Publishing and linking transport data on the Web

Extended version

Julien Plu
UNIVERSITÉ MONTPELLIER 2
Sciences et Techniques
Place Eugène Bataillon
34095 Montpellier cedex 5 - France
julien.plu@etud.univ-montp2.fr

François Scharffe
LIRMM
UMR 5506 - CC477
161 rue Ada
34095 Montpellier Cedex 5 - France
francois.scharffe@lirmm.fr

ABSTRACT

Without Linked Data, transport data is limited to applications exclusively around transport. In this paper, we present a workflow for publishing and linking transport data on the Web. So we will be able to develop transport applications and to add other features which will be created from other datasets. This will be possible because transport data will be linked to these datasets.

We apply this workflow to two datasets: NEPTUNE, a French standard describing a transport line, and Passim, a directory containing relevant information on transport services, in every French city.

Categories and Subject Descriptors

[**Open data, Semantic Web, transport, framework**]:

Keywords

NEPTUNE, Passim, RDF, ontologies, interlinking, publication, conversion, DBPedia, INSEE, Datalift, Linked Data

1. INTRODUCTION

This project to build a framework for publishing and interlinking transport data on the Web was developed in collaboration between LIRMM[1] , Université Montpellier 2[2] and CETE[3] Méditerranée. The feasibility of publishing and interconnecting data transport into linked data on the web will be studied concretely from two data sources: the directory information services Passim and XML files corresponding to the NEPTUNE format describing public transport routes. Open Data, in France, has been booming for the last 2 years.

---

[1]Laboratoire d'Informatique de Robotique et de Microélectronique de Montpellier (`http://www.lirmm.fr`)
[2]`http://www.univ-montp2.fr`
[3]Centre d'Études Techniques de l'Équipement (`http://www.cete-mediterranee.fr`)

Open Data is the publication of public data, free of charge and in open formats so that people who want to use them can do it. This phenomenon is closely linked with the term Linked Data. It refers to a set of best practices for publishing and interlinking structured data on the Web. These best practices were introduced by Tim Berners-Lee and have become known as the Linked Data principles. These principles are 1) use URIs as names for things, 2) use URIs which are dereferenceable, 3) use RDF and SPARQL standards, 4) link URIs between them. So if all data is linked, a huge graph is created, which forms the Web of data (or Semantic Web). The Semantic Web, term introduced by Tim Berners-Lee [4], aims at putting data on the Web in a form that machines can naturally understand, so web content can be treated directly or indirectly by machines. This is done with the help of ontologies. An ontology, according to Tom Gruber [15] is the specification of a conceptualization. A conceptualization is an abstract and simplified view of the world that we want to represent. More simply an ontology represents knowledge as a set of concepts within a domain, and the relationships between those concepts.

1.1 Passim

The directory Passim [14] is published by the CERTU[4]. It identifies and provides a list of information services on French passenger transport and other mobility services. Its content is managed by the CETE Mãľditerranãľe. In practice, the directory is a website which retrieves the relevant services for a city or territory in France, distinguishing between modes or types of transport (car, transit, etc.) and perimeter (urban, departmental, and regional). Services are at least web sites, sometimes including phone services, or mobile applications (in the future it could also be Web services). The directory contains freely reusable data (now published in the open data portal of the French government [17]). This directory could be extended for example with references to other datasets. Its format is published in CSV[5] and is represented as follows (translated):

Sheet number;Service Name;Coverage service;Region;
Department;City;Modes of transport;Type of service;Network accessibility for disabled person;Land informations;Website;
Website accessibility for disabled person;Information points ;Remark;Comments;Sms;Mobile application;List of cities covered

---

[4]Centre d'Études sur les Réseaux, les Transports et l'Urbanisme
[5]Comma Separated Values (`http://www.passim.info/passim.csv`)

(Postal code);Sheet created;Sheet modified

Each column is separated by the character ';'. The names of these columns are self-explanatory, here is a small example of a CSV line:

```
1;05voyageurs;départementale;Provence-Alpes-Côte d'Azur;
Hautes-Alpes;N/A;Autocar, Covoiturage;Calcul d'itinéraire,
Description du réseau,Horaires;Non;;
http://www.05voyageurs.com;Non;;;;;;;09/06/2010;
04/08/2011
```

A succession of ";" means that the column between the two ";" is empty.

1.2 NEPTUNE

NEPTUNE[6] is first the outcome of the European project TRIDENT and then of French work on the CHOUETTE application. NEPTUNE is a French Standard (NFP-99506) specifying the reference format for data exchange of theoretical transport offers, particularly useful for the development of multimodal information systems. NEPTUNE specifications consist in, on the one hand of a conceptual data model in UML (from the TRIDENT project, based on Transmodel V4.1[7]) concerning the definition of the network (lines, stops) and the theoretical service (races, schedules), and on the other hand of an XSD schema. The exchanged data is in the form of a directory with an XML file per line, each file describing all the information about a transport line (stops, schedules, etc.). NEPTUNE profile is fully compatible with the CHOUETTE software whose development is supported by the Ministry of Transport[8]. NEPTUNE will evolve in the context of works within the European standardization of the Netex[9] project. It should be noted that Google has specified an exchange text format, GTFS, which is the standard for publishing transport data into open data (mainly North American networks). The first NEPTUNE open data were published by the General Council of Gironde[10], others will follow (the General Council of Isère and the Saône et Loire). Here is an example of this format with the modeling of a bus stop:

```
<ChouettePTNetwork>
  <ChouetteLineDescription>
    <StopPoint>
      <objectId>NINOXE:StopPoint:15577811
      </objectId>
      <objectVersion>0</objectVersion>
      <creationTime>2007-12-16T14:26:19.000+01:00
      </creationTime>
      <longitude>5.7949447631835940</longitude>
      <latitude>46.5263907175936000</latitude>
      <longLatType>WGS84</longLatType>
      <containedIn>NINOXE:StopArea:1557779
      </containedIn>
      <name>Cimetiere des Sauvages(A)</name>
    </StopPoint>
  </ChouetteLineDescription>
</ChouettePTNetwork>
```

---

[6] Norme dÉchange Profil Transport collectif Utilisant la Normalisation Européenne (http://www.chouette.mobi)

[7] http://www.transmodel.org/index.html

[8] http://www.chouette.mobi/spip.php?rubrique61

[9] http://www.kizoom.com/standards/netex/

[10] http://www.datalocale.fr/dataset/
liste-lignereguliere-transgironde

Where :

- ChouettePTNetwork: a tag covering all information contained in the file;

- ChouetteLineDescription: a tag containing every descriptions a line may have;

- StopPoint: a tag describing the information a stop may have.

For more details on this format one could consult the standard online[11].

1.3 General publication process

The process of publishing and interlinking a dataset from a non-RDF format into RDF can be done in 4 steps:

- selection or creation of an ontology: this step aims at choosing an existing ontology describing the dataset to publish or eventually create a new one;

- conversion to RDF: this step aims at converting the basic format of the dataset (XML, CSV, ...) in RDF, taking into account the chosen or created ontology;

- publication of RDF data: this step aims at making the data which have been previously converted into RDF available on the Web following the linked-data principles;

- interlinking with other datasets in RDF: this step aims at linking the RDF data with other datasets available elsewhere on the Web.
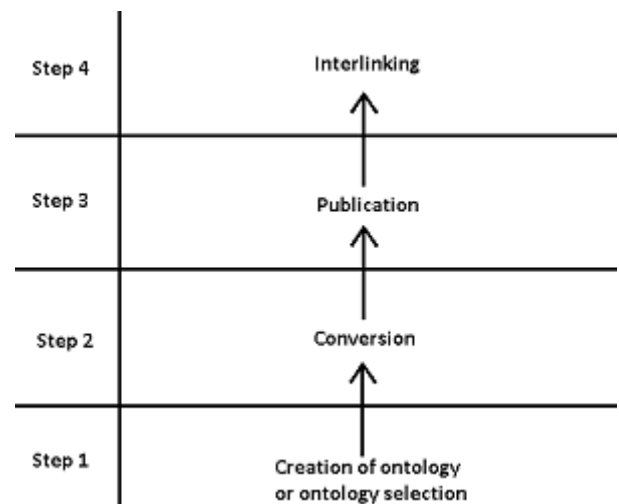
Here is a figure to illustrate these steps:

Figure 1: diagram of the publication process

These steps are more detailed in their respective sections. This process is taken from the ANR project, DataLift[12].

---

[11] http://www.chouette.mobi/IMG/pdf/NF__F_
-Neptune-maj.pdf

[12] http://www.datalift.org

1.4 Structure of the paper

In a first part we will present the two ontologies that were developed in order to describe data of the Passim directory, and NEPTUNE lines. In a second part we will see the means or methods used to convert these data into RDF. In a third part we will describe of the process of publishing the newly converted RDF data. Finally, in a fourth section we will detail the possible interlinking with other datasets.

2. RELATED WORK

In this section we will cover related work in transport data and formats, and data publication processes.

2.1 Open transport data

For the following reasons, transport data publication is popular in the UK:

- they are the pioneers of open data in Europe (culture of public transparency, Crown Copyright, etc.);

- the public transport sector has benefited from a decade of heavy investment in data standardization (Transport Direct project, standard NAPTAN, NPTG, TransXchange, etc.)[13].

A very good example of what can be done is posted on the United Kingdom[14] open data website. A lot of UK public transport data is on the Web, but little is available as linked data. Only a few datasets conform to Semantic Web standards have been identified: 1) the national repository of UK bus stops (NAPTAN), 2) the Bus timetable of Manchester[15], 3) maps of roads, railway stations and airport in the UK[16].

For a better understanding, a package containing all the linked data published by the Government of the United Kingdom is available online[17]. Despite the large number of U.S. transit systems whose tender is published in open data, their translation into linked data has only been the result of a few attempts and seems less advanced than in the United Kingdom.

Other data should follow, including all public transport timetables in the UK this year. There is also another ontology concerning transport, "The Tickets Ontology"[18].

In France, for the past two years, there is a strong interest for publishing public transport data as open data. For now, only three networks are online: Rennes, Nantes and Bordeaux. But with the ever-increasing number of data on the French open data portal there will certainly be more very quickly. From these data published in standard formats (NEPTUNE or GTFS), the technical work necessary to ensure their publication and their interlinking as linked data on the Web should not be too complex once the framework defined in this article will be completed.

---

[13] http://www.dft.gov.uk/public-transportdatastandards/
[14] http://data.gov.uk
[15] http://thedatahub.org/dataset/greater-manchester-bus-timetable-linked-data
[16] http://thedatahub.org/dataset/transport-data-gov-uk
[17] http://thedatahub.org/dataset/data-gov-uk
[18] http://www.heppnetz.de/ontologies/tio/ns

2.2 Data publication process

2.2.1 Selection or creation of an ontology

The following tools are used for the selection of ontologies.

LOV[19] [25]: A RDF dataset identifying vocabularies that describe Semantic Web datasets, as well as the relationships between these vocabularies. Watson [2] and Falcons [10]: two ontologies search engine.

And the following one for the creation of ontologies.

Protégé[20]: an RDFS and OWL ontology editor.

Neologism [3]: a Web application, derived from Drupal for publishing and creating ontologies. The ontologies Passim and NEPTUNE are published with this tool.

2.2.2 Conversion

Just like for the creation and selection of ontologies some tools are available on the Web.

CSV2RDF4LOD[21] : a tool to convert a CSV file in RDF.

GRDDL [16]: a tool to convert a XML file in XSL.

TrIplr[22]: a service generating RDF triples from URI.

D2R Server [11]: a tool converting the contents of a relational database in RDF. D2R is also a web server for publishing the data.

2.2.3 Publication

We considered the following two tools for data publication, although other tools are available.

Virtuoso[23]: a tool to publish RDF data with many other possibilities (mapping databases, personal vocabulary, etc.).

OpenRDF Sesame [9]: a tool to publish RDF data, focusing primarily on the publication and database management functionalities.

This list is not exhaustive. A benchmark study was conducted to compare all these tools [6].

2.2.4 Interlinking

The following tools were considered for interlinking datasets.

Silk [7]: provides a flexible declarative language to specify matching heuristics, which can combine different string comparators that can be digital as well as geographical.

LIMES [20]: implements a fast approach for the discovery of links at large scale, using metric spaces.

Recently another interlinking tool was developed within the project DataLift [13] along with a proposed framework for interlinking [22].

---

[19] Linked Open Vocabularies
[20] http://protege.stanford.edu/
[21] http://logd.tw.rpi.edu/technology/csv2rdf4lod
[22] http://triplr.org/
[23] http://virtuoso.openlinksw.com/

2.2.5 Existing frameworks

Currently there are already some frameworks to publish data in a raw format (CSV, XLS, XML, etc.) in RDF. But they have one (or more) gap(s) relative to what is presented in this article. The three most important are:

The Data Tank[24] : is a framework for publishing data in a RESTful way and easing the task of creating an application and allowing the publication and display of data in CSV, XLS and SHP (Shapefile). However it does not take the XML format, it does not publish data by using an ontology and does not makes interconnection.

LDIF [23]: is a framework for retrieving all the data of interest from several sources and unifying them with a vocabulary by creating a custom URI and when we request this URI that affects multiple datasets simultaneously.

LOD2 Stack[25] : is a suite of tool for data conversion, publication and interlinking. Its disadvantage is that it requires using many tools requiring learning many languages and specification format.

3. DATA PUBLICATION WORKFLOW

3.1 Selection or creation of an ontology

To take advantage of the semantics, we must model information, and for this, we rely on one or many ontologies. If the data is in a domain where there are already well-known ontologies, they are used. Otherwise, we have to create a new ontology, hoping it will be reused by others. Once the ontology created it must be published. Standards for writing ontologies are OWL [19] and RDF Schema [8]. RDF Schema provides basic elements for defining vocabularies or ontologies intended to structure RDF resources [18]: the main RDFS components are integrated into a more expressive (but more complex) ontology language, OWL.

3.1.1 Passim case

Concerning Passim there is no ontology that could have been specialized in this field: information transport. Therefore we have built the Passim ontology[26], linking it to the INSEE[27] ontology for the departments and regions, which originated in the work of DataLift[28]. The Passim ontology contains 4 classes and 18 properties. The list of classes being:

- TransportServiceInformation: this class represents a transport information service;

- Mode: this class represents the different modes of transport covered by the information service;

- Service: this class represents information services;

- Coverage: This class represents the geographic coverage of an information service.

And its schema (Figure 2).

---

[24]http://thedatatank.com/
[25]http://lod2.eu/Welcome.html
[26]http://data.lirmm.fr/ontologies/passim
[27]Institut National de la Statistique et des Études Économiques
[28]http://data.lirmm.fr/ontologies/geo-fr

3.1.2 NEPTUNE case

As for NEPTUNE there are two potentially reusable ontologies:

- the one used by the site of the UK Government, NAPTAN[29];

- the one used by the city of Manchester, TRANSIT[30].

Although these English data formats are "cousins" of the NEPTUNE format, there will be undoubtedly links between vocabularies to do, there were not yet adapted to the NEPTUNE representation, so it was necessary, like for Passim, to develop an ontology for NEPTUNE[31]. Over time (from late 2012 at best), the European standard, Netex (under development) should unify the vocabulary and formats for data exchange in Europe for public transport data. At short term, we cannot link the NEPTUNE data with other Semantic Web data except via the geographic coordinates of stops; our vocabulary seems to diverge somewhat from existing ontologies despite a few similarities, like stops and roads. This ontology can therefore represent all data that may contain the original XML file (stop, route, schedule, etc.). It was particularly connected with the FOAF ontology and INSEE. The NEPTUNE ontology contains 50 classes and 125 properties.

3.2 Conversion to RDF

To make a particular dataset usable, we must publish in a form allowing knowing that the data is related to concepts defined in an ontology. A dataset is described as a list of facts represented as triples (subject predicate value). The Semantic Web standard for this is RDF. It can have different serialization, the most known are:

- RDF/XML: the most used. This format is derived from XML;

- N3: less used than RDF/XML but more human readable;

- Turtle: sub-version of N3, just with less possibilities but same syntax;

- N-Triples: sub-version of Turtle. The most readable among the possible serialization but can be very huge because it is very expressive;

- JSON-LD: not very used because it is young but more readable than RDF/XML. This serialization is derived from JSON.

3.2.1 Passim case

The exported CSV data from the Passim website have been converted to RDF with the tool provided by DataLift to convert the CSV into gross RDF without considering any ontology. Once the RDF file provided by the tool, the only thing left to do was to convert this RDF file into another RDF file so that the data meets the Passim ontology. For this we must use the SPARQL [21] language and specifically the CONSTRUCT queries allowing to query RDF and producing RDF but modeled as desired, that is to say, in our case by taking into account the Passim ontology.

---

[29]http://transport.data.gov.uk/def/naptan/
[30]http://vocab.org/transit/terms/.html
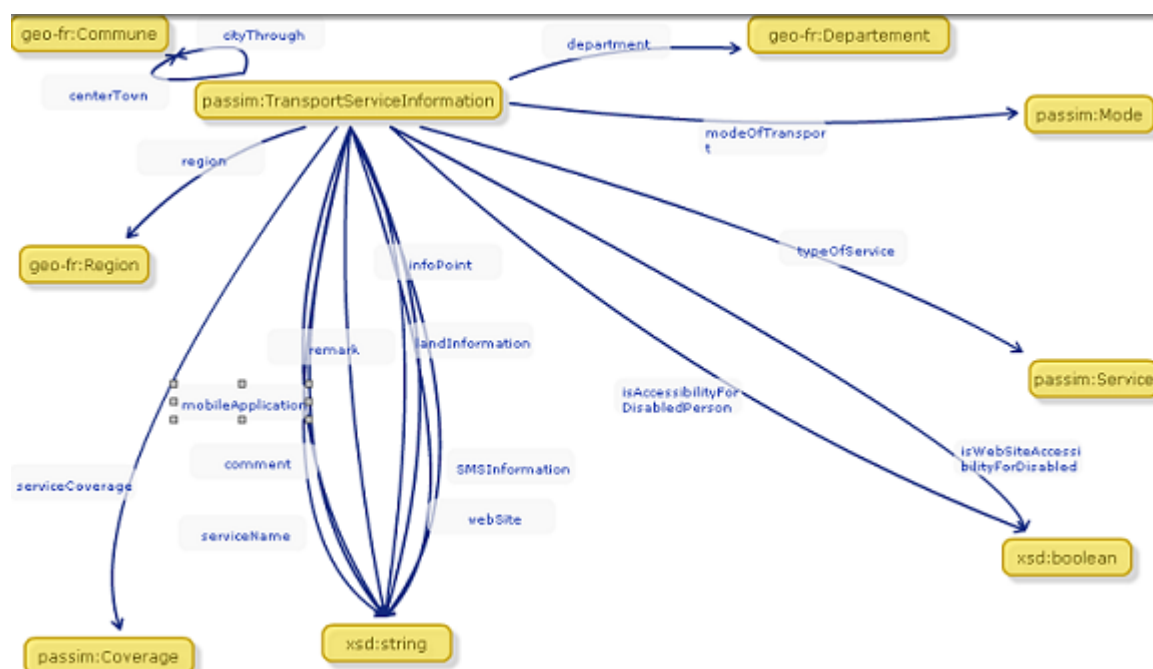[31]http://data.lirmm.fr/ontologies/neptune

Figure 2: diagram of the Passim ontology

3.2.2 NEPTUNE case

Since NEPTUNE was an XML format we use another approach. We read and understood the corresponding standard and created a complex XSL transformation[32]. This sheet can then be run from the DataLift platform[33] to automate the conversion.

3.3 Publication of RDF data

Once in RDF format, we must publish these data (after ensuring that the ontologies which our data refers to are published on the Web). For this, we can:

- either simply publish the RDF file on the Web, i.e. make the file accessible from a browser by a URL;

- or publish them via a SPARQL endpoint. A SPARQL endpoint allows agents (human or machines) to query a knowledge base via the SPARQL query language.

The publication also includes the production of metadata allowing to list these data in various Semantic Web search engines, as Sindice [24]. For this there are certain prerogatives to be respected:

- using a vocabulary for describing data called VoID [1];

- creating a semantic sitemap [12], similar to a classic website sitemap, but adapted to the needs of Semantic Web search engines;

- creating a package on Data Hub. The Data Hub (formerly CKAN[34]) for referes public datasets on the Web;

- manually referencing our dataset on Sindice or other search engines;

- provide URI dereferencing. A description of each resource will be provided if asked over the HTTP protocol;

- having a server to operate content negotiation (sending the corresponding page according to the agent of the request, HTML version if it is a human, or RDF version if it is a machine).

3.3.1 Passim and NEPTUNE case

Passim data when converted to RDF were published on a SPARQL endpoint[35], thus allowing to perform queries on this dataset such as the list of cities served by the company line TaM[36]:

```
SELECT DISTINCT ?city WHERE {
        ?s passim:serviceName ?o .
        ?s passim:cityThrough ?city .
        FILTER (?o = "TaM")
}
```

This dataset was also released on CKAN[37].

For NEPTUNE, only data from the Bordeaux city are available[38]. Also allowing to perform queries on the city of Bordeaux buses such as the names of all the bus stops in the city:

[32] http://www.assembla.com/code/neptunetordf/subversion/nodes/trunk/neptune.xsl?rev=4
[33] http://data.lirmm.fr/datalift
[34] Comprehensive Knowledge Archive Network (http://ckan.org/)
[35] http://data.lirmm.fr/openrdf-workbench/repositories/Passim/query
[36] Transports de l'Agglomération de Montpellier
[37] http://thedatahub.org/dataset/passim
[38] http://data.lirmm.fr/openrdf-workbench/repositories/Neptune/query

```
SELECT DISTINCT ?name WHERE {
        ?stop a neptune:StopPoint .
            ?stop neptune:name ?name .
}
```

3.4 Interlinking

In order to be connected to other Web datasets, the resources in the dataset need to be linked to equivalent resources in other datasets. Interlinking is used to retrieve all data from datasets which are connected to our.

What does it mean? In the Semantic Web (and also in the Web in general), a big problem is linking between them all the data talking about the same thing and linking these data correctly and surely is one of the big problems of the Semantic Web. More there are linked data on the Web of data and more it is important to link them. Then that will be easier to find a lot of information of what we seek.

3.4.1 Passim and NEPTUNE case

For Passim it is possible to link cities names, departments and regions with DBPedia [5] or the INSEE[39] or even Geonames[40]. DBPedia is the Semantic Web version of Wikipedia, more specifically this knowledge base contains all the information which are in the infoboxes in Wikipedia. Geonames is a large knowledge base containing information on cities, departments and French regions, but also other cities and countries worldwide.

4. APPLICATION

With all these data we can imagine a useful application such as for example an application which, when you take a train or plane ticket for a city, lists all the activities that one can have in this city (visits, concerts (with information about singers or groups), shows, restaurants, ...), how to get there by public transports, the opportunity to share them on Twitter or Facebook, see their location with Google Earth, add the dates in your Google Calendar to remember what you have chosen to do and even know the current weather. Figure 3 details datasets used in this application.

The design and operation of the application will not be discussed; emphasis is mainly put on the use of these data and the connections between the RDF datasets. Here is how the datasets used are interconnected between themselves (Figure 4).

As already explained NEPTUNE is not connected with other datasets. The big advantage is that when querying, for example, Seevl, the Seevl resources can be simultaneously retrieved from DBPedia and Last.fm datasets. This operation is more difficult to realize with datasets that are not in RDF. These datasets each have their usefulness:

- Seevl: information about musicians, bands, songs, artists proposal, etc.;

- Last.fm: musical events, concerts, etc.;

- DBPedia: large variety of domains;

- Linked GeoData: RDF content of OpenStreetMap;

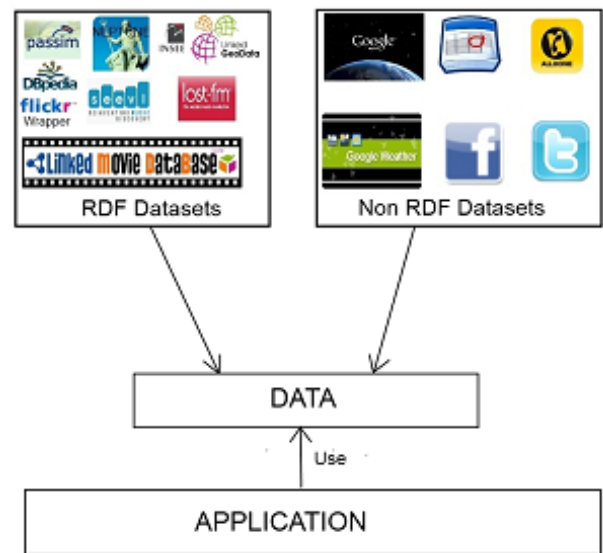[39]http://rdf.insee.fr/geo/
[40]http://www.geonames.org/

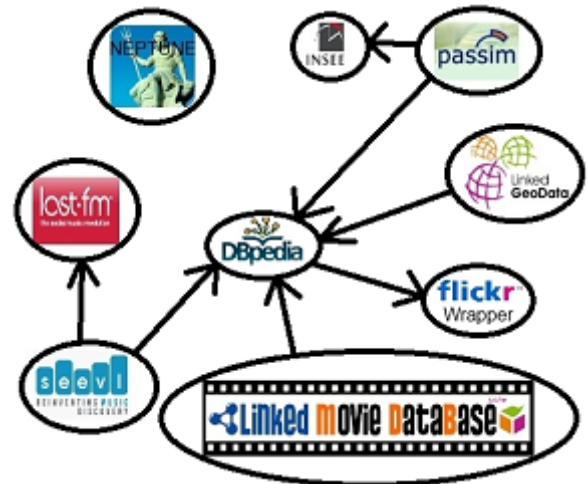
Figure 3: datasets of this application

Figure 4: Interconnection scheme between RDF data of the application

- Flickr Wrapper: RDF wrapper of the Flickr API;

- Linked Movie Database: information about actors, films, directors, etc.;

- Google Earth: visualization of 3D buildings, satellite images, cities, streets, etc.;

- Google Weather: know the weather at a specific location;

- Google Calendar: calendar to add or view our appointments;

- AlloCine: addresses of French cinema, timetable of films that are played.

5. PROBLEMS

This process is easy to implement, useful and we can make beautiful applications. But there are some problems for publishing correctly its data on the Web of data. One of them is, reusing the existing linked data. Reusing linked data involves difficulties that we must keep in mind. For beginning, as in the real world, human user has a degree of interpretation in his usage of the semantic. It is the problem of polysemy: a same data can be describes by many ontologies (for example, according to the context, a car is in the same time a mean of transport, a heritage and an industrial product).

The second one is, the degree of trust that we can have in the data from other datasets. As in the real world too (and also as in the Web of documents), all the published data doesnâĂŹt have the same value (as all the documents doesnâĂŹt have the same value). Measurement data from a laboratory or statistics data from administrations should be used with a good level of trust. However, even these data cannot obviously be considered as an absolute true:

- data depends from context in which they have been produced;

- the definition of any particular size (in an ontology) is not necessarily identical from a dataset to the other;

- there is off course some errors in the data (intentional or not).

Typically, the content of Wikipedia is today one of the main source of content for the Web of data (for the simple reason it is one of the main data source semi-structured and in general free of charge). Yet we know that there are errors in Wikipedia.

Then, there are bound to be contradictions in the Web of data that the tools and applications cannot process (in fact, it is not the role of the final application to check data validity). In the same manner there are also Web pages which have a contradictory content, there are RDF files which express some contradictory facts, or ambiguous (for example, many countries have cities call Montpellier, among other there is France and Quebec).

6. CONCLUSION

Public transport data can be used in countless professional and consumer applications as the data is published and reused, or as web services are available online. For example, a real-time map of all vehicles (trains, subways, buses, trams, taxis) of a city gives the user an overall view and possible delays: this application already exist (Swiss trains[41]), without needing Linked Data.

With this workflow we can develop applications using multiple datasets simultaneously. For example, it will be possible to use generic tools and well-formulated queries, to display restaurants and other activities available around each transit stop (with, if it's a cinema, movie times, if it is a restaurant, the menus and specialties, etc.), or to find tourist transportation routes based on destination and interests of the user.

---

[41] http://swisstrains.ch

We can also make small fun requests, for example, to know how many stops in France are called "Victor Hugo" or how many stops are called after a French writer, and among these stops, which ones are located in a street that bears the same name.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] Keith Alexander, Richard Cyganiak, Michael Hausenblas, and Jun Zhao. Describing linked datasets. In *LDOW*, 2009.

[2] Carlo Allocca, Mathieu D'Aquin, and Enrico Motta. Finding equivalent ontologies in watson. In *ISWC*, 2008.

[3] Cosmin Basca, Stéphane Corlosquet, Richard Cyganiak, Sergio Fernàndez, and Thomas Schandl. Neologism: easy vocabulary publishing. In *SFSW*, 2008.

[4] Tim Berners-Lee. *Weaving the Web: The Past, Present and Future of the World Wide Web by its Inventor*. Harper San Francisco, 2000.

[5] Christian Bizer, Jens Lehmann, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak, and Sebastian Hellmann. Dbpedia - a crystallization point for the web of data. *Web Semantics: Science, Services and Agents on the World Wide Webl*, 7:154–165, September 2009.

[6] Christian Bizer and Andreas Schultz. Benchmarking the performance of storage systems that expose sparql endpoints. In *SSWS*, 2008.

[7] Christian Bizer, Julius Volz, Georgi Kobilarov, and Martin Gaedke. Silk - a link discovery framework for the web of data. In *LDOW*, 2009.

[8] Dan Brickley and R.V Guha. *RDF Schema*. W3C Recommendation, 2004.

[9] Jeen Broekstra, Arjohn Kampman, and Frank van Harmelen. Sesame: a generic architecture for storing and querying rdf and rdf schema. In *ISWC*, 2002.

[10] Gong Cheng, Weiyi Ge, and Yuzhong Qu. Falcons: searching and browsing entities on the semantic web. In *WWW*, 2008.

[11] Richard Cyganiak and Christian Bizer. D2r server - publishing relational databases on the semantic web. In *ISWC*, 2006.

[12] Richard Cyganiak, Holger Stenzhorn, Renaud Delbru, Stefan Decker, and Giovanni Tummarello. Semantic sitemaps: Efficient and flexible access to datasets on the semantic web. In *ESWC*, 2008.

[13] Jérôme Euzenat, Nathalie Abadie, Bénédicte Bucher, Zhengjie Fan, Houda Khrouf, Michael Luger, François Scharffe, and Raphaël Troncy. Dataset interlinking module. Technical report, Datalift, 2011.

[14] Patrick Gendre and Didier Goudergues. Annuaire passim. Vérification PASSIM Région PACA 4ème trimestre, 2007.

[15] Thomas R. Gruber. Towards principles for the design of ontologies used for knowledge sharing. In *Formal*

ontology in conceptual analysis and knowledge representation. Kluwer Academic Publishers, 1993.

[16] Harry Halpin and Ian Davis. *GRDDL Primer*. W3C Recommendation, 2007.

[17] Yannick Maignien. Data.gouv.fr : de l'ouverture des données à l'ouverture des possibles. *International Web Journal*, 2012.

[18] Frank Manola and Eric Miller. *RDF Primer*. W3C Recommendation, 2004.

[19] Deborah McGuinness and Frank van Harmelen. *OWL Web Ontology Language*. W3C Recommendation, 2004.

[20] Axel-Cyrille Ngonga Ngomo and Sören Auer. Limes - a time-efficient approach for large-scale link discovery on the web of data. In *WWW*, 2011.

[21] Eric Prud'hommeaux and Andy Seaborne. *SPARQL Query Language for RDF*. W3C Recommendation, 2008.

[22] François Scharffe and Jérôme Euzenat. Melinda: an interlinking framework for the web of data. Technical report, Datalift, 2011.

[23] Andreas Schultz, Andrea Matteini, Robert Isele, Christian Bizer, and Christian Becker. Ldif - linked data integration framework. In *ISWC*, 2011.

[24] Giovanni Tummarello, Renaud Delbru, and Eyal Oren. Sindice.com: Weaving the open linked data. In *ISWC*, 2007.

[25] Bernard Vatant, Lise Rozat, Pierre-Yves Vandenbussche, Bénédicte Bucher, Nathalie Abadie, and Fabien Gandon. Méthodes et indicateurs pour la sélection d'ontologies fiables et utilisables. Technical report, Datalift, 2011.