

## History-Informationen in der OSM Datenbank, Reloaded

Peter Körner

Jede Änderung an jedem Objekt wird gespeichert und ist öffentlich. Doch wie kommt man an diese Daten heran? Welche Quellen gibt es und wofür sind sie geeignet? Welche Probleme ergeben sich beim Auswerten dieser Informationen und wie kann man diese überwinden? Welche tollen und nützlichen Anwendungen lassen sich mit diesen Daten bauen und wie können solche Daten der Community helfen?

### ***Begriffsklärung:***

In der Diskussion über die Vergangenheit in OpenStreetMap kommt es häufiger zu Missverständnissen. Die Meisten rühren daher, dass es verschiedene Zeitlinien gibt, die miteinander nicht verbunden werden können. Insbesondere muss zwischen folgenden „Vergangenheiten“ unterschieden werden:

- Die echte Vergangenheit manifestiert sich beispielsweise in der Frage, wann ein Geschäft eröffnet hat und wann es wieder schloss
- Die Vergangenheit des OpenStreetMap-Projektes welche sich beispielsweise auf Änderungen in der Gestaltung der Karte auf [openstreetmap.org](http://openstreetmap.org) bezieht
- Vergangene Versionen von Daten die in Form von Versionen als Teil der OpenStreetMap-Datenbank betrachtet werden

Der Vortrag bezieht sich auf die letzte Art von Vergangenheit.

### ***Fragestellungen***

Für die Kartographen des Projektes stehen Fragen nach der direkten Vergangenheit der bearbeiteten Daten im Vordergrund. Dies sind beispielsweise „Wer hat ... zuletzt bearbeitet?“, „Wie sah ... am ... aus?“ und in letzter Instanz „Wie kann ich ... wieder rückgängig machen?“.

Gleichzeitig sind jedoch auch andere Personengruppen an den vergangenen Versionen der Daten interessiert. Insbesondere können die Informationen von Mathematikern und Statistikern ausgewertet werden, um das Verhalten und das Wachstum der Community Zeit- und Ortsbezogen zu analysieren. Ebenso interessant ist die Datenbasis für genaue Berechnungen im Zusammenhang mit dem Lizenzwechsel und den damit verbundenen Maßnahmen [1].

### ***Notwendigkeit***

Warum, außer aus wissenschaftlicher Neugier, könnten solche Betrachtungen interessant sein? Für die Funktionsweise einer Community in der Größenordnung von OpenStreetMap und für gemeinschaftlich gepflegte Datenbestände in dieser Größenordnung gibt es nicht allzu viele Vergleichsmöglichkeiten. Gerne wird die Wikipedia heran gezogen, was auch in Fra-

## Körner: History-Informationen

gen der Communitybildung und –Pfllege gut funktioniert, allerdings beim Umgang mit Vandalismus oder Uneinigkeit schnell an seine Grenzen stößt. Hintergrund der Vandalismusbekämpfung bei der Wikipedia ist die Grundregel, dass „kaputt machen“ genau so einfach geht, wie „wieder reparieren“. Dieses zurückspielen auf einen alten Stand kostet ist nur einige Klicks und steht sehr vielen Mitgliedern offen. Dadurch können Manipulationen und mutwillige Zerstörungen schnell und auf viele Schultern verteilt behoben werden.

Bei OpenStreetMap ist dies nicht so. Das Zurücknehmen von Änderungen ist ein komplizierter Prozess, der nur wenigen „eingeweihten“ zugänglich ist. Oft sind die Änderungen nur komplett zurücknehmbar, wobei nicht selten wesentliche Teile der Arbeit anderer mit vernichtet werden. Manchmal kann der Zerstörung auch nur durch Handarbeit entgegen gewirkt werden.

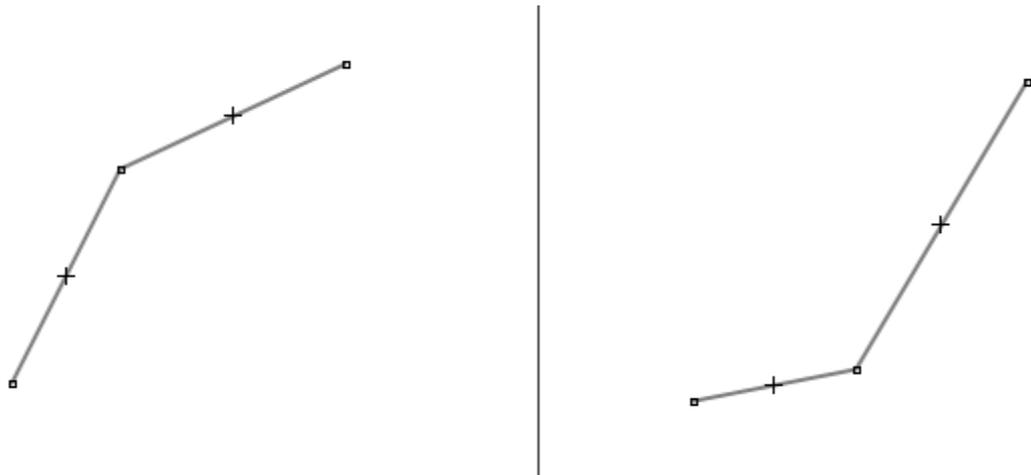
Hier können bessere Methoden in der Verarbeitung der Vergangenheit eingreifen und jedem Community-Mitglied diese Möglichkeiten der Qualitätskontrolle zugänglich machen.

Ein weiterer, wenig erforschter aber viel diskutierter Aspekt, welcher sich mit Hilfe der Auswertung der Datenvergangenheit untermauern ließe, ist die Auswirkung von Importen auf das Communityleben. Aus den Versionsinformationen lässt sich nicht nur Anzahl der Bearbeitungen von importierten Daten mit den Charakteristiken von manuell erfassten Informationen vergleichen, es ist ebenso möglich die Zu- und Abwanderung von Mappern oder die Verlagerung bestimmter Nutzergruppen auf andere Teile des Landes oder zu anderen Arten von Objekten zu quantifizieren. Auch das Verhalten von Neulingen in Regionen mit sowie in solchen ohne Importe könnte verglichen werden.

Dies seien nur zwei Beispiele für die praktische Anwendung der Vergangenheitsinformationen der OpenStreetMap-Datenbank, denen viele noch ungenannte Möglichkeiten gegenüber stehen.

## **Schema**

Das Datenschema, dem alle Datenquellen rund um OpenStreetMap folgen, beinhaltet die Elemente Node (Stütz- und Datenpunkt), Way (Verbindung von zwei oder mehr Punkten) und Relation (Verbindung von einer oder mehreren Nodes oder Ways), jeweils mit angehängten Schlüssel-Wert-Paaren. Für jedes Element wird eine eigene und unabhängige Versionsgeschichte geführt. Wichtig ist hier zu beachten, dass nur Nodes geographische Informationen codieren. Ein Way enthält Referenzen auf Nodes, ohne deren Koordinaten zu reflektieren. In der Auswertung der Versionsgeschichte ergibt sich daher ein Problem, welches sich an einem Beispiel erläutern lässt.



Auf der linken Seite sieht man einen Way mit drei Nodes als Stützpunkte. In der rechten Hälfte wurde die mittlere Node nach unten verschoben. Aus Sicht der Datenbank wurde die Koordinate der Node verändert, weshalb dieser nach der Änderung in einer neuen Version vorliegt. Der Way wurde hingegen, aus Sicht des Datenmodells, nicht verändert.

Würde dem Way hingegen ein Stützpunkt hinzu gefügt oder würden seine Key-Value Informationen (beispielsweise sein Name) geändert, hätte diese Änderung er eine neue Version begründet; die Nodes wären in diesem Beispiel nicht betroffen gewesen.

Zusammenfassend lässt sich sagen, dass Geometrieänderungen in den Nodes abgebildet werden, während Änderungen an den Eigenschaften eines Ways auch in dessen Versionsgeschichte gespeichert werden. Um also beispielsweise die Frage „Wie sah ... am ... aus?“ beantworten zu können, muss beides betrachtet werden.

### ***Weitere Probleme***

Neben den Bürden durch das Datenbankschema begegnen einem Interessierten noch weitere Probleme. Beispielsweise sind die zu verarbeitenden Datenmengen enorm. Ein kompletter Auszug der OpenStreetMap-Datenbank mit allen Versionen belegt derzeit 32 GB [2].

Einige andere Probleme wie eine unklare Lage bezüglich des Datenformats sowie das Fehlen von entsprechenden Bibliotheken zum Lesen und Schreiben dieser Daten haben sich im letzten Jahr verbessert, hier fehlt es aber noch bei der Adaption der Datenformate für verschiedene Programmiersprachen.

### ***Tools & Bibliotheken***

Bei der Suche nach Programmen zum Verarbeiten der Versionsinformationen kommt man derzeit am Jochen Topfs Framework „osmium“ nicht vorbei [3]. Vor allem wird hier eine in C++ geschriebene Bibliothek bereitgestellt, mit welcher ein Programmierer ohne Mühe die gesamte Bandbreite von verwendeten Datenformaten lesen, verarbeiten und schreiben kann. Daneben ist auch das Erstellen von Zwischenformaten (Shapefiles, CSV-Dateien) möglich.

Für Nicht-Programmierer hält „osmium“ einen Strauß von Beispielprogrammen z.B. zum Konvertieren zwischen den verschiedenen Datenformaten bereit. Mit „osmjs“ wird eine Möglichkeit angeboten, die Verarbeitung in Javascript statt in C++ durchzuführen, wodurch für viele ein einfacherer Einstieg in die Verarbeitung von OpenStreetMap-Daten möglich wird.

## Körner: History-Informationen

Eines der Werkzeuge, welche die Arbeit mit Versionsinformationen im vergangenen Jahr stark voran gebracht haben, ist der „OpenStreetMap History Splitter“ [4]. Mithilfe dieses Schneidwerkzeuges lassen sich aus den gewaltigen Datenmengen eines Planeten-Dumps handliche Stücke in geographischer Abgrenzung erzeugen, welche aufgrund der verringerten Dateigröße ohne Weiteres auf handelsüblichen Notebooks oder PCs in angenehmer Zeit verarbeitet werden können.

Ein weiteres Tool aus dem letzten Jahr ist der „OpenStreetMap History Renderer“ [5], welcher es mit einem eigenen Datenbankschema samt Importer ermöglicht, aus einer einzelnen Datenbank Karten für jeden beliebigen Zeitpunkt zu erstellen. Dieses Programm stellt jedoch auch für viele weitere Auswertungen einen guten Ansatzpunkt dar, weil sich die erstellte Datenbank gleichermaßen auch für viele andere Abfragen eignet. In jedem Fall kann der Importer als Beispiel für die Auswertung und Aufbereitung der Datenhistorien dienen.

## Datenquellen

Neben dem unregelmäßig erscheinenden kompletten Export aller Daten inklusiver der Vergangenheitsinformation [2] gibt es wenigstens zwei Quellen für vorgeschchnittene Extrakte [6]. Desweiteren enthalten auch die minütlich, stündlich und täglich publizierten Diff-Dateien [7] vollständige Versionsinformationen, mit welchen eine Datenbank oder ein Extrakt auf dem neusten Stand gehalten werden könnte. Hier mangelt es aber zum Zeitpunkt des Erscheinens an geeigneten Werkzeugen.

Kontakt zum Autor:

Peter Körner  
Gustav-Heinemann Straße 1  
55239 Gau-Odernheim  
0152/5167822  
osm@mazdermind.de

## Verweise & Links

[1] Vortrag „Die neue OpenStreetMap-Lizenz“ von Frederik Ramm auf der FOSSGIS 2011

[2] Full-Planet-Dump vom 13.02.2012 auf <http://planet.osm.org>

[3] Jochen Topfs Framework „osmium“ <https://github.com/joto/osmium>

[4] Peter Körners „OpenStreetMap History Splitter“ <https://github.com/MaZderMind/osm-history-splitter>

[5] Peter Körners „OpenStreetMap History Renderer“ <https://github.com/MaZderMind/osm-history-renderer>

[6] geographische Extrakte <http://ftp5.gwdg.de/pub/misc/openstreetmap/osm-full-history-extracts/> und <http://odbl.poole.ch/extracts/>

[7] vollständige Diffs der Datenbank <http://planet.osm.org/minute-replicate/>, <http://planet.osm.org/hour-replicate/> und <http://planet.osm.org/day-replicate/>