. Different APIs have been used to access the data. Currently (in 2017), the *OSM API* and the *Overpass API* can be considered as de facto standards for accessing the data stored in the official instance of the OSM database, or a mirror. In addition, dumps of the OSM database are offered by *Planet OSM* (http://planet.openstreetmap.org). These world-wide database dumps are complemented by regional extracts, which are available, for example, on *Geofabrik downloads* (http://download.geofabrik.de).

OSM changesets. Changes in the OSM database are grouped into changesets, which are collections of changes in the OSM database. Such a change can be the addition or deletion of a node, a way, or a relation, that is, an OSM object, and it can be the addition, modification, or deletion of a tag of an object. Every changeset also contains information about whom did edit and in which timespan the edits have been performed. In addition, information about the used editor and the source, as well as other relevant information, are optionally part of the changeset. A changeset provides information about the circumstances under which data were modified, but potentially also about the intention of the user and the data quality of the added data. *Planet OSM* provides weekly dumps of the changesets.

The OSM wiki. Without a community of volunteers, VGI cannot exist. Data and information about how to interpret the data, in particular the tags (that is what we will call the folksonomy of OSM), would not be collected; the collected data would not be maintained; tools would not be developed; and the use of the data would not be promoted. These activities are documented in the OSM wiki (http://wiki.openstreetmap.org) as a means of communication in the community. When someone wants to participate in the community-driven process, for example, by contributing to the OSM database, he or she may refer to the wiki to understand the current progress, which help is needed, ongoing decision processes, and the current folksonomy. Many pages in the OSM wiki exist in

several languages, which we will call language versions in the following. The language versions are often direct translations, often shortened to the most important information or country specific information. The OSM wiki contains a documentation of the folksonomy of the OSM data, in particular of many concepts used for tagging OSM objects [8].

Additional sources. Many information around the OSM database, the OSM changesets, and the OSM wiki have been created as part of the community effort to create and use OSM data. Most of this additional information is statistical information created by aggregating existing information, but also complementing datasets exist. The software discussed in this article uses, supplementary to the data sources discussed above, statistical data from *OSMStats* (http://blackadder.dev.openstreetmap.org/OSMStats/) about the history of the OSM database. In addition, data about the usage of certain tags in the OSM database are retrieved from *Tag History* (http://taghistory.raifer.tech). Finally, the OSM specific data are complemented by data about the Earth, in particular about the boundaries of the continents. Such data are provided by the *Natural Earth* project (http://naturalearthdata.com). The files have even been converted to the GeoJSON formatted (http://github.com/nvkelso/natural-earth-vector).

Conceptual overview

VGI is created by a community process, and the data can only be interpreted and used if this community process is well understood. As the data and the community process are very heterogeneous, it is necessary to examine different aspects of this process in detail. A thorough understanding of these aspects renders the interpretation of the data possible, in particular, with regards to the heterogeneity of the data and their folksonomy. The technology that we present in this article aims at fostering such examination of the data by a larger number of scientists, which can be achieved by storing the data at a persistent location and making them available under an open license. The advantages of open software and open data in science have been examined by many studies, which can, for example, be seen by a review paper published by von Krogh et al. [35], but also in other articles, among them by Murray-Rust [36], and Uhlir and Schröder [37]. Among these advantages are the transparency of the methods and the results, and the traceability and potential replication of the results [38]. It can be hoped for that the sharing of the data facilitates their broader use.

The infrastructure discussed in this article structures the process of analyzing aspects of the OSM project into two steps (Fig. 1). In the first step, data from the data sources are retrieved, potentially merged, filtered, and finally aggregated, in order to provide datasets that highlight one particular aspect of the OSM project. The resulting data are published in a public Git (http://git-scm.com) repository. The data can hence be used by many researchers without the need to mine the original data sources again, which would otherwise strongly increase the load of the original servers. To be able to reuse the mined data is of particular importance for the OSM Wiki server, because a multiplicity of pages have to be retrieved several times otherwise. In a second step, the data from the repository can be analysed in detail. As the data and the software to produce the data are open, the analysis results can easily be replicated. In addition, the examination of exactly the same dataset (and hence the same aspect of OSM) increases the reproducibility of the analyses, which becomes of particular interest if analyses are conducted at different points in time.

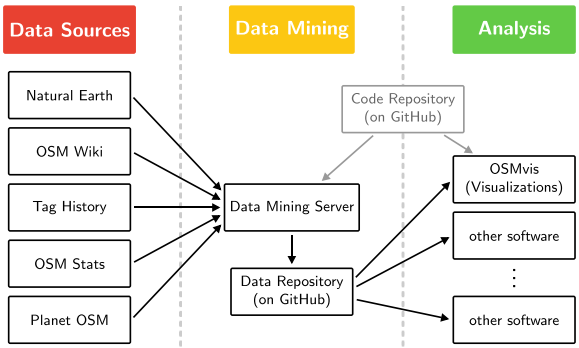

Fig. 1 Technical overview

The data sources have already been discussed in the last section. In the next section, we will discuss the data mining software and potential analyses in more detail.

Technologies and resulting datasets

Different data sources are mined in the context of the discussed infrastructure with the aim to retrieve information about several aspects of the OSM project, its database, and its community. An overview of the different services generating the new datasets an be found in Table 1.

For instance, one of the resulting datasets collects statistical information about all the words used in various languages in the documentation of the OSM Wiki. Statistical data about the history of the documentation of the OSM folksonomy are stored as another resulting dataset. Please note that these two resulting datasets both use the OSM wiki as their primary source, while combining the OSM wiki and the tag history into one dataset renders possible to relate the usage of tags in the OSM dataset and the documentation of the folksonomy (Table 1).

Table 1 Data mining

Data source	Resulting dataset	Programming language	Frequency	Description	OSMvis
Natural Earth	naturalearth_ne_110m_admin_0_countries.topojson	ECMAScript 6	Monthly	Compression by topojson*	OSM Changes Map
OSM wiki†	osm-tags-word-frequency-wiki.json	Haskell	Monthly	Collect statistical information about the words used in different language versions of the documentation of the folksonomy	OSM Tags Word Frequency Wiki
OSM wiki‡	osm-tags-history-wiki.json	Haskell	Monthly	Collect statistical information about the history of the documentation of the folksonomy	OSM Tags Wiki History
OSM wiki‡, Tag History	osm-tags-wiki-vs-osmdata.json	ECMAScript 6	Monthly	Relate the usage of tags in the OSM dataset and the documentation of the folksonomy	First Documentation in the Wiki vs. First Use in the Database
OSM Stats	osmstats.json	ECMAScript 6	Daily	Download only	OSM Changes per Day
Planet OSM	osm-node-changes-per-area.json	Haskell	Daily	Collect and aggregate statistical information about OSM changesets	OSM Changes Map

*http://github.com/topojson/topojson
†http://wiki.openstreetmap.org/wiki/Map_Features and linked pages, as well as corresponding pages in other languages
‡http://wiki.openstreetmap.org/wiki/Map_Features and linked pages, English language version only

Some of the proposed services mine the data on a daily basis, while others perform monthly. This is due to the fact that some data change very often and hence need to be analysed on a more regular basis. The OSM changeset is an example of such data. The information in the OSM wiki, in contrast, does not change on a daily or even weekly basis, and retrieving all sites of the OSM wiki is slow and increases the load of the corresponding OSM wiki server. Therefore, the service mines the OSM wiki on a monthly schedule. For the ease of use, a common format (Fig. 2) has been established, and the datasets are offered in a repository, from which users can easily download the requested datasets. Even different versions from different points in time are available. Among the advantages of such a repository are the provided metadata information, such as license information, a description of the data, and more importantly, the temporal information about when the mining was performed.

```
{
  "dataTimestamp":"2017-05-17T00:00:00Z",
  "dataDescription":"Statistics on node
     changes of OpenStreetMap data",
  "dataSource":"OpenStreetMap project,
     <a href=\"http://opendatacommons.org/
     licenses/odbl/\" target=\"_blank
     \">ODbL</a>",
  "dataUrl":"https://planet.openstreetmap
    .org/replication/day/000/001/708.osc.gz",
  ...
}
```

Fig. 2 Format of a resulting dataset file, exemplified at osm-node-changes-per-area.json

The data mining is performed by using different programming languages, depending on which language is most suitable. As a 'default', a recent variant of Javascript, ECMAScript, is used due to its widely adoption in the programming community. It is easy to read and write, and it handles JSON files efficiently. Other programming languages have their own beneficiaries, depending on the analysis that needs to be performed. Haskell, a purely functional programming language, is, for example, used due to its efficiency in complex data aggregation. The data format JSON has been chosen as the data exchange format due to its simplicity and feasibility to be used within web applications and web processing services.

The discussed infrastructure creates datasets that shed light on very different aspects of OSM. These datasets can be used to analyse the OSM project and the mapping behaviour of the contributors. Statistical analysis, visual analytics, and other approaches can be used to make sense of the data. As the datasets are publicly available, it is hoped for that several applications, websites, and services will take advantage of these datasets. In the next section, we discuss OSMvis, a project that visualizes these datasets with the aim of fostering explorative analyses.

Results and discussion of use cases

This section is dedicated to a critical discussion of the described server infrastructure and the resulting datasets. Hereby, we discuss the merits and limitations of the server infrastructure, in particular with respect to the long-term perspective. In subsequent subsections, we then discuss exemplary use cases. These use cases – a number of visualizations – are meant to demonstrate how the resulting data can be used in a simple and straight-forward way. These visualizations are part of the project *OSMvis*[1], a

collection of visualizations related to the OSM project. The aim of OSMvis is to generate insights about the OSM community and its mapping behaviour as well as about the OSM data themselves by a visual exploration of the datasets. The usefulness of such visualizations, in particular when combining different aspects of the data from different datasets, has been demonstrated by Mocnik et al. [8].

Merits and limitations

The discussed infrastructure aims at fostering reproducible research about VGI in general, and the OSM project in particular. It approaches the question of how to bundle the efforts of data mining, of understanding licenses, and of how to archive the mined data. Such advantages come, however, with limitations. This section aims at critically set the discussed infrastructure into context in order to recognize its merits and limitations.

Technical infrastructure. The data repository is one of the central aspects of the discussed infrastructure, which is used to store and archive the data. Currently, the data repository is hosted by the company GitHub. This is, however, not a real limitation, because GitHub is essentially based on Git (git-scm.com), an open source versioning control system. In case GitHub becomes unserviceable, for example, when it changes its terms of service, offers different services than before, or even shuts down completely, it can easily be replaced by other servers that run Git and offer similar graphical user interfaces. Hosted infrastructure like the one provided by GitHub usually restricts the size of the repository and the contained files. If the data mining process results in larger files, it might become necessary to host the files on similar systems that are self-hosted.

The data mining server is currently hosted at and maintained by Heidelberg University. While this ensures maintenance in the short and medium term, there might be the need of adapting this approach in future times. As an example, several data mining servers might be used. These servers might be hosted at different places and maintained by different organizations, despite their data all being included into the same repository. Even the repository itself could be maintained by different organizations. Another option would be to host the data in different repositories, which all conform to the same organizational principles but are yet hosted and maintained by different organizations. As the software is publicly available under an open license, there are no real obstacles for exploring and implementing such collaborative approaches of data mining and data repositories.

Data mining. The discussed infrastructure has different strengths. One of the most important strengths is to render possible the combination of different datasets, which all are about the OSM project and its manifold aspects. Datasets created by the community, for example, the ones described in the section "Data sources", serve for different purposes and are thus not always meant to be formally analyzed or combined. Accordingly, it needs some effort to relate the OSM wiki to data from the OSM database. In addition, meta data is often only published by the OSM community if it serves for the purpose of improving OSM data or renders new applications possible. The infrastructure discussed in this article aims, in contrast, for bundling data mining activities that focus on the scientific understanding of the entire OSM project and the principles behind VGI. Thereby, the infrastructure renders the following advantages: First, the mined datasets are available publicly without the need to mine the data on one's own. Secondly, the data is not only available at the point in time when being mined but also stored and made permanently available, which, thirdly, renders possible to run the same experiments and visualizations with the same results at a later point in time and thus affords reproducible research. These advantages are critical in case of VGI due to several reasons: First, very different aspects of VGI projects need to be incorporated in holistic approaches for examining and analyzing VGI projects. Secondly, these datasets about VGI projects are very different in their nature, and they expose a high variety and heterogeneity. Thirdly, the hardware resources used by VGI projects are often limited due to the voluntary nature of the project, potentially leading to a critical overuse when extensively performing data mining. For instance, the costs of crawling the OSM wiki as well as the server load of the corresponding server of the OSM foundation are reduced.

The infrastructure depends on many data sources (Fig. 1). These data sources have their own data formats, and they can be accessed in different ways, for example, as files on a http or a ftp server, as text on html websites, APIs, etc. Such data formats and interfaces usually change over time, and data sources might become inaccessible over time. The risk of such issues becomes larger the more data sources are involved. As the data mining is performed only once by the data mining server (and not by every scientist on their own), a quick adaption of the infrastructure to such changes is needed. On the other side, the common container format and the unchanging way of how the data is archived in the data repository allows to hide these issues from the scientist by removing the need to adapt the analyses and visualizations themselves to new data formats or methods of accessing the data. This, in turn, enables reproducible research, despite the very rapid evolvement and improvement of VGI tools and infrastructure, because current and historical data are both available and share a common data structure.

Reproducability. The infrastructure focusses on producing datasets that can easily be stored and archived. While this allows to perform analyses easily and in a reproducible way, extensive adaptions of the data mining process to the particular use case are only possible in parts: principle adaptions can be done in the data mining process, and minor adaptions can be implemented by filtering the dataset before analysing or visualizing it. This limitation is strongly related to the data being mined only once, which shifts the required processing power in the data mining step from the scientist who analyses or visualizes the OSM project for gaining general insights to the group or community that runs the data mining server. In addition, the dataset with the same name, which refers to the point in time and the used algorithms, always contains the same information – the data is only mined once and remains unchanged in the following.

Reproducibility is the effect of several factors, among them the public availability of the data under a shared license; software and workflows that are shared publicly; and the review of the software and workflows [39]. The proposed infrastructure is able to ensure the first two factors – public availability of the data, the software, and the workflows. Thereby, it seems to be important that the data are stored independently of any university or research group in order to guarantee the data to be available in the long term. Storing the data in a version control system prevents, in addition, single files from being modified or deleted: all additions, modifications, or deletions of the data become traceable. However, the infrastructure does not provide a formal review process. It is though hoped that the use of the data by different researchers leads to an informal review process, which is also common to many other collaborative projects hosted on GitHub or similar platforms. In summary, the proposed infrastructure is able to ensure many of the aspects necessary to foster reproducible research.

In the next sections, we discuss several use cases that demonstrate the use of the mined datasets. These use cases shall demonstrate how useful information can easily be visualized and how these visualizations adapt to the infrastructure.

Use case: how do contributors and their contributions spatially relate?

Users from all over the world contribute to OSM by describing their own environment, or the environment in other parts of the world. Before a user can contribute, he or she has to observe the environment in some way, for example, by walking around and visually observing the environment; by recording GPS tracks, or manually writing down GPS waypoints; by tracing and interpreting aerial images; or by incorporating additional information sources, such as information from websites, tablets and signs, and from encyclopaedias. The gained or collected knowledge is stored into the database by either manually editing a small number of objects in some editor; by importing GPS tracks; by tracing aerial images; or by importing datasets. These different methods of observing the environment and adding the data to the database are creating very different data – data with different spatial precision; data about different features; data referring to different concepts; and data of different quality.

Some of the methods to collect knowledge about the environment and incorporate it into the database are only applicable if one is at the place that is to be mapped, while other methods are even applicable if one is somewhere else on Earth. Also the personal motivation to contribute is influencing the way one contributes: the personal interest in having one's own environment mapped may lead to adding local knowledge, while the intention to provide humanitarian aid by helping to map the environment in a crisis region may lead to mapping regions far away. To get a first idea of which people map their own environment, which areas are primarily mapped by people from other parts of the Earth, and which motivation these people may have, we ask: *At which time of the day do users edit OSM data at which place on Earth?*

In Fig. 3, the changes of the OSM data on 2 April 2018 are depicted based on where they have been happening. A change which affects many objects is represented by a large disk, while a change affecting only some objects is represented by a small disk. The point in time from which the changes are depicted can, in the online version, be chosen by a time slider, which ranges from 0 a. m. to 12 p. m. When the position of the time slider is changed, the changes for the corresponding point in time are depicted. When the reference time, in this case UTC, is passing, the twilight zones move on Earth. The visualization depicts three lines which contain all places on Earth at which mean solar time equals 8 a.m., 12 a.m. , or 16 p.m. respectively. These lines are straight and not curved because the Mercator projection is used.

The visualization of the dataset **`osm-node-changes-per-area.json`** reveals several patterns in the collective behaviour of OSM users. Most contributions are about places in Europe, and other continents are much less edited. Changes of OSM data at most places occur mostly during day-time; before 8 a. m., only very little data are edited. This may facilitate the potential interpretation of a high number of OSM data contributions initiated by direct observations, in particular in Europe, where this day and night pattern is particularly distinctive. On other continents than Europe, this pattern is still present but not as distinctive, that is, the number of contributions about North America during night-times is about as high

Fig. 3 World map depicting changes in the OSM database. Every change of nodes in the data is depicted as a disk, and the number of changes at that particular point in place and time is encoded by the size of the disk. Changeset data © OpenStreetMap contributors (cf. www.openstreetmap.org/copyright). **a** 2018-04-02, 6:00 a.m. (UTC) **b** 2018-04-02, 10:45 a.m. (UTC)

as the number of contributions about Europe during night-times, but during day-times, there are many more contributions about Europe. This behaviour suggests that the number of users contributing to a foreign country or continent is higher in North America than in Europe. It can, in any case, be assumed that the relative number of visual observations leading to changes in OSM data is much higher in Europe compared to other continents. Extensive changes of a huge amount of data at neighbouring locations may indicate that the changes have been coordinated in some way. A possible interpretation of a high number of contributions about Mozambique during local night-time on 28 December 2016 may, for example, be the result of a mapping event dedicated to the very region. These conclusions demonstrate that the visualization is capable of efficiently communicating some important patterns in the collective user behaviour, explaining some aspects of how OSM data are added and modified.

Use case: temporal patterns in mapping activities

Mapping activities depend on various factors and vary thus over time. Good weather may facilitate outdoor mapping activities, bad weather the use of aerial images. Also the introduction of new tools as well as promoting activities have an influence on mapping activities. Important and global events potentially change the users habits, and may, in turn, also influence OSM mapping activities. While many potential influences on mapping activities exist, their superposition impedes the traceability of the actual influences. Temporal patterns in the number of changes and the number of new users may shed light on which influences have to be taken a closer look at. This leads to the question: *Which temporal patterns exist in the history of the changes of OSM data?*

A calendar heat map depicting the number of new OSM objects, users, and similar values is a visualization technique that is suitable to approach the question of temporal patterns. It can, for example, be applied to the dataset `osmstats.json`. Each cell of the raster which is used in the calendar heat map in Fig. 4 corresponds to a day in the history of OSM, and the cells are grouped by the temporal units of a week, a month, and a year. This arrangement of the cells facilitates the visual emergence of weekly, monthly, and annual patterns in the data. Such a technique of a calendar heat map was already used by Wickling et al. [40]. In contrast to the cartographic heat maps which were used by Trame et al. [15] and Roick et al. [16, 17], the calendar heat maps in Fig. 4 are not cartographic maps but general diagrams.

The visualization affirms some well known facts, for example, that there are far more nodes than ways and far more ways than relations, but it also reveals further insights about their temporal trend. Since nodes, ways, and relations, in short OSM objects, have been introduced – nodes and ways existed from the beginning, but relations have first been introduced on 7 October 2007 – their number is incessantly growing. Also this growth of the number of objects is increasing over time.

Several temporal patterns can be observed. The OSM data editing activity has been increased for one to three month long periods during the year. These periods of editing activity can often be observed in the second quarter of the year, but occur also in other ones. This trend has become more distinct over the years, in particular in 2013–2016. The annual pattern is superimposed by a weekly pattern: the activity of adding nodes and ways seems to be slightly less subject to variance on Saturdays and Sundays. During the annual phases of high activity, the activity at weekends increases less than at weekdays; and during phases of normal or little activity, the activity is, at least in some periods, slightly stronger at weekends. The weekly pattern as well as the daily fluctuation is much less visible as the annual pattern.

It is not always clear which reasons cause the phases at which there is only little activity compared to other periods. Even single days occur, at which the activity is strongly decreased. On 25 December 2015, for example, only very few nodes have been added to the OSM database, and Christmas might be a good explanation. Only very few new edges, compared to other days, have been added to the database on the weekend of the 13 and 14 August 2016, which was the middle weekend of the 2016 Summer Olympics (5–21 August). Different factors might be explanations for other dates, for example, bad weather conditions that affect an entire continent, or changes in the available editors.

The number of new users also exhibits an annual pattern of periods in which an increasing growth can be observed. A weekly pattern of less new users at weekends is present as well. There is, however, no observable effect of the number of new users on the number of new objects. In particular, the annual pattern of users is not correlated to the annual pattern of new objects. As can be seen by these examples, conclusions about the user behaviour and its influence on the mapping process can be drawn from the visualizations discussed in the previous and in this section.

Use case: visualizing the folksonomy of OSM

The locations of the nodes in the OSM database are complemented by semantic information. A collection of tags, each of them consisting of a key and a value, is stored to every node, way, or relation to capture the meaning of the object. These tags describe concepts, which often are fuzzy. Such fuzziness is an inherent property of geographic concepts as has been pointed out by Bennet [41].

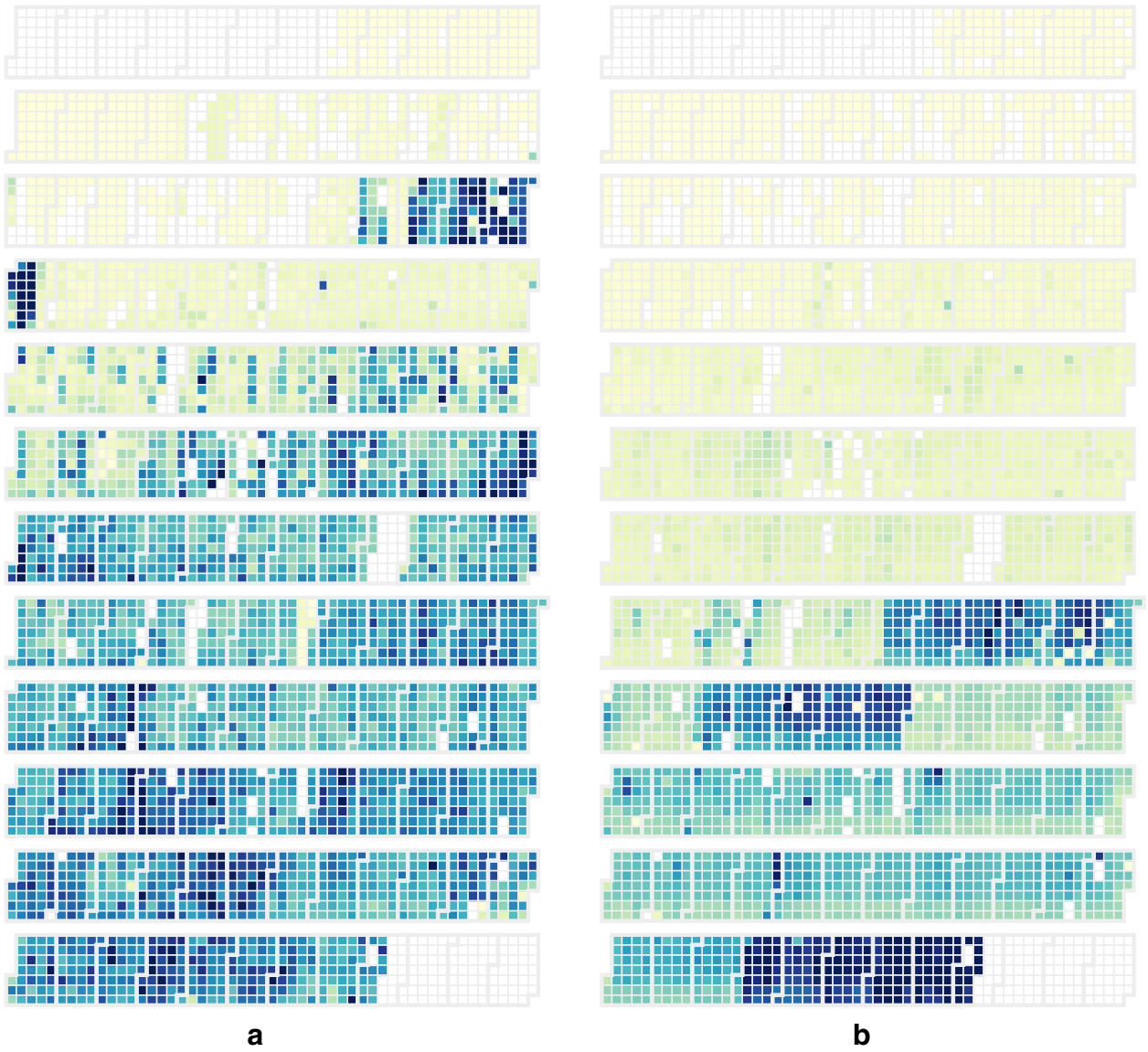

Fig. 4 Calendar heat map with temporally arranged cells. Yellow colour indicates a low number of nodes/users per day, blue colour indicates a high number. In case no data were available for a day, the cell is left white. Data from *OSMStats* © OpenStreetMap contributors (cf. www.openstreetmap.org/copyright). **a** New nodes per day **b** New users per day

Coexisting and fuzzy concepts of a forrest, for example, exist, because it is not clear whether a clearing is part of a forest, how many trees are needed to be called a forest, and how dense the trees need to be to be called a forest [41]. The concepts used for OSM tags may accordingly be fuzzy and also be changing over time. The description of the concepts, and also the consensus among the users of what is referred to by a certain concept, is of varying nature for different tags and for different language versions alike. The OSM folksonomy is, compared with an ontology, weak and not structured by the means of more complex relations.

The description of the concepts used for OSM data, that is, the description of the underlaying folksonomy, potentially has a large influence on which concepts and which tags are used when users contribute. If a concept is, for example, not documented, neither in the OSM wiki nor provided in the list of tags or keys of an editor, it may be used much less than well documented concepts. As the OSM wiki and tools to edit OSM data have a potentially strong influence on the creation process and the use of OSM data, it is of interest to understand the folksonomy and its description. In this section, visualization techniques are discussed to approach this comprehension in

respect to coexisting concepts, and the temporal evolution of the folksonomy.

Which types of terms are used to describe the folksonomy in the OSM wiki?

Different types of ontologies and concepts exist. Some concepts are descriptive, for example, by describing a zebra crossing as white stripes that are arranged in parallel and with equal distance in between, such that the optical illusion of alternating black and white stripes is created. Other concepts describe objects by the function they have in their environment, for example, a zebra crossing as a place where pedestrians can safely cross a street. The OSM wiki pages describe the concepts of the tags and values differently, by referring to very different features (functional and non-functional ones), but also with different precision, a differing number of examples, and with different length. All these properties of the description of the concepts influence how users collect and formalize information. We thus ask: *Which words play an important role in the descriptions of the folksonomy underlaying OSM data, and which differences exist between the different language versions?*

Word clouds are a visualization technique commonly used to get an overview of one or more texts. In Fig. 5, we use word clouds to analyse the dataset **`osm-tags-word-frequency-wiki.json`** with the aim of getting a first overview of which linguistic concepts are referred to in the descriptions of the concepts of OSM tags in the OSM wiki. The word cloud depict the word frequency in the descriptions of the tags in the language versions English and German respectively (Fig. 5a and c). Some words are used frequently on many pages, while other words are only used very frequently on one page. The word clouds in Fig. 5b and d again depict the word frequency, but the occurrence of a word is only counted once per tag, that is, once per key-value combination, to compensate for the effect of a word frequently occurring in the description of one tag only.

The visualizations reveal some differences between the language versions of the wiki. In the English version, the words "tagging" and "tagged" are very prominent, while in the German version, the word "mappen" (meaning "to map") is much more prominent. (The visualizations in Fig. 5 only depict words consisting of at least 5 characters to filter out many filler words, but even when showing words of length 3, the word "tag" is much more popular than the word "map" in the English version.) This difference shows that the English version refers more often to the tagging process itself, while, even when describing the tagging process, the German version refers to the entire mapping process, which is more than the tagging process. The high frequency of these terms is, among others, due to their frequent occurrence in headings. The English version refers frequently to the concept of a "place". In the German version, however, there does not seem to be used any corresponding concept. On the other side, the German version refers to "Öffnungszeiten" (meaning "opening hours"), "Telefonnummer" (meaning "telephone number"), and "Betreiber" (meaning "Operator") more often than the English version, which might indicate that formal and descriptive information is referred to more often.

The English version of the OSM wiki refers to many concepts which might allow for some fuzziness in the tagging process: "often", "usually", "possible", "typically", "recommended", "common", "indicate", and "sometimes". The word cloud depicting the German version, however, only contains the words "meist" (meaning "mostly") and "optional" (meaning "optional"/"optionally"). This might indicate that the existing fuzziness of the concepts and the observations is part of the descriptions in the English version. The reason behind these differences in word frequencies and its potential effect on the mapping process and the resulting data may be subject of future research.

The history of the folksonomy and its effect on data quality

There is no fixed taxonomy of OSM data, because many people contribute to the dataset in a joint effort. Accordingly, the taxonomy is changing over time. As the taxonomy is created by its use in the data rather than by documentation efforts, it is referred to as a folksonomy [8]. The quality of OSM data is strongly interlinked with the evolution of the folksonomy. An understanding of how the folksonomy evolves over time can thus provide useful insights in how the data quality changes over time. The OSMvis visualizations, which are able to provide such insights, have already been discussed in literature [8] and are thus rather shortly summarized in this section.

The dataset **`osm-tags-wiki-vs-osmdata.json`** contains for all relevant tags the points in time when tags have been used and when they have been documented. A corresponding visualization has been published by Mocnik et al. [8] (Fig. 6). It compares, for each tag, its 100th use in the OSM dataset to its first use in the documentation in the OSM wiki. As can be seen, most of the tags have been documented after their first usages, which corroborates that the taxonomy is a folksonomy.

The OSM folksonomy becomes richer and more fine-grained over time, its granularity becomes more uniform, and its scope was growing until recently. The conceptual quality of OSM has, for example been discussed by Ballatore et al. [6] and Ali et al. [42]. The discussed dimensions of conceptual quality include the accuracy, the granularity, the completeness, the consistency, the compliance, and

Fig. 5 Word clouds depicting the word frequency in the descriptions of tags in the OSM wiki. Only words of minimum length 5 are depicted. Data from the OSM wiki © OpenStreetMap contributors (cf. http://wiki.openstreetmap.org/wiki/Wiki_content_license). **a** English version of the OSM wiki **b** English version of the OSM wiki, occurence only counted once per tag **c** German version of the OSM wiki **d** German version of the OSM wiki, occurence only counted once per tag

the richness of the data. These dimensions are interrelated to temporal aspects. Data are added, and the folksonomy becomes more complete and more consistent over time. An examination of single tags is though needed in order to understand the folksonomy in greater detail, in particular, with respect to temporal changes of single tags. Mocnik et al. [8] have proposed a visualization, which interactively visualizes the tags contained in the dataset in more detail (Fig. 7).

Conclusion

Volunteered geographic information (VGI) differs from many other geographic information by being collected by a group of volunteers and by potentially being more heterogeneous. An important part of VGI is the typically coordinated effort to create and maintain the data. The interpretation and analysis of VGI data is thus only possible when considering the social process that leads to their creation. Typically, VGI data are studied without

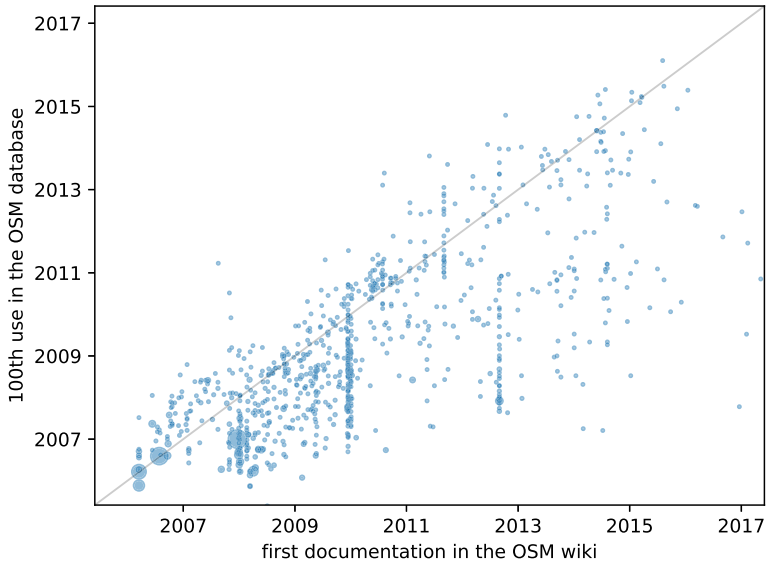

Fig. 6 Comparison of the 100th use of a tag in the OSM database and its first documentation in the OSM wiki. Each blue disk represents a tag. The size of the disk reflects how frequently the tag is used in the OSM database. Only tags that are used at least 1,000 times in the data and that are documented in the OSM wiki are included, tags with value **"*"** are excluded. Data from the OSM database/wiki © OpenStreetMap contributors (cf. http://openstreetmap.org/copyright); image and caption by Mocnik et al. [8], CC BY 4.0

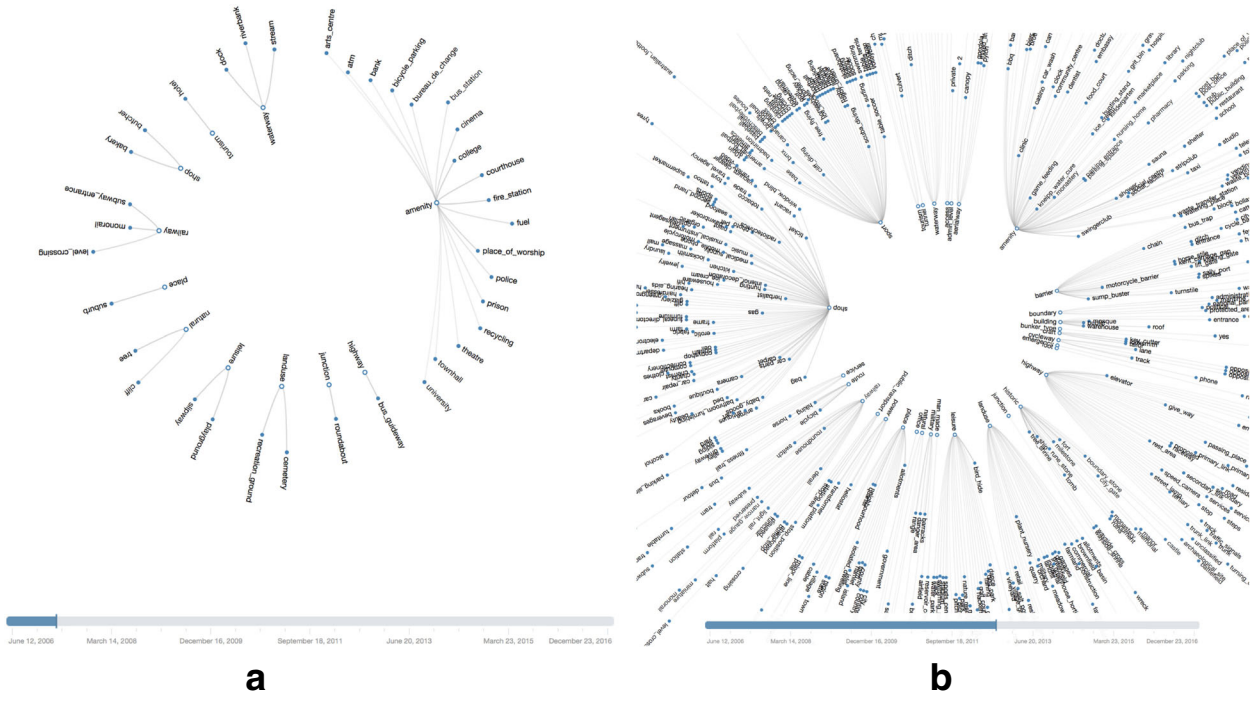

Fig. 7 Visualization technique for the documentation of the folksonomy in the OSM wiki. The nodes of the inner circle refer to the documented keys, while the nodes around, to the corresponding values. The longer a value exists, the more it moves away from the origin. Data from the OSM wiki © OpenStreetMap contributors (cf. http://wiki.openstreetmap.org/wiki/Wiki_content_license). Image and caption by Mocnik et al. [8], CC BY 4.0. **a** 2007 **b** 2012

or in combination with their creation process. A holistic understanding is though to be established and data quality can, as a result, not be examined thoroughly, because a more complete understanding of which factors influence the emergence of the data is still missing.

In this article, we have described an infrastructure for collecting and mining data about the OSM project. This infrastructure does not restrict to certain aspects, like the OSM dataset, but aims at collecting many available information. The resulting datasets are offered under an open license in order to foster their analysis and use. Each dataset also contains a copyright notice and a link to the original datasets that were used for data mining, as well as a description. It is hoped for that the collection of datasets fosters projects that aim at a holistic understanding of the OSM project rather than an understanding of the OSM dataset only.

The discussed visualizations use datasets that were created by the described infrastructure. They offer insights into basic but yet important aspects of the OSM project and offer a way to explore the OSM project from different points of view and in a more holistical way. The visualizations have, in parts, been discussed in the context of data quality, but a more detailed analysis may be needed to gain a deeper understanding of the influence on data quality. In particular, more language versions of the OSM wiki may be compared and the influence of the differing descriptions of concepts in the language versions on the actual data may be explored in greater detail. A visualization of correlations between the modification and introduction of new concepts, and properties of the data at the corresponding points in time, may provide further insights about how data quality is influenced by the folksonomy and its granularity in particular.

Many more aspects of OSM-related data can be visualized by using methods from the field of information visualization. The number of new objects may, for example, be visually compared to large events and weather data. Also nodes and ways in a certain area may be compared with nodes and ways in another area, by visualizing their possibly systematically differing properties, for example, by the use of parallel coordinates. Data related to OSM may also be a good starting point for developing new visualization techniques that incorporate spatial and non-spatial data by combining maps with non-spatial visualizations. The described software and the resulting datasets can serve as a starting point for these and further research directions.

Availability and requirements

Project name: OSMvis Data
Project home page: http://github.com/giscience/osm-vis-data
Operating system: Unix-like systems
Programming language: Haskell, ECMAScript 6, bash scripts
Other requirements: none
License: GPL-3
Any restrictions to use by non-academics: see license

Endnote

[1] http://osm-vis.geog.uni-heidelberg.de

Abbreviations

VGI: Volunteered geographic information; OSM: OpenStreetMap

Acknowledgments

The authors would like to thank Martin Raifer for his helpful comments on the manuscript.

Funding

This research has been supported by the DFG project *A framework for measuring the fitness for purpose of OpenStreetMap data based on intrinsic quality indicators* (FA 1189/3-1).

Availability of data and materials

– code repository for data mining and for visualization: http://github.com/mocnik-science/osm-vis
– data repository: http://github.com/giscience/osm-vis-data
– visualization: http://osm-vis.geog.uni-heidelberg.de.

Authors' contributions

FBM developed the visualizations. AZ and FBM jointly developed the idea of the visualization presented in Fig. 7. FBM and AM wrote the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Received: 26 November 2017 Accepted: 17 March 2018
Published online: 28 May 2018

References

1. Goodchild MF. Citizens as sensors: the world of volunteered geography. GeoJournal. 2007;69:211–21.
2. See L, Mooney P, Foody GM, Bastin L, Comber A, Estima J, Fritz S, Kerle N, Jiang B, Laakso M, Liu H-Y, Milčinski G, Nikšič M, Painho M, Pődör A, Olteanu-Raimond A-M, Rutzinger M. Crowdsourcing, citizen science or volunteered geographic information? The current state of crowdsourced geographic information. ISPRS Int J Geo-Inf. 2016; 5(5). https://doi.org/10.3390/ijgi5050055.
3. Haklay M, Weber P. OpenStreetMap: user-generated street maps. IEEE Pervasive Comput. 2008;7(4):12–8. https://doi.org/10.1109/MPRV.2008.80.
4. Mooney P, Minghini M. A review of OpenStreetMap data. In: Foody GM, See L, Fritz S, Mooney P, Olteanu-Raimond A-M, Fonte CC, Antoniou V, editors. Mapping and the Citizen Sensor. London: Ubiquity Press. 2017. p. 37–59. https://doi.org/10.5334/bbf.c.
5. Mocnik F-B, Mobasheri A, Griesbaum L, Eckle M, Jacobs C, Klonner C. A grounding-based ontology of data quality measures. J Spat Inf Sci. 2018; 16. https://doi.org/10.5311/JOSIS.2018.16.360.
6. Ballatore A, Zipf A. A conceptual quality framework for volunteered geographic information. In: Proceedings of the 12th Conference on Spatial Information Theory (COSIT). 2015. p. 89–107. https://doi.org/10.1007/978-3-319-23374-1_5.

7. Senaratne H, Mobasheri A, Ali AL, Capineri C, Haklay M. A review of volunteered geographic information quality assessment methods. Int J Geogr Inf Sci. 2017;31(1):139–67. https://doi.org/10.1080/13658816.2016.1189556.
8. Mocnik F-B, Zipf A, Raifer M. The OpenStreetMap folksonomy and its evolution. Geo-Spat Inform Sci. 2017;20(3):219–30. https://doi.org/10.1080/10095020.2017.1368193.
9. Mooney P, Corcoran P. How social is OpenStreetMap? In: Proceedings of the 15th AGILE Conference on Geographic Information Science. 2012. p. 282–287.
10. Ballatore A, Mooney P. Conceptualising the geographic world: the cimensions of negotiation in crowdsourced cartography. Int J Geogr Inf Sci. 2015;29(12):2310–27. https://doi.org/10.1080/13658816.2015.1076825.
11. Gröchenig S, Brunauer R, Rehrl K. Estimating completeness of VGI datasets by analyzing community activity over time periods. In: Proceedings of the 17th AGILE Conference on Geographic Information Science. 2014. p. 3–18. https://doi.org/10.1007/978-3-319-03611-3_1.
12. Neis P, Zielstra D, Zipf A. Comparison of volunteered geographic information data contributions and community development for selected world regions. Future Internet. 2013;5(2):282–300. https://doi.org/10.3390/fi5020282.
13. Mooney P, Corcoran P. The annotation process in OpenStreetMap. Transactions in GIS. 2012;16(4):561–79. https://doi.org/10.1111/j.1467-9671.2012.01306.x.
14. Mooney P, Corcoran P. Who are the contributors to OpenStreetMap and what do they do? In: Proceedings of the 20th Annual GIS Research UK (GISRUK). 2012.
15. Trame J, Keßler C. Exploring the lineage of volunteered geographic information with heat maps. GeoViz. 2011.
16. Roick O, Hagenauer J, Zipf A. OSMatrix – grid-based analysis and visualization of OpenStreetMap. In: Proceedings of the 1st European State of the Map Conference (SOTM-EU). 2011.
17. Roick O, Loos L, Zipf A. A technical framework for visualizing spatio-temporal quality metrics of volunteered geographic information. In: Proceedings of the Conference Geoinformatik. 2012.
18. Haklay M. How good is volunteered goegraphical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. Environ Plan B. 2010;37(4):682–703. https://doi.org/10.1068/b35097.
19. Neis P, Zielstra D, Zipf A. The street network evolution of crowdsourced maps: OpenStreetMap in Germany 2007–2011. Future Internet. 2012;4:1–21. https://doi.org/10.3390/fi4010001.
20. Girres J-F, Touya G. Quality assessment of the French OpenStreetMap dataset. Trans GIS. 2010;14(4):435–59. https://doi.org/10.1111/j.1467-9671.2010.01203.x.
21. Camboim SP, Bravo JVM, Sluter CR. An investigation into the completeness of, and the updates to, OpenStreetMap data in a heterogeneous area in Brazil. ISPRS Int J Geo-Inf. 2015;4:1366–88. https://doi.org/10.3390/ijgi4031366.
22. Forghani M, Delavar MR. A quality study of the OpenStreetMap dataset for Tehran. ISPRS Int J Geo-Inf. 2014;3(2):750–63. https://doi.org/10.3390/ijgi3020750.
23. Eckle M, de Albuquerque JP. Quality assessment of remote mapping in OpenStreetMap for disaster management purposes. In: Proceedings of the 12th International Conference on Information Systems for Crisis Response and Management (ISCRAM). 2015.
24. Zipf A, Mobasheri A, Rousell A, Hahmann S. Crowdsourcing for individual needs – the case of routing and navigation for mobility-impaired persons. In: Capineri C, Haklay M, Huang H, Antoniou V, Kettunen J, Ostermann F, Purves R, editors. European Handbook of Crowdsourced Geographic Information. London: Ubiquity Press; 2016. p. 325–337. https://doi.org/10.5334/bax.x.
25. Rousell A, Zipf A. Towards a landmark-based pedestrian navigation service using OSM data. ISPRS Int J Geo-Inf. 2017;6(64). https://doi.org/10.3390/ijgi6030064.
26. Bakillah M, Liang S, Mobasheri A, Arsanjani JJ, Zipf A. Fine-resolution population mapping using OpenStreetMap points-of-interest. Int J Geogr Inf Sci. 2014;28(9):1940–63. https://doi.org/10.1080/13658816.2014.909045.
27. Keßler C, de Groot RTA. Trust as a proxy measure for the quality of volunteered geographic information in the case of OpenStreetMap. In: Proceedings of the 16th AGILE Conference on Geographic Information Science. 2013. p. 21–37. https://doi.org/10.1007/978-3-319-00615-4_2.
28. Arsanjani JJ, Barron C, Bakillah M, Helbich M. Assessing the quality of OpenStreetMap contributors together with their contributions. In: Proceedings of the 16th AGILE Conference on Geographic Information Science. 2013.
29. Rehrl K, Gröchenig S, Hochmair H, Leitinger S, Steinmann R, Wagner A. A conceptual model for analyzing contribution patterns in the context of VGI. In: Krisp JM, editor. Progress in Location-Based Services. Berlin: Springer; 2013. p. 373–388. https://doi.org/10.1007/978-3-642-34203-5_21.
30. Rehrl K, Gröchenig S. A framework for data-centric analysis of mapping activity in the context of volunteered geographic information. ISPRS Int J Geo-Inf. 2016;5(37). https://doi.org/10.3390/ijgi5030037.
31. Hashemi P, Abbaspour RA. Assessment of logical consistency in OpenStreetMap based on the spatial similarity concept. In: Arsanjani JJ, Zipf A, Mooney P, Helbich M, editors. OpenStreetMap in GIScience. Experiences, Research, and Applications. Heidelberg: Springer; 2015. p. 19–36. https://doi.org/10.1007/978-3-319-14280-7_2.
32. Vandecasteele A, Devillers R. Improving volunteered geographic information quality using a tag recommender system: the case of OpenStreetMap. In: Arsanjani JJ, Zipf A, Mooney P, Helbich M, editors. OpenStreetMap in GIScience. Experiences, Research, and Applications. Heidelberg: Springer; 2015. p. 59–80. https://doi.org/10.1007/978-3-319-14280-7_4.
33. Mooney P, Corcoran P, Winstanley AC. Towards quality metrics for OpenStreetMap. In: Proceedings of the 18th ACM SIGSPATIAL International Symposium on Advances in Geographic Information Systems. 2010. p. 514–517.
34. Barron C, Neis P, Zipf A. A comprehensive framework for intrinsic OpenStreetMap quality analysis. Trans GIS. 2014;18(6):877–95. https://doi.org/10.1111/tgis.12073.
35. von Krogh G, von Hippel E. The promise of research on open source software. Manag Sci. 2006;52(7):975–83. https://doi.org/10.1287/mnsc.1060.0560.
36. Murray-Rust P. Open data in science. Ser Rev. 2008;34(1):52–64. https://doi.org/10.1080/00987913.2008.10765152.
37. Uhlir PF, Schröder P. Open data for global science. Data Sci J. 2007;6:36–53. https://doi.org/10.2481/dsj.6.OD36.
38. Nosek BA, Alter G, Banks GC, Boorsbom D, Bowman SD, Breckler SJ, Buck S, Chambers CD, Chin G, Christensen G, Contestabile M, A D, Eich E, Freese J, Glennerster R, Goroff D, Green DP, Hesse B, Humphreys M, Ishiyama J, Karlan D, Kraut A, Lupia A, Mabry P, Madon T, Malhotra N, Mayo-Wilson E, McNutt M, Miguel E, Levy Paluck E, Simonsohn U, Soderberg C, Spellman BA, Turitto J, VandenBos G, Vazire S, Wagenmakers EJ, Wilson R, Yarkoni T. Promoting an open research culture. Science. 2015;348(6242):1422–5. https://doi.org/10.1126/science.aab2374.
39. Singleton AD, Spielman S, Brunsdon C. Establishing a framework for open geographic information science. Int J Geogr Inf Sci. 2016;30(8):1507–21. https://doi.org/10.1080/13658816.2015.1137579.
40. Wicklin R, Allison R. Congestion in the sky. Visualizing domestic airline traffic with SAS software. ASA Statistical Computing and Graphics Data Expo. 2009.
41. Bennett B. What is a forest? On the vagueness of certain geographic concepts. Topoi. 2001;20(2):189–201.
42. Ali AL, Sirilertworakul N, Zipf A, Mobasheri A. Guided classification system for conceptual overlapping classes in OpenStreetMap. ISPRS Int J Geo-Inf. 2016;5(87). https://doi.org/10.3390/ijgi5060087.