

# osmPti2mmds – Erstellung von multimodalen Datensets aus OpenStreetMap und ÖPNV - Informationen

Nikolaus Krismer<sup>1</sup>, Doris Silbernagl<sup>1</sup>, Johann Gamper<sup>2</sup>, Günther Specht<sup>1</sup>

<sup>1</sup>Institut für Informatik; Abteilung DBIS; Universität Innsbruck (UIBK) – nikolaus.krismer@uibk.ac.at

<sup>2</sup>Fakultät für Informatik; Freie Universität Bozen-Bolzano (FUB)

**Zusammenfassung:** Mit Hilfe von OpenStreetMap-Daten und Informationen des ÖPNV generiert das in dieser Arbeit vorgestellte Werkzeug „osmPti2mmds“ Datensets, die zu multimodalen Analysen verwendet werden können. Dabei werden GTFS- oder VDV452-Archivdateien so mit Datenexporten aus OpenStreetMap verknüpft, dass damit beispielsweise multimodale Wegefindungen und Erreichbarkeitsanalysen durchgeführt werden können. Bei der Datenset-Erstellung werden Ankunfts- und Abfahrtszeiten des ÖPNV berücksichtigt, so dass im Unterschied zu unimodalen Analysen ein bedeutender Mehrwert entsteht. Prototypisch werden die durch osmPti2mmds erstellten Datensets in der Webapplikation „IsoMap“ zur Verfügung gestellt, mit der multimodale Isochrone berechnet werden können.

**Schlüsselwörter:** Datensetgenerierung, Open Source, OpenStreetMap (OSM), ÖPNV

**Abstract:** *Using data from OpenStreetMap and information about the public transportation the tool “osmPti2mmds” presented in this paper creates datasets that can be used to carry out multimodal analyses. As part of this process GTFS or VDV452 archive files are linked to data from OpenStreetMap in such a way that multimodal routing and reachability analyses are possible. By using this combined information the arrival- and departure-times of the public transportation system is taken into account. Thereby, a significant improvement in contrast to unimodal analyses is achieved. A prototypical usage for the datasets generated by osmPti2mmds is done by the web-application “IsoMap” that allows the computation of reachability analyses using multimodal isochrones.*

**Keywords:** *dataset generation, open source, OpenStreetMap (osm), public transport*

## 1 Motivation und Stand der Technik

Im Rahmen von Wegefindungen, Erreichbarkeitsanalysen und anderen geografischen Berechnungen spielen die zugrunde liegenden Daten eine zentrale Rolle. So ist es kaum verwunderlich, dass sich in den letzten Jahren mehrere Werkzeuge etabliert haben, um Datensets zu erstellen, die die Grundlage für die Durchführung solcher Berechnungen sind. Viele dieser Werkzeuge verwenden öffentlich verfügbare Daten aus OpenStreetMap, da es sich um einfach zugängliche, qualitativ hochwertige und in den meisten Fällen auch aktuelle Daten handelt (HAKLAY 2008, NEIS ET AL. 2010). Zwei bekannte Vertreter solcher Applikationen sind osm2pgrouting und osm2po (OSM2PGROUTING 2016, MOELLER 2016).

Eine neuere Entwicklung im Umfeld der angesprochenen Berechnungen beschäftigt sich mit der Tatsache, dass häufig multimodale Analysen notwendig sind. Erfolgt eine Wegefindung über eine große Entfernung, so kann das beste Ergebnis für einen Benutzer in der Kombination mehrerer Verkehrsmittel liegen. Es ist also nicht zwangswise jener Weg, der ausschließlich mit einem Auto zurückgelegt wird. Bei einem Wochenendtrip in die Hauptstadt des Nachbarlandes kann etwa eine Kombination von (Miet-)Auto und Zug sinnvoll sein. Multimodalität stellt also neue Anforderungen an Datensets und damit auch an die zur Erstellung eingesetzten Werkzeuge selbst. Einerseits müssen pro Verkehrsmittel unter-

schiedliche Rahmenbedingungen beachtet werden, wie etwa Geschwindigkeiten oder Fahrverbote. Andererseits spielen Fahrpläne und die darin festgehaltenen Abfahrts- und Ankunftszeiten des ÖPNV eine wichtige Rolle. Wird der erste Punkt schon seit einiger Zeit von vielen der so genannten „Routing-Engines“, etwa Graphhopper und OSRM, umgesetzt, so stellt der zweite Punkt im Rahmen von Open-Source Anwendungen ein größeres Problem dar (GRAPHHOPPER 2016, OSRM 2016). Um dieses im Rahmen einer frei verfügbaren Software zu lösen, wurde die Applikation „osmPti2mmds“ entwickelt, die im vorliegenden Artikel vorgestellt wird.

Diese Arbeit teilt sich in folgende Kapitel auf: In Abschnitt 2 wird der von osmPti2mmds umgesetzte Prozess beschrieben. Dabei wird auf den internen Ablauf der Applikation von der Datenextraktion bis hin zur finalen Datenspeicherung eingegangen, sowie eine mögliche Verwendung der Daten im Rahmen der Berechnung von Isochronen aufgezeigt. In Abschnitt 3 wird die Performanz anhand mehrerer erstellter Datensets untersucht. Den Abschluss dieser Arbeit bildet Kapitel 4, in dem neben einer kurzen Zusammenfassung mögliche weitere Entwicklungen des vorgestellten Werkzeuges diskutiert werden.

## 2 Umsetzung

Die Funktionsweise von osmPti2mmds basiert auf einem Prozess, der sich in mehrere (Teil-)schritte unterteilt. Im ersten Schritt werden Informationen über den Straßengraphen aus OpenStreetMap (OSM 2016) geladen. Dies erfolgt durch das Herunterladen von Datenexports, die von Geofabrik zur Verfügung gestellt werden (GEOFABRIK 2016). Falls für die betrachtete Region kein exakt passender Export verfügbar ist, wird der am Besten passende geladen und mit Hilfe von Osmosis die betrachtete Region angepasst (OSMOSIS 2016). Anschließend wird mit Hilfe von osm2po ein routingfähiges Netzwerk aus den OSM-Daten erstellt (MOELLER C. 2016). Dieses wird anschließend in einer mit PostGIS erweiterten PostgreSQL Datenbank abgelegt. Steht keine Information über den ÖPNV der betrachteten Region zur Verfügung, beendet osmPti2mmds an dieser Stelle mit der Erstellung eines Datenbankabbildes seine Arbeit. Selbst in solch einem Fall ist garantiert, dass das entstandene Datenset für unimodale Wegefindungen verwendet werden kann, dann jedoch ohne Berücksichtigung des öffentlichen Personennahverkehrs.

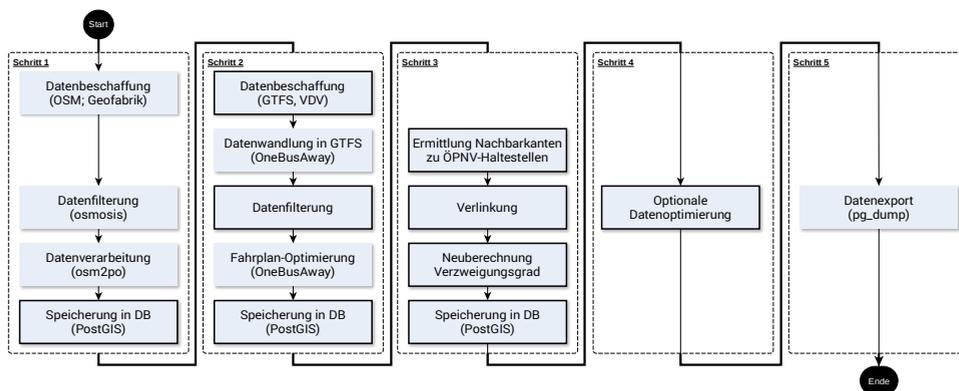
Sind Fahrpläne und Haltestelleninformationen über VDV452 oder GTFS online verfügbar und in osmPti2mmds referenziert, so werden diese im zweiten Prozessschritt geladen und aufbereitet (VDV 2016, GTFS 2016). Im Rahmen dieser Aufbereitung ist der erste Teilschritt, alle Daten in das GTFS-Format zu wandeln, sofern sie nicht schon in diesem vorliegen. Dadurch wird eine Vereinheitlichung des weiteren Prozesses erreicht. Anschließend erfolgt, analog zu den Daten des Straßennetzwerkes, eine Einschränkung auf die betrachtete Region und eine Optimierung der Fahrpläne. Diese Vorverarbeitungsschritte erfolgen zum Teil mit Werkzeugen, die von OneBusAway auf Github unter einer OpenSource-Lizenz frei zur Verfügung gestellt werden (ONEBUSAWAY 2016). Abschließend werden die Informationen in die selbe PostGIS-Datenbank gespeichert, in der schon das Straßennetzwerk abgelegt wurde. Hierbei werden nicht alle Bestandteile aus dem zur Verfügung gestellten GTFS-Dateien verwendet, es werden etwa Tarifinformationen verworfen.

Im dritten Schritt werden die Daten über das Straßennetzwerk und jene über den ÖPNV miteinander verknüpft. Diesen als „Linking“ bezeichneten Schritt realisiert osmPti2mmds

mit Hilfe einer speziell dafür entwickelten Java-Applikation. Dabei erfolgt eine Suche der nächstgelegenen Straßen, um die Haltestellen des ÖPNV damit zu verbinden. Danach werden Änderungen der Verzweigungsgrade im Netzwerk genauso berücksichtigt wie die Speicherung der Haltestellen und der erzeugten „Links“ (als Graphknoten bzw. -kanten) an sich.

Ein optionaler vierter Schritt kann dazu genutzt werden, um die bereits miteinander verknüpften Daten weiterzubearbeiten. Hier können beispielsweise Spatial Reference Identifiers (SRID) konvertiert und Vorberechnungen durchgeführt werden.

Den abschließenden Schritt stellt der Export der Datensets aus der Datenbank dar. Das erstellte Datenbankabbild wird mit Hilfe von gzip (GNU zip) komprimiert, bevor er schließlich im Rahmen einer Archivdatei abgespeichert wird. Diese fünf Hauptschritte, sowie deren Teilschritte, sind in Abbildung 1 graphisch dargestellt, wobei die eingerahmten Teilschritte durch selbst entwickelte Programme umgesetzt sind und die nicht gerahmten Prozessschritte mit Hilfe von vorhandenen open-source Werkzeugen durchgeführt werden:



**Abb. 1:** Prozesslogik von osmPti2mmds

Die in den Schritten eins und zwei durchgeführte Regionseinschränkung erfolgt ebenfalls mit Hilfe von Daten aus OpenStreetMap. Dazu werden Grenzen, welche durch die Tags „type=boundary“ und „boundary=administrative“ markiert sind, unter Verwendung der Overpass-API exportiert. Das Programm „osmtogeojson“ und eine selbst entwickelte Java-Klasse Namens „PolygonGeneratorOSM“, wandeln die durch die API gelieferten XML-Daten in GeoJSON- und Poly-Dateien um (OSMTOGEOJSON 2016). Dies ermöglicht eine Regionsbeschränkung sowohl mit osmosis (in Schritt 1), als auch in Java (in Schritt 2).

## 2.1 Implementierung

Der besprochene Prozess wurde im Rahmen einer Java-Applikation und eines Bash-Skriptes implementiert. Alle entstandenen Quellcodes werden im Rahmen der MIT-Lizenz als Git-Repository zur Verfügung gestellt (OSMPti2MMDs-SOURCE 2016). Durch diese Lizenzierung sind auch Anpassungen und Erweiterungen möglich, welche die Erstellung von Datensets mit Hilfe von aktualisierten ÖPNV-Informationen ermöglichen. Bei der Berechnung der multimodalen Datensets kommen neben Oracle Java 8 und einer Bash-Umgebung weitere Werkzeuge wie gradle, osmosis, osm2po, PostgreSQL 9.5 und PostGIS 2.2 zum Einsatz (GRADLE 2016).

Zur Vereinfachung der Datenset-Erstellung wird neben den Quellen auch eine Vagrant-Beschreibungsdatei bereitgestellt, die das Anlegen einer virtuellen Maschine ermöglicht (OSMPti2MMDS-VM 2016, VAGRANT 2016). Obwohl osmPti2mmds eine Linux-Umgebung voraussetzt, wird es somit möglich, Datensets aus OpenStreetMap und aus Informationen des ÖPNV unter anderen Betriebssystemen zu erstellen.

## 2.2 Verwendung bei Isochron-Berechnungen

Die durch osmPti2mmds generierten Datensets können zur Berechnung multimodaler Isochrone verwendet werden. Dazu erfolgt im optionalen Prozess-Schritt vier von osmPti2mmds eine Strukturänderung der gespeicherten ÖPNV-Fahrplandaten und Vorberechnungen hinsichtlich der Anzahl der erreichbaren Nachbarknoten (sowohl unimodal als auch multimodal) und zur Abschätzung der lokalen Knotendichte, um die Erstellung von Isochronen mit Hilfe von Range-Queries beschleunigen zu können.

Die als „IsoMap“ bezeichnete Demonstrations-Webapplikation greift auf die Datensets direkt in der mit PostGIS erweiterten PostgreSQL Datenbank zu und realisiert damit die Berechnung von Isochronen (ISOMAP 2016). Zur Berechnung können innerhalb von IsoMap mehrere verschiedene Algorithmen zum Einsatz kommen. Momentan stehen fünf Algorithmen zur Verfügung, die an der Freien Universität Bozen-Bolzano implementiert wurden und derzeit in Kooperation mit der Universität Innsbruck weiterentwickelt werden (Gamper et al. 2012, KRISMER ET. AL. 2014). Auf der selben Webseite sind auch die verwendeten Datensets als Archivdateien hinterlegt. Diese können in jede PostgreSQL ab Version 9.3 in Verbindung mit PostGIS ab Version 2.0 importiert werden und stehen unter der Open Data Commons Open Database License (OdbL) v1.0 zur Verfügung.

## 3 Evaluierung

Insgesamt werden zehn Datensets generiert und untersucht. Darunter befinden sich neben zwei unimodalen Datensets für die Städte Innsbruck und Salzburg (da in diesen Regionen keine Informationen über den ÖPNV als VDV452- oder GTFS-Datei zur Verfügung stehen), acht multimodale Datensets. Für die Stadt Bozen wird eine Konvertierung von VDV452 nach GTFS durchgeführt. Alle anderen Informationen stehen direkt als GTFS auf der Website des jeweiligen städtischen Verkehrsbetriebs zur Verfügung bzw. werden diese beim Datenset „GreatBritain“ von der Homepage des grenzübergreifenden Verkehrsverbundes „GB-Rails“ geladen.

Als Testsystem kam ein Lenovo T440s mit einer Intel i7-4600U Dual-Core CPU mit 2.1 Ghz (im Turbo Modus bis zu 3.3 Ghz), zwölf GiB Arbeitsspeicher und einer Samsung Pro 850 Solid-State Drive mit 512 GB zum Einsatz. Die bei der Erstellung der Datensets erhobenen statistische Details finden sich in Tabelle 1. Die dort gelistete Berechnungsdauer berücksichtigt alle notwendigen Berechnungsschritte im Rahmen des osmPti2mmds-Prozesses, so wie er in Abschnitt 2 besprochen wurde (Berechnungszeit „BZ“), beinhaltet aber nicht die Zeit für den Download der Daten (bei Verwendung einer 30MBit-Internetleitung Downloadzeit „DZ“), sowie für das Ausschneiden der betrachteten Region aus den OSM-Daten (Ausschnittszeit „AZ“). Der optionale vierte Schritt zur etwaigen Umwandlung und Optimierung der Daten wurde bei den Berechnungszeiten nicht durchgeführt, da diese für

die Verwendung im Rahmen von Wegefindungen und Erreichbarkeitsanalysen nicht zwingend notwendig ist.

**Tabelle 1:** Statistiken zu den erstellten Datensets

Name	DB-Größe (in MiB)	# Knoten	# Kanten	# ÖPNV – Fahrten	Dauer (in sec)		
					BZ	DZ	AZ
Berlin	851	524.279	189.800	2.377.784	565	858	1.556
Bozen	21	16.459	6.237	49.354	43	333	650
Cuneo	7	10.661	3.878	7.276	24	348	641
GreatBritain	5.260	12.981.910	5.262.838	1.072.244	4.261	318	976
Innsbruck	10	25.519	9.805	0	12	237	257
Rom	1.127	308.790	112.679	3.518.613	509	345	727
Salzburg	18	46.734	18.435	0	17	237	258
San Francisco, CA	250	96.878	34.791	739.233	209	187	316
Turin	210	80.557	26.926	668.768	174	337	647
Washington, D.C.	307	115.835	39.002	977.240	228	19	27

## 4 Diskussion

Das Erstellen von Datensets, die in weiterer Folge zur Berechnung von multimodalen Wegefindungen und Erreichbarkeitsanalysen genutzt werden können, stellt eine nicht zu unterschätzende Aufgabe dar. Zwar bietet OpenStreetMap frei verfügbare Daten an, jedoch müssen diese für die Nutzung hinsichtlich der zuvor erwähnten Anwendungen weiterverarbeitet und mit anderen Informationen verknüpft werden. Es gibt zwar Anbieter und zahlreiche frei verfügbare Werkzeuge, die einige dieser Aufgaben übernehmen können, bei einigen anderen, wie etwa bei der Verknüpfung von Straßen- und ÖPNV-Netzwerken, ist eine Eigenimplementierung jedoch unvermeidbar.

Im Rahmen der ÖPNV-Daten ist ein weiteres Problem in der Datenbeschaffung selbst zu sehen. Nicht für alle Regionen, die von Interesse für Analysen sein können, sind frei verfügbare Daten in einem dafür geeigneten Format verfügbar. In diesen Gebieten bleibt derzeit nur die unimodale Analyse. Unter bestimmten Voraussetzungen kann zwar multimodale Information aus OpenStreetMap extrahiert werden (etwa Haltestellen und Umstiegspunkte), da aber genaue Fahrzeiten sowie deren Taktung nicht vorhanden sind, müssen diese geschätzt oder komplett vernachlässigt werden. Eine realistische Erreichbarkeitsanalyse unter Berücksichtigung der Erreichbarkeitsverteilung ist in solchen Regionen nicht möglich (KRISMER et. al. 2015).

Mit `osmPti2mmds` wurde ein frei verfügbares Projekt geschaffen, welches bei der Erstellung von Datensets aus OSM-Daten und – falls verfügbar – auch von Informationen über den ÖPNV hilfreich ist. Dies wurde im Rahmen einer Evaluierung mit Hilfe der ebenfalls frei verfügbaren Webapplikation „IsoMap“ aufgezeigt.

Mögliche Verbesserungen des vorgestellten Werkzeuges sind in der Performance-Optimierung zu sehen. Eine relativ gut umsetzbare Steigerung der Performance würde durch Parallelisierung erreicht werden. Der Import des Straßennetzwerkes aus OSM und das Einfügen

der Bushaltestellen aus GTFS-Dateien können unabhängig voneinander erfolgen und sind deshalb parallelisierbar.

Die Datensets selbst können im Hinblick auf mehrere Kriterien verbessert werden. So ist noch zu evaluieren, ob bei Verwendung von anderen SRIDs kleinere Datenmengen beziehungsweise performantere Analysen möglich sind. Obwohl bei der Umetzung von IsoMap ein erster Vorstoß durch Umwandlung in SRID EPSG:3857 gemacht wurde, könnten andere SRIDs Geschwindigkeitsvorteile mit sich bringen. In Zukunft ist geplant Höheninformationen in die Datensets aufzunehmen und bei den Berechnungen zu berücksichtigen.

## 5 Literatur

- GAMPER J., BOEHLER M.H., INNEREBNER M. (2012), Scalable Computation of Isochrones with Network Expiration. In Proc. of SSDBM-12, pages 526-543, June 25-27, 2012, Chania/Crete/ Greece
- GEOFABRIK (2016), <http://www.geofabrik.de> (01.2016)
- Gradle (2016), <http://www.gradle.org> (01.2016)
- GRAPHHOPPER (2016), <https://graphhopper.com> (01.2016)
- GTFS (2016), <https://developers.google.com/transit/gtfs> (01.2016)
- HAKLAY M. (2008), How good is OpenStreetMap information? A comparative study of OpenStreetMap and Ordnance Survey datasets for London and the rest of England. Environment & Planning B: Planning and Design
- ISOMAP (2016), <http://dbis-isochrone.uibk.ac.at> (01.2016)
- KRISMER N., SPECHT G., & GAMPER J. (2014), Incremental calculation of isochrones regarding duration. GvDB 2014, Ritten/Italien
- KRISMER N., GAMPER J., SPECHT G. (2015), Reachability calculation based on average isochrones regarding time distribution, AGIT 2015, Salzburg/Austria
- NEIS P., ZIELSTRA D., ZIPF A., STRUCK A. (2010), Empirische Untersuchungen zur Datenqualität von OpenStreetMap - Erfahrungen aus zwei Jahren Betrieb mehrerer OSM-Online-Dienste. AGIT 2010. Salzburg/Austria
- MOELLER, C. (2016), <http://osm2po.de> (01.2016)
- ONEBUSAWAY (2016), <https://github.com/OneBusAway> (01.2016)
- OSM (2016), <https://www.openstreetmap.org/> (01.2016)
- OSM2PGROUTING (2016), <http://pgrouting.org/docs/tools/osm2pgrouting.html> (01.2016)
- OSMOSIS (2016), <https://wiki.openstreetmap.org/wiki/DE:Osmosis> (01.2016)
- OSMPTI2MMDS-SOURCE (2016), <https://git.uibk.ac.at/dbis-isochrone/isochrone-datamodel> (01.2016)
- OSMPTI2MMDS-VM (2016), <https://git.uibk.ac.at/dbis-isochrone/isochrone-vm> (01.2016)
- OSRM (2016), <https://project-osrm.org> (01.2016)
- OSMTOGEOJSON (2016), <https://github.com/tyrasd/osmtogeojson>, (01.2016)
- VAGRANT (2016), <https://www.vagrantup.com> (01.2016)
- VDV (2016), <https://www.vdv.de/oepnv-datenmodell.aspx> (01.2016)