Toni Erdmann

### Abstract

"PTNA - Public Transport Network Analysis" ermöglicht eine Soll-Ist-Analyse für ÖPNV-Linien in OpenStreetMap [1]. Soll-Daten bestehen aus einer CSV-Liste von ÖPNV-Linien, jeweils mindestens aus der Liniennummer und dem Fahrzeugtyp. Diese Informationen werden Ist-Analyse mit den OSM-Daten verglichen. Zusätzlich erfolgt eine Fehleranalyse der einzelnen ÖPNV-Relationen in OSM.

PTNA sollte auf der FOSSGIS 2020 in Freiburg vorgestellt werden. Wegen Corona gibt es lediglich einen Beitrag im Tagungsband [1]. Dieser Beitrag stellt die wichtigsten Erweiterungen an PTNA seit März 2020 vor: eine **Analyse von GTFS-Daten**, sowie einen **Vergleich von OSM-Daten mit GTFS-Daten**. "GTFS" steht für: "General Transit Feed Specification", "GTFS is a community-driven open standard for rider-facing transit information." [2][3][4].

### **GTFS-Daten**

### Import von GTFS-Daten

GTFS-Daten müssen bezüglich ihrer Lizenz mit OSM kompatibel sein. CC0 ist ideal, für CC-BY-4.0 gilt das zum Beispiel nur, wenn die Form der "Namensnennung" konkreter genannt wird. Für einige GTFS-Daten liegen Angaben seitens der Eigentümer der GTFS-Daten vor, die die Namensnennung im Verzeichnis der Beitragenden im OSM-Wiki [5] als ausreichend bestätigen.

PTNA prüft GTFS-Quellen regelmäßig auf neue Versionen der Daten. Es wird in der Regel nur die erste im Monat verfügbare Version importiert. Auf Anfrage kann ein Import angestoßen werden.

### Aufbereitung von GTFS-Daten

Beim Import der GTFS-Daten wird aus den diversen CSV-Dateien des Datensatzes eine sqlite3-Datenbank aufgebaut. Ein zweiter, wichtiger Schritt ist die Reduktion der Datenmenge durch Aggregierung, aus OSM-Sicht, redundanter Informationen. Das ist im Wesentlichen die Zusammenfassung identische Trips mit der selben Reihenfolge von Haltestellen aber unterschiedlichen Abfahrtzeiten zu einem repräsentativen Trip.

# **Analyse von GTFS-Daten**

Um die Qualität und Verwendbarkeit der GTFS-Daten für OSM zu bestimmen werden von PTNA diverse Analysen durchgeführt:

1) Ist ein Trip eine Teilroute eines anderen Trips bezüglich der Haltestellen

- 2) Besteht ein Trip nur aus 2 Haltestellen (Ausnahme: Seilbahnen, Fähren, ...)
- 3) Haben die erste und zweite Haltestelle den selben Namen, die selbe stop\_id
- 4) Haben die vorletzte und letzte Haltestelle den selben Namen, die selbe stop\_id
- 5) Unterscheiden sich zwei Trips nur durch die shape id der Fahrstrecke
- 6) Unterscheiden sich zwei Trips nur durch die stop id der Haltestellen
- 7) Beträgt die Fahrzeit eines Trips 0 Sekunden

# Normalisierung von Namen im GTFS-Datensatz

In GTFS-Datensätzen sind die Namen von Haltestellen häufig abgekürzt. In OSM hingegen gilt die Regel: Namen werden ausgeschrieben. PTNA führt daher für den Deutschsprachigen Raum eine Normalisierung/Expansion von Namen für die GTFS-Daten durch, z.B.:

```
    "Friedenstr."
```

```
de_DE und de_AT: "Friedenstraße"de_CH: "Friedenstrasse"
```

- "W.-Heisenberg-W."
  - "Werner-Heisenberg-Weg"
- "Bruck Abzw."

```
de_DE und de_CH: "Bruck Abzweig"de_AT: "Bruck Abzweigung"
```

Die Eigentümer der GTFS-Datensätze sind beim Abkürzen sehr kreativ: PTNA kann bezüglich der Normalisierung der Namen nie vollständig sein.

# Darstellung der GTFS-Daten auf der PTNA Webseite

Auf der Einstiegsseite von PTNA [11] ist ...

# Vergleich von OSM-Daten mit GTFS-Daten

**Begrifflichkeiten: GTFS versus OSM** 

GTFS und OSM nutzen unterschiedliche Begriffe für Objekte des ÖPNV. Tabelle 1 zeigt einen Überblick über die von PTNA verwendeten Elemente und deren Entsprechungen. Für OSM gilt dabei die so genannte "Public Transport Version 2" ("PTv2") [7].

PTNA: Qualitätssicherung für ÖPNV-Linien in OpenStreetMap

GTFS Begriff	OSM Begriff (PTv2)	Anmerkung		
feed	network:guid	Viele GTFS-Datensätze werden von den Verkehrsverbünden herausgegeben. PTNA vergibt generische feed-Namen [5], die in OSM als network:guid verwendet werden können:		
		- network = Münchner Ver- kehrs-und Tarifverbund		
		- network:short = MVV		
		- network:guid = DE-BY-MVV		
route	route_master	Fasst alle Fahrten einer ÖPNV- Linie zusammen.		
route_id		Eindeuige ID einer route.		
trip	route	Ein GTFS <i>trip</i> entspricht einer einzelnen Fahrt zu einer bestimmten Uhrzeit, definiert durch die angefahrenen Stops und optionaler Fahrstrecke (shape).		
		Eine OSM <i>route</i> beschreibt alle Fahrten, die die selben Stops anfahren plus deren Fahrstrecke.		
trip_id		Eindeutige ID eines trips.		
shape		In GTFS: die eigentliche Fahr- strecke (nicht immer vorhan- den).		
		In OSM ist das die Kette der highway/railway/ member einer route Relation (ohne 'role').		
shape_id		Eindeutige ID eines shapes		
route_short_name	ref	Das OSM tag <i>ref</i> einer <i>route masterlroute</i> Relation.		
stop	public_transport=platform	Ein <i>stop</i> entspricht der neben dem Fahrweg liegenden <i>platform</i> .		
stop_id	ref:IFOPT,	Eindeutige ID eines stops,		

		nicht immer identisch mit ref:IFOPT,
	public_transport=stop_position	Den Begriff Halteposition des Fahrzeugs gibt es in GTFS nicht.
agency_name	operator	Bei manchen GTFS-Daten kann der Name der <i>agency</i> als OSM <i>operator</i> einer routermaster/route Relation genutzt werden.

Tabelle 1: Vergleich GTFS vs. OSM Begriffe

Für den Vergleich von GTFS- und OSM-Daten können die aufbereiteten und analysierten GTFS-Daten auf verschiedene Arten genutzt werden.

# Visueller Vergleich

Hierbei erfolgt der Vergleich eines GTFS-trip und einer OSM-route auf getrennten Karten

- Keinerlei Änderungen an bestehenden OSM-Daten notwendig
- Mühsam in der Anwendung. Es werden zwei Browserfenster benötigt.
  - Die Analyse der GTFS-Daten muss über den GTFS-Einstieg der PTNA-Seite [9], der Liste der Länder (Flaggen), der Liste der Länderspezifischen GTFS-feed, die GTFS-route und den GTFS-trip gefunden werden
  - Die Analyse der OSM-route-Daten muss über den PTNA-Link im PTNA-Report angestoßen und in einem zweiten Fenster angezeigt werden. Alternativ kann die Analyse unter Verwendung der OSM-Relation-ID in der Adresszeile des Browsers genutzt werden [10]

# Nutzung von GTFS-Daten in CSV-Daten im OSM-Wiki

Es erfolgt eine Definition des GTFS-feed und der GTFS-route\_id in den SOLL-Daten = CSV-Liste im OSM-Wiki.

- Änderungen lediglich an Daten im OSM-Wiki, nicht an OSM-Daten
- Die erstmalige Erstellung kann durch PTNA unterstützt werden
  - Export der GTFS-routen als CSV-Liste
- Wartung hält sich in Grenzen, wenn die GTFS-route\_id stabil bleibt und weil GTFS-trip ids nicht enthalten sind
- Es erscheint pro ÖPNV-Linie ein GTFS-Link und ein Vergleiche-Icon im PTNA-Report
- Beispiel für CSV-Daten im OSM-Wiki für den Bus 210 im Münchner Verkehrsund Tarifverbund:

210; bus; ; Brunnthal, Zusestraße; Neuperlach Süd (U/S); Verkehrsbetrieb Ettenhuber GmbH; DE-BY-MVV; mvv-19-210-1

# Nutzung von GTFS-Daten in OSM route\_master / route Relationen

Die Definition des GTFS-feed, der GTFS-route\_id bzw. GTFS-trip\_id erfolgt im OSM-Datensatz der route\_master- bzw. route-Relation als IST-Daten [8].

- Änderungen an OSM-Daten
- Aufwändig in der Erstellung
- Gegebenenfalls aufwändig in der Wartung
  - Das hängt von der Stabilität der GTFS-id-Daten über mehrere Versionen hinweg ab (route\_id, trip\_id, shape\_id, stop\_id)
- Erstellung und Wartung kann durch den vorangegangenen Abschnitt (GTFS-Daten in CSV-Liste im OSM-Wiki) und den dort möglichen Vergleich erleichtert werden
- Es erscheint für jede so getaggte Relation ein GTFS-Link und ein Vergleiche-Icon im PTNA-Report
- Beispiel für die tags in einer OSM-route-Relation des Busses 210:
  - o qtfs:feed=DE-BY-MVV
  - o gtfs:route\_id=mvv-19-210-1
  - $\circ \quad \texttt{gtfs:trip\_id:sample=mvv-19-210-1-1-1-H-0-MoTuWeThFrSaSu-119-344}$
  - Dieser (Beispiel-)Bus mit dieser GTFS-trip\_id f\u00e4hrt Montag Sonntag um 06:48 (Stand: 2025-01-28)

Abbildung 1 zeigt einen Auszug aus einem PTNA-Report. Der Auszug korrespondiert mit dem folgenden Eintrag in den CSV-Daten im OSM-Wiki:

210; bus; ; Brunnthal, Zusestraße; Neuperlach Süd (U/S); Verkehrsbetrieb Ettenhuber GmbH; DE-BY-MVV; mvv-19-210-1

Name (name=)	Typ (type=)	Relation (id=)	PTv	Fehler	Anmerkungen	
210	Von: Brunnthal, Zusestraße; Nach: Neuperlach Süd (U/S); Betreiber: Verkehrsbetrieb Ettenhuber GmbH; GTFS, 👣					
Bus 210	route_master	67811 (iD, JOSM, UNROLL, GTFS, (1))			'check_date' = '2024-12-24'	
Bus 210: Brunnthal, Zusestraße => Neuperlach Süd (S/U)		1549761 (iD, JOSM, Relatify, PTNA, GTFS, (1))	2		'check_date' = '2024-12-24'	
Bus 210: Neuperlach Süd (S/U) => Brunnthal, Zusestraße		1549762 (iD, JOSM, Relatify, PTNA, GTFS, (1)	2		'check_date' = '2024-12-24'	

Abbildung 1: PTNA-Report für den Bus 210

Für den Vergleich von GTFS- und OSM-Daten finden sich hier 3 verschiedene Links:

- Ein **PTNA**-Link in der dritten Spalte führt zu einer Analyse der OSM-*route*-Relation, mit Karte und Details der einzelnen Elemente der Relation (Abbildung 2).
- Ein **GTFS**-Link in der Kopfzeile außen rechts führt zu einer Analyse der GTFSroute mit einem Überblick über die zugehörigen GTFS-*trips*. Die genutzten GTFS-Daten werden der CSV-Liste im OSM-Wiki entnommen.
- Ein **Vergleiche**-Icon **1** in der Kopfzeile außen rechts führt zu einem Vergleich der GTFS-*route* mit dem OSM-*route\_master*. Die genutzten GTFS-Daten werden der CSV-Liste im OSM-Wiki entnommen (Abbildung 4).
- Ein **GTFS**-Link in der dritten Spalte einer OSM-*route\_master*-Relation führt zu einer Analyse der GTFS-*route* mit einem Überblick über die zugehörigen GTFS-*trips*. Die genutzten GTFS-Daten werden den *gtfs:feed* und *gtfs:route\_id* Angaben der *route\_master*-Relation entnommen.
- Ein **Vergleiche**-Icon **1** in der dritten Spalte einer OSM-*route\_master*-Relation führt zu einem Vergleich der GTFS-*route* mit dem OSM-*route\_master*. Die genutzten GTFS-Daten werden den *gtfs:feed* und *gtfs:route\_id* Angaben der *route\_master*-Relation entnommen.
- Ein GTFS-Link in der dritten Spalte einer OSM-route-Relation führt zu einer Analyse des GTFS-trips, mit Karte und Datails zum Trip. Die genutzten GTFS-Daten werden den gtfs:feed und gtfs:trip\_id bzw. gtfs:trip\_id:sample Angaben der route-Relation entnommen (Abbildung 3).
- Ein **Vergleiche**-Icon **1** in der dritten Spalte einer OSM-*route*-Relation führt zu einem Vergleich des GTFS-*trips* mit der OSM-*route*. Die genutzten GTFS-Daten werden den *gtfs:feed* und *gtfs:trip\_id* bzw. *gtfs:trip\_id:sample* Angaben der *route*-Relation entnommen (Abbildung 5).

# Visueller Vergleich

Die Links zu **PTNA** und **GTFS** führen für GTFS-*trips* und OSM-*routes* zu identisch aufgebauten Karten auf den Webseiten. Ein visueller Vergleich kann hiermit erfolgen.

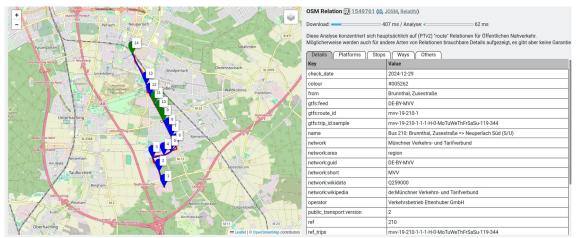


Abbildung 2 Analyse einer OSM-route-Relation durch PTNA



Abbildung 3 Analyse eines GTFS-trips durch PTNA

# Vergleich von GTFS- mit OSM-Daten durch Berechnung von score-Werten

Der Link bei dem **Vergleiche-Icon** für eine **OSM-route\_master**-Relation führt zu dem Vergleich GTFS-*route* versus OSM-*route\_master*. Der Vergleich erfolgt in Tabellenform mit so genannten **score**-Werten für jede Kombination GTFS-*trip*/OSM-*route*: Je kleiner der **score**-Wert, desto besser die Übereinstimmung der beiden Datensätze. Ist ein **score**-Wert in fetter Schrift abgebildet, so stimmen *gtfs:feed* und *gtfs:trip\_id:sample* in der OSM-*route*-Relation mit der GTFS-*trip\_id* überein.

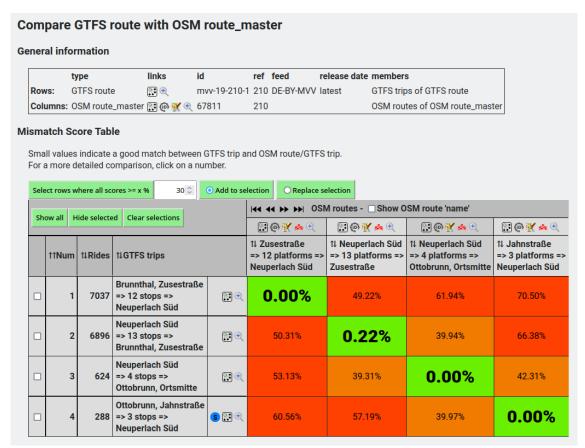


Abbildung 4 Vergleich von GTFS-route mit OSM-route\_master in Tabellenform

Ein Klick auf einen **score**-Wert führt zum Vergleich eines GTFS-*trips* mit einer OSMroute (nächster Abschnitt).

Der Link bei dem **Vergleiche-Icon** für eine **OSM-***route*-Relation führt zu dem Vergleich GTFS-*trip* versus OSM-*route*. Der Vergleich erfolgt auf einer Karte und in Tabellenform mit den **score**-Werten für einzelne Metriken: je kleiner der **score**-Wert, desto besser die Übereinstimmung. Beim Scrollen nach unten wird dann ein Vergleich der Haltestellen in Tabellenform sichtbar.

### Positiver Fall:

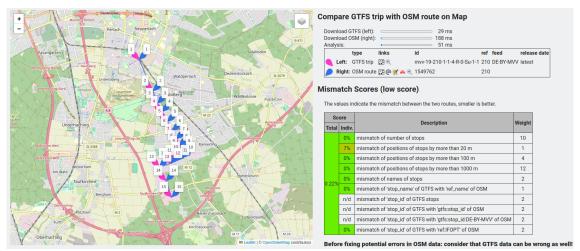


Abbildung 5 Vergleich von GTFS-trip mit OSM-route auf Karte und score-Tabelle: guter Match



Abbildung 6 Vergleich von GTFS-trip mit OSM-route mit Haltestellen-Tabelle: guter Match

### Negativer Fall:

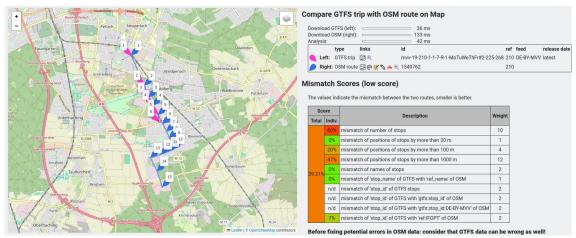


Abbildung 7 Vergleich von GTFS-trip mit OSM-route auf Karte und **score**-Tabelle: schlechter Match

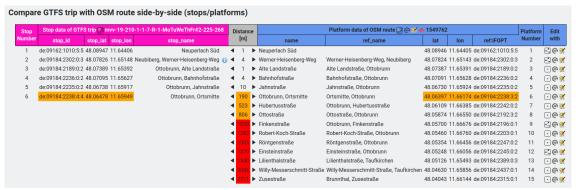


Abbildung 8 Vergleich von GTFS-trip mit OSM-route mit Haltestellen-Tabelle: schlechter Match

# **Anhang**

Auf die Angabe und Beispiele mit IDs von realen GTFS- oder OSM-Daten wird hier bewusst verzichtet. Diese können bis zum Zeitpunkt der Konferenz schon wieder veraltet sein.

Für den "Hardcore"-Vergleich von GTFS- mit GTFS-Daten aus unterschiedlichen Versionen gibt es eine einfache Webseite in der man an entsprechenden Stellen die IDs von GTFS-Daten eingeben kann [11]. Über das Vergleiche-Icon in der letzten Spalte der Tabelle einer Länderspezifischen GTFS-Übersicht lässt sich ebenfalls ein Vergleich anstoßen, z.B.: was hat sich zwischen zwei GTFS-Versionen geändert?

Für PTNA stehen noch einige Ideen im Raum:

- Ausgabe der Ergebnisse (score-Werte) des Vergleichs von GTFS- und OSM-Daten im PTNA-Report
  - basierend auf den CSV-Daten im OSM-Wiki
  - basierend auf den *qtfs:\** tags in den OSM-Relationen
- Einbau einer "sidebar", über die Fehlermeldung im PTNA-Report ein- und ausgeblendet werden können
  - Analog zu KeepRight [12] und PTSA [13]
- Erweiterung von Statistiken
  - Entwicklung der Anzahl Fehler einer Kategorie über einen Zeitraum
- Re-Design der sqlite3-Datenbankschemas für GTFS-Daten
  - o man lernt nie aus

### Kontakt zum Autor:

Antonius "Toni" Erdmann Friedenstraße 21 D-85521 Ottobrunn +49 89 6094219 osm-ToniE@web.de

Diskussionskanal in der OSM-Community:

"PTNA: news for Public Transport Network Analysis" [14]

### Literatur

- [1] Toni Erdmann, "PTNA: Qualitätssicherung für ÖPNV in OpenStreetMap", FOSSGIS 2020 Tagungsband, Seite 109, <a href="https://www.fossgis-konferenz.de/2020/data/FOSSGIS Tagungs-band\_2020.pdf">https://www.fossgis-konferenz.de/2020/data/FOSSGIS Tagungs-band\_2020.pdf</a>
- [2] "GTFS", "General Transit Feed Specification", <a href="https://gtfs.org/">https://gtfs.org/</a>
- [3] "GTFS reference", <a href="https://gtfs.org/documentation/schedule/reference/">https://gtfs.org/documentation/schedule/reference/</a>
- [4] "GTFS best practices", <a href="https://gtfs.org/documentation/schedule/schedule-best-practices/">https://gtfs.org/documentation/schedule/schedule-best-practices/</a>
- [5] "Codes for identifying the principal <u>subdivisions</u> (e.g., <u>provinces</u> or states) of all <u>countries</u> coded in ISO 3166-1", <u>https://en.wikipedia.org/wiki/ISO 3166-2</u>

- [6] "OSM contributors", <a href="https://wiki.openstreetmap.org/wiki/Contributors">https://wiki.openstreetmap.org/wiki/Contributors</a>
- [7] "Proposal:Public Transport", approved version, <a href="https://wiki.openstreetmap.org/w/index.php?">https://wiki.openstreetmap.org/w/index.php?</a>
  title=Proposed\_features/Public\_Transport&oldid=625726
- [8] "GTFS", Spezifikation der gtfs:\* tags im OSM-Wiki, https://wiki.openstreetmap.org/wiki/GTFS
- [9] "PTNA", "Public Transport Network Analysis", <a href="https://ptna.openstreetmap.de">https://ptna.openstreetmap.de</a>
- [10] Beispiel für die Analyse einer Route-Relation durch PTNA, <a href="https://ptna.openstreetmap.de/relation.php?id=1549761&lang=de/">https://ptna.openstreetmap.de/relation.php?id=1549761&lang=de/</a>
- [11] Vergleich von GTFS-Daten mit GTFS-Daten, <a href="https://ptna.openstreetmap.de/gtfs/compare.-php">https://ptna.openstreetmap.de/gtfs/compare.-php</a>
- [12] "KeepRight", <a href="https://keepright.at/report\_map.php?lang=de">https://keepright.at/report\_map.php?lang=de</a>
- [12] "PTSA", "Public Transport Stop Analysis", <a href="https://gauss.whz.de/ptsa/#3/50.72/12.50">https://gauss.whz.de/ptsa/#3/50.72/12.50</a>
- [14] OpenStreetMap Community Forum, "PTNA: news for Public Transport Network Analysis": <a href="https://c.osm.org/t/ptna-news-for-public-transport-network-analysis/8383/">https://c.osm.org/t/ptna-news-for-public-transport-network-analysis/8383/</a>
- [15] GitHub-URL für PTNAs Komponenten, https://github.com/osm-ToniE/