Towards interoperable traffic data sources

Mikko Rinne
Aalto University
Scope

• Survey of open data interfaces and datasets for land person traffic and transportation
  – Self-empowered pedestrians, bikers and drivers
  – Public transportation users
  – Excluded: Air travel, sea travel and goods transport
• Formats and Protocols
• Examples of use
• Identifiers
• Linked open data in traffic
• Summary & Conclusions
Formats and Protocols
International Standards

European Committee for Standardization (CEN) Technical Committee (TC) 278

- SG6: IFOPT
- SG7: SIRI
- WG3 Public Transport
- WG4 Traffic and Traveller Information
- DATEX II
- WG8 Road Traffic Data

ISO / TC 204: Intelligent transport systems

- WG8 Public Transport
- WG10: Traveller Information Systems
- WG9: Integrated transport information, management and control

ITS in your pocket
Proven solutions driving user services

10th ITS EUROPEAN CONGRESS
Helsinki, Finland | 16-19 June 2014
CEN / TC 278 Specifications

• Service Interface for Real-Time Information (SIRI)
  – Exchange of information about public transport services and vehicles in distributed systems
  – E.g. Vehicle Monitoring (VM), Stop (ST) and General Messaging (GM) Services
  – Messages exchanged either as XML documents with http-post or using SOAP

• Identification of Fixed Objects in Public Transport (IFOPT)
  – Main fixed objects related to public transport
  – E.g. Stop place model

• DATEX II
  – DATEX II aims to attract all actors in the dynamic traffic and travel information sector
  – E.g. traffic elements, operator actions, measured data and variable message signs
Google GTFS and realtime

- **General Transit Feed Specification (GTFS)**
  - Set of specifications by Google for route, timetable and infrastructure information
  - A “feed” consists of a set of CSV text files collected into a zip-compressed package, made available in the net
  - Each file addresses a certain aspect of the information, e.g. stops, routes, trips and stop times
  - Extremely popular; over 700 transit agencies listed*) in Nov 2013, list was incomplete

- **GTFS realtime**
  - Extension of GTFS supporting realtime updates to fleet positions
  - Based on Google-proprietary “protocol buffer” format
  - Trip updates, service alerts, vehicle positions

Open Street Map

- Open Street Map (OSM) currently consists of over 2 billion nodes kept up-to-date by 1.4 million active contributors.
- Has recommendations for tagging of both fixed public transportation infrastructure and public transportation routes.
  - Offers a comprehensive set of tags for public transport stops (23 vs. e.g. 12 parameters in GTFS).
- Flipside of flexibility: Applications need to be prepared for inconsistencies and changing recommendations and practices.
- OSM wiki lists 11 applications using OSM public transport data.
Examples of Open Traffic Data
Road Conditions and Traffic

• Digitraffic by the Finnish Transport Agency
  – Dynamic real-time fluency and travel time
  – Current road weather station data
  – WS-I based WSDL specification of SOAP-messaging

• Incident and roadwork information
  – RSS
  – FTP downloads of XML-files in national and DATEX II formats

• No other freely available traffic data discovered, but e.g. Google and TomTom make their data available through web services and mobile applications
Public Transportation Info

- Early SIRI implementations available in Tampere (VM, ST, GM) and New York (VM, SM)
- UK stop information is NaPTAN (earlier version of IFOPT)
- HAFAS
  - The timetable information system of German traffic, transport and logistics software company HaCon
  - Users in 25 countries, e.g. Deutsche Bahn in Germany and SNCF in France
  - Supports multiple interfaces including SIRI but API documentation hard to find
- Finnish API:s mostly local solutions
  - Nation-wide connection search and bus stop timetable API using HTTP get requests, responses in XML
  - Helsinki regional transport offers API:s for journey planning, live fleet tracking, service disruptions and stop-specific next departures
  - National railway provides information in RSS feeds
Crowd-Sourcing Vehicle Positions

- So far mostly used by car navigation companies to obtain traffic fluency data
- Research experiments made in areas, where no positioning data from buses is available
  - Public transportation passengers using a mobile application to semi-automatically report the position of the buses to fellow passengers.
Identifiers

• No dominant international scheme has emerged
  – IFOPT and DATEX II recommending to pre-fix identifiers with IANA top level domain (ISO 3166-1) country codes
  – IFOPT recognizes the needs for both unique system-level identifiers and human Usable short identifiers
  – NaPTAN identifies stops with a 12-character “AtchoCode” and 7- or 8-digit “NaPtanCode”
  – International union of railroads (UIC) has uses a 7-digit station identifier beginning with a 2-digit country code
• Finnish Digitraffic uses local schemes based on Finnish road numbering and addresses
• GTFS defines no identifiers, but includes fields also for URI:s of things
Example Identifiers in Finnish Public Transport API:s

<table>
<thead>
<tr>
<th>Name</th>
<th>Source</th>
<th>Scope</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>StationId</td>
<td>matka.fi, HSL</td>
<td>National</td>
<td>7-digit numeric code</td>
</tr>
<tr>
<td>CountryId</td>
<td>matka.fi, HSL</td>
<td>Global, IANA</td>
<td>&quot;fi&quot; for Finland.</td>
</tr>
<tr>
<td>CityId</td>
<td>matka.fi, HSL</td>
<td>National</td>
<td>Optional, defines city or county.</td>
</tr>
<tr>
<td>GlobalId</td>
<td>matka.fi, HSL</td>
<td>National</td>
<td>&quot;Digistop&quot;-id. National database collecting information on bus stops.</td>
</tr>
<tr>
<td>JORE stop code</td>
<td>HSL (Live)</td>
<td>Capital area</td>
<td>A 7-digit stop code listed in the Helsinki region register of public transportation.</td>
</tr>
<tr>
<td>Metropolitan area number</td>
<td>HSL</td>
<td>Capital area</td>
<td>An older 4-, 5- or 6-digit stop identifier with a four-digit stop number and 0-2 character city prefix.</td>
</tr>
<tr>
<td>JORE route</td>
<td>HSL</td>
<td>Capital area</td>
<td>Public transportation routes: Seven characters.</td>
</tr>
<tr>
<td>Route</td>
<td>HSL Live</td>
<td>Capital area</td>
<td>Four- or five-digit identifiers with examples given as &quot;1052V.1&quot; and &quot;1064&quot;</td>
</tr>
<tr>
<td>Vehicle id</td>
<td>HSL Live</td>
<td>Capital area</td>
<td>Vehicle identification as e.g. &quot;CEENG1074300245&quot;</td>
</tr>
<tr>
<td>Departure id</td>
<td>HSL Next Departures</td>
<td>Capital area</td>
<td>An 8-digit number.</td>
</tr>
</tbody>
</table>
## Linked Open Traffic Data

<table>
<thead>
<tr>
<th>Stars:*)</th>
<th>Definition:</th>
<th>Status:</th>
</tr>
</thead>
<tbody>
<tr>
<td>★</td>
<td>Available on the web with an open licence</td>
<td>Not always fully openly available, most providers require a separate license agreement. All surveyed licenses except Digiroad were free-of-charge.</td>
</tr>
<tr>
<td>★★</td>
<td>Available as machine-readable structured data</td>
<td>Everything is machine-readable. Some RSS-feeds provide only unstructured text.</td>
</tr>
<tr>
<td>★★★</td>
<td>as above plus non-proprietary format (e.g. CSV instead of excel)</td>
<td>The “protocol buffers” in GTFS-realtime are a proprietary format. Specifications are available, but processing is easier using tools from Google. Otherwise non-proprietary formats, mostly based on XML. Some using commonly agreed XML schemas.</td>
</tr>
<tr>
<td>★★★★</td>
<td>as above plus use open standards from W3C (RDF and SPARQL) to identify things so that they can be referenced</td>
<td>Nobody is providing data in RDF. GTFS supports URIs for agencies, routes and stops, but original purpose is to use them for web pages.</td>
</tr>
<tr>
<td>★★★★★</td>
<td>as above plus link the data to other people’s data to provide context</td>
<td>Some interfaces are referencing XML namespace definitions, but clear cross-compatibility of concepts by e.g. pointing to globally available ontologies was not detected between any two systems. None of the data directly links to any other data. OWL-ontologies on Open Street Map are available.</td>
</tr>
</tbody>
</table>

*) [http://www.w3.org/DesignIssues/LinkedData.html](http://www.w3.org/DesignIssues/LinkedData.html)
Traffic Linked Open Data Experiments

• Trial service in Madrid\(^1\)
  - Local bus transport authority EMT follows a similar scheme as the providers in Finland: XML documents using a locally defined schema
  - The trial follows linked open data principles
    • URI:s given to both buses and bus stops
    • RDF used as the data format both for static and streaming data
  - *EMT Live Services* are converted into a SPARQL stream and static data is provided by a SPARQL endpoint
  - A tablet application has been created for data consumption demonstration.

• Smart city service in Dublin\(^2\)
  - Main target of the service city traffic monitoring and administration
  - Built on linked data principles using RDF and SPARQL, extended with proprietary streaming operators
  - Traffic-related information includes bus, bikes, air quality, ambient noise, pedestrian counts, traffic cameras and weather

---


Traffic Data Ontologies

• Three ontologies to provide vocabularies for OSM were discovered
• OSMOnto
  – Appears to be a one-time effort created for a research paper
  – Available as a downloadable OWL-file, but not as a browsable description
• OSM Semantic Network (OSN)
  – Looks up-to-date with OSM recommendations on public transport infrastructure
  – Browser-friendly access also available
• LinkedGeoData
  – By the Leipzig University to make OpenStreetMap data available as an RDF knowledge base according to the Linked Data principles
  – Using popular ontologies such as rdfs, foaf and skos in addition to an own linkedgeodata.org ontology
  – No references to OSN within linkedgeodata were discovered.
Conclusions

• Almost all current interface-formats are based on XML
  – JSON, RSS, proprietary CSV also appear
• Main protocol http, in older systems ftp
• New European standards DATEX II, SIRI and IFOPT created by CEN/TC 278 are not widely available in open interfaces yet
• OSM enjoys huge support and also contains public transport infrastructure information, but suffers from internal fragmentation
• For public transportation GTFS by Google is simple and has over 700 traffic operators listed
• Finnish open interfaces provide data on roads, traffic and public transport, but are practically all based on different specifications
• Traffic identifier situation is still very fragmented
  – Agreement to use ISO 3166-1 country prefixes
• Linked open data principles absent from current solutions