

Qualitätsmanagement bei OpenStreetMap

Jochen Topf

jochen.topf@geofabrik.de

Paper aus dem Tagungsband zur FOSSGIS 2009 (17. bis 19. März 2009, Hannover)

Die Erfassung von Geodaten für ein Projekt beginnt klassischerweise mit einer Analyse, welche Daten man braucht. Man wird festlegen, welche Struktur die Daten haben sollen, eine Klassifikation der Geodaten festlegen, Attribute vorgeben und so weiter. Passend zur geplanten Nutzung der Daten kann man festlegen, welchen Qualitätskriterien die Daten genügen müssen: Wie vollständig müssen sie sein? Welche Fehler in der Lagegenauigkeit oder Attributierung können toleriert werden?

Ganz anders ist der Ansatz beim OpenStreetMap-Projekt [1, 2]. Es gibt keine Vorgaben, wie genau eine geographische Position ermittelt werden muss, keine abschließende Liste von zu erfassenden Features oder deren Attributen. Die einzige Vorgabe ist die grobe Richtung: Wir wollen eine weltweite Karte schaffen. Darunter versteht natürlich jeder etwas anderes, und trotzdem hat OpenStreetMap etwas geschaffen, dass sich sehen lassen kann.

Hintergrund

OpenStreetMap geht von ganz anderen Voraussetzungen aus als ein typische GIS-Projekt:

1. Es gibt keine im vornherein festgelegte Nutzung der Geodaten. Stattdessen sollen die Daten so flexibel wie möglich genutzt werden können. Dadurch ist es sehr schwierig, allgemeine Qualitätskriterien festzulegen.
2. Jeder Beteiligte hat eine andere Vorstellung davon, was er erreichen will. Natürlich gibt es auch viele Übereinstimmungen, zum Beispiel sehen wohl alle ein, dass es wichtig ist, die Straßengeometrien und -Typen zu erfassen. Aber welche Typen es denn nun genau geben soll und ob es wichtiger ist, Briefkästen zu erfassen oder Geschwindigkeitsbeschränkungen oder Waldflächen, da gehen die Meinungen schon weit auseinander.
3. Es steht eine fast unbegrenzt erscheinende Menge an „kostenloser Arbeitskraft“ zur Verfügung. Die OSM-Aktiven „Mapper“ beteiligen sich aus Spaß am Projekt. Solange man es schafft, sie zu motivieren, tun sie die Arbeit gerne und kostenlos. Schafft man keine geeigneten Anreize, wird auch keine Arbeit getan.

Nur vor diesem Hintergrund kann man verstehen, wie OSM funktioniert und wie Qualitätskontrolle und -verbesserung funktionieren kann.

Motivation der Mapper

Die „Mapper“ (OSM-Aktiven) werden von den verschiedensten Motivationen getrieben. Manche wollen bestimmte Kartendaten haben, die anderweitig nicht erhältlich sind. Viele haben einfach Spass am Datensammeln oder finden Karten irgendwie interessant. Aber egal, welche Gründe sie zu OSM geführt haben, ein ganze entscheidendes Kriterium für ihre kontinuierliche Mitarbeit ist der sichtbare Erfolg. Wenn sie nicht sehen können, was mit ihren Daten passiert, dann verlieren sie schnell den Spass.

Der typische Erfolg ist natürlich die Sichtbarkeit der Eingaben auf der Karte. Werden einige Geodaten, die der Mapper eingeben hat, auf der Karte nicht sichtbar, wird er die weitere Eingabe solcher Daten vernachlässigen und eher solche Daten eingeben und verbessern, bei denen er seine Fortschritte auf der Karte verfolgen kann.

Alle Mapper orientieren sich dabei in erster Linie natürlich an den auf der Projektwebseite sichtbaren Karten. Dazu kommen aber inzwischen auch mehr und mehr Spezialkarten:

Die Cyclemap [3] hebt zum Beispiel Fahrradrouten hervor. Die ÖPNV-Karte [4] zeigt Bus- und Bahn- und U-Bahnlinien an. Inzwischen erscheinen immer mehr solcher Karten, die sich auf bestimmte thematische Bereiche konzentrieren. Da jeder die OpenStreetMap-Daten bekommen kann und die Software zur Erstellung von Karten typischerweise Open Source ist, ist die Hürde relativ klein, eine Karte nach den eigenen Interessen zu erstellen. Solche Karten sind dann ein mächtiger Anreiz für andere, die Daten auch so zu erfassen, dass sie auf der Karte erscheinen.

Entstehen eines Mapping-Konsens

In einem klassischen GIS-Projekt wird typischerweise von einer Person oder einer kleinen Gruppe von Personen festgelegt, welche Daten wie zu erfassen sind. Eine abstrakte Beschreibung der Daten und ihrer Modellierung ermöglicht es dann einer größeren Gruppe, die eigentliche Erfassung nach den festgelegten Regeln durchzuführen. Die gleiche abstrakte Beschreibung ermöglicht dem Nutzer die Interpretation der Daten und damit ihre sachgerechte Nutzung.

Bei OpenStreetMap gibt es keine kleine Gruppe, die vorher alles festgelegt hat. Jeder Mapper ist im Prinzip gleich und kann erfassen und nutzen, was und wie er will. Es gibt a priori keine abstrakte Beschreibung der Daten, sie ergibt sich erst als Nebenprodukt der Arbeit der OSM-Aktiven. Dabei kann man zwei Gruppen unterscheiden:

Auf der einen Seite stehen die Mapper, die die Daten erfassen oder importieren. Sie entscheiden, welche Features sie erfassen und wie sie sie attributieren („taggen“). Da sie gerne wollen, dass „ihre“ Daten auf der Karte erscheinen, lassen sie sich bei der Entscheidung, was sie wie erfassen, davon leiten, was wie auf der Karte erscheinen wird.

Auf der anderen Seite stehen die Programmierer und Designer der Karten. Sie entscheiden, welche Features mit welchen Attributen (Tags) auf ihren Karten wie erscheinen. Da sie gerne wollen, dass möglichst viele Daten richtig auf ihrer Karte erscheinen, lassen sie sich bei der Entscheidung, was sie wie darstellen, davon leiten, was wie erfaßt wurde.

Jeder schaut also auf das, was die andere Partei macht, und versucht gleichzeitig seine Agenda durchzusetzen, versucht also vor allem die Daten auf die Karte zu bekommen, die ihm wichtig sind. So entsteht in einer Art „Verhandlung“ ein Konsens, was denn wie zu erfassen ist und wie es dann dargestellt oder anderweitig genutzt wird. Der Prozess kommt ohne a priori-Festlegungen aus, er funktioniert zum großen Teil sogar, ohne dass der Mapper und der Nutzer der Daten je direkt miteinander reden müssen. Die „Verhandlung“ findet typischerweise nicht wirklich im Sinne einer Aussprache statt, sie ergibt sich einfach aus den Handlungen der Beteiligten, die gegenseitig auf die Handlungen ihres Gegenübers reagieren. Der Vorgang ähnelt daher der biologischen Evolution, in der Lebewesen und Umwelt, Jäger und Beute sich gegenseitig beeinflussen und dadurch gemeinsam weiterentwickeln.

Natürlich wäre das ganze sehr ineffizient, wenn der Mapper ständig schauen muss, was wie auf vielen verschiedenen Karte erscheint und der Kartendesigner ständig durch die Daten schauen muss, um zu sehen, was er wie darstellen könnte. Es gibt daher im OpenStreetMap-Wiki eine Dokumentation, was wie zu erfassen ist. Diese Dokumentation ist aber lückenhaft und notorisch veraltet. Sie taugt nur in Ansätzen als abstrakte Beschreibung.

Wir haben häufig das Bild im Kopf, dass es eine perfekte abstrakte Beschreibung der Welt gibt, die wir nur noch entwickeln müssen. Danach läßt sich dann alles erfassen. In der Praxis ist es aber trotz aller Aufforderungen „nicht für den Renderer zu mappen“, also bei der Erfassung der Daten nicht nur die Visualisierung der Daten im Auge zu haben, sondern ein abstraktes, theoretisches Modell, doch so, dass man ständig auf das Ergebnis schießt und seine Arbeit danach anpasst.

Will man den oben beschriebenen Vorgang der „Verhandlung“ oder „Evolution“ beeinflussen, um seine Ziele durchzusetzen, zum Beispiel weil man bestimmte Daten für eine Spezialkarte haben möchte, so kann man auf beiden Seiten der Medaille eingreifen: Entweder erfaßt oder importiert man die Daten in einem für die eigenen Zwecke geeigneten Schema und wartet darauf, dass ein Kartendesigner die Daten umsetzt oder man macht selbst eine Karte, die Daten in diesem Schema anzeigt und wartet darauf, dass die Mapper die passenden Daten sammeln und erfassen.

Nicht-Karten-Anwendungen

In den letzten Abschnitten war ständig von Karten die Rede. Karten sind aber natürlich nur das sichtbarste Produkt der OSM-Datensammlung. Geodaten sind für viele weitere Zwecke nutzbar, insbesondere natürlich für die automatisierte Navigation, die in den letzten Jahren so wichtig geworden ist. Aber es sind viele weitere Anwendungen denkbar.

Damit der oben erklärte Mechanismus funktionieren kann, müssen diese Anwendungen aber für die Mapper

sichtbar werden. Sieht der Mapper nicht, was mit seinen Daten passiert, kann er sie auch nicht richtig erfassen. Und das ist in den meisten Fällen mit einer Karte eben am einfachsten zu erledigen. Aber auch jede andere Möglichkeit, dem Mapper zu zeigen, wie seine Arbeit genutzt wird, kommt in Frage.

Qualitätskontrolle und Verbesserung der Qualität

Vor dem beschriebenen Hintergrund ist jede Art der Qualitätskontrolle natürlich schwierig. Es gibt keine festen Kriterien, nach denen so eine Kontrolle stattfinden könnte. Und ohne eine Messung der Qualität kann auch keine Verbesserung stattfinden.

Aber auch hier hilft der verteilte „evolutionäre“ Ansatz uns wieder: Jeder, der die Daten nutzen will, kann für sich Qualitätskriterien definieren. Und wenn er in der Lage ist, die Qualität der Daten nach diesen Kriterien zu messen und zu visualisieren, so ergibt sich dadurch unter Umständen ein Ansatz für die Mapper, die Daten nach diesen Kriterien zu erfassen.

Ein Beispiel soll diesen Ansatz demonstrieren: Es ist sicher wichtig, dass für alle Straßen, die einen Namen haben, dieser Name auch erfasst wird. Nun können wir nicht sicher sein, ob eine Straße überhaupt einen Namen hat, aber zumindest bei Wohnstraßen können wir davon ausgehen, dass das der Fall ist. Wir wollen uns daher auf diesen Fall konzentrieren. Wir definieren also die Forderung „Alle Wohnstraßen (in OSM sind das die mit dem Tag „highway=residential“ oder „highway=living_street“) müssen einen Namen haben“. Der Anteil solcher Straßen, die einen Namen haben, gibt uns eine Maßzahl, wie gut dieses Kriterium erfüllt ist¹.

Wollen wir die Qualität der OSM-Daten in diesem Punkt erhöhen, so müssen wir das Problem visualisieren, um dem Mapper zu ermöglichen, die Daten nachzuerfassen. Und tatsächlich ist das in diesem Fall auch verschiedentlich geschehen. Es gibt mehrere „No-Name-Karten“, die Straßen ohne Namen besonders hervorgehoben darstellen.

Wettbewerb

Es gibt noch einen weiteren Aspekt, der in anderen GIS-Projekten keine solche Rolle spielt: Für die Mapper ist OSM ein Hobbyprojekt und in gewisser Weise auch eine Art Spiel. Bei dem Spiel „kämpft“ man nicht nur gegen die Wirklichkeit, die man möglichst gut erfassen will, sondern auch gegen die anderen Mit-Mapper. Ein gesunder Wettbewerb, bei dem jeder gerne „seinen“ geographische Bereich möglichst gut abgedeckt haben möchte, gibt weiteren Anreiz, Daten geeignet zu erfassen. Wenn man sieht, dass auf der „No-Name-Karte“ des Nachbarortes keine Straßen mehr „rot“ sind, dann steigt der Anreiz, auch im eigenen Ort für „Ordnung“ zu sorgen.

Werkzeuge

Über die Zeit wurden für OpenStreetMap verschiedene Werkzeuge entwickelt, die zur Qualitätsverbesserung beitragen sollen. Sie helfen Probleme zu finden, zu visualisieren oder zu beseitigen. Im folgenden werden einige solcher Tools vorgestellt.

OpenStreetBugs

Nicht jeder will sich in OSM einarbeiten und lernen, wie er selbst Daten erfassen oder korrigieren kann. Aber auch der „Mann auf der Straße“ kann natürlich Informationen beitragen, wo etwas auf der Karte fehlt oder fehlerhaft ist. Für ihn gibt es OpenStreetBugs [5], eine Webanwendung, über die jeder durch einen Klick auf die Karte Fehlerstellen markieren und mit einem Kommentar versehen kann. OSMer können diese Informationen dann nehmen und die zugrundeliegenden Daten kontrollieren und korrigieren. Diese Art der Qualitätskontrolle ist zwar sehr unsystematisch, aber dafür einfach und von jedem einzusetzen.

¹ Im Februar 2009 hatten von ca. 2 Mio. Wohnstraßen in Europa etwa 1,5 Mio einen Namen, also etwa 75%.

Maplint

Im Januar 2007 entstand das Programm Maplint [6] des Autors. Es war einer der ersten Versuche, systematisch Fehler und Inkonsistenzen in den Daten zu entdecken und zu visualisieren. Maplint liest OSM-Dateien, erkennt bestimmte Probleme und gibt die OSM-Datei angereichert mit „Fehlermeldungen“ wieder aus. Über eine spezielle Konfiguration für den Kartenrenderer Osmarender können diese dann in Karten umgesetzt werden, in denen Fehler besonders hervorgehoben sind. Maplint erkennt zum Beispiel Geometrien ohne jegliche Tags (Attribute), Autobahnen ohne Nummer, Kirchen ohne angegebene Religion, Wohnstraßen ohne Namen und anderes. Maplint wird auch heute noch eingesetzt, ist aber wegen der technologischen Basis (es basiert auf XSLT-Skripten) zu langsam in der Benutzung und schwierig in der Wartung, dass es nicht mehr weiterentwickelt wird.

Validator

Entsprechend der Philosophie von OSM, dass jeder alles mappen kann und es möglichst wenig Einschränkungen gibt, erlaubt der vielfach benutzte OpenStreetMap-Editor JOSM auch die Eingabe von „unsinnigen“ und fehlerhaften Daten. Er führt selbst nur wenige absolut notwendige Konsistenzchecks durch.

Der Validator [7] ist ein Plugin für JOSM, das viele weitere Checks anbietet. Er zeigt Fehler und Inkonsistenzen im Editor in Tabellenform und in der Kartenansicht an. Teilweise kann er Fehler auch auf Knopfdruck direkt korrigieren. Er führt aber niemals automatisch eine Änderung an den Daten durch. So bleibt es immer dem Menschen überlassen, zu entscheiden, ob er vielleicht in diesem Fall doch von der Einschätzung der Maschine abweichen will.

Der Validator erkennt zum Beispiel, wenn sich zwei Straßen kreuzen, ohne dass entweder ein Kreuzungspunkt gesetzt ist oder zumindest eine der Straßen als Brücke oder Tunnel attribuiert ist. Oder er kann erkennen, wenn eine Straße einen sehr ähnlichen Namen hat wie eine andere Straße in der Nähe. Vielleicht handelt es sich ja um einen Tippfehler und es sollte sich um die gleiche Straße handeln.

Der Validator ist ein unverzichtbares Hilfsmittel für die Mapper. Er wird immer wieder um neue Tests erweitert.

OSM Inspector

Seit dem Herbst 2008 existiert der OSM Inspector [8], der wie Maplint vom Autor entwickelt wurde und in gewisser Weise eine Art Nachfolger darstellt. Der OSM Inspector ist eine Webanwendung, die einfach von jedem bedient werden kann. In verschiedenen Views (Ansichten) werden thematisch unterschiedliche Spezialkarten angeboten, die jeweils einen Aspekt der OSM-Daten anzeigen und dabei insbesondere auch fehlerhafte oder inkonsistente Daten hervorgehoben darstellen.

So gibt es zum Beispiel einen Address View, in dem Postadressen von Gebäuden und Grundstücken besonders hervorgehoben werden. Fehlt die Straße zu einem Gebäude, das eine Adresse an dieser Straße hat, wird das entsprechend hervorgehoben.

Ein anderer View zeigt Geometriefehler wie sich selbst überschneidende Polygongrenzen (siehe Abb. 1), ein weiterer zeigt möglicherweise falsch geschriebene Tags.

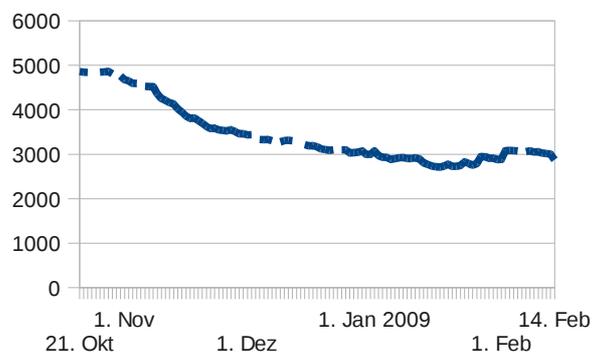


Abb 1: Die Anzahl der Geometriefehler nimmt ab nach Einführung des OSM Inspectors Ende Oktober 2008, obwohl die Anzahl der Geometrien ständig weiter wächst

Straßenlisten

Einige Fehler lassen sich nur durch die Analyse von OSM-Daten erkennen, aber für viele ist ein Rückgriff auf

fremde Daten nötig. Zum Beispiel läßt sich aus dem OSM-Daten alleine natürlich nie feststellen, wie vollständig das Straßennetz ist. Viele Gemeinden sind allerdings bereit, eine Liste aller Straßennamen auf ihrem Gebiet zur Verfügung zu stellen. Diese kann dann gegen eine aus OSM-Daten erzeugte Liste abgeglichen werden. Gibt es Abweichungen, so liegt auf der einen oder anderen Seite ein Fehler vor, dem man nachgehen muss.

Für viele Orte wurden und werden solche Vergleiche durchgeführt und dann mehr oder weniger systematisch durchgegangen. In Hamburg wurde so eine Übereinstimmung von 99,8% mit der offiziellen Straßenliste erreicht [9].

Fazit

Qualitätskontrolle und -verbesserung wird notwendigerweise im OpenStreetMap-Projekt ganz anders angegangen als das in klassischen GIS-Projekten der Fall ist. Es gibt einige Ansätze und Werkzeuge, die in Teilbereichen bereits sehr erfolgreich sind, aber noch weiterentwickelt werden müssen. Eine allgemeine und systematische Qualitätskontrolle findet aber nicht statt.

Literatur und Webseiten

[1] <http://www.openstreetmap.org/>

[2] Frederik Ramm, Jochen Topf: OpenStreetMap: Die freie Weltkarte nutzen und mitgestalten, Lehmanns Media, Berlin, 2. Auflage 2009.

[3] <http://www.opencyclemap.org/>

[4] <http://www.öpnvkarte.de/>

[5] <http://www.openstreetbugs.org/>

[6] <http://wiki.openstreetmap.org/wiki/Maplint>

[7] <http://wiki.openstreetmap.org/wiki/JOSM/Plugins/Validator>

[8] <http://tools.geofabrik.de/osmi/>

[9] <http://www.openstreetmap.de/presse/2008-10-24-hamburg-stat.html>

