

---

This is an electronic reprint of the original article.  
This reprint may differ from the original in pagination and typographic detail.

Rinne, Mikko

## Towards interoperable traffic data sources

*Published in:*

10th ITS European Congress, Helsinki, Finland, June 16-19, 2014

Published: 01/01/2014

*Document Version*

Publisher's PDF, also known as Version of record

*Please cite the original version:*

Rinne, M. (2014). Towards interoperable traffic data sources. In *10th ITS European Congress, Helsinki, Finland, June 16-19, 2014* ERTICO.

---

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

# Towards interoperable traffic data sources

**Mikko Rinne**

Aalto University, PO Box 15400, FI-00076 Aalto, Finland, mikko.rinne@aalto.fi

## **Abstract**

Data related to the mobility of people has been one of the key areas for opening data resources and interfaces. Public availability of information on road weather, traffic and public transportation both as timetables as well as live fleet tracking have enabled new types of applications, especially for mobile platforms. This paper surveys the current status of traffic data availability and interoperability through a review of current specifications for applicable interfaces. These are compared with examples of implementations currently in use. The conformance to linked open data principles is reviewed. Based on the observations, recommendations for future interoperability improvements are given.

**Keywords:** Transport, Open data, Linked data

## **1 Introduction**

The availability of open traffic data is improving at a breathtaking pace. Advances in mobile technology are fuelling the development of mobile applications both for passengers of public transportation and taxis as well as self-empowered commuters like car drivers, cyclists and pedestrians.

The first generation of public traffic information *APIs* (Application Programming Interface) have mostly been implemented by individual transport operators, national traffic agencies, software companies and open-source developers. New solutions are evolving from provider-oriented to consumer-centric management of multimodal trip chains. Information producer and consumer roles may alternate dynamically. Traffic application ecosystem development, from a developer's point-of-view, benefits from common practices and formats across geographical areas and traffic forms.

This survey takes a look at the specifications and systems currently in use, both internationally and nationally. The target is to find out how the different approaches work together for future improvements in compatibility. In light of this target it is also reviewed, how the principles of linked open data are currently taken into account.

The rest of this document is structured as follows: Section 2 looks at traffic-related data exchange formats. Examples of use are categorized by the type of information accessed in Section 3. Cross-compatibility of identifiers is considered in Section 4 and compliance to

linked open data principles in Section 5. The results are summarized and recommendations given in Section 6.

## 2 Data Formats

There is a multitude of formats in use for traffic- and travel-related data. As the specifications only define the transfer or storage formats of the data, not whether any particular content is openly licensed, the formats and specifications reviewed in this section have been selected either based on their focus on traffic-related data or with reverse engineering based on their usage in currently available open or free-license traffic data services.

### 2.1 European Traffic Standards (SIRI, IFOPT, DATEX II)

*European intelligent transport systems*<sup>1</sup> are being standardized by Technical Committee (TC) 278 "Intelligent transport systems" of the *European Committee for Standardization* (CEN = Comité Européen de Normalisation). CEN/TC 278 works in collaboration with *ISO/TC 204*, responsible for global standardisation in the area.

*Service Interface for Real-time Information*<sup>2</sup> (SIRI) is an *XML*-protocol<sup>3</sup> for real-time exchange of information about public transport services and vehicles in distributed systems developed by SG7 of CEN/TC 278 WG3. SIRI messages may be exchanged either as XML documents with *HTTP*<sup>4</sup> POST method or using *SOAP*<sup>5</sup> messages, for which also a *WSDL*-binding (Web Service Definition Language<sup>6</sup>) has been defined. SIRI supports both request/response and publish/subscribe communications. For publish/subscribe the interaction follows recommendations of the *WS-PubSub* (Publish-Subscribe Notification for Web Services<sup>7</sup>) specification. The SIRI schemas are available as free downloads<sup>8</sup>, but the full documents have to be acquired through participating national organizations. Currently SIRI includes the following services:

- **Production Timetable Service (PT):** Information about the expected operation of a transport network for a day in the near future, incorporating any changes.
- **Estimated Timetable Service (ET):** Real-time deviations and control actions

---

<sup>1</sup> <http://www.itsstandards.eu/>

<sup>2</sup> <http://user47094.vs.easily.co.uk/siri/>

<sup>3</sup> <http://www.w3.org/XML/>

<sup>4</sup> <http://www.w3.org/Protocols/>

<sup>5</sup> <http://www.w3.org/TR/soap/>

<sup>6</sup> <http://www.w3.org/TR/wsdl>

<sup>7</sup> <http://www.ibm.com/developerworks/library/specification/ws-pubsub/>

<sup>8</sup> <http://user47094.vs.easily.co.uk/siri/schema/schemas.htm#2.0>

affecting the timetable for the current day. Can be filtered by operator or line.

- **Stop Services (ST):** Current and near future vehicle arrivals and departures at a nominated stop or other monitoring point.
- **Vehicle Monitoring Service (VM):** Current location and expected activities of a particular vehicle.
- **Connection Timetable (CT) and Connection Monitoring (CM) Services:** Intended for real-time monitoring of "guaranteed interchanges".
- **General Messaging Service (GM):** Exchange of general-purpose messages.

The corresponding public transport infrastructure format is defined in the *Identification of Fixed Objects in Public Transport*<sup>9</sup> (IFOPT) specification [2] by SG6 of CEN/TC 278 WG3. IFOPT defines a *stop place model* and identification principles for fixed objects related to public transport, e.g. stop points, stop areas, stations, connection links and entrances. IFOPT is based on an earlier UK specification on *National Public Transport Access Nodes*<sup>10</sup> (NaPTAN).

The *DATEX* standard was originally developed for information exchange between traffic management centres, traffic information centres and service providers. The second generation *DATEX II*<sup>11</sup> has a wider scope, aiming to attract all actors in the dynamic traffic and travel information sector. The work has been done in CEN/TC 278 WG 8 with plans to augment the results to international standards in collaboration with ISO TC204 WG9. Version 2.2 is currently only available to registered members, but v. 2.1 is publically available [4]. Information exchanged with DATEX II is composed of the following *basic elements*:

- **Traffic elements:** Road and traffic-related events not initiated by the traffic operator
- **Operator actions:** Classified as network management, roadworks, roadside assistance and sign settings referring to VMS messages.
- **Impacts:** Information on lane availability and delays.
- **Non-road event information:** Information on events not directly on the road.
- **Elaborated data:** Derived or computed data, e.g. travel times or traffic status
- **Measured data:** Information measured by sensors such as traffic and weather
- **Variable Message Signs (VMS):** Messages displayed on variable electric signs.

*Parking information publication* was added in version 2.1 addressing the exchange of both static and dynamic status information related to parking.

### 2.2 De-facto standards by Google (GTFS)

---

<sup>9</sup> <http://www.dft.gov.uk/naptan/ifoft/>

<sup>10</sup> <http://data.gov.uk/dataset/naptan>

<sup>11</sup> <http://www.datex2.eu/>

*General Transit Feed Specification*<sup>12</sup> (GTFS) covers a set of specifications by Google for route, timetable and infrastructure information. A GTFS "feed" consists of a set of *CSV* (Comma-Separated Values<sup>13</sup>) files collected into a zip-compressed package, made available in the Internet. Each file addresses a certain aspect such as the public transport company, stops, routes, trips, stop times, service exceptions and fare attributes.

GTFS-realtime<sup>14</sup> is an extension of the GTFS specification supporting realtime updates to fleet positions. The supported types of data are vehicle positions, trip updates (delays, cancellations, changed routes) and service alerts. The specified method of delivery is HTTP GET with frequent updates. The GTFS-realtime data exchange format is based on protocol buffers<sup>15</sup>, which are specified by Google for serializing structured data.

### 2.3 Open Internet Formats

The *Open Street Map*<sup>16</sup> (OSM) is a highly successful crowdsourcing project of geographical data, including public transport infrastructure. Based on statistics available from the site (checked 1.11.2013) the database includes over 2 billion nodes kept up-to-date by 1.4 million active contributors. The data formats used in OSM are in constant evolutionary development. As an example the original *public transport recommendation*<sup>17</sup> used the "highway=bus\_stop" tag for saving information on public transportation stops. This was revised by a newer recommendation<sup>18</sup> in 2011, where "public\_transport=stop\_position" describes the place where a bus stops and "public\_transport=platform" the place where passengers wait for and can board the vehicle. Another recommendation has also been worked on<sup>19</sup>, but it hasn't been finalized or approved as of yet. Applications using OSM need to be prepared to cope with both inconsistencies as well as changes of recommendations and practices.

### 2.4 Other Formats

*Web Services Interoperability* (WS-I) Basic Profile<sup>20</sup> [10] defines a set of non-proprietary Web services specifications with references to specifications promoting interoperability.

---

<sup>12</sup> <https://developers.google.com/transit/gtfs/>

<sup>13</sup> <http://www.ietf.org/rfc/rfc4180.txt>

<sup>14</sup> <https://developers.google.com/transit/gtfs-realtime/>

<sup>15</sup> <https://developers.google.com/protocol-buffers/>

<sup>16</sup> <http://www.openstreetmap.org/>

<sup>17</sup> [http://wiki.openstreetmap.org/wiki/Public\\_Transport](http://wiki.openstreetmap.org/wiki/Public_Transport)

<sup>18</sup> [http://wiki.openstreetmap.org/wiki/Proposed\\_features/Public\\_Transport](http://wiki.openstreetmap.org/wiki/Proposed_features/Public_Transport)

<sup>19</sup> [http://wiki.openstreetmap.org/wiki/Proposed\\_features/Stop\\_Area](http://wiki.openstreetmap.org/wiki/Proposed_features/Stop_Area)

<sup>20</sup> <http://www.ws-i.org/profiles/basicprofile-1.1-2004-08-24.html>

Messaging is formatted according to SOAP and service descriptions are in WSDL. Finnish *Digitraffic* services (more details in Section 3.2) are based on WS-I.

*Really Simple Syndication*<sup>21</sup> (RSS) is an XML-based format for content distribution. Various types of traffic and transport data are available as RSS-feeds. The challenge for software development is that RSS-feeds are often provided as unstructured text, which makes parsing specific data items difficult and error-prone.

### 3 Examples of open traffic data in use

Many of the specifications listed in Section 2 are either fairly new, proprietary or originally tied to an individual service. To illustrate the current status this section looks at implementations, where either the data formats presented above or proprietary formats are used to make traffic and transport data openly available. Some of the examples are focused in Finland, serving as a case-study snapshot of the situation.

#### 3.1 Public transport

For public transport infrastructure data GTFS has gained very wide acceptance. Publically available feeds are listed both by Google<sup>22</sup> and a GTFS Data Exchange<sup>23</sup> site (over 700 transit agencies Nov 15th 2013). It should also be noted that cities, whose GTFS feeds are available but require registration, may be absent from the lists.

OSM provides a comprehensive format for crowd-sourced information including both public transportation stops as well as the routes associated with the stops. A large set of tags (23 vs. e.g. 12 parameters in GTFS) is available for describing a public transportation stop, e.g. whether the stop is sheltered or has a bench, or whether the departures board or the advertising are electric. OSM wiki<sup>24</sup> lists 11 applications using OSM public transport data.

Not many open implementations based on SIRI were discovered yet. Vehicle Monitoring (VM) and Stop Monitoring (SM) have been implemented by *MTA Bus Time*<sup>25</sup> (Metropolitan Transportation Authority) in the New York region and *ITS Factory*<sup>26</sup> in the city of Tampere in Finland. ITS Factory also supports general messaging (GM). No IFOPT-formatted infrastructure data sets were discovered, but UK stops in NaPTAN-format are available.

*HAFAS*, the timetable information system of the German traffic, transport and logistics

---

<sup>21</sup> <http://www.rss-specifications.com/rss-specifications.htm>

<sup>22</sup> <https://code.google.com/p/googletransitdatafeed/wiki/PublicFeeds>

<sup>23</sup> [http://www.gtfs-data-exchange.com/agencies#filter\\_official](http://www.gtfs-data-exchange.com/agencies#filter_official)

<sup>24</sup> [http://wiki.openstreetmap.org/wiki/List\\_of\\_OSM\\_based\\_Services#Public\\_Transport](http://wiki.openstreetmap.org/wiki/List_of_OSM_based_Services#Public_Transport)

<sup>25</sup> <http://bustime.mta.info/wiki/Developers/SIRIIntro>

<sup>26</sup> <http://data.itsfactory.fi/>

software company HaCon<sup>27</sup>, lists users in 25 countries, e.g. Deutsche Bahn in Germany and SNCF in France. HAFAS supports also real-time information fusion with timetable data for planned trips, and lists support for *UIC* (International Union of Railroads<sup>28</sup>) and SIRI, among others. An API is available, but only user compilations of pieces of data from the HAFAS systems in different countries<sup>29</sup> were found, so HAFAS is not included as an open interface in the comparison of Section 5.

One way to achieve live tracking of public transport vehicles without installing positioning equipment and real-time data connections is to use crowdsourcing according to [3], where public transportation passengers used a mobile application to report the bus position.

The Finnish Transport Agency is offering their open *matka.fi*<sup>30</sup> API to search trains, domestic air transport, buses between countryside and cities, public transportation in cities and some demand-responsive-transport services. The service is provided using an HTTP GET interface with responses given in XML. A complete dump of the database behind the API is available also in *kalkati.net*<sup>31</sup> XML format.

*Helsinki Regional Transport Authority*<sup>32</sup> (HSL) provides a comprehensive set of APIs, which need a license agreement, but are available free-of-charge. In addition to the timetable GTFS-feed and *kalkati.net* format data dump information is provided as:

- **Journey planner API:** Timetable information via HTTP GET, responses either XML or *JSON* (JavaScript Object Notation<sup>33</sup>).
- **Live:** Real-time fleet information over an implementation-specific interface [6] available from HSL<sup>34</sup>. Both a CSV-based push interface using a socket connection and an HTTP request interface are available.
- **Service disruption info:** Realtime information on service disruptions in the HSL area [7]. The API uses HTTP GET requests and returns answers in XML.
- **Next departures:** A WSDL-defined SOAP-based API of next departures [8].

The Finnish railroad system (VR) offers an open API (trains on the map<sup>35</sup>) providing real-time

---

<sup>27</sup> <http://www.hacon.de/>

<sup>28</sup> <http://www.uic.org/>

<sup>29</sup> <http://www.administrator.de/wissen/hafas-fahrplanauskunft-api-sammlung-177145.html>

<sup>30</sup> <http://developer.matka.fi/pages/en/home.php>

<sup>31</sup> A national project by the Finnish Ministry of Transport and Communications for the communication between public transport operators. Process diagrams and XML schemas (<http://developer.matka.fi/pages/en/kalkati.net-xml-database-dump.php>)

<sup>32</sup> <http://developer.reittiopas.fi/pages/en/home.php>

<sup>33</sup> <http://www.json.org/>

<sup>34</sup> [http://dl.dropboxusercontent.com/u/20567085/Mattersoft%20Live%21%20interface%20description%20v1\\_6.pdf](http://dl.dropboxusercontent.com/u/20567085/Mattersoft%20Live%21%20interface%20description%20v1_6.pdf)

<sup>35</sup> [http://wiki.itsfactory.fi/index.php/Junat\\_kartalla\\_API](http://wiki.itsfactory.fi/index.php/Junat_kartalla_API)

GPS information on train locations. Information is returned as RSS-feeds<sup>36</sup>.

### 3.2 Road Conditions and Traffic Information

Digitraffic<sup>37</sup> [5] by the Finnish Transport Agency (FTA) offers real-time and historical information about traffic on the Finnish main roads. It is offered to developers and organisations free of charge with a license agreement. It is based on WS-I, i.e. WSDL-descriptions of SOAP-messaging. It includes road weather station data, weather camera pictures, dynamic real-time fluency and travel time data for specified road links. Incidents and roadworks are currently available as an RSS-feed<sup>38</sup>, but a structured version only through FTP (File Transfer Protocol<sup>39</sup>) downloads from an "Alert-C" location database. Incidents are delivered as separate XML files in both national and DATEX II formats<sup>40</sup>.

Information on parking is still mostly based on proprietary solutions. IFOPT has a basic parking model, mainly to associate parking facilities with stop areas. Parking is covered by DATEX II and services are being planned, but no working examples were discovered yet.

## 4 Identifiers

As interfaces and data sets are made available, their interoperability is largely determined by the uniformity of the identifiers used. Identifiers referring to the same things should be the same, or at least possible to computationally link to each other. How can the data in these interfaces be interlinked? How is it known, whether the identifiers are compatible?

### 4.1 Identifiers in specifications

IFOPT [2] section 6.9 addresses *stop identification*. No explicit identification scheme is defined, but it is noted that each stop can be associated with different kinds of identifiers:

1. **Unique System Identifiers:** Federated identifiers, can be long and do not need to be human user friendly. May also need to meet technical constraints such as being e.g. a single token from a restricted character set and not too verbose.
2. **Public Labels:** Human-recognizable labels for e.g. stations, platforms, bus stops.
3. **Public Codes:** Easy-to-remember short codes for human input

---

<sup>36</sup> [http://www.vr.fi/fi/index/palvelut/avoin\\_data.html](http://www.vr.fi/fi/index/palvelut/avoin_data.html)

<sup>37</sup> <http://www.infotripla.fi/digitraffic/english/index.html>

<sup>38</sup> <http://www2.liikennevirasto.fi/alk/RSS/liikennetiedote.rss>

<sup>39</sup> <http://www.ietf.org/rfc/rfc959.txt>

<sup>40</sup> [http://www.infotripla.fi/digitraffic/docs/D2-web\\_service-en.pdf](http://www.infotripla.fi/digitraffic/docs/D2-web_service-en.pdf)

In NaPTAN<sup>41</sup> each stop has two identifiers unique within the UK: *AtchoCode*, a 12-character system identifier and *NaPtanCode*, a short (7- or 8-digit) identifier suitable for public plates and easy to enter on a mobile phone. IFOPT also has a normative section (6.12) for stop place identifiers, in which the recommended structure of a unique identifier for Stop Places is *Country Code + Administrative Area Identifier + Stop Place Identifier*. The IANA top level domain country codes<sup>42</sup> (same as ISO 3166-1) are preferred as the country code in IFOPT and DATEX II (*InternationalIdentifier* class).

GTFS- or GTFS-realtime formats do not define any specific format for identifiers (other than all GTFS information being text files), but a number of identifiers are listed in the formats: *agency\_id* and URL<sup>43</sup>, *stop\_id* and URL, *route\_id* and URL, *stop\_code*, *zone\_id*, *trip\_id*, *block\_id*, *fare\_id* and *vehicle\_id*. Likewise the stop identifiers in OSM depend on the input of the contributors entering the stops. The OSM specifications also mention the UIC name and UIC reference<sup>44</sup> as tags, but they are officially abandoned. Some train stops e.g. in Germany are tagged with *uic\_ref* and *uic\_name* tags in OSM.

Identifiers encountered in the set of surveyed Finnish interfaces are summarized in Table 1. The identifier space is mostly populated by service-specific codes rooted in national standards and conventions. No specification references an ontology or XML vocabulary. The only references to international standards are "CountryId" using an IANA-compliant identifier and the use of codes from the *Trident* project for vehicle types. The work on a national stop code system has been started (Digistop), but it doesn't have full coverage. In many cases individual services use overlapping numbering schemes (matka.fi, HSL journey planner). Even services of the same provider (HSL journey planner, HSL live) may use different but overlapping identifiers.

**Table 1: Identifiers in Finnish traffic-related interfaces**

Name:	Source:	Scope:	Description
TNO	Digitraffic	National	National road number + address range, unique road info site
Site + Link	Digitraffic	Service-specific	Measurement site number + link between numbers
"road weather station"	Digitraffic	Service-specific	Four-digit number + string (including road type, road number, station name), no international specification
StationId	matka.fi, HSL	National (kalkati.net)	7-digit numeric code
CountryId	matka.fi, HSL	Global, IANA	"fi" for Finland.

<sup>41</sup> <http://www.dft.gov.uk/naptan/overview.htm>

<sup>42</sup> [http://en.wikipedia.org/wiki/List\\_of\\_Internet\\_top-level\\_domains](http://en.wikipedia.org/wiki/List_of_Internet_top-level_domains)

<sup>43</sup> <http://www.ietf.org/rfc/rfc3986.txt>

<sup>44</sup> [http://wiki.openstreetmap.org/wiki/Proposed\\_features/UIC\\_Reference](http://wiki.openstreetmap.org/wiki/Proposed_features/UIC_Reference)

## Towards Interoperable Traffic Data Sources

Name:	Source:	Scope:	Description
CityId	matka.fi, HSL	National	Optional, defines city or county.
GlobalId	matka.fi, HSL	National	"Digistop" <sup>45</sup> -id. National database collecting information on bus stops.
JORE stop code	HSL (Live)	Capital area	A 7-digit stop code listed in the Helsinki region register of public transportation <sup>46</sup> .
Metropolitan area number	HSL	Capital area	An older 4-, 5- or 6-digit stop identifier with a four-digit stop number and 0-2 character city prefix.
JORE route	HSL	Capital area	Public transportation routes: Seven characters.
Route	HSL Live	Capital area	Four- or five-digit identifiers with examples given as "1052V.1" and "1064"
Vehicle id	HSL Live	Capital area	Vehicle identification as e.g. "CEENG1074300245"
Departure id	HSL Next Departures	Capital area	An 8-digit number.
Vehicle type	Kalkati.net	International	A set of codes from the Trident project.
Train identifier	VR Trains on The Map	National	Train specifier used in local timetables, e.g. "P274"
Train station ID	VR ToTM	National	Abbreviation of the city name.

## 5 Linked Open Traffic Data

The concept of linked open data<sup>47</sup> was introduced by Tim Berners-Lee in 2006 and complemented in 2010 by the *five-star rating system*, which in Table 2 is compared with the current status of the surveyed traffic interfaces.

**Table 2: Five-star linked data definitions compared to the current status of traffic data**

Stars:	Definition:	Status:
★	Available on the web with an open licence.	Most providers require a separate license agreement. All surveyed licenses were free-of-charge.
★★	Available as machine-readable structured data.	Everything is machine-readable. Some RSS-feeds provide only unstructured text.
★★★	as above plus	The "protocol buffers" in GTFS-realtime are a

<sup>45</sup> <http://www.digistop.net>

<sup>46</sup> <http://www.cgi.fi/asiakasesimerkit/helsingin-seudun-liikenteen-joukkoliikennerekisteri-jore>

<sup>47</sup> <http://www.w3.org/DesignIssues/LinkedData.html>

	non-proprietary format (e.g. CSV instead of excel)	proprietary format. Specifications are available, but processing is easier using tools from Google. Otherwise non-proprietary formats, mostly based on XML. Some use commonly agreed XML schemas.
★★★★	as above plus use open standards from W3C (RDF and SPARQL) to identify things so that they can be referenced	Nobody is providing data in RDF. GTFS supports URIs for agencies, routes and stops, but original purpose is to use them for web pages rather than additional machine-readable data.
★★★★★	as above plus link the data to other people's data to provide context	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.

A typical interface today uses XML, JSON or CSV notation instead of *RDF* (Resource Description Framework<sup>48</sup>). There are no widely used and commonly agreed ontologies applied on these interfaces. Therefore traffic data interfaces mainly reach three stars, with some uncertainty also on the first (separate license agreements = not fully open), second (some data only available as unstructured RSS-feeds) and third (e.g. GTFS-realtime) stars.

Some recent research papers are documenting trials to convert provider-specific XML formats to linked open data. In [1] XML documents with local schema provided by EMT<sup>49</sup> were transformed to RDF and *URIs*<sup>50</sup> were assigned to both buses and bus stops. In [9] a smart city service was created in Dublin based on open information sources using RDF and extended SPARQL.

Three approaches to define globally available ontologies on OSM were discovered: *OSMonto*<sup>51</sup>, *OSM Semantic Network (OSN)*<sup>52</sup> and *LinkedGeoData*<sup>53</sup>.

## 6 Summary and Conclusions

A lot of open transport data is already available, with more coming all the time. The

<sup>48</sup> <http://www.w3.org/RDF/>

<sup>49</sup> <http://www.emtmadrid.es/>

<sup>50</sup> <http://www.ietf.org/rfc/rfc3986.txt>

<sup>51</sup> <http://wiki.openstreetmap.org/wiki/OSMonto>

<sup>52</sup> <http://wiki.openstreetmap.org/wiki/OSMSemanticNetwork>

<sup>53</sup> <http://linkedgeo.org/>

*OpenCities* project maintains a catalogue of urban transport data sources<sup>54</sup>. This study has looked at the current status of open data. The survey split itself naturally into two main categories: data formats used to publish open traffic data and examples of open data sources for traffic purposes.

Almost all interfaces are based on XML, with occasional appearances of JSON, RSS and proprietary CSV-based formats. The main protocol is HTTP, with some appearances of FTP in older systems. For live fleet tracking a push-based interface would be better than frequent HTTP polling, but only one interface (HSL Live using TCP socket) provided true streaming.

CEN/TC 278 has created pan-European ITS specifications for road traffic data as well as public transport fleet and infrastructure, but few openly available interfaces were discovered as of yet. Several de-facto candidates for public transportation information are available. Google Maps integration and simplicity are sharply raising the popularity of the GTFS formats among traffic operators and independent software developers. OSM, on the other hand, enjoys a large base of contributors for updating data and building community tools.

The situation with identifiers is fuzzy at this time: For example in Finland transport operators are using proprietary systems with overlapping schemes used for historical reasons both between and within operator services. A national public transportation stop, route or vehicle identification scheme with full coverage doesn't exist. Internationally there are fragmented attempts with *UIC references* of the railroad union, *HAFAS id* and an agreement in CEN/TC 278 to prefix identifiers with the IANA (ISO 3166-1) country identifiers. Identifier allocation should be federated, and therefore agreement on an international prefix and a suitable number of digits should be enough to return the problem to national level.

Linked open data principles are largely absent from the interfaces and datasets available today. An important step towards that goal would be to define new specifications together with, and partially in the form of accompanying ontologies or vocabularies, where the concepts could be formally specified. When common ontologies are referenced, programs can automatically identify two identifiers being the same and perform reasoning based on specified relations of concepts. After vocabularies are shared, the use of one of the RDF serializations (including XML) would be a small step.

Despite the challenges there is a major ongoing movement with both authorities and application developers to make traffic and transport data more openly available. It should also be in everyone's interest to promote wider use of international approaches, opening larger markets for application development and improving the scope of traffic applications in international travel. Important choices between candidates for next generation traffic and transport interfaces need to be made, and identifier compatibility coupled with linked open data principles are one possibility to improve the situation.

---

<sup>54</sup> <http://data.opencities.net/group/transport>

## 7 Acknowledgments

The work has been carried out in SPIRE and TrafficSense projects funded by Tekes and Aalto University. The guidance of Dr. Seppo Törmä and Dr. Esko Nuutila is gratefully acknowledged.

## 8 References

1. Calbimonte, J. et al.: Demo paper : Tablet-based visualization of transportation data in Madrid using SPARQLStream. Semantic Sensor Networks Workshop 2013. p. 4 (2013).
2. CEN/TC 278: Identification of Fixed Objects in Public Transport (TC 278 WI 00278207), (2008).
3. Corsar, D. et al.: GetThere : A Rural Passenger Information System Utilising Linked Data & Citizen Sensing. Demo in International Semantic Web Conference 2013. pp. 1–4 (2013).
4. European Commission: Datex II v2.1 User Guide, (2012).
5. gofore: Digitraffic Web Services v. 1.1, (2013).
6. Mattersoft: Mattersoft Live! Interface Description for Helsinki Region Transport v. 1.6, (2012).
7. Seasam Group: Poikkeusinfo, XML-rajapinnan kuvaus, rajapinnan versio 2, [http://developer.reittiopas.fi/media/Poikkeusinfo\\_XML\\_rajapinta\\_V2\\_2\\_01.pdf](http://developer.reittiopas.fi/media/Poikkeusinfo_XML_rajapinta_V2_2_01.pdf) [Retrieved Apr 11 2014], (2010).
8. Seasam House Oy: Omat Lähdöt ohjelmointirajapinta: Versio 1.01, [http://developer.reittiopas.fi/media/Omat\\_Lahdot\\_v101.pdf](http://developer.reittiopas.fi/media/Omat_Lahdot_v101.pdf) [Retrieved Apr 11 2014], (2010).
9. Tallevi-diotallevi, S. et al.: Real-time Urban Monitoring in Dublin using Semantic and Stream Technologies. Proceedings of the 12th International Semantic Web Conference. p. 16 , Sydney, Australia (2013).
10. Web Services Interoperability Organization: WS-I Basic Profile Version 1.1, (2004).