

Projet Transversal Valorisation Innovation EDREDON

Entrepot de Données Routières Diffusion Ouverte Nationale

Avancement fin 2014

Date

12 Décembre 2014

Auteurs :	
Relecteurs	Patrick GENDRE (DTerMed)

Résumé :

Ce document présente l'état d'avancement du projet Transversal Innovation du CEREMA portant sur les l'ouverture des données de trafic routier et de déplacements, ainsi que sur le test de solutions de recueil de traces de mobilité par mobile.

Le projet a démarré en octobre 2014 et doit se dérouler en 2015 et 2016.

Le périmètre du projet est le suivant :

- le volet **entrepot** porte sur des données déjà publiées sous licence open data (on ne traite pas de la question de l'ouverture des données, qui est discutée en amont par chaque producteur de données publiques), sur la fonction entrepot (et donc ni sur le recueil de données en amont, ni sur les analyses en aval), sur l'ensemble des données publiques de trafic routier (même si le périmètre pourra être élargi ultérieurement à d'autres données de mobilité et de déplacement).

- le volet **traces** porte sur les outils de recueil de traces à partir de données mobiles : test d'une application android développée par Inria Lille, état des lieux sur le sujet, impact de ce type de données sur les méthodes d'enquêtes ménage déplacements DTTV (« EMD CERTU »), et inclura bien sûr une réflexion sur la publication de ce type de données (y compris sous licence open data) et leur réutilisation, avec des recommandations concrètes issues du retour d'expériences des développements effectués à Lille.

EDREDON fédère des équipes des Directions Techniques TV (ex-CERTU) et ITM (ex-SETRA) et des directions Territoriales (ex-CETE) à Bordeaux, Aix-en-Provence, Metz, Nantes, et Lille. A partir de 2015, les autres Directions Territoriales (Ile-de-France, Lyon, Rouen) participeront également.

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.

Mots clés :

données ouvertes, données routières, trafic routier, open data

Diffusion :

[publique \(licence Creative Commons CC-by-nd\)](#)

Nombre de pages :

8 pages

Table des matières

1.Contexte.....	3
2.Premiers résultats 2014.....	3
1.Présentation de l'existant.....	3
2.Recensement des données (plates-formes open data) et autres sources.....	3
3.Analyse des données.....	4
4.Contacts et Analyse des besoins:.....	4
5.Analyse des outils.....	4
6.Ouverture d'un site web.....	4
7.Application de recueil de traces.....	4
3.Perspectives pour 2015.....	5

1. Contexte

Le projet EDREDON fait partie des actions d'innovation retenues en 2014 pour être financées en interne par le Cerema, il porte sur l'innovation dans les données transport. Le projet comprend 2 volets:

- entrepôts de données de trafic open data (d'où son titre qui signifie « Entrepôt de données routières et déplacements engrangées pour leur diffusion ouverte nationale »);

- traces de mobilité : développement d'un outil d'analyse de traces mobiles de déplacements multimodaux, en partenariat avec l'INRIA Lille.

Le périmètre du projet est le suivant :

- le volet **entrepot** porte sur des données déjà publiées sous licence open data (on ne traite pas de la question de l'ouverture des données, qui est discutée en amont par chaque producteur de données publiques), sur la fonction entrepot (et donc ni sur le recueil de données en amont, ni sur les analyses en aval), sur l'ensemble des données publiques de trafic routier (même si le périmètre pourra être élargi ultérieurement à d'autres données de mobilité et de déplacement).

- le volet **traces** porte sur les outils de recueil de traces à partir de données mobiles : test d'une application android développée par Inria Lille, état des lieux sur le sujet, impact de ce type de données sur les méthodes d'enquêtes ménage déplacements DTTV (« EMD CERTU »), et inclura bien sûr une réflexion sur la publication de ce type de données (y compris sous licence open data) et leur réutilisation, avec des recommandations concrètes issues du retour d'expériences des développements effectués à Lille.

EDREDON fédère des équipes des Directions Techniques TV (ex-CERTU) et ITM (ex-SETRA) et des directions Territoriales (ex-CETE) à Bordeaux, Aix-en-Provence, Metz, Nantes, et Lille. A partir de 2015, les autres Directions Territoriales (Ile-de-France, Lyon, Rouen) participeront également.

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.

Dans l'attente de l'ouverture du site du projet, les informations relatives au projet sont publiées sur cette page :<http://mim.cete-aix.fr/spip.php?article373>

puis sur cette page <http://mim.cete-aix.fr/spip.php?article380>

2. Premiers résultats 2014

Le lancement a eu lieu le 6/10/14.

1. Présentation de l'existant

Les participants ont présenté leurs travaux qui peuvent servir de base réutilisable pour le projet : SITL (DterEst), BD Siredo (DterMed), IRIS (DterOuest), ISIDOR / Tipi trafic, premier état des lieux européen (DtectIM), PIGMA et Open data Aquitaine + Toulouse et projet SERRES (DterSO)

2. Recensement des données (plates-formes open data) et autres sources

La DTecTV (Certu) a réalisé une étude sur l'open data en 2014 (<http://www.certu.fr/open-data-les-travaux-du-cerema-a1219.html>) et recense dans l'annuaire les données open data transport:

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

Les collectivités suivantes ont été identifiées comme publiant des données routières sous licence open data sont:

Toulouse, Nantes, Bordeaux (CUB), Lyon, Montpellier, Strasbourg (CUS), CG44, CG71, CG92, CG27, CG76

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 voirie mais pas de données de trafic. Il faudrait néanmoins les contacter.

La DIRIF publie également des données de comptage depuis 2008.

Le type de données publiées peut être regroupé en 3 grandes classes :

- a minima des données historiques, le plus souvent des moyennes annuelles TMJA par des CG: **Comptages**

- ce qui implique par ailleurs de fournir une description des tronçons routiers sur lesquels s'appliquent les données: **Réseau routier**

- pour certaines grandes villes, des données **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).

Il est apparu que les plates-formes régionales d'information géographique (Pigma, Crige...) n'apportent pas une grande aide à très court terme pour recenser les sources de données, qui sont encore peu nombreuses et peuvent être identifiées « à la main ». A moyen terme en revanche, elles seront certainement un relais indispensable. En effet, un mécanisme de catalogage automatique par data.gouv.fr de toutes les données cataloguées dans les plates-formes régionales via le standard Inspire est rendu possible en filtrant uniquement les données open data (exemple: <http://afigeo.datacircl.es/services/53889fc3280d0d8d3c2ede68/datasets?q=trafic&opendata&offset=0> récupère les données de trafic open data publiées en Aquitaine). Néanmoins l'indexation des données de trafic n'étant pas faite de manière identique d'une région à l'autre, le recensement sur data.gouv.fr reste incomplet.

3. Analyse des données

La DTer Est a fait des premiers tests pour récupérer des données temps réel (de Strasbourg) et les importer puis les archiver dans une base de données PostgreSQL/postgis, similaire à ce qui est fait pour le projet SITL.

La prochaine étape serait de republier les données archivées ou certains indicateurs statistiques qui en seraient extraits. La difficulté principale consiste à définir un modèle de données de cible pour la publication des données : que recommande-t-on pour la publication de données de référentiel routier, de données de comptage? Il existe des recommandations nationales pour le ministère mais elles doivent sans doute être adaptées pour simplifier la réutilisation dans un contexte open data où les utilisateurs ne sont pas forcément du métier. Il faut sans doute par ailleurs publier sous plusieurs formats.

La DTer Ouest travaille sur les données Bretagne-Pays de Loire et va prendre contact avec les gestionnaires de trafic de Nantes, Rennes et Brest, et rencontre régulièrement les CG de sa zone.

La DTer Sud-Ouest a commencé à analyser les données de Bordeaux et d'Aquitaine et va reconstruire ses interlocuteurs Bordelais.

La DTerMed a demandé un compte sur la plate-forme smartData du Grand Lyon et contacté Montpellier, ainsi que les conseils généraux de PACA et LR.

4. Contacts et Analyse des besoins:

(collectivités, MEDDE, institutionnels, bureaux d'études, prestataires et éditeurs)

Pour l'instant nous avons commencé à identifier des contacts, rédigé une grille d'entretien et contacté quelques personnes, par mail ou lors de réunions régionales, il y a plusieurs expressions d'intérêt mais encore eu aucun retour concret (sauf l'entretien avec Pigma en Aquitaine). La base de contacts reste à construire.

5. Analyse des outils

La veille a été très limitée à ce jour, et donc à démarrer dès janvier 2015: recensement de solutions, tests techniques, contacts avec des entreprises compétentes.

Pour l'instant seul CKAN a été identifié, pour la fonction de catalogage des données, et pressenti comme pouvant avoir un intérêt pour recenser les données trafic/routières. CKAN est un logiciel libre très largement utilisé pour réaliser des sites open data. Le logiciel a été installé et est accessible pour l'instant sur Intranet.

Plusieurs sources de données open data ont été importées sur notre CKAN, en utilisant la possibilité de CKAN d'importer des descriptions de jeux de données depuis des fichiers Calc ou autres.

<http://ckan.cete-mediterranee.i2/fr/dataset>

Il faudra aussi regarder les outils d'extraction / calcul et de publication des données.

6. Ouverture d'un site web

Les échanges avec la direction informatique et de Communication du Cerema ont permis de rattacher à une solution standard de gestion de contenu web (Spip/Giseh) qu'il est prévu de déployer pour le Cerema à partir de janvier 2015. La structure du site sera en miroir des livrables et des actions envisagées, sans doute avec 5 rubriques :

- présentation du projet
- volet Entrepôt : -> données (une fiche par producteur ou jeu de données, puis un dossier par type de données : réseau routier, comptages, temps réel), besoins des acteurs, solutions
- volet Traces de déplacements

Les rubriques seront alimentées au fil de l'eau, et seront consolidées progressivement par des livrables plus aboutis.

7. Application de recueil de traces

- Contexte et analyse: Analyse de l'existant : expérimentations en cours (Rhone-Alpes, IDF, Mobi-Lise...), Développement d'un argumentaire, Orientations pour le prototype, architecture

La Dter Nord-Picardie a procédé à des tests de l'application développée par INRIA Lille et rendra compte des premiers résultats début 2015.

3. Perspectives pour 2015

Pour le 1er semestre, les actions à mener sont les suivantes:

- capitalisation : continuer à travailler avec les données existantes (sachant que de nouvelles données sont publiées en open data à peu près chaque mois):

- constituer une **base de contacts**

- **interroger tous les types d'acteurs** concernés

- rédiger **une fiche par source** (site/acteur publiant des données): expression et analyse des besoins, utilisation des données, propositions d'améliorations

- faire des **recommandations par type de données : référentiel, comptages, temps réel**

- prototypage : au-delà des premiers tests sur les données existantes, il faut commencer à construire des solutions (prototypes). L'analyse initiale des besoins faite pour le lancement montre que les données publiées sont assez hétérogènes (conditions d'accès, formats), parfois pas assez décrites ou qualifiées (problème de localisants, difficultés de récupération des données temps réel lorsqu'elles sont publiées).

- **veille sur les outils** : catalogage, outils d'extraction / transformation, publication

- ouvrir le catalogue **CKAN Edredon** et publier une fiche descriptive associée à chaque jeu de données

- - ouverture sur le web de **données republiées sur la base de l'outil SITL**

- ...

L'année 2015 doit permettre de démontrer une ou deux applications concrètes. Exemple : observatoire national des données de trafic, archivage de données temps réel...

Au 2ème semestre, on pourra commencer à réfléchir à des scénarios de pérennisation:

- au plus, le Cerema pourrait héberger toutes les données publiques routières de France, mais cela semble bien ambitieux, même si on se limite aux seules données open data;

- on peut imaginer des scénarios intermédiaires où on hébergerait uniquement les données de certains gestionnaires, notamment celles des DIR, ou uniquement les données d'un certain type (référentiels, trafics annuels...);

- même si le Cerema n'héberge pas les données, cette prestation pourrait être sous-traitée mais pilotée par le Cerema;

- par ailleurs, le Cerema peut aussi développer et maintenir des outils (scripts de qualification et transformation des données), qui seraient mis à la disposition de la communauté technique;

- enfin, a minima, le Cerema ambitionne d'animer et pérenniser la communauté technique des gestionnaires de données routières publiques, urbaines et inter-urbaines.

Sur la base de cette réflexion, la 2ème année du projet (2016) doit permettre de consolider les développements et chercher à pérenniser ce qui peut l'être.

Le tableau ci-dessous indique les contributions envisagées, qui restent à préciser très largement.

DterMed:	6 hm	Patrick Gendre, Yves Bonin, Silvio Rousic, Bernard Rongione, Marie-Amélie Horvath, Vincent Robin, Jean-Philippe Tenaud	coordination générale, lot 2 évaluation des développements sur l'open data existants; prototypage contacts avec les plates-formes régionales et open data, IGN, OSM, Etalab
----------	------	---	---

DTerSO:	6 hm	Marie Gadrat, ...	description de l'existant Pigma, Serres, évaluation des développements sur ces plates-formes prototype
DTerOuest:	2 hm	Daniel Grégoire, Guylaine Piron	description de l'existant Bretagne / Pays de Loire prototype
DterEst:	2 hm	Eric Klein, Laurent Taraschini	description de l'existant pour l'Alsace, volet qualification de données, du volet achat / développements informatiques prototype
DTerNP:	4 hm	Thomas Durlin, Mathieu Rabaud, Patrick Palmier	pilote du volet « Traces multimodales » (lot 3), contribution le cas échéant sur entrepôt existant de données à Lille
DTecITM:	4 hm	Romain Sevestre, Etienne Hombourger, Romain Bouilly	responsable lot 1 analyse des besoins, description de l'existant RRN, recommandations techniques
DTecTV:	1 hm	Bernard Allouche Tristan Guilloux, Laurent Jardinier Laurent Chevereau	analyse des besoins, intégration de la multimodalité, volet géomatique, communication open data méthodologie d'enquêtes par mobile
DTecIDF	1 hm	Ludovic Maria	analyse de l'existant et des besoins
DTerNC	1 hm	Eric Violette	analyse de l'existant et des besoins
DTerCE	1 hm	Frédéric Murard	analyse de l'existant et des besoins

ANNEXES.

Grille pour l'analyse des données trafic open data .

Description des données

- décrire l'ensemble des données sur le thème transport pour la collectivité / la plate-forme, même si on ne travaillera ensuite qu'avec les données trafic : licences, conditions d'accès et d'utilisation (besoin de créer un compte, API, etc.)

Contacts avec les acteurs :

- responsable de la publication (web mestrre de la plate-forme publique open data)
- producteur(s) des données
- alimenter un fichier XL de nos contacts (le seul qui ne sera pas diffusable sur internet car nominatif!)

Les interroger sur : (si besoin on pourra écrire un e-mail type de contact)

- comment s'est passée l'ouverture des données : décision, mise en oeuvre (quelle plate-forme, qui l'opère, combien ça coûte), processus (chaîne de traitement depuis le recueil des données jusqu'à la publication)
- quels retours des utilisateurs ?
- quelles leçons, problèmes restent à résoudre
- intérêt pour le projet Edredon, quelles attentes éventuelles, souhait d'être tenu informé
- prochaines étapes dans la publication open data

Tests sur les données :

- décrire de bout en bout comment réutiliser les données publiées sur 1/2/3 exemples : création d'une carte, calcul d'une nouvelle (exemple : calcul de la capacité à partir des Q,V), utilisation de l'API quand l'accès aux données l'implique

Conséquences pour Edredon

- 1ère analyse issue des entretiens et des tests : qu'est-ce qui manque actuellement et qui serait pertinent de développer dans Edredon : outils, données (reformatées etc.), recommandations...

proposition de mail pour les contacts

bonjour,

dans le cadre de sa démarche interne d'innovation, le Cerema lance une action sur l'innovation dans les données sur les déplacements, ciblée notamment sur les entrepôts de données de trafic routier (lien page web): l'objectif est de mieux connaître les attentes des producteurs et réutilisateurs, d'analyser les problèmes techniques qui restent à résoudre et de proposer des solutions pour faciliter la réutilisation de ce type de données sur le territoire.

Nous avons recensé votre portail xxx comme l'un des portails open data existant en France qui publie des données routières et je souhaite en savoir plus sur l'organisation qui permet de publier ces données, les réutilisations qui sont faites des données, les points à améliorer d'après vous, et les prochaines évolutions prévues. Seriez vous disponible pour un entretien au téléphone d'une petite demi-heure?

En vous remerciant d'avance.

Meilleures salutations,

Première liste de jeux de données référencées dans CKAN:

titre	url	auteur
Comptage CUB 2013	http://data.lacub.fr/data.php?themes=8&layer=72	CUB

Accidents Corporels France	https://www.data.gouv.fr/fr/datasets/base-de-donnees-accidents-corporels-de-la-circulation-sur-6-annees/	Ministère de l'Intérieur
Comptages Loire Atlantique	http://data.loire-atlantique.fr/donnees/detail/trafic-des-routes-departementales-de-loire-atlantique/	CG44
Réseau RD Loire Atlantique	http://data.loire-atlantique.fr/donnees/detail/routes-departementales-de-loire-atlantique/	CG44
Réseau routier Nantes Métropole	http://data.nantes.fr/donnees/detail/localisation-des-troncons-routiers-de-nantes-metropole/	Nantes
Temps parcours Nantes Métropole	http://data.nantes.fr/donnees/fonctionnement-de-lapi/gettempsparcours/	Nantes
Traficolor Nantes Métropole	http://data.nantes.fr/donnees/fonctionnement-de-lapi/getfluiditeaxesroutiers/	Nantes
Indices circulation routière sur le RRN	http://www.setra.developpement-durable.gouv.fr/indice-de-circulation-sur-le-a4263.html	DTecITM
Comptages Hauts-de-Seine	http://opendata.hauts-de-seine.net/jeu-de-donnees/comptages-routiers-sur-la-voirie-departementale	CG92
Comptages Montpellier	http://opendata.montpelliernumerique.fr/Comptage-vehicules-particuliers	Montpellier
Réseau routier Grand Lyon	http://smartdata.grandlyon.com/localisation/tronfon-web-criter/	GrandLyon
Comptages Grand Lyon	http://smartdata.grandlyon.com/transport/trafic-historique/	GrandLyon
Traficolor Grand Lyon	http://smartdata.grandlyon.com/localisation/etat-du-trafic-temps-rfel/	GrandLyon
Evenements TR Grand Lyon	http://smartdata.grandlyon.com/localisation/evfnement-routier-temps-rfel/	GrandLyon
Traficolor Strasbourg	http://www.strasbourg.eu/ma-situation/professionnel/open-data/donnees/mobilite-transport-open-data/trafic-routier-cus	CUS
Réseau routier Strasbourg	http://www.strasbourg.eu/ma-situation/professionnel/open-data/donnees/mobilite-transport-open-data/trafic-routier-cus	CUS
Evenements TR Strasbourg	http://www.strasbourg.eu/ma-situation/professionnel/open-data/donnees/mobilite-transport-open-data/trafic-routier-cus	CUS
Réseau routier RD Saone-et-Loire	http://www.opendata71.fr/thematiques/informations-routieres/referentiel-des-routes-departementales	CG71