

Projets Transversaux Veille Innovation

EDREDON : Entrepôt de données routières et déplacements engrangés pour leur diffusion ouverte nationale

Bilan 1er semestre 2015 et propositions d'orientation pour 2015-2016

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:

- le volet **entrepôt** 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 entrepôt (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) vont également participer.

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>

2. Premiers résultats

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 (DterITM), Pigma / open data Aquitaine, Toulouse / projet Serres (DterSO), projets nationaux et premier état des lieux européen (DteciTM).

Il s'est agi de faire le tour de table des participants, en vue d'identifier des solutions possibles d'entrepôt de données de trafic. C'est IRIS qui apparaît comme le mieux placé et avec lequel des expérimentations seront faites au cours de l'été 2015.

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

Un site web utilisant le logiciel CKAN a été mis en place pour recenser les données routières open data sont:

<https://ckan.cerema-dtm.fr/>

Les données ont été regroupées 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).

Le catalogue est limité mais fonctionne bien, et répond à un besoin au sens où il n'en existait pas; il faut le compléter avec toutes les sources de données connues. Des contacts ont été établis avec Etalab, qui gère le portail data.gouv.fr. L'idée serait de travailler avec eux sur le moissonnage systématique des données routières open data au niveau national.

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. Compte tenu de 2 départs début 2015, il ne leur est plus possible de s'investir sur le projet en 2015.

La DTer Ouest travaille sur les données Bretagne-Pays de Loire.

La DTer Sud-Ouest a commencé à analyser les données de Bordeaux et d'Aquitaine.

La DTerMed a récupéré les données de plate-forme Data du Grand Lyon.

Il est prévu de partir de ces analyses pour:

- produire une documentation pratique (tutoriels) de l'utilisation concrète qui peut être faite de ces données (production de cartes, etc.)
- proposer des améliorations (feedback vers les producteurs de données)
- proposer un ou des format(s) cible(s) de publication pour les données de comptage, de description du réseau, et si possible pour le temps réel

Ces documents seront produits d'ici le 31/12.

4. Contacts et Analyse des besoins:

Nous avons commencé à identifier des contacts (collectivités, MEDDE, institutionnels, bureaux d'études, prestataires et éditeurs), 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. La base de contacts reste à construire. Des entretiens ont été réalisés avec des comme la DIRIF, le CG92, la plate-forme Pigma en Aquitaine, le club Trafic Bretagne-Pays de Loire, club EIS Paca-LR.

Cette action est à poursuivre et à amplifier. La DteclTM n'a pas encore pris contact au niveau national avec la Centrale (DGITM; CGDD), les départements (ADF, ADSTD), ou les métropoles (club PC urbains), ni a fortiori au niveau EU (compte tenu du départ de R Sevestre de la DterITM, nous n'aurons sans doute pas d'état des lieux européens, ce qui semble pourtant important pour un travail national). Aucun contact non plus avec les bureaux d'études et éditeurs de logiciels.

5. Analyse des outils

Pour l'instant deux outils ont été identifiés:

- CKAN pour la fonction de catalogage des données, utilisé pour notre catalogue des 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.
- IRIS, développé par la DTerOuest, est utilisé pour mutualiser les données de comptage en Bretagne et Pays-de-Loire, et donnant accès à des outils de requête et d'édition de statistiques aux partenaires via un site extranet. Les développements en cours permettront l'ouverture des données sur internet, pour les partenaires qui ont décidé l'ouverture de leurs données. IRIS est proposé pour devenir l'outil de mutualisation 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, en cours d'élaboration. Le CG44 pourrait être le 1er signataire.

6. Ouverture d'un site web

Les échanges avec la direction informatique et de Communication du Cerema ont conclu que le site du projet serait sous Spip/Giseh, sans attendre la solution de gestion de contenu web qu'il est prévu de déployer pour le Cerema. Le site est alimenté au fil de l'eau, essentiellement par la DterMed.

7. Application de recueil de traces

Pour l'instant aucun document n'a été produit pour le projet, par manque de disponibilité de l'équipe (en sous-effectif).

3. Organisation

1. Pilotage

La proposition comprend 2 volets assez voir complètement indépendants:

- entrepôt de données open data
- outil de recueil de traces de déplacements multimodaux

Le Référent du projet est Richard PASQUET, directeur DterSO. Il rend compte de l'avancement du projet au CODIR. Ce projet ne nécessite pas de mettre en place un comité de pilotage complet.

Le projet est coordonné par la DterMed, le volet « recueil et analyse de traces » est géré par la DterNP.

2. Fonctionnement

Décentralisé, échange par mail et téléphone; Point d'avancement mensuel, 1 visio par trimestre

Pour que le projet prenne de l'ampleur et permette de véritables développements opérationnels, il faudrait pouvoir disposer d'un chef de projet à temps plein sur ce sujet, de préférence au sein d'une direction technique.

3. Budget indicatif et contributions

En 2014, une enveloppe de 96 k€ a été affectée en 2014: sur un budget de 70, environ 50 k€ ont été consommés au 31/12.

Pour 2015, une enveloppe de 384 k€ a été créée, le budget actuel est de 217k€ (les Dter Est, NP, CE, et la DtecTV n'ont pas créé encore d'affaire 2015) et le consommé au 30/6 est de 45 k€. Par ailleurs, il était souhaitable de pouvoir encadrer des vacataires et des stagiaires, voire de sous-traiter des développements sur ce projet : nous avons demandé 2 mois de vacation par Dter et 40 k€ pour des développements, mais rien obtenu.

4. Perspectives

L'année 2015 doit permettre de démontrer une ou deux applications concrètes. Exemple : observatoire national des données de trafic, archivage de données temps réel, mise en place d'un entrepôt régional (ex. Poitou-Charentes)...

Dès la rentrée il faudra proposer des scénarios de pérennisation et « fiche produit Cerema ».

L'orientation qui apparaît après 9 mois est :

- de se concentrer sur les données de comptage (et le référentiel routier sous-jacent)
- de mettre en avant la mutualisation de données (partenariale, avec conventions) dont l'open data serait seulement un sous-produit. Cela élargit le périmètre du projet donc pose un problème de ressources, mais correspond mieux à une attente « solvable » et donc à l'idée d'une « offre » Cerema.

L'offre porterait sur :

- 1) la mise en oeuvre de BD de trafic (comptages) mutualisées au niveau régional, pour les territoires qui le souhaitent (et potentiellement France entière, d'un point de vue technique)**
- 2) le catalogage national des sources de données trafic/comptages**
- 3) des recommandations sur les formats de données pour les échanges et l'open data**
- 4) le cas échéant la publication des données des DIR**
- 5) le cas échéant 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 budget indicatif pourrait être un peu inférieur à celui prévu en 2015 (384 k€) dont on peut douter qu'il soit consommé, en revanche il semble difficile de produire des réalisations concrètes sans sous-traitance ou au moins crédits vacataires/stagiaires l'an prochain.

Suites souhaitables du projet en 2016 :

- pour les directions d'administration centrale : la DtecITM est en contact avec **DIT/GRT** (et devait assurer la liaison avec la Centrale en général et des organismes type ADSTD, Cotita etc.). Il est possible qu'une décision soit prise quant à l'ouverture des données à la rentrée, auquel cas la Centrale pourrait être demandeuse d'un appui technique du Cerema pour mettre en oeuvre cette ouverture. En revanche de manière plus large sur les données de trafic hors RRN, le sujet pourrait intéresser **le CGDD au titre du SOES**.

- le développement d'une « offre » sur les entrepôts de données (y compris partenariats régionaux) pourrait relever des 5% 'service public'.