Share-PSI 2.0 Workshop

Uses of open data within government for innovation and efficiency

Call For Participation
Meeting Samos - July2014

Proposal of

OpenDataFrance
French regional authorities Association for Opendata movement

Subject:
Open Traffic Information Standard & Experimentation for Enhanced Services

In this contribution, we :
- develop the feedback (REX) of French regional authorities about Opendata (1),
- present organization that was established en France to cope with Opendata weakness (2),
- analyse the importance of data's standardization on national or global scale and the particular case of «traffic information» (3),
- finally propose actions for regional authority in France (4)

1 - Feedback in France

The year 2010 was the starting point for open access of public data in France, beginning in the city of Rennes, joined from that time by the Government which opened the inter-ministerial portal data.gouv.fr in 2011 and forty regional authorities: large cities (8/10), middle-sized town (10), small municipalities (10), general councils (10/98) and regional councils (4/25).
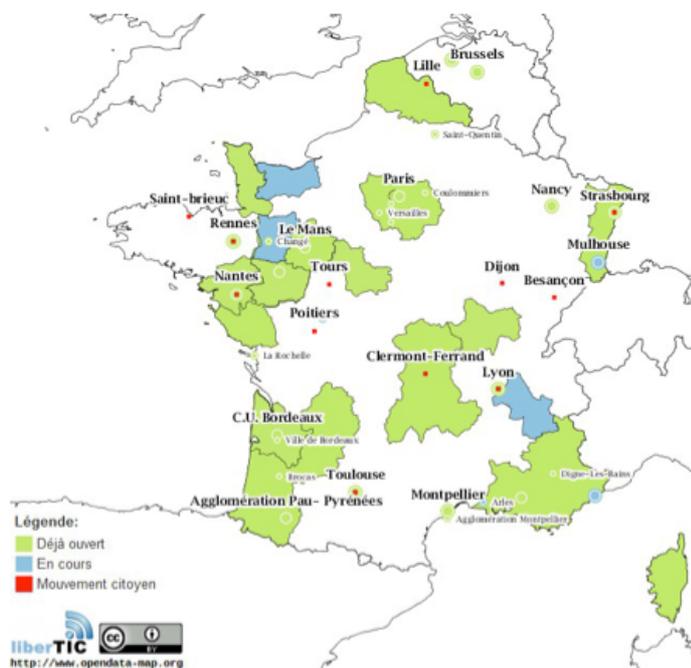
Private stakeholders, business or associative movements also publish data under the regime of opendata, as for instance the association OpenStreetMap France.

However, apart from a few subtle openings such as the SNCF (Train/Ile-de-France), JCDecaux (Shared bike Service in Paris and Lyon), and Keolis (Rennes's public transportation services) it is rare to find industrial players that sincerely and significantly took part in this process. These data are usually linked to delegated services and produced in the framework public services. The weaknesses of contract did not always enable the regional authorities to get back data produced in this context.

After a two years period of development of opendata in France, we are able to make a few inferences:

Positive aspects:

- generalization of the process in many regional authorities ;
- maturity of basic elements : the process, licenses, the description format, convergence of platform (design and offer) ;
- leverage effect on the digital ecosystem : innovations, opening and sharing culture :
  - o good communication toward the public (non-specialized and specialized media) ;
  - o several calls for projects and reusability support systems : a great deal of reusing applications developed in each sector ;
  - o opendata are reused internally (de-compartmentalization of services), between regional authorities, by the general public (several thousands of visits per month) ;
- good national activity, cooperation between institutions (regional authorities, government, representatives or legislatives agencies), visibility of the France and its opendata governance on a global scale.

Negative aspects:
- unsustainable and insufficient quality services
  - o economic system weakened by the local dimension of the support ;
  - o **non-portability of the applications due to the non standardization of data** ;
  - o large amount of usage without direct commercial outlet :
    - ▪ social and solidarity economy,
    - ▪ mobile application and competitive market,
    - ▪ more oriented on social value than economical value
  - o important lack of 'general' data
    - ▪ other non "opendata" regional authorities
    - ▪ other para-public agencies and services (Health, Education, Transport, etc.
    - ▪ insufficient private operators data
      - • in Public Service's Delegation (contractual matter)
      - • other market sector (culture and competition)
- not enough involvement of the non-specialized public (geek or militant)
  - o lack of relevant data;
  - o scattering of sources (portal), format and animation;
  - o lack of expertise in term of data handling (tools and process);

In short, a resulting disappointment due to high expectations ("Eldorado of data").

2 - Reaction of regional authorities to this partial success

Every authority came to the same conclusion concerning the difficulties experienced and the challenges they had to meet to fulfill expectations.

Therefore, in mid-2013 the former decided to create OpenDataFrance (www.opendatafrance.net). An association gathering authorities that are interested in the open access of public data, in order to federate and take collective actions for:
- data's standardization
- writing of law article
- implementation of saver-tools, sharing, animations and communication;
- representation of this movement during public meeting or conferences;
- …

A closer tie with the governmental team occurs in early 2013 and enabled the coordination of the action on the local and national scale.

3 - Data's standardization: Priority of transport data

Data set's standardization is the main action of OpenDataFrance of the year 2014.
OpenDataFrance (ODF) conducted an audit in order to know which data were the most used, the most asked, the services that consume the more data, the services that are the most required by the population. Transport data and mobility services are number one in every regional authorities.

Consequently, ODF took interest to the state-of-the-art, the best process, the experts and the resources available on that topic.

3.1 State-of-the-art

(source CEREMA France, ex CETE/CERTU)

Data concerning the mobility are very diversified: different types of mobility (car, public transportation, bike, walking..), different kind of data (descriptive, historical, real time predictive...), different data producer with diverging strategies and interests (authorities, private players, users...).
Open Data promoted, via the opening of public mobility data, the emergence of services that public players would not have been able to set up on their own, because of the lack of time, skill, or simply because mobility is a domain in which reactivity and data's update prevail if one want to offer relevant services. Reactivity can be optimized thanks to the feedback given by the travelers themselves (crowdsourcing), enabling an update of databases in live.
As far as mobility is concerned, even more so than in other domain, we must share with citizens and learn how to handle "traveler's resources" in order to improve services related to it.

3.2 Heterogeneity of transport data format.

The first difficulty appears when it comes to the choice of shared data's format. At the beginning there was no real step backwards concerning this issue, nowadays we have no choice but to note that there exists a huge diversity of data's format. Consequently, it involves several other difficulties in advance: it can be relatively difficult to use these data, their weight, the compatibility, the matching of different set of data, etc.
For instance, according to urban public transport only, we can have to deal with simple format (such as CVS, XLS, TAB), but also to format used in geographical system of information (SHP, KML / KMZ, Mapinfo which requires a sufficient knowledge about GIS), emergent and incomplete format (GTFS) or even more complex format that are difficult to handle for users, even non-neophytes (TRIDENT, NEPTUNE).

If the choice in term of data is a political one, this reflects necessarily a will to limit the use of data to protect leader operators.

It also brings the issue of the standardization weakness of services and applications. Furthermore if the choice of the format is made by default with a bad knowledge of the range of format and/ or

potential use of datas, feedback from user community can lead regional authorities to look back on given format. Therefore, several authorities which have not set TRIDENT as the only format, widen their availability to diverse format and simple one (for example, GTFS in Rennes: http://data.keolis-rennes.com/fr/les-donnees/donnees-telechargeables.html)

The updating shared data is another important issue. As this update can be easy in the heart information service's administration (often in charge of OpenData process), it can become more complex with other services and partners, because it requires the set up of specific process and a continuous dialog to find information.

The release of a data and their updates rise another problem which is the storage capacity and information hosting. This issue occurs both for real time data and their programming interface (API i.e. Application Programming Interface) and for the recording/archiving data.

3.3 Analysis of the different types of transport data

In the following, we propose to point the different set of mobility data encountered, by distinguishing:
- urban and non-urban public transport data (standard data and real-time data);
- bike and cycling path data (range on real-time data);
- vehicle event data and public road data;
- parking data;
- data related to other ways of transportation.

Set of standard public transport data

Main data concerning public transportation shared by local authority and/or AOT are about the geographical description and qualification of transportation network services.

For collective transportation networks, stops, lines with their open data platforms attributes, and schedules are mostly found. Only a handful of collectivities keep their update history available online. Number of passengers, prices, etc. can also be found online.

The main data formats used are the most commonly used because of their simplicity:
- Shapefile (SHP): standard format on Geographic Information System software;
- Comma-Separated Values (CSV) or XLS: Excel type spreadsheet forma;
- General Transit Feed Specifications (GTFS): Standard format to communicate public transportation schedule and the topography of the network (stops and lines spotted on a map);
- Keyhole Markup Language (KML/KMZ): Language based on XML formalism and destined to organize the display geospatial data on GIS (and Google Earth).

Other formats can be found even though they are less used: XML, Trident, JSON, DXF. As a rule, these last are much less user-friendly.
Real time data case: Real time shared public transportation are really rare, but when they are, it is through API.

Bike and cycling path

Most of the data related to bike that are posted online by regional authorities, deals with cycling path (geographical objects) and urban furniture (bike parking). Most of the time conurbation and municipality that have a "bike share service" shared data and statistics. They also offer real-time data if the service provider in charge of it suggest the idea (the community hosts or references those data).

Other less commonly used data appears such as the pricing or offer of service.

The main data formats are:

- Keyhole Markup Language (KML / KMZ);
- Shapefile (SHP);
- Comma-separated values (CSV) and XLS Format;
- JavaScript Object Notation (JSON): generic textual data format, that derives from the notation of ECMAScript language. It enables a structured representation of the information as XML.
- Application Programming Interface (API), a programming interface that treats real-time data.

Parking data

Several data related to car parking (or other means of transportation such as camper, taxis, buses..) are shared. Most of the shared information are related to public parking, public road, parking for person with disabilities and reduced mobility, Park & Ride. We recently started to collect data about newly developed services such as carpooling or car sharing. Finally real-time data are not an obligation, even though many authorities started provide the former on their Open Data platform.

Data related to parking places for camper, taxis and buses or delivery area, are clearly less abundant than those of car park services.

The main data formats are:

- Keyhole Markup Language (KML/KMZ);
- Shapefile (SHP);
- Comma-separated values (CSV) and XLS Format;
- JavaScript Object Notation (JSON).

Case of public road data

Proficiency in term of public road is broken down among government and regional authorities.

On national scale, the transport minister provides media and operator with traffic information so that they can develop value-added services. It concerns traffic and road event data (vehicle accident, traffic jam...). These data are essentially related to national road network, only the non-franchised part, mostly for huge conurbation structuring road and for some particular case of non-concessioned motorway. Government also provides data of certain department. (This website www.diffusion-numerique.info-routiere.gouv.fr describes the data provided and explain how to interpret them).

Data collected from the national road network are not directly given to general public.

Collectivities mostly release data concerning the road network's description (street description and nomenclature).

Data on accidents (number, severity and location of the accident) and the average annual traffic are broadcast by several communities. The real-time data are much less frequent. Data communities are generally made available in digital formats widely used in the Open Data community. The main data formats used are:

- Keyhole Markup Language (KML/KMZ);
- Shapefile (SHP);
- Comma-separated values (CSV) and XLS Format;

In contrast, traffic data released by the authorities are not formatted according to international standards regulations in the field of traffic information. Data from different communities are heterogeneous, not in terms of standard computers, which are widely shared, but in terms of formats datasets and fields that compose them. This lack of harmonization between different platforms Open Data is likely a strong technical obstacle for developers who wish to create applications nationwide.

The data relating to other means of travel

Some platforms Open Data broadcast some additional data sets for other means of transport: delivery areas, hiking trails or information on ferries and river transport ... The broadcast data are static. The main data formats used are:
- Keyhole Markup Language (KML/KMZ);
- Shapefile (SHP);
- Comma-separated values (CSV) and XLS Format;
- JavaScript Object Notation (JSON).

4 - Conclusion and Proposals

The case of traffic information data is relevant of the current lack of harmonization: Data released by the government are presented by "owners" (or private) attributes and are not formatted according to international standards in the field of information road. It is therefore not possible to easily reuse to data released by various suppliers.

Many European workgroups deepen standardization and inter operatibility formats :
CEN Workgroups :
- SG1: Data Communication on Vehicles
- SG3: Traveller Information – Visually Impaired Persons (TI-VIP)
- SG4: Reference Data Model (Transmodel)
- SG5 : Interoperable Fare Management (IFM)
- SG6: Identification of Fixed Objects for Public Transport (IFOPT)
- SG7: Service Interfaces for Real Time Interchange (SIRI)
- SG8: Distributed Journey Planning (DJPS)
- SG9: Network and Timetable Exchange (NeTEx)

Some in the OpenData community propose standards for the harmonization of national standards formats or European.

Without pooling regional OpenData approaches, some believe that governments should take over these standards and support local authorities wishing to start or improve their platform to find the best possible synergy between the various approaches and ensure homogeneity of data released. Interdepartmental structure Etalab (data.gouv.fr) in France will take up this subject in 2014, this is one of the items of the Action Plan France24 for the implementation of the G8 OpenData agreement.

4.1 Technical aspects

The most fruitful path would be to propose as basis the **GTFS "standard"** (General Transit Feed Specification) which is inherently compatible OpenData format structuring for the description of static resources (network diagram) and dynamic (time). Although this format is still relatively incomplete, it offers the advantage of being an international reference, easy to use, with many examples of reuses.
The more it is used and commonly accepted, the more it will enrich and improve to fulfill expectation of operators and developers. In a 2.0 perspective, it is a "viral" format that enables experimentation, feedback and large assimilation by reusers.

Complex and confidential, the "French" size Netpune used by several administrations and French operators is not designed for OpenData approach (proprietary format, etc.) because it is very complex to reuse. There are similarly other specific standard in other European countries: TransXChange (GB), 452 VDV (Germany), Noptis (Scandinavia), Bison (NL).
Nevertheless, it is still possible to use those standards for internal and technical purposes but likely not for General Pubic.

NETEX (Exchange Network), supported by the European Committee for Standardization, could be the convergence between GTFS and other standards providing complementary elements such as, for example, tariff information. If NetEx is an interesting approach which combine advantages of previous formats, right now, it is still a very complex standard to implement and reuse with rare example of implementation, and it is not design to meet worldwide operators or major players applications.
XML based standards have the major drawback to be verbose, specific tools would be designed to make it simpler and easier access these descriptive data models.

4.2 Governance

It is clear that while standards are beginning to appear, the main difficulty is rather due to the political will of public operators and the brakes or obstacles of private actors.
Important work is to be undertaken to educate, motivate, involve and supervise public and private actors in this opendata data transport culture.

We can see many public actors (national or international institutions), semi-public (center of study, research or standardization) or nonprofit associations (think-tank ...) working on this issue with competence and energy.

Regular exchanges with the actors already involved in such a movement should be implemented:
- okfn.org works on the subject and has created a working group on active transportation: http://transport.okfn.org. It would be appropriate to approach them as an "expert" under Share-PSI,
- The European Commission and probably each government (including Etalab France) are committed to support, promote and regulate the opening of transport data OpenData,
- Other stakeholders: public and private Centers of Studies (CEREMA in France, the Netherlands Waag, etc.).
- ePSI Platform has also seized about this for many years and has published such a manifesto for open data transport 2012: http://fr.scribd.com/doc/111890372/Helsinki-Open-Transport-Data-Manifesto

It will be noticed in particular several major proposals in the manifesto ePSI :
- For public actors: Need to publish data in standard and opendata formats, with zero or marginal cost. Leverage upon private operators for these data to be published when they

collect or produce them for public service. Assume experimental phases over the data to promote reuses by developers;

- For private operators: engagement in opendata process assuming there are numerous opportunities for development and innovation. Collaborate with ecosystems to get the maximum benefit from this opening;
- For Regulators (governments and European commission): promote the development of opendata culture. Act as a prescriber, regulator, legislator and controller with respect to any actors under their responsibility. Encourage the emergence of standard, ecosystems and any communication specific data reuse;
- For developers and re-users: to put pressure on public and private actors, be proactive with proposal.

Share-PSI could be a channel for dissemination and awareness of these many works.

It could also report and enhance experiments carried out by its members, particularly inside OpenDataFrance (i.e. Opendata involved French authorities), wishing to progress on data and mobility services and candidate of carrying a large nationwide experimentation.