Using Onestop IDs as a “Crosswalk” Between Transit Agencies and Data Sources

Drew Dara-Abrams
Example IDs from GTFS feeds

<table>
<thead>
<tr>
<th>agency_id</th>
<th>agency_name</th>
<th>agency_url</th>
<th>agency_timezone</th>
<th>agency_lang</th>
</tr>
</thead>
<tbody>
<tr>
<td>SFMTA</td>
<td>San Francisco Municipal Transportation Agency</td>
<td><a href="http://www.sfmta.com">http://www.sfmta.com</a></td>
<td>America/Los_Angeles</td>
<td>en</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>stop_id</th>
<th>stop_name</th>
<th>stop_desc</th>
<th>stop_lat</th>
<th>stop_lon</th>
<th>zone_id</th>
<th>stop_url</th>
<th>location_type</th>
<th>parent_station</th>
<th>stop_timezone</th>
<th>wheelchair_boarding</th>
</tr>
</thead>
<tbody>
<tr>
<td>5727</td>
<td>Metro Civic Center Station/Downtn</td>
<td></td>
<td>37.778542</td>
<td>-122.414813</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>route_id</th>
<th>agency_id</th>
<th>route_short_name</th>
<th>route_long_name</th>
<th>route_desc</th>
<th>route_type</th>
<th>route_url</th>
<th>route_color</th>
<th>route_text_color</th>
</tr>
</thead>
<tbody>
<tr>
<td>1093</td>
<td>SFMTA</td>
<td>N</td>
<td>JUDAH</td>
<td></td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>agency_id</th>
<th>agency_name</th>
<th>agency_url</th>
<th>agency_timezone</th>
<th>agency_lang</th>
</tr>
</thead>
<tbody>
<tr>
<td>BART</td>
<td>Bay Area Rapid Transit</td>
<td><a href="http://www.bart.gov">http://www.bart.gov</a></td>
<td>America/Los_Angeles</td>
<td>en</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>stop_id</th>
<th>stop_name</th>
<th>stop_desc</th>
<th>stop_lat</th>
<th>stop_lon</th>
<th>zone_id</th>
<th>stop_url</th>
<th>location_type</th>
<th>parent_station</th>
<th>stop_timezone</th>
<th>wheelchair_boarding</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIVC</td>
<td>Civic Center/UN Plaza</td>
<td></td>
<td>37.779528</td>
<td>-122.413756</td>
<td>CIVC</td>
<td><a href="http://www.bart.gov/stations/CIVC/">http://www.bart.gov/stations/CIVC/</a></td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>route_id</th>
<th>agency_id</th>
<th>route_short_name</th>
<th>route_long_name</th>
<th>route_desc</th>
<th>route_type</th>
<th>route_url</th>
<th>route_color</th>
<th>route_text_color</th>
</tr>
</thead>
<tbody>
<tr>
<td>07</td>
<td>BART</td>
<td></td>
<td>Richmond - Daly City/Millbrae</td>
<td></td>
<td>1</td>
<td><a href="http://www.bart.gov/schedules/bylineresults?route=7,ff0000">http://www.bart.gov/schedules/bylineresults?route=7,ff0000</a>,</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
How many GTFS entity IDs could there be?

6,000+
TRANSPORTATION AUTHORITIES
To connect users with the information they need, we’ve worked with thousands of transportation authorities to get transit routes on Google Maps.

2,500,000
TRAIN STATIONS, BUS STOPS, FERRY TERMINALS, AND MORE
An ideal ID scheme for GTFS entities would be:

- globally unique
- able to join across feeds
- stable across feed versions
- machine readable
- machine writable
- human readable
- human writable
- efficient to index for searches (geographic and textual)
- distributed in its architecture, for use by many producers and consumers
Existing options for IDs across sources

**feeds → web URL**
http://www.bart.gov/dev/schedules/google_transit.zip

**agencies → US National Transit Database ID** (766 total as of last September)
3030 Washington Metropolitan Area Transit Authority

**stops → UK National Public Transport Access Node Database** (~450,000)
Moor Street Queensway bus stop in Birmingham
AtcoCode: 43000206801
NaptanCode: nwmajgaw
Transitland’s Onestop ID scheme

entity | geohash | name

o-9q9-BART
Onestop ID for feeds and operators

- **feeds**: f-9q9-BART
- **operators**: o-9q9-BART
- **stops**: s-9q8yymt-CivicCenter~UNplaza, s-9q8yymt-CivicCenter~UNplaza>8thSt, s-9q8yymt-CivicCenter~UNplaza<2
- **routes**: r-9q9n-fremont~dalycity
Onestop ID for stops

**feeds**
- f-9q9-BART

**operators**
- o-9q9-BART

**stops**
- s-9q8yymt-CivicCenter~UNplaza
- s-9q8yymt-CivicCenter~UNplaza>8thSt
- s-9q8yymt-CivicCenter~UNplaza<2

**routes**
- r-9q9n-fremont~dalycity
Onestop ID for routes

**feeds**
- f-9q9-BART

**operators**
- o-9q9-BART

**stops**
- s-9q8yymt-CivicCenter~UNplaza
- s-9q8yymt-CivicCenter~UNplaza>8thSt
- s-9q8yymt-CivicCenter~UNplaza<2

**routes**
- r-9q9n-fremont~dalycity
Onestop ID to represent relationships

- f-9q9-BART
  - o-9q9-BART
    - r-...
    - r-9q9n-fremont-dalycity
  - STOP
    - s-9q8yymt-CivicCenter~UNplaza
  - <!-- FEED -->

- f-9q8y-SFMTA
  - o-9q8y-SFMTA
    - r-9q8yv-n-judah
    - r-...
  - <!-- OPERATOR -->
  - p-9q8y-SFMTA
  - <!-- ROUTE -->
  - --> STOP
  - --> s-9q8yymt-CivicCenter~UNplaza
live demos
We welcome your involvement

Read more about Transitland
https://transit.land

Try the Playground data explorer
https://transit.land/playground

Try the Transitland Datastore API
https://github.com/transitland/transitland-datastore

Comment on the Onestop ID scheme
https://github.com/transitland/onestop-id-scheme

Follow us on Twitter
@transitland

Or drop me a note
drew@mapzen.com
Geohashes

https://mapzen.com/blog/geohashes-and-you
http://mapzen.github.io/leaflet-spatial-prefix-tree/
Onestop ID has been developed against 765 transit agencies and 14 countries.