Public-OSM Partnership (POP) – A Pilot Study

Stefan Keller and Kailing Peng

March 17, 2022 (Version 13)

Impressum

Authors

Stefan Keller and Kailing Peng, Institute for Software, OST Eastern Switzerland University of Applied Sciences, Campus Rapperswil

Contact

Prof. Stefan Keller, Institute for Software OST, www.ost.ch/ifs , stefan.keller@ost.ch

Project and study members

- Stefan Keller, Head of Institute for Software, OST Eastern Switzerland University of Applied Sciences, Campus Rapperswil
- Kailing Peng, Assistant at Institute for Software, OST Eastern Switzerland University of Applied Sciences, Campus Rapperswil
- Christian Nüssli, Applikationsverantwortlicher ELZ, Protection & Rescue, City of Zurich
- Stefan Oderbolz, Open Data Specialist, Open Data Zurich, City of Zurich
- Romedi Filli, Head of Geoinformatics, Office for Geoinformation, Canton Schaffhausen
- Simone Würsch, Fachstellenleiterin, Office for Spatial Development, Canton Zurich
- Stéphane Malta e Sousa, Project and Development Manager, Department of Territorial Development, Canton of Neuchâtel

Acknowledgement

This document with project number 20-17 was produced in 2021 thanks to the kind support of the dedicated funds of the National Spatial Data Infrastructure (Nationalen Geodateninfrastruktur, NGDI). Many thanks go also to all project and study members.

Citation proposal (APA style)

Keller Stefan & Peng Kailing (2022) Public-OSM Partnership (POP) – A Pilot Study. A technical paper from the Institute for Software, OST Eastern Switzerland University of Applied Sciences. Webaccess: http://eprints.ost.ch/ (tba.)

Table of Contents

1. Overview and Objectives

OpenStreetMap (OSM) is a free, editable map of the whole world that is being built by volunteers largely from scratch and released with an open-content license. Over last decades, OSM has grew up to the largest platform for geometric data worldwide. Numerous applications have explored creative ways to take advantage of its data, which means the "fitness-for-use" quality and software basis (tools, libraries) are correspondingly mature and advanced.

However, the use of OSM in government agencies is still new. Similar to the Public-Private Partnership, this Public-OSM Partnership (POP) is a partnership of community representatives working together with citizens and association representatives of the crowdsourcing project OpenStreetMap (OSM) to fulfill a public task by coordinating and optimizing their resources and activities.

To achieve this study goal, our study members need to know how data is processed in government well, also have certain mapping experience on OSM, or keen to exploring the potential of OSM in scope of government use. The two main target groups of this pilot study are defined by the main data owners: blue-light organizations (Schutz & Rettung Stadt Zürich) and GIS and cadastre offices (GIS office Kt. Zürich).

The objectives are

1. Clarification of important licensing issues when monitoring, comparing or transferring OSM data (with ODbL licence)
2. Identification of typical processes and evaluation of software tools for monitoring OSM data.
3. Identification of typical processes and evaluation of software tools for the alignment and quality assurance of OSM data.
4. Identification of typical processes and evaluation of Software tools for the transfer of OSM data.

2. OpenStreetMap (OSM)

2.1 About OSM

OpenStreetMap (OSM) was founded in 2004, encouraging the growth, development and distribution of free geospatial data and provide geospatial data for anybody to use and share. As demonstrations of community mapping effort, OSM progressed quite spectacularly over the years, achieving many mapping milestones.

The OpenStreetMap License allows free access to map images and all of its underlying map data. It uses wiki-style collaborative editing software to support its development, which means that OSM maps will always keep growing bigger and better. Extensive software (leaflet, overpass API, Overpass Turbo) and map editing tools (iD, JOSM, StreetComplete) development work is taking this mapping project in many different directions. Individuals, governments and commercial companies have already begun putting this data to use, and in many countries, for many uses, OSM is a viable alternative to other map providers and promote new and interesting uses of geo-data.

2.2 The ODbL and other Data Licenses

The “sine qua non” of any data incorporated into OpenStreetMap is that the data’s license must be compatible. OpenStreetMap data is distributed under the Open Database License (ODbL). All contributors also agree to a set of Contributor Terms. All contributed data must be compatible with these. In practice, this means that the data must be licensed without restrictions on reuse except for some attribution requirements (see below).

Your data may already be licensed under a popular data license. Some licenses are already compatible with OSM’s Open Database License, meaning that you can legally incorporate the data into OSM without further waivers. Others may require you to provide a waiver of incompatible terms. The most common licenses are:

- Creative Commons Attribution (CC-BY 4.0): This license is almost compatible with OSM licensing, but there is one small technical incompatibility, and attribution needs to be clarified. In practice, this means your organization should provide a waiver for use of the CC-BY-licensed data in OSM. See https://blog.openstreetmap.org/2017/03/17/use-of-ccby-data/ for details.

- Creative Commons Zero (CC0) / ODC Public Domain Dedication & License / US Federal Public Domain: Data licensed this way can be included in OSM without further license considerations.

- Open Database License (ODbL): This is the license used by OSM. As such it is compatible, but you should document that your data was initially made available under the ODbL.

- Open Government License UK: OGL-licensed data can be included in OSM, subject to attribution and derived data issues (see below). You should document that your data was initially made available under the OGL.

- Other Creative Commons and ‘open data’ licenses: These licenses are unlikely to be compatible with OSM’s licensing. To make your data admissible for OSM, you should ‘dual-license’ it – i.e., also offer it under one of the above compatible licenses.

Common Licensing Issues:

Many licenses require attribution of the original data source. OSM itself provides attribution of third-party sources on a wiki page (http://wiki.osm.org/Contributors) and, for the very largest sources, at https://www.openstreetmap.org/copyright.

It is impractical to attribute the thousands of data sources on the map itself. As such, if your organization requires that attribution is always visible when interacting with the data, it will not be suitable for OSM. You may need to consider an explicit waiver to take account of this point, particularly if your data is currently licensed under a Creative Commons Attribution license.

Several ‘open’ licenses, including some in the Creative Commons family, have a prohibition on commercial use. This prohibition is incompatible with OSM. Although the OSM Foundation itself is a non-profit organization, OSM data is used by many commercial organizations.

Some ‘open’ licenses have additional clauses that require the user to indemnify the data provider, or to refrain from derogatory treatment, and so on. These additional clauses are incompatible with OSM.

The OSM Foundation is not allowed to change the ODbL without explicit consent of the contributors – whereas it’s obvious that getting consent from about 8 million contributors is difficult to achieve.

Future Liability:

Once the data is in OSM, you are no longer liable for it. OSM's license and Contributor Terms expressly disclaim contributors' liability and offer no warranty. You will therefore not be liable for any changes that OSM contributors may make to the data.

You should be confident that there are no additional rights in the data other than those held by your organization. This is most likely to be an issue if the data was created with reference to a copyrighted map; by looking up addresses from a copyrighted database; or by using a commercial dataset.

For example, data created by tracing over Google Maps, or by using the Google Maps API, is not acceptable for use in OSM. You would need to recreate it using OSM as a base map. Similarly, data built on a commercial dataset such as those offered by TomTom or HERE will not be acceptable.

UK public sector organizations may have data created with reference to maps from the Ordnance Survey. The OS has a special 'Presumption to Publish' process for these cases. You should follow the guide at https://www.ordnancesurvey.co.uk/business-government/licensing-agreements/presumption-to-publish-form .

2.3 A Non Inaugurated View on OSM

In order to get an outside view n OSM in the preliminary stage of the study, we carried out interview with some project members. We used the SWOT (Strengths, Weaknesses, Opportunities, and Threats) structure fpr structuring the interviews.

Note: The POP "idea" is a mutual partnership of government and OSM. However, this study focusses on the benefits of OSM for government, so it is about how to integrate OSM data for use by the government. The interview participants are representatives of GIS and Open Data experts from Swiss government originations. Many interviewees are new to OSM, the insight the interviewees shared are based on their personal or job experiences that neither is declaimed as factual events nor represents any conclusion from this study. Opinions or wishes which did not align with this study focus have been separated and are answered in chapter 2.4 "Integrating OGD into OSM".

As a board member of Swiss OpenStreetMap Association and employee at IFS Institute for Software, Raphael Das Gupta came from an active mapper's view and commented on the SWOT points raised above, and they are appended to each session. We would also attach the interview script of Raphael Das Gupta as Appendix A.

Perceived strengths:

- Great scalability and worldwide free access. Open source under free licenses. Immediate update results.
- The platform is not capital-driven and privacy preserved. It helps with bias correction.
- Numerous Points of Interests (POI) with rich geo-data and vector tiles can be extracted as different layers for fast combination or filtered for different purpose of usages.
- A good supplement to official data, especially for non-mandatory to-be-collected datasets or for some simplified datasets.
- The world's largest GIS community is supported by a mass of passionate and devoted mappers.

Comments on strengths:

OSM has indeed proven its scalability, regarding various aspects. The flexibility of its "schema-less" data schema helped it to expand into any topic mappable by vector geodata, far beyond just the initial name-sake streets. It has also scaled in terms of sheer data amounts, registered users and active mappers.

OSM data is indeed available under a free license (ODbL, see https://osm.org/copyright). But as the core "product" of OSM is that data, it is considered an *open data* project, not a free / libre / open-source software (FLOSS) project, even though most of the community-created software ecosystem around it is indeed FLOSS, too.

While the project itself is mostly community-driven, neither the license terms nor the community are in general averse to (though sometimes sceptical of) commercial use of OSM data or commercial contributions to it, as long as they stick to the license terms and other rules and guidelines. Commercial entities regularly interacting with OSM are even considered part of the OSM community.

But OSM indeed doesn't suffer from biases that projects controlled by a single commercial entity often do, that may only map those areas well, where they expect enough customer demand. And the perception that contributing to OSM isn't just free work helping a single for-profit company, but a service to the public, surely has helped growth and adoption.

Perceived weaknesses:

- Data accuracy and reliability are crucial for government's consideration. The trust for crowd-sourced data like OSM needs to be proved and gained.
- The data model lacks standards, Data is not homogeneous enough.
- Data vandalism caused by far no big damage, but we cannot prevent it from happening.
- The data consumption from OSM requires a high level of pre-knowledge, to extract specific dataset from OSM and sort them out for proper use is quite challenging.

Comments on weaknesses:

I'd claim that the OSM community has already proven its trustworthiness. That sometimes inaccurate or false data seeps into OSM is unavoidable with OSM, but it is usually quickly spotted and rectified, especially for "important" features.

The various ways of mapping the same kind of features or similar "facts" can indeed be a challenge or even a problem for consumers of OSM data, especially for further processing. But that isn't an issue unique to OSM: If governmental data (which are often limit to the respective individual country) without prior unification is stitched together for global coverage, I wouldn't expect it to be more homogeneous than OSM.

As already written, data vandalism isn't usually a big problem to OSM. Depending on how (and how often) they ingest OSM data, it can though be a problem for certain OSM data consumers.

It's true that knowledge and experience are needed to use OSM data effectively. This is no different than with other geodata, but the problem here is that expertise acquired with other kinds of geodata can only partially be applied to working with OSM due different concepts.

Perceived opportunities:

- Sharing open data makes data collection more efficient. It is meaningful to avoid the same work like cadastral surveys being done twice by OSM and governments.
- From past experience, the local Swiss community is glad to offer their assistance and cooperation when they are better informed and involved for project ideas.
- For the government, it is an effective way to raise the awareness and gain the trust of local talents through projects concerning citizen's daily live. It strengthens the authority-and-citizen bond.
- Open data is a worldwide trend, governments in many countries are increasingly investing in this movement and OSM is winning with its unique edges in this field.

Comments on opportunities: Agreed.

Perceived threats:

- Potential usage will be a mixing of official and crowdsourced data. Data user might not recognize the difference.
- To discover a feasible approach to ensure a smooth workflow between two parties (OSM and authorities) could be tricky when we take cultural barriers and mentality differences into account.
- The platform is non-commercial owned but paid contributions are happening and it may lead the users to biased information, instead of the fairest for their best interest.
- The value of the data platform is built upon volunteering mappers. There is no guarantee that the platform will always remain active and trustable.

Comments on threats:

I'm not sure how much of a problem it would be if users cannot readily recognize what parts of the data are "official" (probably: surveyed and provided by the government itself of by its contractors) vs. crowdsourced in usages where both are combined. There are of course use-cases where one has to know for sure which data is authoritative. But I'd imagine that in these use-cases, only authoritative data would be used anyway (which would probably be only a subset of "official" data) and not a mixture of authoritative and non-authoritative data, whether "official", crowd-sourced of other origin.

While the different approaches to get things done (or let them happen) between government offices and the OSM communities and the individuals involved therein can indeed be a challenge, I believe this is one that can be overcome when there is willingness and the required openness on both sides. A hindrance might be business models imposed on some offices by governments, budgets or the state at large, which can put up incentives against open data and against cooperation with (perceived) "competitors" such as OSM.

OSM isn't completely unbiased. No data set is. How its bias(es) compared to different data sets or data sources is probably difficult to evaluate. While paid contribution could lead to additional biases that would be seen as problematic by the OSM community itself, I don't see why these biases in particular would be a more relevant problem for governments using or involved in OSM than every bias is for them.

While nobody can guarantee that OSM will be available (and continue to be updated and trustworthy) forever, I'd consider it "too big to fail" now; or rather, too important for enough parties to suddenly disappear. If the OpenStreetMap Foundation (OSFM) shepherding the project in a quite hands-off / let-the-community-figure-things-out way were to unexpectedly implode, the free license would allow other parties to continue the project without it, and there would probably be enough incentive to do so, because so many businesses, NGOs and governments already depend to various degrees on the project.

Most common perceived interests and needs:

- To collect specific dataset which is not in authorative data bank like detailed building, infrastructure, routing, ground cover information and labels.
- Legal clarification regarding the mixed-use of data, e.g. to combine official data of Switzerland and OSM for Germany to get a uniformed map.
- combine this theoretical study with the practice to discover more potential fit-for-use cases on government scope and some convincing prototypes that can demonstrate a reliable working method with the community

2.4 Integrating OGD into OSM

During this study period, members showed interest in how the data sharing in both directions works, to form a sustainable workflow. Since this study is focused on integrating OSM data for government use, with consent of the authors who carried out a study named "Data in OpenStreetMap integrating - a guideline for data owners" (translated) (Hitz & Stürmer 2021), we summarized the recommendations from this study report to cater this interest. See also Fairhurst, R. (2020).

This is entirely in keeping with the basic mission of government agencies and federally related businesses to add value for the public. Publishing data can generate immediate benefits by enabling a specific use case to be realized. But even without a specific idea of how it will be used, it can make sense to include data in OSM. The idea of Open Data thrives, among other things, on data owners inviting interested members of the public to use the data they make available in new ways.

1) "Data owners should publish their data on OpenStreetMap to increase the visibility and usability of the data."
2) "Data owners should publish their data on OpenStreetMap even if there is no concrete idea about a possible benefit. This benefit will be found by the interested public."
3) "Data owners should choose a license for their data that enables the use for OpenStreetMap or set terms of use that explicitly consider the use of the data for OpenStreetMap. The easiest way to achieve this is to choose the Open Database License (ODbL 1.0)."
4) "Data owners should contact the Swiss OpenStreetMap Association at an early stage to plan the procedure regarding data integration in consultation with the OpenStreetMap community."
5) "Data owners should follow a managed approach for the integration of data into OpenStreetMap."
6) "Data owners should translate their own (identification) reference system into OSM tags and, if necessary, define new keys in consultation with the Swiss OpenStreetMap Association."
7) "Data owners should regularly monitor and update their data in/to OpenStreetMap and define a process for this."
8) "Data owners should build up internal competencies for OpenStreetMap and designate a central point of contact for OpenStreetMap issues both externally and internally."

Some of those recommendation are in common with ours at the end of this report.

3. Integrating OSM Data into OGD

3.1 Requiremements of the ODbL

The section before considered licensing requirements for data contributed to OSM. What are the restrictions placed on data downloaded from OSM?

The Open Database License has two main requirements. First, anyone using OSM data must attribute (credit) OSM. For maps displayed on-screen, this is usually taken to be a credit visible in the corner of the map like “”. There exist guidelines about attribution: see https://wiki.osmfoundation.org/wiki/Licence/Attribution_Guidelines .

Secondly, when OSM data is mixed with another dataset, that data must be available on terms compatible with OSM’s. This is sometimes known as a ‘share alike’ or ‘copyleft’ clause and prevents (for example) another map supplier taking OSM’s footpath data and combining with its own road data, without contributing the road data to OSM. However, although mixing data has this stipulation (a ‘Derivative Database’), independently overlaying different datasets does not (a ‘Collective Database’). The subject is reasonably complex and there is guidance on the OSM Foundation website. In the field of geo-information, ISO principles and guidelines are widely adopted for quality assessment. The

updated standard (ISO 19157, 2013) defines the following data quality elements: completeness, logical consistency, positional accuracy, temporal quality, thematic accuracy and usability. These six elements are described in the following main chapter.

3.2 Practical Options

What can a public authority do if it wants to integrate OSM data into it's OGD and keep its own license or terms of use?

3.2.1 Option 1. Non-public use or internal use

There's nothing special to do in case of non-public use or internal use of database and product (ODbL paragraph/no. 4.5 b and c).

Two main usage variants, where OSM data are published together with the own ones - always with appropriate attribution, of course: Extract from OSM and Collective Database:

3.2.2 Option 2. Extract from OSM

Extract from OSM as a geographic and/or thematic extract, "layer", service.

This is the most obvious and common use case. In order to learn more about such "horizontal map layers" see
https://wiki.osmfoundation.org/wiki/Licence/Community_Guidelines/Horizontal_Map_Layers_-_Guideline .

3.2.3 Option 3. Collective database

ODbL share-alike does not apply to collective databases, but only to the corresponding parts (ODbL paragraph/no. 4.5 a).

The more demanding but possible and very interesting use case from the two. See figure 1. In order to learn more about collective databases see
https://wiki.osmfoundation.org/wiki/Licence/Community_Guidelines/Collective_Database_Guideline_Guideline .

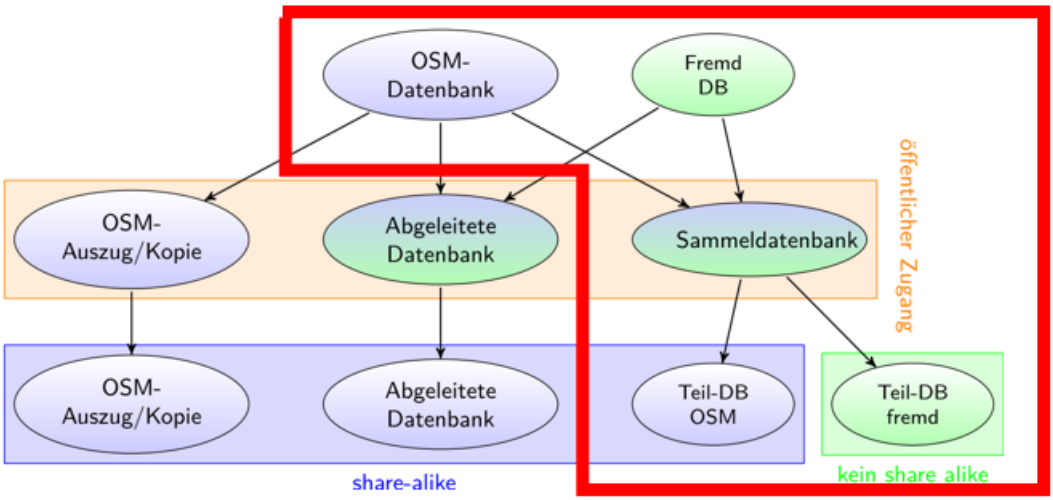


Figure 1 Option 3 - ODbL share-alike does not apply to collective databases, but only to the corresponding parts

3.2.4 Option 4. OSM as data indicator

Another variant that is possible from the authorities' point of view (i.e. share-alike does not apply: Authorities could use OSM data as a change indicator for their data. Cf. "non-public use of the database, internal ue". => A very reasonable use case.

3.2.5 Option 5. Mixed database situation

Finally, there's a variant where clarifications and consultations are needed: "Mixed" attributes. This is about redistribution of all or a substantial part of the data => Here one must check and investigate if this is a case of a Collective Database ("Sammeldatenbank") or a Derived Database (see option 6).

3.2.6 Option 6: Derived database

When elements are taken directly into a mixed layer or dataset, and when these are then to be published and distributed under a license then the share-alike from ODbL applies to OGD ("Derived Database") and the OGD has to be released under the ODbL. => This is a very unlikely use case, since authorities do well to separate their data from the "non-official".

So as you can see, there are many options how a public authority can integrate OSM data into it's OGD while keeping its own license or terms of use: It's any option except option 6 "derived database".

In any case note that the OSMF is sympathetic to the use of OSM data (like Google, among others - although probably for other reasons). Note also that in Switzerland there are only contracts, no "sui generis" database right as in other European countries.

4. Data Quality of OSM

4.1 Definitions of Data Quality

In the field of geo-information, ISO principles and guidelines are widely adopted for quality assessment. The updated standard (ISO 19157, 2013) defines the following data quality elements: completeness, logical consistency, positional accuracy, temporal quality, thematic accuracy and usability. These six elements are described as follows:

- Completeness refers to the presence or absence of objects, of their attributes and of relationships compared to the product's specification and some reference dataset.
- Logical consistency refers to the degree of adherence to logical rules of data structure, attribution and relationships as described in a product's specifications.
- Positional accuracy as a data quality element consists of two data quality sub-elements, namely horizontal accuracy and vertical accuracy.
- Temporal quality refers to the quality of the temporal attributes, such as date of collection, date of publication, update frequency, last update or temporal validity (also referred to as currency), and also to relationships between the temporal validity of objects.
- Thematic accuracy refers to the accuracy of classes or thematic tags associated with specific locations or objects placed in geographic space, such as classes assigned to pixels in a land cover map or tags assigned to a vector-encoded entity, e.g., a highway, river, building or green area.

- Usability (or fitness-for-use) refers to the external quality of a dataset and is focused on the needs of the user. Usability acts as a complementary element by linking both user requirements and data quality measures to check whether the data for a specific application can be used.

Based on experience with spatial quality projects (specifying criteria and auditing datasets) Vullings et al. (2015) have defined a framework (Figure 2) and illustrated a fitness-for-use approach with study cases in their research (2015).

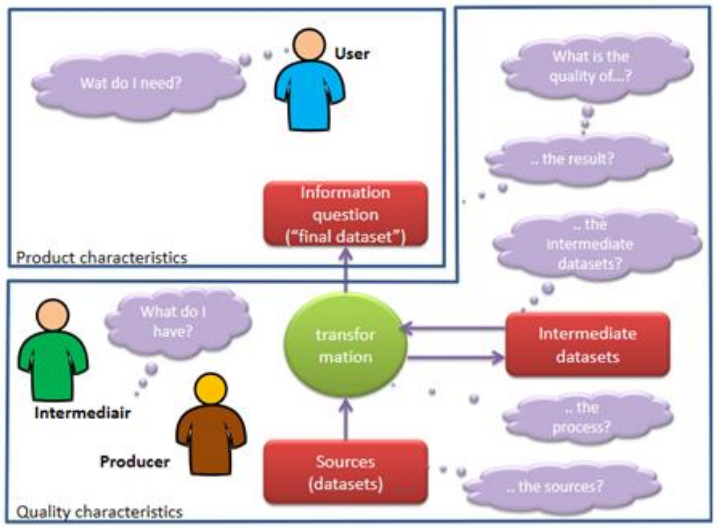


Figure 2: Quality Assessment Methods for Crowdsourced Data

4.2 OSM and Data Quality in Research

To gain a global view of how data quality has been discussed in the recent 10 years and how government data is involved respectively, we did a literature research "OSM and Quality in Research" and shortlisted 51 interesting candidate articles. For better visibility we split the table into two parts: Appendix B contains the articles title, geographic region authors investigated and which role the government data plays in their research or use case. Appendix C is arranged in the same order as B, with providing more information for an easy look-up – name of the main author, year of publication and online access link.

As mentioned in the last chapter, the ISO principle is adopted internationally as geospatial data standard and it is used almost in all the research literatures as guideline for evaluating data qualities. To be able to tell if something is good enough or not, we always need a reference to compare to. The most ideal reference would compare to official data. For instance, Haklay (2010) compared the OSM road network with Ordnance Survey Meridian 2 for England and Kounadi (2011) with the Hellenic Military Geographical Service (HMGS) dataset for Athens, Greece. But it is also often mentioned in the literatures that official data are not easily reachable. As another alternative, instead of comparing OSM with a ground truth reference dataset, authors traced the history of data of their study area and compared the up-to-date status with its previous performance to measure the enhancement and modification of data accuracy.

Some of the studies from the list are exclusively dedicated to OSM data, some are in name of VGI data but the main representative or even only rearch object is still OSM data. The most popular dataset are land cover, road network, building (address, house number etc.), routing and navigation, and general polygon's geometry.

The strongest assertion that volunteer and crowdsourced data, which includes OSM, has at least a fitness-for-use comes from the paper of Lewandowski & Specht (2015): "Collectively, these data suggest that volunteer data are not consistently more variable than expert data".

For different purposes different aspects of OSM data are used for individual geodata quality assessment, major data categories are:

- Map-based data: This is the most popular topic as it concerns all VGI sources are include geometries as points, lines, and polygons – all basic elements to design a map.
- Image-base data: It is generated where contributors take pictures of a particular geographic object or surrounding with any handheld device and attach a geospatial reference to it. This sort of data are desirable for environmental monitoring, pedestrian navigation and human trajectory analysis for creating geographical gazetteers.
- Text-based data: where people contribute geographic information in the form of text and disseminate data to the public, sometime even in near real-time basis like crisis mapping. This data category is interesting for purposes like detect disease spreading, event detection, or for science and research.

4.3 Assessing the Quality of OSM data

Data quality elements – especially completeness and positional accuracy – are usually assessed by comparing an object with its counterpart in reference data, which are considered to represent the (ground) 'truth'. This assessment requires the existence of reference data with similar characteristics and a valid timeframe to make the comparison.

The assessment of thematic accuracy in crowdsourced data, especially VGI, may be performed using this traditional approach, where the information is compared to reference data, e.g., satellite imagery or authoritative data.

VGI has made it possible for a much wider group of contributors to create and share geographical information. Contributors with little experience and expertise of geospatial data, might have contributed to the perception of the unreliability of this data source. There are several ways to classify the quality assessment methods, but the following categories are commonly mentioned in literature with different titles:

- Comparing data against “authoritative” spatial data.
- Rules and patterns learnt by experts and/or statistics (machine learning) for checking the objects. This is also called an intrinsic approach.
- Gate keep and weight users’ entries (e.g. with respect to their experiences, expertise, proximity, number of their entries, history and change sets).

In order to help the adoption of crowdsourced geospatial data Severinsen et al. (2019) proposed a formulaic model to address VGI quality issues by quantifying trust in VGI with a ‘VGTrust’ model. This model assesses information about a data author, and the spatial and temporal trust associated with the data they create, to produce an overall rating metric.

The project "osmcross" from Keller (2020, see figure 3) investigates the possibilities to estimate the spatial completeness (coverage) of selected Point-of-Interest (POI) in OpenStreetMap data. He is using an intrinsic aph whiproacch does not require a reference dataset and therefore is applicable worldwide. Keller's approach tries to estimate the completeness by learning the correlation of one POI class to other POI classes of objects using machine learning. This is a regression in the first place which is then turned into a classification with roughly three fuzzy defined variables, e.g. coverage is "good", "ok", "poor").

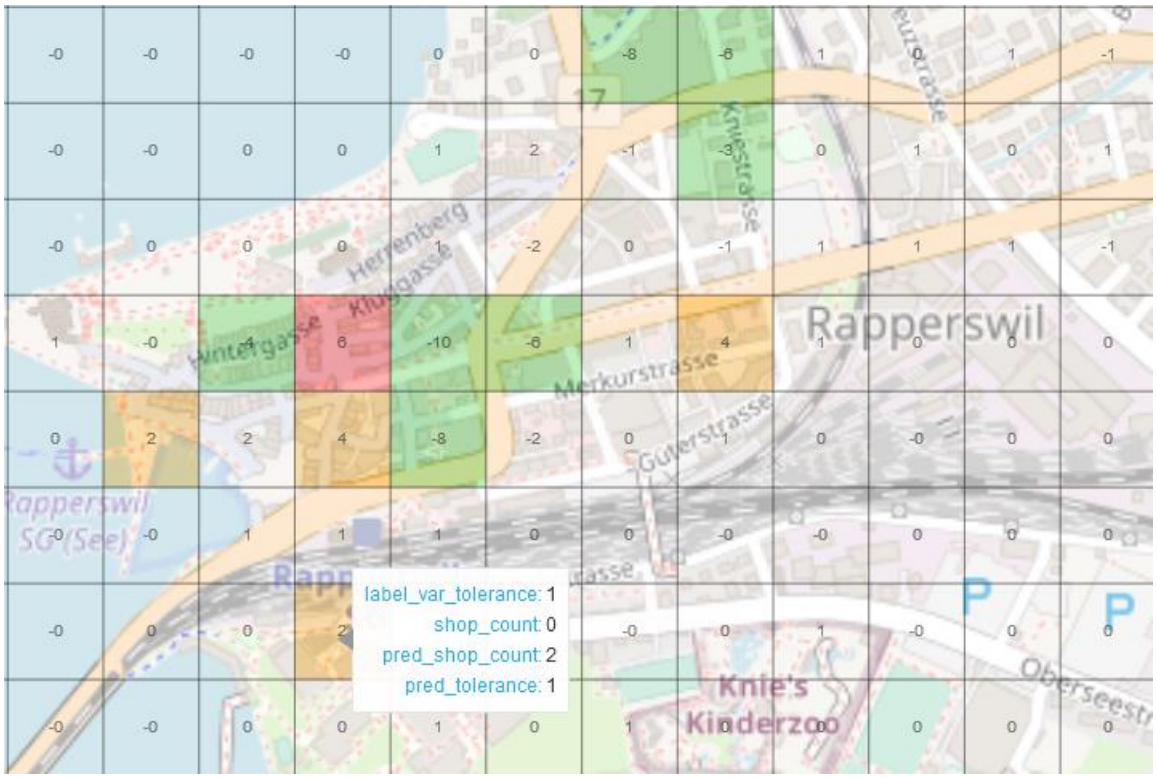


Figure 3: Screenshot of osmcross (mapcomplete/poicomplete).

Here are some more selected free online web tools to assess the data quality of OpenStreetMap data:

- ohsome quality analyst (OQT) - Get OpenStreetMap Data Quality for Specific Regions and Topics: https://oqt.ohsome.org/
- OpenStreetMap Analytics - density of buildings, amenities etc. (by HOTOSM): https://osm-analytics.org
- Osmose - Quality assurance tools available to detect issues in OpenStreetMap data: http://osmose.openstreetmap.fr/de/map/ (https://wiki.osm.org/Osmose).

4.4 Monitoring OSM Data

Spam and unwanted data in OSM can be operation mistakes or intended vandalism. A project like OSM uses various ways of dealing with the problems and many sites are set up for documentation, comments, communication, reporting, etc.

Detecting and spoting vandalism

JOSM History window opens the OpenStreetMap changeset page and loads details on how many objects were added/modified/deleted. Users can also see the changeset comment, which may contain more self-reported details from the user, as well as information on the editor and sources used.

OpenStreetMap Changeset Analyzer (OSMCha) is an advanced changeset explorer, it stores the metadata of the changesets and allows searching through them. It helps mappers to analyze and review data changes to OSM and also helps to investigate suspicious changesets. Changesets review requested by new mappers, changesets with "revert" in the comment which are tagged for review and lager data imports/ bots will be presented here for further evaluation. Since inexperienced mappers are more inclined to make mistakes or bad edits at the stage of getting a passing familiarity, "new contributors feed" is exclusively set up for quality control and their feed can be further evaluated.

OSM Hall Monitor reads diff files to look for large edits, skewed ratios of deletions to total edits or modifications to total edits, or changesets with modifications across the map; also, can watch specific

users or objects, with notification available; smart geospatial comparison like objects of certain shapes is a function to be added on in the future

OpenStreetMap Analytic Difference Engine (also called "OSM Analytic Tracker" or OSMAT for short) continually analyses activities in a given region (depending on the level of activity, ranging up to small countries) and presents tag changes in fairly easy to digest lists, allowing experienced contributors to spot and react to mapping mistakes in their region within minutes. It also allows opening "diffmaps" for quick overviews of geometric changes.

Osmose detects a very wide range of issue types and categorizes issues into filterable themes with severity levels. By expanding the information panel which is represented with an "i" symbol button, documentation about currently selected marker the section are Titles, Details, Correction aids, pitfalls to Avoid, source code are available for reference.

Reporting spam

Every mapping platform for OSM also provides a help forum, when notes/entry/comment with spam content are spotted, they can be reported via the report link. Beyond that, administrators will block spam users.

Data working group: For accusations of copyright infringement, imports, and serious Disputes and Vandalism in the OSM data the data working group should be mailed. "I've seen a problem; what should I do?" provides detailed guidelines.

Fixing and processing errors and spam

The JOSM Reverter Plugin is a convenient way to revert changes in OpenStreetMap. After every revert it is advisable to upload the changes by sharing a changeset comment on the reason for the reverting action.

Osmose allows direct correction of issues through an integrated tags editor through OSM account login. It offers some other interesting features "testmode by username"(correct own mistakes or race with co-worker) and "analyzer relationship"(check if the errors have been corrected after sending a modification request).

Data working group reverts changesets that are too large to be downloaded from the API without a timeout, data has been added to the database that is not license-compatible with OSM, as well as boundary or language disputes.

See also "Approaches and tools for integrating data into OSM" in Appendix D.

5. Case Studies of OSM Data in OGD Applications

The case studies have been chosen to show the diverse existing and potential applications of OSM used by governements. The studies are focusing on the main applications of geospatial data, namely base maps, geocoding (i.e. searching geographic locations) and routing. Case Study 4 "Use of OSM in the Administration of other Countries" overlaps with these applications.

There are many other applications with OSM not covered here, like POI (www.openpoimap.org)

5.1 Case Study 1 – Monitoring

Key function: A tool to monitor data changes in their concerning regions and able to spot suspicious updates of objects on their watching list.

Primary actor: Operators in Schutz und Rettung Zürich (SRZ)

Solution: Webapp to monitor OSM data used in SRZ's daily operations. System https://srzedi.srz.borsnet.ch/

Scenario: A change set contributed by a OSM mapper and it crossed the watching area of SRZ. The tool marks the coverage of this change set on map and shows a brief list view of the set. If any watching objects are overlapped in this set, operator opens the link and check the details of changed object. During the process of data validation, operator change the monitoring status to "in process" to avoid other team members working on the same task. When the change set is evaluated as a bad or suspicious contribution, the operator will revert the object on OSM and close the task by changing status to "closed".

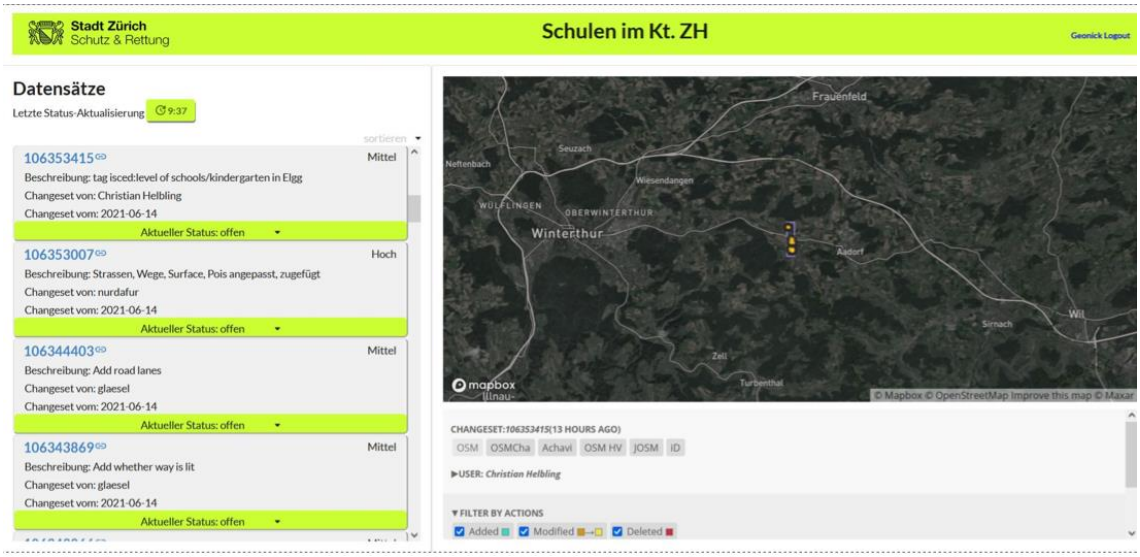


Figure 3: Example: Result of monitoring app with filter "Schools" in "Canton Zürich" (Source: own)

5.2 Case Study 2 – OSM Data for Maps

5.2.1 Canton of Zürich

Key function: Use OSM data to reproduce a visually identical base map for Canton Zürich which shows the advanced feature of OSM.

Primary actor: Public and internal map users

Solution: OSM data and official map data for and from a browser app maps.zh.ch. System: A map product developed on QGIS 3, target scale is 1:5000.

Scenario: The map product inherits the styling of government map; it will not affect the user experience. On this map users can easily view the objects with updated details provided by OSM. The geographic information is longer blocked due to the limitation of administration coverage, the details of neighbor canton Schaffhausen and neighbor country Germany are also available on the same map in consistent style.

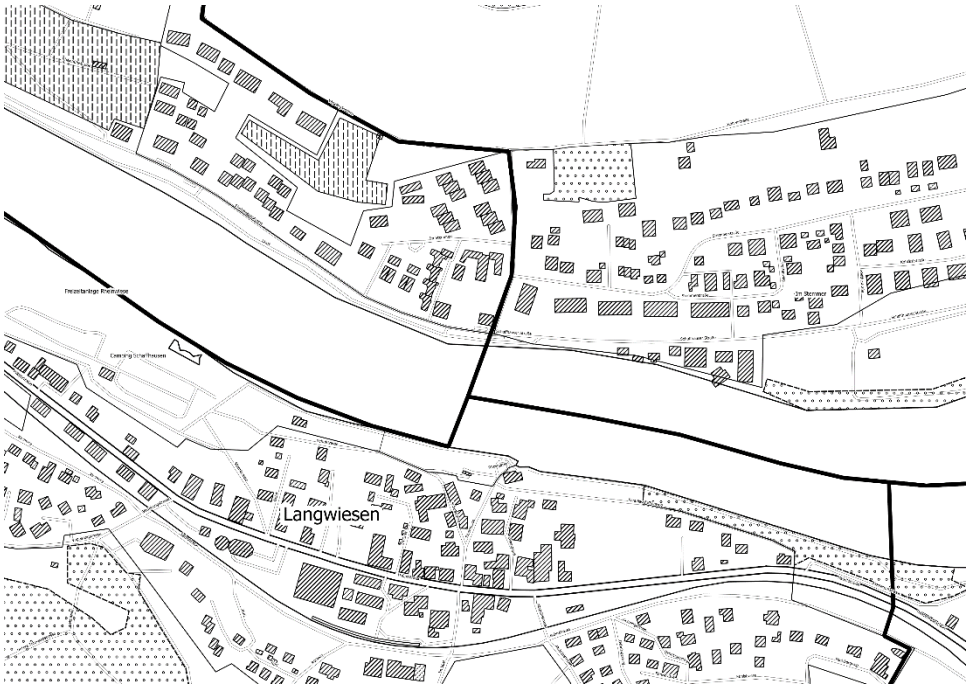


Figure 4: Map made with OSM data and styled with QGIS for the overview map Canton of Zürich, region Langwiesen (ZH). (Source: own)

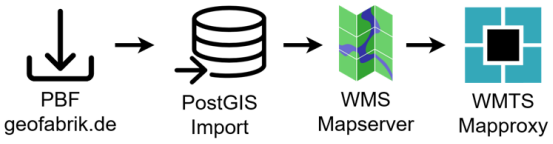
5.2.2 Cantons of Neuchâtel and Vaud

Key function: Use OSM data to produce a base map.

Primary actor: Registered map users

Solution: OSM data and official map data for the services of Canton Neuchâtel and Canton Vaud.
System: A map product.

Scenario: Below is the workflow of importing OSM data and use it as a base map on the geoportal of Conton Neuchâtel.



Result: https://sitn.ne.ch/s/TtOQs

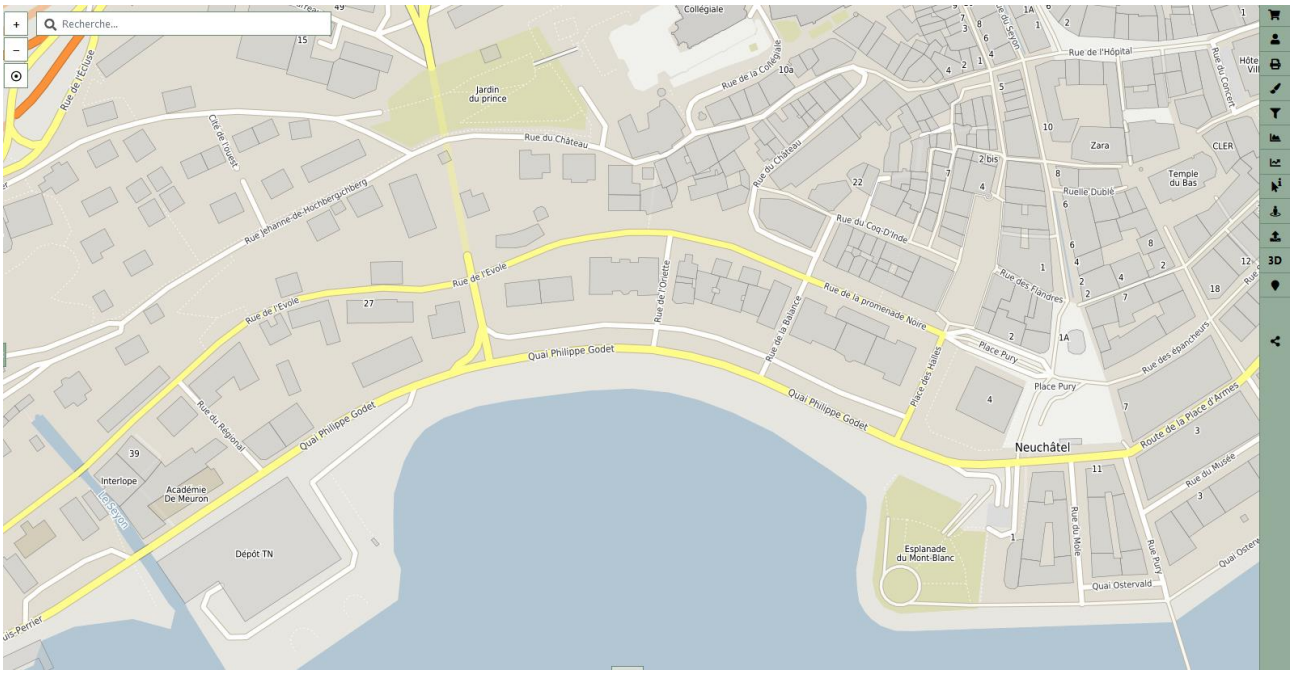


Figure 5: Map display of Neuchâtel (scale 1:2000)

Similarly, Canton of Vaud uses the same source and software stack for their OSM base map "Fonds de plan ASIT". According to the project coordinator Mr. Xavier Mérour, on top of the OSM data, official data are added to ASIT with a weekly update, while the overall map and OSM import are updated on a monthly basis. (System: https://viageo.ch/catalogue/donnee/300051).

5.3 Case Study 3 – Routing Applications based on OSM Data

5.3.1 Canton of Schaffhausen

Key function: Use OSM data to make routing suggestions for all non-public travelling methods. This routing service also provides access to locations that are temporarily set up for events.

Primary actors: Map users

Solution: OSM data and geoportal of Schaffhausen. System: https://map.geo.sh.ch/geoportal/Routing

Scenario: A user of Schaffhausen geoportal does not need additional applications to get routing instructions when travelling within the canton. Routing calculation will take the road condition of other cantons, which cut Schaffhausen into separate pieces, into consideration before presenting the suggestions. When there is temporary open-air event, user can find necessary information like where the gates, toilets, food & beverage stalls are during the event period.

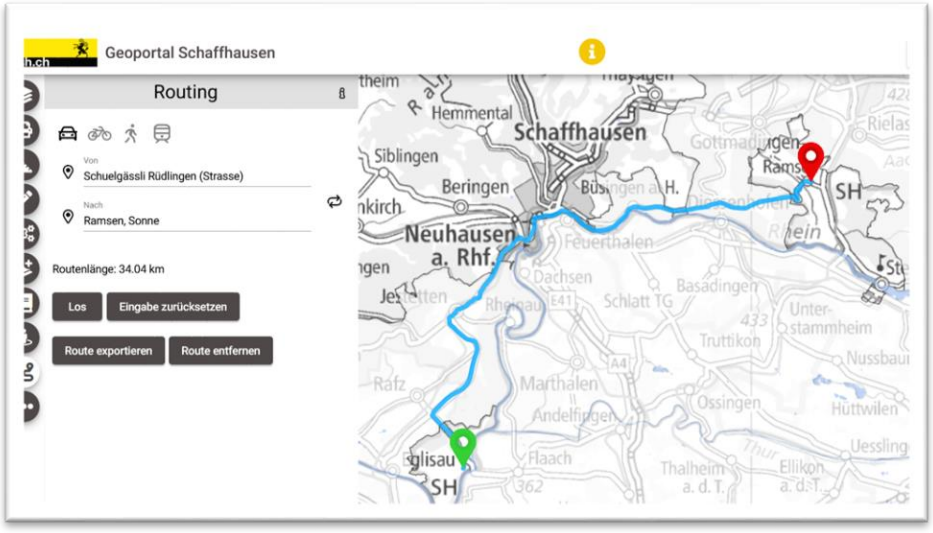


Figure 2: Routing result of driving crossing Schaffhausen via Germany. (Source: Canton of Schaffhausen)

5.3.2 Canton of Neuchâtel

Key function: Auto-mapping function to locate the nearest fire station and provide optimal routing suggestions for fire units to reach the spot within 15-18 min

Primary actors: Fire stations and intervention sectors of Canton Neuchâtel

Solution: OSM data. System: A testing map application

Scenario: When a fire breakout is reported, the application locates the isochrones zones of the fire spot and selects the nearest fire unit which can arrive at the spot with least travelling time.

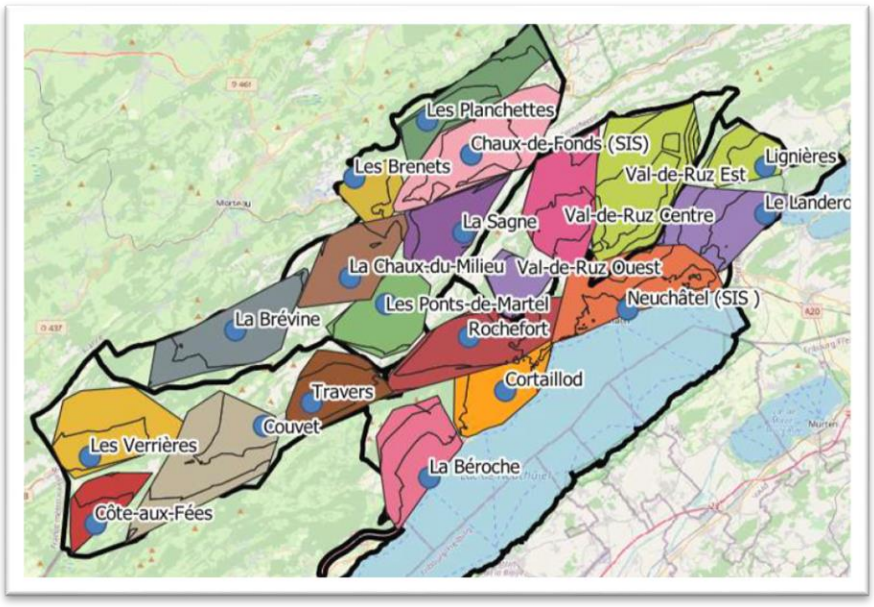


Figure 3 Example: Isochrone zones within the designated reaching time. (Source: Canton de Neuchâtel)

5.4 Case Study 4 – Use of OSM in the Administration of other Countries

Goals: Documentation research on topic "OpenStreetMap for government in Europe ".

Coming with a bibliometric review, Fernandes et al. have investigated 37 publications and relevant context in the integration process, along with a deep study on how VGI data are used in European National Mapping Agencies (NMAs).

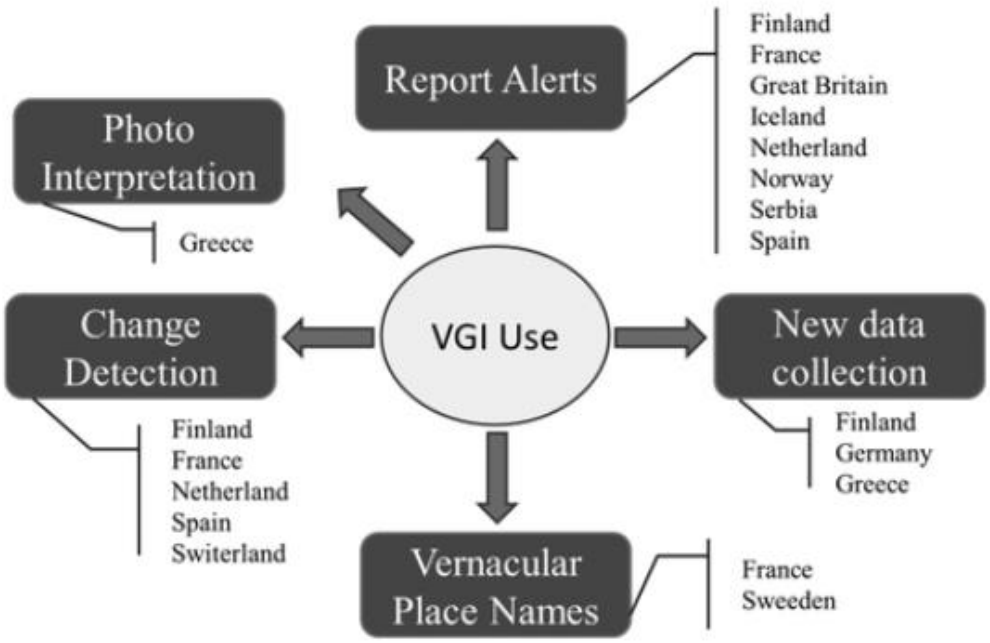


Figure 4: The use of VGI in the European NMAs (Olteanu-Raimond, et al., 2017)

According to Fernandes et al. in a first action government can open up its data (so called Open Government Data, OGD). This encourages transparency in government activities and reduces the cost of data sharing between government agencies, encourages innovation in civic services, thereby supporting economic development.

A second possibility that is being mentioned by Fernandes et al. is the creation of a method that proposes indicators to evaluate the outdated authoritative mapping OGD. By systematizing a method, it is possible to verify which areas of mapping need updating. The proposal deals with a targeted update, without the need for new mapping.

Finally, a third possibility for the government is to use VGI data like OpenStreetMap for several reasons: First to monitor updates in urban space, evaluates which areas are growing. Second to compare and integrate VGI data with authoritative mapping. Both create an opportunity for government institutions to get updated, validated and new information.

Other than academic research, there is also an "OpenStreetMap for Government" wiki page available for people to present their implementation Case Study in this field. We narrowed the geographic range down to European Solution: See figure 5.

OpenStreetMap for Government Use in Europe

Category	Country	Authority	Project
Migrate to OSM	Austria	Vorarlberg	Cycle routes
QA, direct editing	Belgium	City of Antwerp	Improving official road registry with OSM data
QA, community building	Belgium	Trage Wegen VZW	Base registry of "slow roads" Flanders + Brussels
QA, direct editing	Belgium	Brussels Capital Region	Getting official bike route data to work through OSM
Mapmaking	Belgium	Commune de Brugellette/ Champs Libres	Outreach, data collection
Data mashups	France	Etalab	Address database and services
Migrate to OSM	France	SNCF	Indoor mapping in OSM
QA, direct editing	Finland	Helsinki Region Transport (HSL)	Use OSM as the base for all our maps
Data generation	Germany	Regionalverband Ruhr	Stadtplanwerk 2.0 (city maps 2.0)
Quality Assurance	Lithuania	GIS-Centras, Lithuania	Bi-directional error/update detection
QA, direct editing	Norway	Entur AS on behalf of Ministry of Transport and Communications, Norway	Completion of road network, and footpaths in connections with public transport stops
QA	UK	Ordnance Survey	Validating official road data with OSM
Data generation	UK	Thinkwhere	Data generation and webmapping by the government

Figure 5 Source: https://wiki.openstreetmap.org/wiki/OpenStreetMap_for_Government

Sarretta (2021) has recently shared a study of integrating OSM and government data. His study "Towards the Integration of Authoritative and OpenStreetMap Geospatial Datasets in Support of the European Strategy for Data" aims to be the first step towards a broad assessment of the enablers and barriers of integration authoritative datasets from European NMAs with datasets from OSM and to provide a preliminary set of recommendations on interoperability matters. To achieve this goal, the study carried out an experiment based on Free and Open-Source Software for Geospatial (FOSS4G) to test the integration of country-wide address datasets from a European NMA and the OSM project, which is also highly possible the first integration implementation action between OSM and authoritative datasets at a national level.

The first and remarkable experiment in the study took place in Finland, it integrated the Finnish national address datasets from OSM with the National Land Survey of Finland and generated an NLS-OSM integrated dataset out of both. The comparing result proved the OSM data is geographically unevenly distributed and the data density in urban areas is higher than in rural areas, which is also not a surprising find-out. Despite that some street or city names were misspelled, Osm indeed contributes more detailed or up-to-date information.

As part of the study goal, other than discovering what is technically possible, the author also looked into licenses issues, while the lack of license compatibility might indeed represent a serious obstacle to the actual use of the integrated datasets – this applies in both directions. Most governmental data come under the Creative Commons Attribution 4.0 International license (CC-BY 4.0), CC-BY 4.0 and ODbl are not fully compatible. Importing CC-BY 4.0 data in OSM requires additional explicit permissions from licensors ("waiver"). And including OSM data in some datasets requires to release the integrated data under ODbL. In addition to legal interoperability, organizational interoperability both within and across organizations (including governments and OSM communities) will be the key to make data integration a common, standardized and policy-enabled process rather than an isolated and ad-hoc exercise.

Finally, an application should be mentioned that nicely demonstrates the possibilities of crowdsourcing, especially of OSM: The Castle-Dossier-Map (https://castle-map.infs.ch/): It shows castles of Switzerland and neighboring countries (Liechtenstein, Austria, Germany, France, Italy).

6. Conclusions

These have been the objectives and its related chapters where this report gave answers:

1. Clarification of important licensing issues when monitoring, comparing or transferring OSM data (with ODbL licence)

2. Identification of typical processes and evaluation of software tools for monitoring OSM data.

3. Identification of typical processes and evaluation of software tools for the alignment and quality assurance of OSM data.

4. Identification of typical processes and evaluation of software tools for the transfer of OSM data.

All in all we think, that all objectives have been fulfilled and answered. The study scope originally only includes clarifying data licensing issues, identifying typical processes and evaluating tools, software/ product delivery was not in the scope. As there are no existing products to illustrate the work process or to scale the "fitness-for-use" level for our study, actual products/applications were developed for Case Study 1, 2 and 3, in corresponding to the requirement of each case.

Complementarity of OSM and OGD: OSM shares many similarities with Wikipedia. But OSM is also different from Wikipedia as OSM is rather complementing official data (Heuel 2012).

This document tries to show which approaches and tools are available for organizations, especially public administrations, to integrate OSM into their own data.

Some important approaches are:

- Option 2. "Extract from OSM"
- Option 3. "Collective Database"
- Option 4. "OSM as data indicator".

Some important tools are:

- (Changeset) Monitoring Tools.
- Conflation Tools in JOSM, as webapplication (OSMConflator) or using desktop GIS like QGIS.
- More tools are typically based on specific own Python data pipelines.

7. Recommendations

7.1 Recommendations to the OSM community

Enhancing the Quality - Focused Mapping in OSM: Inspired by members during interview as many wished to get some hands-on experience on collecting the dataset for objects of their interest. Although neither software nor product was in the study scope, we found it a very interesting to have a platform to publish mapping details on project base. Therefore, we managed to pull some manpower from IFS institute and developed an addition webpage to present the mapping data. So, the "Project of the Month Switzerland" action was born. https://potm.osm.ch is a complete and multilingual application with a nice graphic view that shows the introduction of ongoing project, mapping activity level by mappers on daily basis and the growth rate of the total mapped objects. This page is not just platform to present pure statistics, it also serves as a platform of playful interaction, for example, a project initiator could set some incentives to gamify the mapping behaviors and encourage more contributions. Shortlisted projects from Project of the Month Switzerland (https://wiki.openstreetmap.org/wiki/EN:Project_of_the_month_Switzerland) will be presented there.

1. The mappers should get even more involved in the SOSM association (for example, become a member of SOSM).
2. The mappers should participate even more in thematic mapping projects (for example in the project of the month CH).

7.2 Recommendations to Governments

We strongly support the recommendation from the study Hitz & Stürmer (2021) that public authorities should build up internal competencies for OpenStreetMap and designate a central point of contact for OpenStreetMap issues both externally and internally.

A recommendation on how to share styles with QGIS: In QGIS, the sharing of display models (styles) and the direct application to geospatial data of the same (vector) data structure is not yet satisfactorily solved. Especially with many vector data sources QGIS 3.20 ‘out-of-the-box’ is not sufficient. Hence this article, whose goal is to show solutions to this. Strictly speaking, it is about recommendations for authors of display models (styles) with QGIS so that they in turn can instruct the users of their display models. For general use we recommend the QGIS plugin “Layer Style Loader” as a solution. It fulfills most of the requirements. Technic specifications and implementation details are published here: https://md.coredump.ch/s/7QoNzCG_w#

1. The authorities should use OpenStreetMap data where it makes sense. This is allowed without having to change their own license if they know the options.
2. Authorities should contact the SOSM association if they have questions about the use of OpenStreetMap data.
3. Authorities should clarify whether engagement with OpenStreetMap is beneficial to them, such as becoming a member of SOSM.
4. Authorities should build internal capacity for OpenStreetMap. (For example, participate in a mapathon/mapping party).
5. Authorities should designate a central point of contact for questions about OpenStreetMap for both external and internal inquiries (see also the University of Bern's "Leitfaden").

Those who are not sure which tool to take, how to proceed, or which approach is now the most appropriate can seek advice from the local OSM community or OSM experts.

8. Bibliography

Barbier, G., Zafarani, R., Gao, H., Fung, G., & Liu, H. (2012). Maximizing benefits from crowdsourced data. Computational and Mathematical Organization Theory, 18(3), 257-279. URL: https://link.springer.com/content/pdf/10.1007/s10588-012-9121-2.pdf (accessed 2021-10-25)

Basiri A., M. Haklay, G. Foody & P. Mooney (2019) Crowdsourced geospatial data quality: challenges and future directions, International Journal ofGeographical Information Science, 33:8, 1588-1593, DOI: 10.1080/13658816.2019.1593422. URL: https://doi.org/10.1080/13658816.2019.1593422 (accessed 2021-10-25)

De Medeiros, G. F., Degrossi, L. C., & Holanda, M. (2020, November). QualiOSM: Improving Data Quality in the Collaborative Mapping Tool OpenStreetMap. In Proceedings of the XXI Brazilian Symposium on Geoinformatics GEOINFO, São José dos Campos, Brazil (pp. 10-21). URL: https://github.com/gmedeiros93/josm/tree/master/josm/plugins/Quali_OSM (accessed 2021-10-25)

Fairhurst, R. (2020): 'Providing data to OpenStreetMap - a new guide for data owners'. Web link: https://blog.openstreetmap.org/2020/07/22/providing-data-to-openstreetmap-a-new-guide-for-data-owners/ (accessed 2020-11-15)

Fernandes, V. O., Elias, E. N., & Zipf, A. (2020). Integration of Authoritative and Volunteered Geographic Information for Updating Urban Mapping: Challenges and Potentials. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 43, 261-268. URL: https://www.researchgate.net/profile/Elias-Elias-5/publication/343837728_INTEGRATION_OF_AUTHORITATIVE_AND_VOLUNTEERED_GEOGRAPHIC_INFORMATION_FOR_UPDATING_URBAN_MAPPING_CHALLENGES_AND_POTENTIALS/links/5f443741a6fdcccc43fb1832/INTEGRATION-OF-AUTHORITATIVE-AND-VOLUNTEERED-GEOGRAPHIC-INFORMATION-FOR-UPDATING-URBAN-MAPPING-CHALLENGES-AND-POTENTIALS.pdf (accessed 2021-10-25)

Fonte, C C, Antoniou, V, Bastin, L, Estima, J, Arsanjani, J J, Bayas, J-C L, See, L and Vatseva, R. 2017. Assessing VGI Data Quality. URL: https://www.researchgate.net/publication/319632519_Assessing_VGI_Data_Quality (accessed 2021-10-25)

Girres, J. F. & Touya, G. (2010). Quality assessment of the French OpenStreetMap dataset. Transactions in GIS, 14(4), 435-459. URL: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9671.2010.01203.x (accessed 2021-10-25)

Haklay, M., Antoniou, V., Basiouka, S., Soden, R., & Mooney, P. (2014). Crowdsourced geographic information use in government. World Bank Publications.

Heuel S. (2012). "Ist OpenStreetMap das Wikipedia für Karten?" Blog Post 2012-03-26. URL https://digital.ebp.ch/2012/03/26/ist-openstreetmap-das-wikipedia-fur-karten/ (accessed 2021-10-25)

Hitz-Gamper B. & Stürmer M. (2021) Daten in OpenStreetMap intergrieren – ein Leitfaden für Dateninhaber. Forschungsstelle Digitale Nachhaltigkeit, Universität Bern. DOI/URL http://dx.doi.org/10.48350/159438 (accessed 2021-10-25)

Keller S. (2020). osmcross - Estimating quality aspects of OpenStreetMap. Institute for Software, OST. Web access: https://www.ifs.hsr.ch/index.php?id=19673&L=4 (accessed 2021-05-11)

Keller S. (2021). Darstellungsmodelle (Styles) teilen mit QGIS. URL: https://md.coredump.ch/s/7QoNzCG_w# (accessed 2021-10-25)

Lewandowski, E., & Specht, H. (2015). Influence of volunteer and project characteristics on data quality of biological surveys. Conservation biology, 29(3), 713-723. DOI: 10.1111/cobi.12481. URL https://doi.org/10.1111/cobi.12481 , https://www.jstor.org/stable/24483103 (accessed 2021-10-23).

Mooney, P., Corcoran, P., & Winstanley, A. C. (2010). Towards quality metrics for OpenStreetMap. In Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information

systems (pp. 514-517). URL: https://dl.acm.org/doi/pdf/10.1145/1869790.1869875 (accessed 2021-10-25)

OpenStreetMap. URL https://en.wikipedia.org/wiki/OpenStreetMap (accessed 2021-10-25)

Sarretta A. (2021). Towards the integration of authoritative and OpenStreetMap geospatial datasets in support of the European strategy for data. URL: https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLVI-4-W2-2021/159/2021/ (accessed 2021-10-25)

Severinsen, J., de Roiste, M., Reitsma, F., & Hartato, E. (2019). VGTrust: measuring trust for volunteered geographic information. International Journal of Geographical Information Science, 33(8), 1683-1701. URL: https://www.tandfonline.com/doi/pdf/10.1080/13658816.2019.1572893 (accessed 2021-10-25)

Taştan, H., & Altan, M. O. (2011). Spatial Data Quality. *Proceedings of 3rd Turkish-German Joint Geodetic Days*. URL: https://www.academia.edu/download/61090326/c87acedf5ca753220191101-121841-56mcp8.pdf (accessed 2021-10-25)

Vullings, W., Bulens, J. D., Rip, F. I., Boss, M., Meijer, M., Hazeu, G., & Storm, M. (2015). Spatial data quality: What do you mean. In Proceedings of 18th AGILE International Conference on Geographic Information Science, Lisbon, Portugal (pp. 9-12). URL: https://agile-online.org/conference_paper/cds/agile_2015/shortpapers/87/87_Paper_in_PDF.pdf (accessed 2021-10-25)

Zielstra, D., & Zipf, A. (2010). Quantitative studies on the data quality of OpenStreetMap in Germany. In Proceedings of GIScience (Vol. 2010, No. 3). URL: https://www.researchgate.net/profile/Alexander-Zipf/publication/267989860_Quantitative_Studies_on_the_Data_Quality_of_OpenStreetMap_in_Germany/links/54d99a590cf25013d0426ba0/Quantitative-Studies-on-the-Data-Quality-of-OpenStreetMap-in-Germany.pdf (accessed 2021-10-25)

Zscheile F. (2020), "Open Database License", FOSSGIS – BKG Workshop. Oktober 2020.

Zscheile F. (2021) "Quellen- und Lizenzangabe(n) – Vorgaben der Open Database Licence (ODbL)" FOSSGIS 2021 online, Juni 2021. URL https://www.fossgis-konferenz.de/2021/sessions/FKBDAC.php (accessed 2021-10-25)

9. APPENDIX A: SWOT Interview Script Raphael Das Gupta

Strengths

Q: As a mapper with long-time experience, what credits would you give to OSM?

In general, OSM has a similar effect for geoinformation as Wikipedia has for encyclopedic knowledge: It empowers users to add and modify data for topics that they are interested in, which might be specific types of features or data needed for applications relevant to the respective user. That means users can influence what is being mapped and ensure that it includes what they care about. Using any of the various editors available, a user can fix an error, update outdated information or add missing stuff without having to jump through any administrative or organizational hoops, other than creating an account.

This is quite different from contributing to most government data, where you'd first have to find out whether contributions are accepted at all, who's responsible for accepting them or where to file them, and then wait for the contributions to be reviewed, adopted and a new version of the dataset published.

Also, OSM contributors aren't limited to types of information (e.g., feature types) the dataset maintainers envisioned in advance. If a mapper is interested in a topic suitable for OSM that isn't covered yet, they can discuss and define how to map related information, document it on the OSM wiki and then proceed to mapping that content, too.

Q: What makes people contribute to OSM?

Apart from enjoying the activity of mapping itself or being interested in that and wanting to be active in that field, many contributors want to have better free data for their own use or make the world a better place by creating and improving this vast free dataset, that is used by many different projects, including programs by humanitarian organizations.

Q: Why and how is OSM data used? Why is it important?

I see two main reasons for OSM being preferred over other data sources:

OSM is free (non-proprietary), and thus available without having to negotiate contracts first. (Though the terms of its license apply, of course, but the requirements imposed there are easy to meet for anyone willing.) This is in sharp contrast to many proprietary data sets, which might exclude certain uses or not contract with you at all, and to proprietary services, where the underlying data might not even be directly available to you.

OSM has data covering the whole Earth, as a single unified dataset. While OSM mapping conventions can vary between countries, this is often still much better (and more comparable between regions) than having to integrate data from authorities of different nations, which might differ grossly in what they survey and how they represent it in their published data.

These two factors together mean (nearly) universal availability of geodata with (nearly) global coverage. OSM is the only vector geodataset including a that wide range of topics and subjects with these qualities, that I'm aware of. Therein lies also the importance of its data, I'd say.

There's many ways how OSM data is being used: Maps of various kinds, navigation (routing), statistics and analysis, even real-world-based virtual scenarios for games, and probably countless applications I don't even know about.

A use-case I'm particularly aware of, because it was the reason that we (IFS Institute for Software) developed OSMaxx, is making maps for regions where either no official ones are available from local authorities (and also none as commercial products) or the available ones are too lacking in quality or coverage / completeness.

Q: As a senior software developer, what do you enjoy about working with OSM data?

From a practical standpoint: OSM provides access to the actual underlying data, not just derived products and services like a visual map to explore, routing, maybe some statistics etc. This availability of the raw data enables applications that wouldn't be possible only built upon such derived products and services.

From an idealistic standpoint: I like the idea that it is open data under a free license, available for free. It contains no proprietary information and imposes the same copy-left condition onto all creators of derived database products. To my knowledge, it is the only platform granting this level of access equality on a geoinformation dataset of this scale.

As OSM is a single large dataset for (almost) all topics, containing (almost) everything related, data in OSM is more easily discoverable than if one would have to hunt down all the right datasets for the various information one's application needs, and then having to figure out how to combine them correctly.

Q: Why don't we just use local government data or google maps for geo-data?

While Google Maps probably has the most comprehensive data collection for commercial properties like shop names and business hours in some regions, especially economically important ones, this isn't true everywhere and Google Maps is not suitable for all use cases. Some small villages with no shops, scattered habitations with small huts in third counties, don't even exist on Google Maps. And for applications like evacuation, rescue or food supply in humanitarian programs, we need to know where people live and what the geographic condition there (and on the way there) is like, information often lacking in Google Maps for remote areas.

As for government data, we first must consider its coverage. Government agencies usually only collect what they are mandated to collect, with a given budget. With limited resources, government cannot keep a copy of every imaginable dataset while also keeping them properly maintained. Further, different government organizations collect and hold different datasets, filtered for specific applications and trimmed to specific regions. Integrating this data again for a use-case that covers several regions (maybe each regions in several nations) and topics covered by distinct governmental datasets can be more difficult than simply collecting data together in a common platform like OSM targeting a common all-encompassing data schema.

Weaknesses

Q: What are OSM's weaknesses w.r.t. governmental use of the data?

OSM data is not authoritative. When you choose to use it, you don't have anyone on whom to impose responsibility for its accuracy or completeness or whom to blame for any lack thereof. The quality and quantity of OSM data is not the same everywhere: How comprehensive it is depends on the local mappers. Its inhomogeneous level of details makes the data more difficult to consume than a typical single government dataset.

OSM data can be changed by everyone, pseudonymously. When OSM data is used for making policies or deciding on subsidies (and this is known or suspected), this can result in an incentive to put inaccurate or even false data into OSM (or to remove accurate and correct data) to influence governmental decisions to one's own benefit.

Another problem can be vandalism or defacement as well as bona fide errors OSM mappers might make. Usually these are quickly discovered by the OSM community and removed shortly after, at least for important and highly visible objects like city names but often also for less significant ones. Though, a government office offering a map based on OSM data would want to make sure that it never ever shows, say, a racial slur on it. Or if it offers a navigation service, that should never route people through know dangerous or forbidden-access areas. One way to deal with this is to not use live OSM data, but only snapshots taken at specific times and vetted before replacing the previous snapshot. There are also commercial offerings for that, e.g., by Mapbox. Though this, too, can backfire: There's been a prominent name vandalism occurrence that was very quickly rectified on

OSM, but could be seen on Mapbox-based maps for a long time, because the service happened to have taken the snapshot right in the short period of time the false city name was in OSM (and it wasn't caught by Mapbox quality assurance steps) and it took the company a very long time until they published the next snapshot that included the correction.

Yet another aspect is that OSM uses the same data (and on its own website also the same map) for the whole world. One could argue that that is a strength, not a weakness, as there can be only one objective on-the-ground truth, and OSM can be the platform where the worldwide community agrees upon on what that truth probably is. Though there is at least one area, where governments of (some) different nations like to see (and show) different "truths" from one another: Their own boundaries (and maybe those of allies). Paper maps for disputed areas differ depending on where they are sold. Google Maps and some other commercial online maps show different borders, depending on where they're accessed from. OSM doesn't do this, and its option to mark disputed boundaries or areas as, well, disputed (and to show them as such) might not be enough for such a government to want to still use it.

Q: What are OSM's weaknesses from a developer's point of view?

The data model of OSM is vastly different from the one commonly used in GIS. That alone isn't a problem, as OSM's model fits OSM's usecase quite well. But as a consequence (and for historic reasons) the OSM community usually uses tools tailor-made to consume OSM data, often developed independently from the rest of GIS eco-system. That means that these toolchains cannot readily be used on geodata from non-OSM sources, without first bringing that data into OSM's way of representation. Vice versa, typical GIS software cannot usually be used on untransformed OSM data – and transforming it to the common GIS data model often isn't lossless.

This makes data integration and tool reuse more challenging in projects combining OSM and non-OSM geodata than in those that use only one or the other. Having to switch between several toolchains can be quite annoying, time-consuming and at times confusing.

Opportunities

Q: What opportunies do you see for OSM in a long run?

As already mentioned, people usually map in OSM what they themselves care about, or what they themselves use or need on a map or in geodata. This might not only be good for having a dataset of the subjectively most relevant features (which is useful in and of itself), but also for drawing conclusions: By analyzing what is and what isn't mapped in OSM and by comparing that to existing governmental geodata, one can probably get some inspiration on what useful additions to the latter might be.

Such an analysis could even give hints not only on what data is relevant to the citizens, but what infrastructure is important to them. Of course, OSM data or a comparison of OSM with other geodata shouldn't be used as the only source for that (especially if official decisions are to be based on it), because some infrastructure is important despite not being useful on a map (or not being mapped in OSM because it's too difficult or even impossible for laymen, e.g., when not easily accessible by anyone) and because the OSM mappers might not be representative of the overall population. Here, dedicated online feedback platforms about public infrastructure and citizens' needs could be used (and at some places already are). It makes sense to base these platforms on OSM data and OSM maps, too, as that enables feedback about more things than most government maps contain. And if platform users are made aware that it's OSM based and want to give feedback about something not yet contained in OSM, they can add it themselves or at least report it as missing, so the OSM community can add it. This would benefit both OSM, because it could gain new mappers that way, as well as the citizens, because they could give feedback on things in their area important to them and have it included in OSM (and thus many maps) subsequently, if it wasn't already.

A good reason for government to collaborate with OSM and to even get their own data into OSM if suitable, is that the feedback loop with data users and map users (i.e., citizens) can be much tighter than for typical open government data offerings. That can help with quality assurance: When the data

is being used as part of the overall OSM geodata set, errors and mistakes in it will often be more likely to be noticed, and when they are noticed, they are more likely to be reported or even fixed by the user. Why the latter? Because it's easier to do: With traditional government data offerings, the user would first have to find out who is responsible for that kind of data object in this area and figure out the correct way to contact them, e.g., finding and submitting the right feedback form. In contrast, on the OSM website users can simply file a note or even edit and improve the data themselves. Third-party services based on OSM data or OSM maps are encouraged to link directly to that functionality, too, so that it's easily accessible also for their users.

Threats

Q: What do you see as possible threats for the development of OSM?

The source of OSM's strength – the ability of everyone to edit and improve the data – can also be a threat when that data is used for official government purposes, especially when that is known.

Although only surprisingly little data vandalism in OSM is going on and grave cases are usually quickly spotted and rectified, there are still cases where some people try to manipulate the OSM data for possible personal gain. When the augmented reality game "Pokémon Go!" was new and gaining popularity, OSM had an influx of new mappers making edits of questionable quality and veracity. That was because the occurrence of certain game elements in "Pokémon Go!" was made dependent on real-world features, e.g., water-type Pokémons are more likely to be found near (actual) bodies of water, and the rumor spread among players that the game used OSM data to determine where those and other relevant real-world features were.

Some players might have noticed that the game didn't seem to know about some geographic features and that that feature was missing in OSM, too. If they added it to OSM, they improved the completeness of the OSM data.

But some players decided to make up features and enter them in OSM, e.g., changing their surrounding areas on the map to lakes or rivers (where there weren't any in reality), probably thinking they could influence the game in real-time and encounter said water-type Pokémons that way. While the Game did (and maybe still does) indeed use OSM data, that attempt to cheat at the game failed, because the game used only a snapshot of OSM data rather than live access, and because the wrong landscape elements would usually have already been removed by the OSM community when the next snapshot was taken. Nonetheless, these disingenuous edits created a lot of unnecessary clean-up work for the OSM community.

Similar scenarios could happen when there is a more serious conflict of interest. And when it's known that and how government uses OSM data to make decisions or policy, such conflict of interest could indeed arise. The OSM community cannot completely prevent people from trying to influence things by contributing false data. So while it would be much easier for governments to maintain their data only in OSM (in effect not having their own data at all, but "only" contributing to the common OSM dataset) instead of syncing OSM and their own in-house data sets in one or both directions, that's probably not something governments would want (or should) usually do, as for some data, governments need full authoritative control or at least a certain autonomy.

10. APPENDIX B: OSM and Quality in Research - Evaluation

No.	Title	Geographic regions covered	Government data involved
1	A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis	San Francisco (USA), Madrid (Spain), and Yaoundé (Cameroon)	-
2	A Contributor-Focused Intrinsic Quality Assessment of OpenStreetMap in Mozambique Using Unsupervised Machine Learning	Mozambique, Africa	-
3	A grounding-based ontology of data quality measures	-	-
4	A Review of OpenStreetMap Data	-	many from government imported data having an acceptable open data licence allowing the corresponding geodata to be inserted into the OSM database e.g. USA, Canada, Austria
5	A review of volunteered geographic information quality assessment methods	-	authoritative data as quality indicators when exists
6	A Rule-Based Spatial Reasoning Approach for OpenStreetMap Data Quality Enrichment; Case Study of Routing and Navigation	-	-
7	A Systematic Study of OpenStreetMap Data Quality Assessment	Germany, UK & USA	-
8	A systems perspective on volunteered geographic information	-	-
9	Accurate attribute mapping from volunteered geographic information: issues of volunteer quantity and quality	-	-
10	Analysis of collaboration networks in OpenStreetMap through weighted social multigraph mining	-	-
11	Assessing data completeness of VGI through an automated matching procedure for linear data	UK	projecting OSM data to the British National Grid and importing OSM and ITN into a PostGIS database, along with the tessellation information
12	Assessing quality of volunteer crowdsourcing contributions:	-	-

	lessons from the cropland capture game		
13	Assessing the impact of demographic characteristics on spatial error in volunteered geographic information features	Denver,Adams County, Jefferson County of USA	compare to ORNL dataset (Oak Ridge National Laboratory, which was considered as the authoritative source)
14	Assessing the Quality of OpenStreetMap Contributors together with their Contributions	Heidelberg, Germany	official data provided by the Federal Agency for Cartography and Geodesy (BKG),
15	Assessing VGI data quality	-	-
16	Assuring the quality of volunteered geographic information	US, British Isles	more available and accurate than data from US continental states
17	Automated highway tag assessment of OpenStreetMap road networks	London	-
18	Automated quality improvement of road network in OpenStreetMap	Dublin and Waterford, Ireland	-
19	Automatic analysis of positional plausibility for points of interest in OpenStreetMap using coexistence patterns	Paris and Melbourne	-
20	Charting the geographies of crowdsourced information in Greater London	London	study data provided by the Greater London Authority and Nomis under the Open Licence Government Licence
21	Citizens as sensors: the world of volunteered geography	Worldwide	-
22	Comparing the quality of crowdsourced data contributed by expert and non-experts	random locations	-
23	Cooperation or competition–when do people contribute more? A field experiment on gamification of crowdsourcing	Germany	-
24	Crowdsourced geospatial data quality: challenges and future directions	-	-
25	Crowdsourcing geospatial data	Port-au-Prince, Haiti	compare coverage of London area to Ordnance Survey Meridian 2 data
26	Crowdsourcing: it matters who the crowd are. The impacts of between group variations in recording land cover	North and South America	-
27	Do experts or crowd- based models produce more bias?	US	-

	Evidence from Encyclopædia Britannica and Wikipedia		
28	Guided Classification System for Conceptual Overlapping Classes in OpenStreetMap	UK, Germany, Canada and France	Geo-Wiki - where the authoritative data sources are enhanced with open sources and the power of crowdsourcing is used for validation
29	How algorithmic popularity bias hinders or promotes quality	-	-
30	How many volunteers does it take to map an area well? The validity of Linus' law to volunteered geographic information	London	compare to the OrdnanceSurvey Meridian 2 data
31	Machine Learning Framework for the Estimation of Average Speed in Rural Road Networks with OpenStreetMap Data	two regions in Chile and Australia.	-
32	Mining urban land-use patterns from volunteered geographic information by means of genetic algorithms and artificial neural networks	urban areas in Europe	work with osm and Global Monitoring for Environment and Security Urban Atlas (GMESUA) data
33	On data quality assurance and conflation entanglement in crowdsourcing for environmental studies	Flood Inundation in London	-
34	OpenStreetMap Data Quality Assessment via Deep Learning and Remote Sensing Imagery	Las Vegas, Nevada, USA	-
35	OpenStreetMap data quality enrichment through awareness raising and collective action tools—experiences from a European project	4 cities in UK	-
36	Optimization of OpenStreetMap Building Footprints Based on Semantic Information of Oblique UAV Images	Germany	Compare optimized buidling footprint results with reference data from German Authority Topographic-Cartographic Information System(ATKIS)
37	Placing Wikimapia: an exploratory analysis	Moscow	-
38	Quality assessment of OpenStreetMap data using trajectory mining	-	-
39	Quality assessment of the contributed land use information from OpenStreetMap versus authoritative datasets	four German metropolitan areas	the pan-European GMESUA dataset as a reference

40	Quality evaluation of VGI using authoritative data—a comparison with land use data in Southern Germany.	Southern Germany	comparison of the OSM “naturals” dataset with the German authoritative ATKIS DLM dataset
41	staty: Quality Assurance for Public Transit Stations in OpenStreetMap	United Kingdom, the United States, Germany, Switzerland, and Austria.	-
42	Surveys on surveys: limitations and potentialities	Waterloo and Ontario, Canada	-
43	Temporal analysis on contribution inequality in OpenStreetMap: a comparative study for four countries	Germany, France, US, Netherlands	-
44	The ‘cottage effect’ in citizen science? Spatial bias in aquatic monitoring programs	Ontario, Canada	compare the accuracy of the samples against Ministry of Environment scientists, Conservation Authority scientists
45	The impact of biases in the crowdsourced trajectories on the output of data mining processes. Association of Geographic Information Laboratories in Europe (AGILE)	-	-
46	Towards automatic vandalism detection in OpenStreetMap	-	-
47	Tracking Editing Processes in Volunteered Geographic Information: The Case of OpenStreetMap	-	-
48	Using crowdsourced trajectories for automated OSM data entry approach	-	-
49	Using provenance to disambiguate locational references in social network posts	-	-
50	VGTrust: measuring trust for volunteered geographic information	Christchurch, New Zealand	case study was created to assess whether the VGTrust model could be successfully deployed to facilitate geographic data collection by government agencies
51	Why is participation inequality important?	-	-

11. APPENDIX C: OSM and Quality in Research – Literature

1. A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis. Christopher Barron. 2013. https://onlinelibrary.wiley.com/doi/full/10.1111/tgis.12073?

2. A Contributor-Focused Intrinsic Quality Assessment of OpenStreetMap in Mozambique Using Unsupervised Machine Learning. Aphiwe Madubedube. 2021. https://doi.org/10.3390/ijgi10030156

3. A grounding-based ontology of data quality measures. Franz-Benjamin Mocnik. 2017. https://digitalcommons.library.umaine.edu/josis/vol2018/iss16/1/

4. A Review of OpenStreetMap Data. Peter Mooney. 2017. https://library.oapen.org/bitstream/handle/20.500.12657/31138/637890.pdf?sequence=1#page=46

5. A review of volunteered geographic information quality assessment methods. Hansi Senaratne . 2017. https://www.tandfonline.com/doi/10.1080/13658816.2016.1189556

6. A Rule-Based Spatial Reasoning Approach for OpenStreetMap Data Quality Enrichment; Case Study of Routing and Navigation. Amin Mobasheri. 2017. https://www.mdpi.com/1424-8220/17/11/2498

7. A Systematic Study of OpenStreetMap Data Quality Assessment. Sukhjit Singh Sehra. 2014. https://ieeexplore.ieee.org/abstract/document/6822226?

8. A systems perspective on volunteered geographic information. Victoria Fast. 2014. https://www.mdpi.com/2220-9964/3/4/1278/htm

9. Accurate attribute mapping from volunteered geographic information: issues of volunteer quantity and quality. G.M.Foody. 2015. https://doi.org/10.1080/00087041.2015.1108658

10. Analysis of collaboration networks in OpenStreetMap through weighted social multigraph mining. Quy Thy Truong . 2018. https://www.tandfonline.com/doi/full/10.1080/13658816.2018.1556395

11. Assessing data completeness of VGI through an automated matching procedure for linear data. Thomas Koukoletsos. 2012. https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1467-9671.2012.01304.x

12. Assessing quality of volunteer crowdsourcing contributions: lessons from the cropland capture game. Carl F. Salk . 2015. https://www.tandfonline.com/doi/full/10.1080/17538947.2015.1039609

13. Assessing the impact of demographic characteristics on spatial error in volunteered geographic information features. William F. Mullen. 2015. https://link.springer.com/article/10.1007%2Fs10708-014-9564-8

14. Assessing the Quality of OpenStreetMap Contributors together with their Contributions. Jamal Jokar Arsanjani. 2013. https://agile-online.org/conference_paper/cds/agile_2013/short_papers/sp_s4.2_arsanjani.pdf

15. Assessing VGI data quality. Fonte, C C. 2017. https://scholar.google.com/scholar_lookup?hl=en&publication_year=2017&pages=137-163%29.+London%3A+Ubquity+press&author=C.+C.%2C+Antoniou%2C+V.%2C+Bastin%2C+L.%2C+Estima%2C+J.%2C+Arsanjani%2C+J.+J.%2C+Bayas%2C+J.+C.+L.%2C+See%2C+L.+and+Vatseva%2C+R.+Fonte&title=Assessing+VGI+data+quality.+In+Foody%2C+G.%2C+et+al.+%28eds.%29

16. Assuring the quality of volunteered geographic information. Michael F. Goodchild. 2012. https://www.sciencedirect.com/science/article/pii/S2211675312000097/pdfft?md5=c549236a4aa66ad19d575977d38b090d&pid=1-s2.0-S2211675312000097-main.pdf

17. Automated highway tag assessment of OpenStreetMap road networks. Musfira Jilani. 2014. https://dl.acm.org/doi/pdf/10.1145/2666310.2666476

18. Automated quality improvement of road network in OpenStreetMap. Musfira Jilani. 2013. http://flrec.ifas.ufl.edu/geomatics/agile2013/papers/jilani_ACTIVITY_AGILE_2013.pdf

19. Automatic analysis of positional plausibility for points of interest in OpenStreetMap using coexistence patterns. Alireza Kashian. 2017. https://doi.org/10.1080/13658816.2019.1584803

20. Charting the geographies of crowdsourced information in Greater London. Andrea Ballatore. 2018. https://scholar.google.com/scholar_lookup?hl=en&publication_year=2018&author=A.+Ballatore&author=S.+De+Sabbata&title=Charting+the+geographies+of+crowdsourced+information+in+Greater+London

21. Citizens as sensors: the world of volunteered geography. Michael F. Goodchild. 2007. https://link.springer.com/article/10.1007/s10708-007-9111-y

22. Comparing the quality of crowdsourced data contributed by expert and non-experts. Linda See. 2013. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0069958

23. Cooperation or competition–when do people contribute more? A field experiment on gamification of crowdsourcing. BenediktMorschheuser. 2018. https://www.sciencedirect.com/science/article/pii/S1071581918305822?via%3Dihub

24. Crowdsourced geospatial data quality: challenges and future directions. Anahid Basiri. 2019. https://www.tandfonline.com/doi/full/10.1080/13658816.2019.1593422

25. Crowdsourcing geospatial data. ChristianHeipke. 2010. https://www.sciencedirect.com/science/article/pii/S0924271610000602?via%3Dihub

26. Crowdsourcing: it matters who the crowd are. The impacts of between group variations in recording land cover. Comber, A.. 2016. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0158329

27. Do experts or crowd- based models produce more bias? Evidence from Encyclopædia Britannica and Wikipedia. Shane Greenstein . 2017. https://dash.harvard.edu/bitstream/handle/1/41946110/greenstein%2czhu_do-experts-or-crowd.pdf?sequence=1&isAllowed=y

28. Guided Classification System for Conceptual Overlapping Classes in OpenStreetMap. Ahmed Loai Ali. 2016. https://doi.org/10.3390/ijgi5060087

29. How algorithmic popularity bias hinders or promotes quality. Giovanni Luca Ciampaglia. 2018. https://www.nature.com/articles/s41598-018-34203-2

30. How many volunteers does it take to map an area well? The validity of Linus' law to volunteered geographic information. Mordechai Haklay. 2010. https://www.tandfonline.com/doi/pdf/10.1179/000870410X12911304958827?needAccess=true

31. Machine Learning Framework for the Estimation of Average Speed in Rural Road Networks with OpenStreetMap Data. Sina Keller. 2020. https://doi.org/10.3390/ijgi9110638

32. Mining urban land-use patterns from volunteered geographic information by means of genetic algorithms and artificial neural networks. Julian Hagenauer . 2011. https://doi.org/10.1080/13658816.2011.619501

33. On data quality assurance and conflation entanglement in crowdsourcing for environmental studies. Didier G. Leibovici. 2017. https://www.mdpi.com/2220-9964/6/3/78

34. OpenStreetMap Data Quality Assessment via Deep Learning and Remote Sensing Imagery. Xuejing Xie. 2019. https://doi.org/10.1109/ACCESS.2019.2957825

35. OpenStreetMap data quality enrichment through awareness raising and collective action tools—experiences from a European project. Amin Mobasheri. 2017. https://www.tandfonline.com/doi/full/10.1080/10095020.2018.1493817

36. Optimization of OpenStreetMap Building Footprints Based on Semantic Information of Oblique UAV Images. Xiangyu Zhuo. 2018. https://doi.org/10.3390/rs10040624

37. Placing Wikimapia: an exploratory analysis. Andrea Ballatore. 2018. https://www.tandfonline.com/doi/epub/10.1080/13658816.2018.1463441?needAccess=true

38. Quality assessment of OpenStreetMap data using trajectory mining. Anahid Basiri . 2016. https://www.tandfonline.com/doi/full/10.1080/10095020.2016.1151213

39. Quality assessment of the contributed land use information from OpenStreetMap versus authoritative datasets. Jamal Jokar Arsanjani. 2015. https://link.springer.com/chapter/10.1007/978-3-319-14280-7_3

40. Quality evaluation of VGI using authoritative data—a comparison with land use data in Southern Germany.. Helen Dorn. 2015. https://www.mdpi.com/2220-9964/4/3/1657/htm

41. staty: Quality Assurance for Public Transit Stations in OpenStreetMap. Hannah Bast. 2020. https://doi.org/10.1145/3397536.3422342

42. Surveys on surveys: limitations and potentialities. John Goyder . 1986. https://academic.oup.com/poq/article/50/1/27/1821385?login=true

43. Temporal analysis on contribution inequality in OpenStreetMap: a comparative study for four countries. Anran Yang. 2016. https://www.mdpi.com/2220-9964/5/1/5

44. The ‘cottage effect’ in citizen science? Spatial bias in aquatic monitoring programs. Edward E. Millar. 2018. https://www.tandfonline.com/doi/10.1080/13658816.2018.1423686

45. The impact of biases in the crowdsourced trajectories on the output of data mining processes. Association of Geographic Information Laboratories in Europe (AGILE). Anahid Basiri. 2018. http://www.cs.nuim.ie/~pmooney/vgi-alive2018/papers/1.3.pdf

46. Towards automatic vandalism detection in OpenStreetMap. Pascal Neis. 2012. https://www.mdpi.com/2220-9964/1/3/315

47. Tracking Editing Processes in Volunteered Geographic Information: The Case of OpenStreetMap. Carsten Keßler. 2011. https://www.carsten.io/papers/iope2011.pdf

48. Using crowdsourced trajectories for automated OSM data entry approach. Anahid Basiri. 2016. https://www.mdpi.com/1424-8220/16/9/1510

49. Using provenance to disambiguate locational references in social network posts. Thomas Hervey. 2018. https://www.tandfonline.com/doi/full/10.1080/13658816.2018.1459627

50. VGTrust: measuring trust for volunteered geographic information. Jeremy Severinsen . 2019. https://www.tandfonline.com/doi/10.1080/13658816.2019.1572893

51. Why is participation inequality important?. Mordechai Haklay. 2016. https://www.ubiquitypress.com/site/books/10.5334/bax/download/942/#page=46

12. APPENDIX D: Approaches and Tools for Integrating Data into OSM

Introduction

The ever-increasing availability of Open Data allows organization-external actors to interact with the data and create useful applications. More and more public administrations (here called: organizations), realize this and want to integrate their data (here: original data) into OpenStreetMap (OSM). This document is intended to show these organizations what approaches and tools are available to integrate their original data into OSM.

Open Data is non-security and non-personal data that is freely available and usable. By non-personal data, we also mean formerly personal data that has been anonymized.

We assume that people are already familiar with OSM and that licensing issues have been resolved (cf. Fairhurst 2020). If not, it is advisable to refer to the literature and to sonsult LearnOSM, OpenSchoolMaps and the original OSM wiki pages.

- Useful information can also be found in a "Integrating data into OpenStreetMap - a guide for Swiss public authorities and federal enterprises" by the Digital Sustainability Research Unit at the University of Bern.
- For those interested in how organizations are using OpenStreetMap - such as emergency response organizations, blue light organizations (police, fire, and ambulance and rescue services), mobility services, tourism, and GIS agencies, the study "Public-OpenStreetMap Partnership" (POP Study) is recommended (Keller 2021).
- If you want to learn more about the availability of open geodata, take a closer look at the educational course at OST.

In the next chapter “Four Approaches to Integrating Open Data in OSM” are presented. Of these, the ‘managed’ approach is the most important, which is followed by a separate chapter. The rest of this document is dedicated to this managed ’approach with the following three tool’ chapters:

- Data integration coordination tools
- General tools for integrating data into OSM; and
- Actions and tools for targeted integration of data into OSM.

This is followed by two chapters on

- Update of OSM with original data, as well as for
- Monitoring of OSM data and comparison with original data.

Four Approaches to Integrating Open Data into OSM

As Fairhurst writes, there are four approaches an organization can take to integrate their open, and license-compliant original data into OSM:

1. The **'Unmanaged' Approach**, where the organization makes its original data available to the OSM community and invites its members to integrate that original data.
2. The **'Managed' Approach**, where the organization let its original data to be processed into a format that is suitable for use by the available tools.
3. The **'Bulk Import' Approach**: another approach is to insert the original data programmatically. It is not recommended. In any case, these "Import Guidelines" should be followed.

4. The **"Reference" Approach**: Finally, in consultation with the OSM community, you can decide that the original data is not suitable for direct inclusion in OpenStreetMap. That does not mean the data is useless to OSM. Read more about this in Fairhurst (1).

The following is a introduction of the 'Managed' Approach, which was as Approach No.2 mentioned above.

OpenStreetMap is a loose community of mappers who act out of pure self-motivation. This document respects that self-motivation and expects the reader of this document with know-hows to appreciate the same.

The 'Managed' Approach to integrating data into OSM.

With this approach, the organization converts its original data into a data format that is ready for integration into OSM. This means, for example:

- Setting the appropriate OSM tags from the original attributes.
- Remove all data that is already in OSM to avoid duplicate entries in OSM.
- Ensure that no multiple line geometries exist and lines are not overly detailed (< 1m point distance).

This preprocessing can be done by a script written specifically for this purpose. What is difficult to detect automatically are overlaps or topological breaks (note that OSM data is routable).

The data can then be made available to the OSM community for manual iteration ("conflation") - i.e. the processed original data is entered into OSM piece by piece.

There are a few software tools that are used to coordinate users, and there are those that help users integrate the original data into OSM. These tools are described in the next two chapters.

All of the tools are open source. Some of them are free, online usable web applications - mostly with OSM login, others are on-premise web applications or desktop applications and need to be installed first.

Tools for coordinating data integration

The following tools are useful for community or user coordination:

1. **OSM Tasking Manager**: Web application that divides the area of the original data into manageable spatial processing units (grids or polygons). Users select a processing unit and mark it as 'completed'.
   - Website: https://tasks.osm.ch/ (CH) with OSM login.
   - Setup: Hosting by SOSM (or HOTOSM). Deploying the configuration (perimeter as GeoJSON) is relatively simply, but it is usually done by experts.
   - Limitations: To start, one must first interact with the operator and send them the configuration.
2. **MapCraft**: Similar to Tasking Manager, this allows to split tasks into arbitrary areas.
   - Website: https://mapcraft.nanodesu.ru/
   - Setup: Webapp with OSM login.
   - Limitations: Less well known than OSM Tasking Manager.
3. **Damn project** ("Divide and Map. Now.") - Also similar to OSM Tasking Manager.

- Website: https://www.damn-project.org/
- Setup: with a few clicks.
- Limitations: Still a young one-man project (late 2020).

4. **MapRoulette:** This web application gives the user similar, small editing tasks by repeatedly showing the map view with the coordinates of the potential OSM object that needs to be adjusted. The user then decides whether to accept or skip the editing task. If he accepts the task, his OSM editor opens the mentioned location.
   - Website: https://maproulette.org/ with OSM login.
   - Setup: Ready after uploading a GeoJSON with point coordinates, which usually needs to be created by experts.
   - Limitations: Only location is given, no tags; knowledge of proper tags must come from users.

General tools for integrating data into OSM

First of all, we should note that the main user tool is an editor. We have mentioned **JOSM-Editor** in the introduction and embedded **iD-Editor** in OSM.org. There are other editors like Vespucci or edit functions that are built into Maps.me.

Good editors help - and guide(!) - when entering geodata, including assigning tags. This is done with tagging templates (defaults), which are called "presets" in OSM jargon. The tagging templates of the iD-Editor can be found here in the "iD Tagging Schema", those of JOSM here, and those of Vespucci here.

Below is a list of other tools, some of which require special knowledge to use and install (here starting with the tool that can be configured ad-hoc by end users themselves):

1. **QGIS plugin Go2NextFeature3+:** A plugin similar to MapRoulette, where QGIS is a desktop application. It is not suitable for coordinating multiple users like MapRoulette. On the other hand, it has the advantage of being quick to set up, and it allows user to copy and paste multiple attribute values of the original data, i.e. copy multiple tags into the editor at once. Together with the plugin "Lat Lon Tools", with which OSM (or another external map) can be called directly with one click, this results in an efficient workflow.The workflow works is described in "Validating and Integrating Local Datasets in OpenStreetMap using QGIS". There is an introduction to QGIS on OpenSchoolMaps.
   - Website Gitlab: Repository 'Go2NextFeature3+' (further development of Go2NextFeature3).
   - Setup: Immediately usable after installation from repository directly from QGIS. Users can configure themselves with a few clicks.
   - Limitations: Only for "single user" operation.
2. **OSM Conflator (with audit)**: This is a web application and a command line script for basic merging tasks. Script and web application have to be programmed, respectively hosted by user-selves.
   - Website Github: https://github.com/mapsme/osm_conflate, (demo: http://audit.osmz.ru/project/demo).
   - Setup: tbd.
   - Limitations: tbd.
3. **Live Conflation**: This is a web application that provides a GUI for fusion tasks. It has to be self-hosted. The original data is first converted into vector tiles (Mapbox MVT format). By calling the OSM API, the closest and most similar (tags) candidates are presented first. Once the user has merged the desired tags, the changes can be uploaded directly through the OSM API.
   - Website Github: https://github.com/systemed/conflation

- Setup: tbd.
- Limitations: tbd.

4. **JOSM Conflation plugin**: tbd.
   - Website: https://josm.openstreetmap.de/
   - Setup: Installation of the plugin from JOSM.
   - Limitations: For "single user" operation only.
5. **Other tools**: E.g. Cygnus or Hootenanny, see Fairhurst (2020).

Actions and tools for targeted integration of data in OSM.

Given the fact that the most important objects in OpenStreetMap - such as roads and buildings especially in urban areas - are more and more completely captured, the interest of mappers tends to shift. On one hand, this leads to "micromapping," i.e., capturing details (e.g., the color and material of a bench), and on the other hand, to targeted capture.

Actions for targeted data integration

This targeted capturing can take different forms:

- Actions with **event** character, namely "Mapping Parties" (which how it all started!).
- Actions with a **lottery** character, namely "Project of the Month" (see below).
- Actions on a **contractual basis**. This refers to the commissioning of people (employees) to integrate data into OSM. The "Organised Editing Guidelines" must be observed. These "Organised Editing Guidelines" should not be confused with the "Import Guidelines" mentioned at the beginning.

The "Project of the Month (Switzerland)" is a planned monthly action, together with a corresponding web application of the same name for the targeted collection of OSM objects. The web application includes game-like elements with ranking and simple capture statistics: see "Project of the Month" (PotM). The web application developed in France is based on projects such as the German "Focus of the Week" or carries out avtivities regularly from Wikipedia (see e.g. Wiki Loves Switzerland).

Tools for targeted data integration

The first to mention are (mono-)thematic editors, which are called "targeted" because they facilitate the capture of a specific topic. The two best known editors mentioning below deal with businesses details, such as restaurants. Both are free webapps with OSM login:

- **On OSM**: This webapp is very simple: It inserts OSM Notes only (no direct modification in OSM) and it can actually only insert new businesses; existing business entries cannot be updated properly. Website onosm.org.
- **OpenStreetMap My Business (OSMyBiz):** The webapp is inspired by On OSM, it inserts a new node directly in event of a new entry. In event of a update, only one OSM note will be saved, but with an indication of what has changed. The app remembers the transactions of users and inform them when their entry is modified by others. Website OSMyBiz.
- **Others**: e.g. MapContrib.

Following is a list of few existing tools that contain playful elements, so-called "serious gaming" or "gamification":

- **StreetComplete**: Intuitive mobile app.
  - StreetComplete website (documentation).
  - Setup: Download mobile app from Google Play Store.
  - Configuration: Possible.

- Limitations: Mobile phones with Android only.

- **MapComplete**: Manages a single theme, can insert new nodes and update existing ones. MapComplete allows even tech-savvy users to create their own "MapComplete themes" (quests). Here as an example the "Free Artwork Map".
  - MapComplete website.
  - Setup: Free webapp with OSM login.
  - Configuration: Possible (here a JSON-Konfig.-Datei for opening hours during Covid19).
  - Limitations: This project is still very new and it has potential. It updates the version every time when a tag is changed, i.e. if five tags are changed, the version of the OSM object increases by +5.

    NOTE: This webapp is not recommended for the time being, as it increases the version of an OSM object for every question answered (instead of increasing the version only once per object).
- **Others**: e.g. MapSwipe.

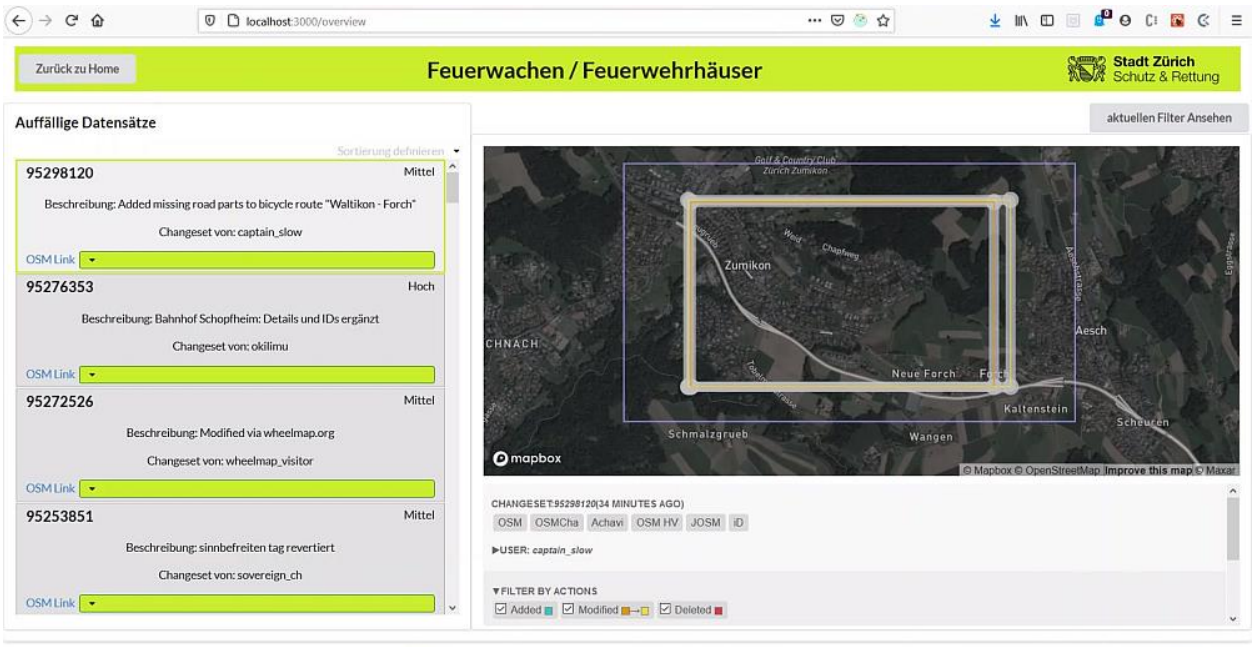
Update of OSM with original data

The "managed" approach discussed here is primarily concerned with one-time integration. Once integrated, it is left to the mappers to update the data where necessary. Experience shows that this works quite well! However, there is no way to tell how often and how quickly the data is updated.

If an organization wants to ensure that its data is updated within a specific period, then they have no choice but to hire people to do the editing with a fee. Please refer to the Organised Editing Guidelines mentioned above.

We recommend every interested authority or organization which wants to have its data entered into OSM, to set up a central **point of contact** ("Point-of-Contact") for external and internal inquiries. Some companies already do it that way.

If the authority or organization follows the 'Managed' Approach, it is advisable to provide **moderators** ("facilitators") who actively accompany and/ or organize actions.



Dialog in the "Targeted Monitoring Tool" (draft) showing conspicuous edits of fire stations within LRC's territory.

Monitoring of OSM data and comparison with original data

Monitoring (filtered) OSM data:

1. **OSMCha**: tbd.
   - Website documentation developed by Mapbox (possibly soon to be run by OSM US Chapter): https://wiki.openstreetmap.org/wiki/OSMCha
   - Setup: Webapp with OSM login.
2. "**Targeted Monitoring Tool**": A webapp for targeted monitoring of OpenStreetMap data based on an extended OSMCha backend. A first prototype was developed by Geometa Lab for SRZ (license: ISC).
   - Website documentation: https://github.com/Schutz-Rettung-Zurich/srz-edi/
   - Setup: tbd.
3. (**OSM History DB of Switzerland** (osmhistorydb-ch): This is a project to implement a geospatial database containing the history (lineage) of OSM data objects, initiated by members of the SOSM community. (License: MIT))
   - Website/documentation/repository: https://github.com/sosm/osmhistorydb-ch.

Matching with original data:

1. **QGIS plugin OSM Data Sync**: A "stand-alone" plugin based on the QGIS desktop application. How the plugin works is described in this document "Synchronizing Local Datasets and OpenStreetMap using QGIS". There is an introduction to QGIS on OpenSchoolMaps.
   - Website: Gitlab 'OSMDataSync'.
   - Setup: Immediately usable by installing it from QGIS. Users can do the configurartion themselves with a few clicks. May need an expert to help with the appropriate overpass query.
   - Limitations: Only for "single user" operation. Suitable for object sets smaller than 5000 (Overpass Query).

Conclusion

This document tries to show which approaches and tools are available for organizations, especially public administrations, to integrate their original data into OSM.

By far the most important tools are universal editors like iD and JOSM etc. Apart from these, the following tools are popular - depending on the purpose and without claim to completeness:

- **MapRoulette** - For targeted integration of original data and updating OSM.
- The QGIS plugin **Go2NextFeature3**+ - For quick, step-by-step/one-by-one comparison of original data and OSM.
- The **OSM Tasking Manager** - for organized and targeted integration of original data into OSM.

New additions include tools that support the "Managed" Approach, such as the "**Project of the Month**". An insightful example of this is the "Projet du mois Septembre 2020" website, created by the French OSM community to capture defibrillators in OSM.

Those who are not sure which tool to take, how to proceed, or which approach is now the most appropriate can seek advice from the local OSM community or OSM experts.