

État des lieux de l'ouverture des données routières en France, propositions d'amélioration

**Auteurs : Vincent ROBIN (Cerema DTer Méditerranée)
Patrick GENDRE (Cerema DTer Méditerranée)**

0. Résumé (Contexte et objectifs)

1. Introduction

Contexte

Les exploitants et les gestionnaires de réseaux routiers ont en charge la gestion du trafic, la sécurité des déplacements avec notamment le maintien de la viabilité et l'information de l'utilisateur. Cela nécessite une bonne connaissance du trafic. Pour cela, les gestionnaires ont équipé les routes de capteurs depuis des décennies, dont les données sont utilisées tant en temps réel (régulation) qu'en temps différé (statistiques et études). Les données en temps différé sont utilisées par une grande diversité d'acteurs, bien plus large que les seuls gestionnaires des réseaux, pour des besoins variés: statistiques, analyse du trafic (global ou particulier: PL, flux touristiques), sécurité, études de projets neufs ou d'aménagements, évaluations économiques, modèles de trafic, dimensionnement des ouvrages et de l'entretien routier, études environnementales (bruit, pollution), exploitation (plans de gestion de trafic, préparation de chantiers)...

Ces utilisations nécessitent des données diverses (des moyennes annuelles aux mesures individuelles de véhicules, comptages temporaires ou permanents, comptages de véhicules spécifiques: PL, 2 roues...). Même en se limitant aux moyens de recueil « traditionnels » (boucles), il y a sans doute au moins 4000 compteurs permanents en France, sans doute environ 200 gestionnaires de données de trafic, sans compter les mainteneurs et les entreprises fournissant des capteurs ou des prestations de recueil de données.

Pour les utilisateurs, l'obtention des données est parfois compliquée (identification des services gestionnaires, compréhension des données disponibles, récupération des données ponctuellement ou périodiquement, avec signature possible d'une convention, reformatage éventuels...). Par ailleurs, certaines données existent et ne sont pas connues d'utilisateurs potentiels.

Pour les gestionnaires qui ont recueilli et stockent les données, la mise à disposition des données peut être chronophage et parfois peu prioritaire.

À un autre niveau, on assiste depuis moins de 10 ans à un phénomène d'ouverture des données publiques, inscrit de plus en plus solidement dans la loi et les réglementations, qui vise à favoriser la réutilisation et l'innovation dans un esprit de transparence.

Plusieurs jeux de données de trafic routier ont été publiés sous licence open data, quoique de manière moins généralisée que d'autres types de données, notamment les données environnementales qui ont été ouvertes les premières.

L'ouverture des données routières, en particulier les comptages routiers, présente l'intérêt d'éviter au gestionnaire de passer du temps à répondre à des demandes particulières (l'utilisateur récupérant les données qui l'intéressent sur un portail), et surtout, permet en principe à un plus grand nombre d'acteurs d'avoir facilement accès aux données et de les intégrer à leurs analyses, cela d'autant plus que les données seront publiées d'une part selon des modalités similaires sur tout le territoire, d'autre part sous des formats simples d'emploi.

Le Cerema établissement public issu de la fusion des 8 CETE et 3 Services Techniques Centraux du réseau technique de l'équipement, a été créé en 2014. Fin 2014 a été mis en place un financement interne de Projets Transversaux de Veille et d'Innovation dans le cadre duquel nous avons proposé de mener une action concernant l'ouverture des données de trafic routier; notre proposition a été retenue

et intitulée « projet EDREDON » (Entrepôt de données routières et déplacements engrangées pour leur diffusion ouverte nationale).

Le projet se concentre sur les données de trafic en temps différé issues de stations de mesure Siredo .

Cet article présente les résultats obtenus en 2015 et les perspectives pour 2016.

Objectifs du projet

Les objectifs du projet sont de

- capitaliser les connaissances
- proposer des solutions pour **faciliter la publication et la réutilisation des données de trafic routier**, (en termes d'outils: catalogues et entrepôts de données mutualisées).

Nous faisons le choix de nous **concentrer sur la mise à disposition des données**, qui nous semble au cœur de la chaîne et peut être améliorée, car elle n'a pas été à notre connaissance traitée explicitement jusqu'à présent. L'open data peut simplifier énormément la publication, et inciter à une homogénéisation des conditions de mise à disposition (réduisant ainsi fortement les coûts de publication et de réutilisation des données, et donc inversement en augmentant largement la valeur d'usage).

A contrario, le projet ne porte pas:

- sur la question de l'ouverture des données, qui est discutée en amont par chaque producteur de données publiques;
- sur le recueil de données (situé en amont de la mise à disposition des données);
- sur les analyses et études réalisées à partir de ces données (situées en aval).

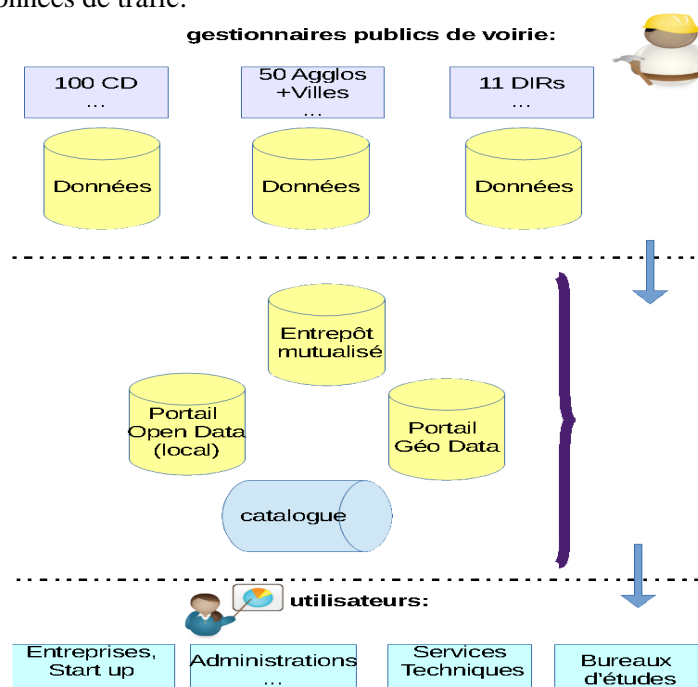
Edredon porte sur des données déjà publiées sous licence open data et sur la fonction entrepôt, prioritairement sur les données publiques de trafic routier en temps différé (le périmètre pourrait être élargi ultérieurement à d'autres données de mobilité et de déplacement).

Le projet fédère des équipes des Directions Techniques TV (ex-CERTU) et ITM (ex-SETRA) et des directions Territoriales (ex-CETE) du Cerema principalement celles de Bordeaux, Aix-en-Provence, Nantes, mais les autres sites sont également impliqués : Île-de-France, Lyon, Rouen, Metz, Lille.

Ce projet vise à développer et capitaliser nos compétences internes, créer et animer une communauté technique dans ce domaine, et à tester et proposer des outils et des services, notamment aux gestionnaires de voirie qui produisent et publient, ou utilisent ces données.

Le site du projet est <http://www.edredon.projet.cerema.fr>

La figure ci-dessous situe notre projet, se concentrant sur la zone intermédiaire entre Producteurs et Utilisateurs des données de trafic:



2. État des lieux

(État de l'art, descriptif de la situation, données existantes)

Ouverture des données de transport et routières

L'ouverture des données publiques est concernée par plusieurs lois ou réglementations récentes ou en cours d'élaboration:

- la loi CADA (1978)
- la directive PSI (2003)
- la directive Inspire (2007)
- indirectement la directive ITS (2010)
- la loi Notre (2015)
- la loi Macron (2015)
- indirectement la loi TECV (2015)
- la loi Lemaire (2016)

L'ouverture des données de transport a été l'objet d'un débat national en France en 2014, qui s'est traduit par un rapport (« Ouverture des données de transport », Francis Jutant, mars 2015) remis au ministre puis quelques semaines plus tard, par le vote de dispositions dans le cadre de la loi Macron. Les données routières ne font pas explicitement partie de ces dispositions, mais au moins pour les routes gérées par des services publics, les données sont de fait des données publiques réutilisables de droit, et de nombreuses collectivités publient des données routières, ainsi que plusieurs pays européens parmi lesquels les Pays-Bas et le Royaume-Uni semblent les plus avancés.

En France, le Cerema a publié une étude sur l'open data dans les transports en 2015 (« [L'Open Data en collectivité à la lumière des données de mobilité](http://www.certu.fr/open-data-les-travaux-du-cerema-a1219.html) », Laurent Chevereau, mars 2015 <http://www.certu.fr/open-data-les-travaux-du-cerema-a1219.html>) et recense dans l'annuaire Passim¹ les données open data transport.

En ce qui concerne les données routières, le projet Edredon a permis de refaire un état des lieux des collectivités publiant des données de trafic sous licence open data: on a recensé Toulouse, Nantes, Bordeaux (CUB), Lyon, Montpellier, Strasbourg (CUS), CD22, CD27, CD44, CD71, CD76, CD92.

D'autres collectivités comme Angers, Pau, PACA, Paris, IDF, NordPDC, Auvergne, La Rochelle, Rennes, Nice, ou Nancy ont une plate-forme open data, publient des données de mobilité/TC ou de voirie mais pas (encore) de données de trafic. Il faudra néanmoins continuer la veille et contacter tous ces acteurs.

Pour le Réseau Routier National, la DIRIF publie également des données de comptage depuis 2008.

On constate également que l'ouverture des données n'est pas une priorité pour les gestionnaires: l'open data est souvent une demande plus générale qui émane de la collectivité. En revanche, il existe une forte demande de partager et mutualiser les données de trafic entre gestionnaires, tant pour leurs besoins propres que pour des besoins transverses : modélisation du trafic, du bruit, de la pollution à des niveaux territoriaux plus larges.

Il faut considérer que les portails de diffusion de données open data ne sont pas forcément confondus avec les entrepôts de données de trafic régionaux et les outils mutualisés entre gestionnaires routiers.

Les données minimales publiées peuvent être regroupées en 3 grandes classes :

- Le réseau routier : avant de fournir des données de comptage, il faut fournir une description des tronçons routiers sur lesquels s'appliquent les données:
- Les comptages : les données de trafic historiques sont le plus souvent des moyennes annuelles TMJA, notamment pour les données des Départements:

¹ <http://petitpois.passim.info/poi/search?ack=&k=&q=open+data&w=&s=active>

- Les données publiées en temps réel : états de trafic (Traficolor), temps de parcours, données de trafic (débits, vitesses, taux d'occupation), événements (le plus complexe, la norme étant DatexII).

Première Analyse

On constate une assez forte hétérogénéité des données (nature des données), ainsi que de la manière dont les données sont décrites (catalogage) et mises à disposition (formats).

3. Résultats

Les résultats ont été publiés sur le site <http://www.edredon.cerema.fr> au cours de l'année 2015.

A priori, la méthode de travail envisagée était de recenser systématiquement les données routières publiées, d'interroger les collectivités et gestionnaires qui publient ces données, de diagnostiquer les points d'amélioration, notamment sur le catalogage, les formats de publication, l'historisation des données, puis de s'appuyer sur les connaissances et outils du Cerema et sur une veille technique pour tester et proposer des améliorations.

Dans le cadre de ce projet interne plutôt exploratoire, il est très difficile pour les équipes de dégager du temps pour s'investir et il n'est pas possible de mener à bien un travail systématique, comme le contact avec les gestionnaires de voirie et les réutilisateurs, ou l'analyse un peu prospective sur les outils (big data, cloud...).

Notre approche a donc été en pratique plus « bottom-up »: nous avons avancé là où nous avons pu, compte tenu des actions en cours, des opportunités et des disponibilités de chacun.

Nous avons principalement avancé sur les 3 axes suivants :

- catalogage de données : mise en place d'un outil CKAN et propositions pour les mots-clés
- entrepôt de données : mise en place de l'application IRIS
- formats : étude des modalités possibles de publication de données du réseau routier national (RRN)

Catalogage

Il n'existe pas actuellement de manière simple de trouver les données de trafic (open data) en France. Dans un premier temps, nous avons donc choisi de créer un répertoire recensant ces données, en s'appuyant sur CKAN, un logiciel libre très largement utilisé pour réaliser des catalogues open data. Le catalogage, manuel, est accessible sur le site <http://www.edredon.cerema.fr/jeux-de-donnees-r25.html>.

Néanmoins, ce fonctionnement n'est pas satisfaisant à moyen terme et devrait être plus automatisé. Les données open data sont publiées sur deux types de plates-formes :

- les plates-formes géographiques régionales, où il est parfois difficile de s'y retrouver car toutes les données n'y sont pas téléchargeables (mais parfois consultables); en revanche ces plates-formes s'appuient sur des standards issus du monde géomatique (notamment liés à la directive Inspire), ce qui facilite grandement leur moissonnage (c'est-à-dire leur interrogation par des robots) et a permis à la mission Etalab qui gère le site data.gouv.fr de développer un outil pour interroger toutes les plates-formes existantes : <https://inspire.data.gouv.fr/>
- les plates-formes open data proprement dites, qui sont beaucoup moins standardisées

Après quelques échanges avec la mission Etalab, nous envisageons à terme le fonctionnement suivant:

- décrire les données directement dans le catalogue de <https://www.data.gouv.fr>, (quitte à maintenir le cas échéant notre catalogue spécialisé sur les données de trafic, moissonner depuis data.gouv.fr et le compléter manuellement là où ce serait nécessaire, par exemple sur les données disponibles sur demande voire payantes et pas sous licence open data)

- inciter les plates-formes à étiqueter avec les mêmes mots-clés (tags) des jeux de données similaires, avec le concours de la mission Etalab, ce qui facilitera grandement la recherche de données; cela reste possible dans la mesure où il n'y a que quelques dizaines de plates-formes qu'il est possible de contacter individuellement. Pour commencer, on propose « points de mesure de trafic routier », « tronçons routiers », « comptages routiers ».

Entrepôt de données mutualisées : l'application IRIS

IRIS est une application développée par le CEREMA DterOuest à Nantes pour mutualiser les données de trafic routier en Bretagne et Pays-de-Loire, qui donne accès à des outils de requête et d'édition de statistiques aux partenaires via un site extranet. Les développements réalisés en 2015 permettent désormais de rendre les données accessibles sur internet, pour les partenaires qui ont décidé l'ouverture de leurs données.

Un des principaux résultats de l'année est que la Direction Territoriale Sud-Ouest, qui cherchait un outil de mutualisation des données de trafic sur sa zone d'action, a décidé d'utiliser IRIS, et va progressivement archiver dans IRIS les données de ses partenaires (DIR et CD); la Charente est le premier département.

IRIS est envisagé comme outil de mutualisation partenariale des données de trafic du Cerema, et de publication open data pour les partenaires qui le souhaitent, sur la base d'une convention entre le Cerema et chaque collectivité. Le Conseil Départemental de Loire-Atlantique est le 1er signataire d'une telle convention; des discussions sont ouvertes avec des collectivités urbaines, comme Rennes Métropole.

L'application est accessible à l'adresse suivante <https://dreal.applis-bretagne.fr/iris/>

Formats des données

Il est bien sûr souhaitable que les données open data soient diffusées de manière homogène sur tout le territoire, cela facilitera à la fois leur publication et leur réutilisation. Deux formats principaux coexistent, le Datex II pour l'échange de données et les fichiers sous forme de tableur issus des logiciels de traitement.

Datex II

Datex II est la norme européenne d'échanges de données de trafic. C'est un format XML (spécifié dans un schéma) pour échanger des données entre centres d'information routière.

Datex II est depuis un règlement² publié en décembre 2014 par la Commission Européenne comme format pour la fourniture de données de trafic aux opérateurs d'information dans le cadre de la directive ITS. Il sera certainement aussi le format de référence pour la publication de données historiques.

Les formats d'échanges sont spécifiés dans les 2 documents suivants :

- [Les échanges de données pour l'exploitation de la route. Présentation de Datex II version 1.0 - Mars 2012](#)
- [Échange de données de trafic en Datex II - Décembre 2013](#)

Néanmoins il s'agit d'un format surtout conçu pour les échanges entre centres de gestion de trafic, et/ou avec les grands opérateurs d'information routière.

Le format Datex II n'est donc pas forcément le seul format à préconiser pour la publication de données dans un contexte open data où les réutilisateurs ne sont pas, pour la grande majorité, des professionnels de l'information routière.

D'ailleurs, à ce jour, aucune donnée de trafic open data en France n'est publiée en Datex II (depuis 2013, date où le libre accès aux informations routières du RRN a été fermé). A noter néanmoins que Datex II est utilisé par exemple aux Pays-Bas pour la publication des données de trafic de l'entrepôt national de données de trafic.

² http://ec.europa.eu/transport/themes/its/news/2014-12-18-rtti_en.htm

Fichiers de mesures

Les gestionnaires de voirie utilisent un format texte pour fournir les données de trafic, issues des logiciels Mélodie et Route +. Ces fichiers contiennent des informations techniques utiles au gestionnaire mais pas forcément à l'utilisateur final et sont formatés sous forme de tableaux .

Format cible

Le format de choix pour le réutilisateur « de base » est a priori un format texte tabulé, de type CSV, avec une ligne par mesure et des colonnes pour identifier le point de mesure, l'horodate, et chaque type de donnée, facile à ouvrir dans un tableur, ou dans un SIG (s'il est doté de colonnes pour les coordonnées géographiques).

À partir des données existantes sur le réseau national, dont le Cerema assure l'archivage depuis de nombreuses années, nous avons produit

- des fichiers géoréférencés en X,Y (type Shapefile) localisant les points de mesure et les tronçons routiers associés;
- des fichiers CSV contenant les données (en commençant par les débits tous véhicules horaires et journaliers)
-

Dès que sera pris la décision d'ouverture des données sur le RRN, il sera possible rapidement de publier les données de trafic en temps différé, sur la base du format proposé tests de production de fichiers CSV à partir des données du RRN.

4. Perspectives

Le travail mené en 2015 a permis de confirmer l'intérêt de l'open data pour faire le lien entre les producteurs et les utilisateurs de données de trafic, valoriser ces données, et d'investir sur le sujet au niveau national, de manière transverse. Le Cerema est compétent et légitime pour animer une telle action.

L'idée serait de pouvoir animer une communauté technique autour de la donnée de trafic, pour les producteurs et les utilisateurs.

En termes de site web, on envisage 4 aspects :

- un wiki expliquant ce que sont les données de trafic, comment elles sont constituées, quels sont les standards, où les trouver, comment les utiliser, etc.
- un catalogue de données recensant également les données non open data (par exemple l'outil IRIS en tant qu'outil partenarial y compris pour des données non encore ouvertes)
- une veille technique (sur les données, logiciels, études, marchés, agenda)
- un outil, par exemple une simple liste de diffusion mail, souvent plus efficace qu'un forum, soit limitée aux services publics, soit étendue aux bureaux d'études, industriels

Néanmoins le programme de travail 2016 n'inclut pas encore une action « données de trafic » ou plus large « données »; les actions sur les données sont en général attachées à un besoin particulier (utilisation pour les modèles, données du réseau routier national). L'année 2016 devra donc être consacrée à continuer les travaux engagés, à moyens constants, et par ailleurs afin de bien définir une action plus large qui serait menée à partir de 2017.

Ainsi, les actions prévues pour 2016 sont de :

- compléter le recensement des données, en coopération avec Etalab
- finaliser les propositions de formats de publication des données de trafic
- élargir la couverture territoriale de l'application IRIS, pérenniser son hébergement, étendre les données aux comptages temporaires et établir des conventions de partenariat avec les collectivités gestionnaires
- faire évoluer le site web initial vers un site de type www.trafic-routier.donnees.cerema.fr

Au-delà des actions en cours, il serait pertinent de voir comment tirer parti des innovations telles que:

- le cloud : les coûts d'hébergement de données et d'applications se sont effondrés.
- le big data : l'accès à des outils de traitement massif de données est beaucoup plus facile qu'il y a quelques années, notamment grâce à la diffusion de logiciels libres, en notant que le terme big data est à relativiser : par exemple, avec 4000 points de mesure en France, une année de mesures débit 6 minute représentent 350 millions de mesures qui tiendraient aisément sur une clé USB
- le collaboratif : un autre aspect intéressant à explorer se situe dans l'amélioration collaborative des données: par exemple, les stations de comptage de trafic pourraient être cartographiées dans OpenStreetMap, à l'instar des feux par exemple.

Enfin, si la communauté « données de trafic » fonctionne et suscite une adhésion, il faudra voir comment l'élargir à d'autres types de données (temps de parcours, enquêtes de trafic, etc.).

5. Conclusions

La mutualisation des données publiques de trafic routier s'inscrit forcément dans une durée assez longue. En pratique, la priorité est sans doute d'arriver à susciter et animer une communauté technique sur le sujet des données routières et de trafic, d'harmoniser la mise à disposition, puis d'avancer pas à pas.

Il est important de préciser que nous n'avons évidemment pas forcément à héberger toutes les données, ni les moyens en interne pour produire l'ensemble des données, informations et outils utiles à la communauté « données de trafic routier », notre objectif est clairement d'arriver à associer les collectivités gestionnaires de voirie bien sûr, mais aussi sociétés de services ou entreprises, qui chacun peuvent contribuer à leur niveau.

Le CEREMA souhaite recueillir l'expression d'intérêt des collectivités gestionnaires de voirie et recherche des partenaires intéressés.

Références.

www.edredon.cerema.fr

ckan.cerema-dtm.fr

lois Macron, TECV, Notre, Lemaire

Directive ITS + règlement délégué, Directive PSI, Directive Inspire...

<http://www.certu-catalogue.fr/l-open-data-en-collectivite-a-la-lumiere-des-donnees-de-mobilite.html>

<http://www.infra-transports-materiaux.cerema.fr/panorama-des-systemes-de-recueil-de-donnees-de-a5424.html>

Orateurs

* **Vincent Robin** est chargé d'études en recueil de données et système d'information trafic au service CTA de la Direction territoriale Méditerranée du Cerema, Spécialiste du domaine «transports durables, sécurité, inter-modalité et mobilité».

* **Patrick Gendre** est chargé de mission Information Multimodale au CEREMA Méditerranée à Aix-en-Provence. Il a occupé différents postes au sein du réseau technique Cerema dans le domaine de l'information transport/trafic et les transports intelligents depuis 20 ans.

<http://mim.cete-aix.fr/>