

Rapport

CETE
Méditerranée

Accès aux données décrivant l'offre TC

État des lieux

Ressources, territoires et habitats
Énergie et climat
Prévention des risques
Développement durable
Infrastructures, transports et mer

**Présent
pour
l'avenir**



Centre d'Etudes Techniques de l'Équipement Méditerranée

Accès aux données décrivant l'offre de Transport Collectif état des lieux

date : 15 janvier 2010

auteurs : CETE Méditerranée

responsable de l'étude : Patrick Gendre, DCEDI

participants : Xavier Campens, Didier Danflous, ALR

résumé de l'étude :

Les données décrivant l'offre théorique de transport collectif sont de plus en plus fournies par les Exploitants de réseaux aux collectivités Autorités Organisatrices de transport (AOT) en vue notamment de constituer des Bases de Données, des Systèmes d'Information Multimodale (SIM) et des services d'information aux usagers.

Au-delà des SIM, les données d'offre TC sont également utiles pour des études en temps différé (modélisation, analyses d'accessibilité, cartes, etc.), et pour créer d'autres services d'information.

Or il apparaît qu'en général, les partenaires d'un SIM ont convenu de mettre à disposition leurs données en vue d'alimenter le site internet d'information usagers créé en même temps que la Base de Données regroupant les offres TC de tous les partenaires, mais pas pour d'autres utilisations.

L'objectif de cette étude est de faire un état des lieux de la situation en matière de diffusion de ces données d'offre théorique TC, et de suggérer, sur cette base, quelques propositions. Il s'agit en quelque sorte de poser le problème, sachant que la ou les solutions impliqueront des compétences juridiques que ne possède pas le CETE, et évidemment des décisions des collectivités (ou des transporteurs).

Le document s'adresse aux personnes intéressées par l'information transport en général, AO et collectivités, services de l'état, mais aussi transporteurs, opérateurs de services d'information ou bureaux d'études.

Le travail s'appuie sur un questionnaire adressé aux AOT pour faire un 1er bilan et propose quelques pistes pour progresser tant au niveau local que national (définition des droits patrimoniaux et des objectifs de diffusion, licences types, diffusion brute aux services d'étude, expérimentations de libre mise à disposition...).

Les attentes des usagers et des utilisateurs sont de plus en plus importantes ; il semble donc essentiel de clarifier si possible dès 2010 les conditions pratiques d'utilisation des données d'offre théorique TC, d'autant plus que la question se pose aussi pour d'autres données (perturbation, horaires TC temps réel, vélos en libre-service, information routière, etc.).

zone géographique :

nombre de pages : 15

n° d'affaire : 09C000208

maître d'ouvrage : CERTU/DD/PST (M. Yannick DENIS)



SOMMAIRE

Table des matières

1 INTRODUCTION.....	4
1.1 Contexte.....	4
1.1.1 Objectifs.....	4
1.1.2 Démarche.....	4
2 GÉNÉRALITÉS.....	5
2.1.1 éléments juridiques	5
2.1.2 La question de l'ouverture des données.....	5
3 POINT DE VUE DES AO ET AUTRES FOURNISSEURS DE DONNÉES.....	7
3.1 retours du questionnaire.....	7
3.2 en résumé.....	9
3.3 Cas de la SNCF.....	10
4 PERSPECTIVES.....	11
5 ANNEXES.....	12
5.1 Questionnaire.....	12
5.1.1 – contexte.....	12
5.1.2 - organisme.....	12
5.1.3 - pratiques actuelles.....	12
5.1.4 - attentes.....	13
5.1.5 – autres points.....	13
5.2 Références.....	14
5.2.1 études et documents.....	14
5.2.2 web.....	14
5.3 personnes interrogées.....	15
5.4 Glossaire.....	15

1 Introduction

1.1 Contexte

Les données décrivant l'offre théorique de transport collectif sont de plus en plus fournies par les Exploitants de réseaux aux collectivités Autorités Organisatrices de transport en vue notamment de constituer des Bases de Données, des Systèmes d'Information Multimodale (SIM) et des services d'information aux usagers.

Les Autorités Organisatrices ont créé un groupe 'SIM' hébergé par le GART qui se réunit environ 4 à 5 fois par an. Il est co-animé par le GART et la Région Pays de Loire. Le CETE participe à ce groupe.

Au-delà des SIM, les données d'offre TC sont également utiles pour des études en temps différé (modélisation, analyses d'accessibilité, cartes, etc.), entre autres pour les services de l'état et donc les CETE. Lors d'échanges récents sur cette question, on a pu identifier qu'au moins les CETE Normandie-Centre, Nord-Picardie, Ouest, Lyon et Est utilisaient ou avaient besoin à court terme d'utiliser des bases de données d'offre en vue d'études de diagnostic déplacements ou autres modèles multimodaux.

Par ailleurs, les données d'offre TC des SIM peuvent être utilisées pour créer d'autres services d'information, au-delà des territoires couverts par chaque SIM (certains sites web existent déjà, comme itransports ou Google Transit), et, au niveau national, le ministère du développement durable est en train de créer une Agence de l'Information Multimodale (AFIMB) en vue de favoriser l'interopérabilité entre les SIM et l'émergence de services d'information couvrant tout le territoire.

Or il apparaît qu'en général, les partenaires d'un SIM ont convenu de mettre à disposition leurs données en vue d'alimenter le site internet d'information usagers créé en même temps que la Base de Données regroupant les offres TC de tous les partenaires, mais pas pour d'autres utilisations.

1.1.1 Objectifs

L'objectif est de faire un état des lieux de la situation en matière de diffusion de ces données d'offre théorique TC, et de suggérer, sur cette base, quelques propositions. Il s'agit en quelque sorte de poser le problème, sachant que la ou les solutions impliqueront des compétences juridiques que ne possède pas le CETE, et évidemment des décisions des collectivités (ou des transporteurs). L'idée est de travailler de manière informelle mais "rapide" (6 mois). Ensuite, de toutes façons, la production par chaque AO de documents ayant valeur juridique (licences...) sera longue, a fortiori au niveau des partenaires de SIM, qui impliquera une concertation.

L'étude n'a donc pas vocation à apporter des solutions (juridiques notamment), simplement à produire un état des lieux (avec quelques références utiles), une analyse des besoins et des pistes pour progresser. Elle s'adresse aux personnes intéressées par l'information transport en général, AO et collectivités, services de l'état, mais aussi transporteurs, opérateurs de services d'information ou bureaux d'études.

1.1.2 Démarche

Le travail effectué est le suivant:

- recensement bibliographique
- élaboration et diffusion questionnaire aux AO concernant la diffusion des données TC
- description de la situation en matière d'accès aux données TC
- synthèse et propositions

2 Généralités

2.1.1 éléments juridiques

La PREDIM avait produit des notes relatives aux données TC, rédigées par Urba2000 ; ces notes concluaient au caractère public de ces données. Elles s'appuyaient pour cela sur les différents textes existants et récemment mis à jour :

- la Loi d'Orientation sur les Transports Intérieurs garantit le droit au transport et le droit à l'information transport ; la loi SRU la complète par l'obligation de mettre en place des SIM sur les périmètres de transport urbain ;
- plusieurs textes de niveau européen (directive) et national (loi, ordonnance, circulaire, décret).

Nous renvoyons à ces notes pour plus d'explications.

La direction affaires juridiques du ministère (MEEDDM) suit également ce dossier.

La Commission d'Accès aux Documents Administratifs a répondu aux questions posées par la Région Pays de Loire concernant les données Destineo. Cet avis est consultable sur internet également. La CADA considère que "Les données contenues dans la base Destineo, site d'information sur les transports publics de la région des Pays de la Loire, sont élaborées ou détenues dans le cadre de la mission de transport public de personnes, qui constitue un service public industriel et commercial. Dès lors, elles ne constituent pas des informations publiques au sens de la loi du 17 juillet 1978, et ne sont pas régies par le chapitre II de cette loi relatif à la réutilisation de ces informations". Ce point reste néanmoins discutable et n'a pas forcément de conséquence pratique sur les conditions de mise à disposition des données.

D'autres organismes publics peuvent être concernés par le sujet, notamment l'Agence du Patrimoine Immatériel de l'état (APIE) qui a produit dans le cadre du plan 'France numérique 2012' des licences-types permettant la réutilisation des données publiques, ou la future Autorité de Régulation des Activités Ferroviaires (ARAF) qui doit être créée en décembre 2009, et qui pourrait avoir un mot à dire sur les conditions de diffusion des horaires de train.

2.1.2 La question de l'ouverture des données

Pourquoi ouvrir ses données ?

Même si les collectivités et les exploitants mettent en place localement des services et systèmes d'information aux usagers, notamment sur l'internet, et si la qualité de l'information de « première main » est reconnue, l'essentiel de l'audience est néanmoins captée par d'autres médias (web, mobiles, presse, radio, télé, etc.) qui eux aussi diffusent de l'information trafic/déplacement. Il faut bien avoir à l'esprit que la chaîne d'information comporte plusieurs maillons, de l'exploitant à l'utilisateur en passant par des bases de données (référentiels) et des serveurs d'information, et qu'en principe chaque collectivité et chaque gestionnaire a intérêt à démultiplier les canaux pour toucher les usagers et valoriser « son » information transport.

Plus largement, le point clé nous semble être celui de l'ouverture des données et des informations, nécessaire pour diffuser une information cohérente de qualité, mais aussi pour une définition cohérente des politiques de gestion de déplacements, et pour une exploitation coordonnée des réseaux.

Les maîtres d'ouvrage prennent d'ailleurs conscience que l'information sur l'offre multimodale peut aussi être un outil de régulation, et pas seulement en temps réel, juste avant ou pendant un déplacement : l'information multimodale permet aussi de fournir des conseils ou des solutions d'itinéraires prenant en compte la congestion récurrente et les besoins de gestion des déplacements, ainsi que des outils d'analyse de l'offre en temps différé, pour les techniciens comme pour les usagers.

Un accord des partenaires définissant les principes de mutualisation des données semble un préalable essentiel à la mise en place d'outils communs pour la gestion multimodale des déplacements, accord qui est néanmoins plus difficile à obtenir en pratique que la définition d'objectifs commun de réduction de la congestion ou d'amélioration de la sécurité.

Les conditions du succès d'une démarche de gestion multimodale des déplacements sont donc :

- 1- Le partage et l'ouverture des données
- 2- La mise à disposition d'outils mutualisés (bases de données, sites internet, modèles...)
- 3- L'implication de l'utilisateur

Nous renvoyons sur ce sujet à d'autres études (y compris celles du CETE, voir annexe).

L'expérience anglo-saxonne

En pratique, la libre diffusion des données de transport public sur l'internet commence à s'observer aux Etats-Unis, et dans une moindre mesure au Royaume-Uni.

Aux Etats-Unis, plusieurs collectivités ont choisi de mettre les données en ligne sur le web (en général au format XML GTFS de Google, plus répandu outre-Atlantique, similaire à notre format Trident français, en un peu plus simple). Le site de référence pour comprendre ce qu'apporte cette ouverture des données est le portail des services d'info TC www.citygoround.org qui recense les sites créés en utilisant les données mises à disposition par 90 agences TC américaines (il en reste plus de 600 qui n'ont pas encore ouvert leurs données).

Par exemple, la politique du MassDoT, ministère des transports de l'état du Massachussets (comprenant Boston), a évolué en 2009. Les données sont désormais mises à la disposition du public et des développeurs informatiques, et l'agence veut maintenant "motiver" ces développeurs afin qu'ils utilisent ces données pour en construire de nouvelles applications et services. Leur démarche est assez bien documentée sur Internet, tout comme celle de San Francisco ou Portland. La licence d'utilisation est minimale, dans le sens que le fournisseur des données ne s'engage à rien en termes de qualité des données, et où le fournisseur doit simplement citer la source et ne pas modifier les données (en vue de se prémunir contre d'éventuels recours des utilisateurs : clauses de non-responsabilité, "disclaimer" en anglais).

Au Royaume-Uni, l'initiative "show us a better way" s'inscrit dans le cadre de la politique de l'information du gouvernement britannique, qui vise à rendre plus accessible et plus lisible l'information publique et surtout à en encourager sa réutilisation tant par le public que par le secteur privé. Mi juin 2008, le gouvernement britannique lançait un concours d'idées ouvert à tous, sur les usages innovants qui pourraient être faits de l'information publique numérique. Les candidats devaient décrire leur idée, expliquer à qui s'adressait leur service, comment il fonctionnait et de quelles données publiques (non nominatives) le service avait besoin. Plus de 450 compétiteurs ont participé au concours. Le domaine des transports a suscité de très nombreuses réponses. Le bilan est très positif en termes de capacité à mobiliser la créativité et le savoir-faire des usagers, à inventer des nouveaux services, mais plus nuancé en termes de valeur ajoutée réelle de ces services.

Le gouvernement britannique a récemment confié une mission à Tim Berners-Lee, l'inventeur du World Wide Web en matière de diffusion des données publiques en général et transport en particulier, jugeant en particulier que la diffusion de l'information via le portail national Transport Direct n'était pas satisfaisante.

En résumé, cette politique (qui commence en général toujours par une phase d'expérimentation très pragmatique, et parfois ne va pas plus loin) s'explique par :

- une 'tradition' de diffusion des données publiques des pays anglo-saxons (et scandinaves) ;
- l'idée de mettre en commun les données concernant l'offre multimodale ;
- une volonté d'innover ("web 2.0") ;
- un objectif d'économie : à la limite, l'autorité publique peut se passer de financer le site internet, ou en tout cas se limiter à un site fournissant l'information de base et laisser des tiers développer des sites pour mobiles ou fournissant des informations ou des fonctionnalités complémentaires.

En France, on peut donner l'exemple des données du Vélib de la Ville de Paris opéré par JCDecaux : au départ des programmeurs 'malins' allaient directement aspirer les données du site, puis JCDecaux a fini par leur mettre à disposition un service web leur permettant d'accéder à ces infos en XML ; la libération est partielle car le service n'est pas officiel et on a aucune information sur la licence d'usage (cf. <http://www.bortzmeyer.org/velib-rest.html>) ; cet accès a permis la naissance de très nombreuses applications web/widget/applis mobiles, etc.

Comme la FING dans sa démarche Villes 2.0, il est souhaitable à notre avis de faire le même genre d'expérimentation en France, pour répondre aux 3 principaux cas d'utilisation des données :

- utilisation « participative » par des usagers (mashups...) ;
- utilisation « ad hoc » pour des études par des collectivités ou autres structures y compris associatives (peu d'exigences en terme de format et de mise à jour) ;
- opérateurs de service d'information (Google, Moviken, Orange, etc.) en web service.

3 Point de vue des AO et autres fournisseurs de données

3.1 retours du questionnaire

Le CETE a diffusé via le groupe SIM du GART un petit questionnaire concernant les conditions de mise à disposition des données d'offre TC par les AO.

Nous avons obtenu plusieurs réponses, dont voici quelques exemples significatifs :

Conseil Général de l'Isère : Le CG est propriétaire des données du réseau départemental Transisère et les gère dans l'outil Pegase (y compris les scolaires). Une convention a été signée avec les AOTU, qui mettent à disposition du SIM Itinère (le site internet vient d'ouvrir en septembre 2009), ainsi qu'avec la SNCF, à divers formats (Excel, Trident, Hastus...). Techniquement, le SIM Itinère peut ensuite réexporter ces données au format Trident notamment. Ainsi, les données horaires du groupement de commandes qui réunit le CG et les AOTU iséroises hors Grenoble ont été utilisées pour le système billettique (définition des itinéraires, horaires de prise de service pour chaque course) Le CG38 réfléchit aux conditions de mise à disposition de ses données routières et TC ; une étude pourrait être lancée dans cet objectif.

Le CG38 fait remarquer qu'ils ne sont jamais sollicités pour leur données horaires, à part par les autres AO (ni par exemple par l'agence d'urbanisme, le CETE ou autre service d'état, ni par des sociétés privées).

Le CG38 va lancer très prochainement une étude de faisabilité d'interconnexion des SIM Multitud, GMCD et Itinère.

Conseil Général de la Gironde : un 'SIV33' est en cours de réalisation dans le cadre d'un marché confié à la société Cityway ; la mise à disposition des données par web service est prévue. Le CG33 est propriétaire des données, y compris positions des arrêts et tracés de lignes. Des échanges avec les données du réseau TCU de Bordeaux sont prévus (les 2 calculs d'itinéraire interopéreront).

Il est envisagé de travailler sur une convention précisant les conditions de mise à disposition des données.

Conseil Général d'Indre-et-Loire : Le CG37 contribue au SIM de la Région Centre, avec laquelle il a signé une convention. Les horaires sont gérés avec l'outil Rapido et fournis sous ce format au SIM. Le géo-référencement des arrêts est en cours. Il n'y a pas encore de stratégie de diffusion particulière, à part la mise à disposition du SIM Centre.

Conseil Général d'Eure-et-Loir : La participation du CG28 au projet SIM jv-malin de la région Centre (ouvert le 4/12/09) a accéléré la réalisation d'une campagne complète de géolocalisation des points d'arrêts des lignes régulières (soit près de 1 300 points d'arrêts). Aujourd'hui une procédure de mise à jour des données géolocalisées est en cours de réflexion. En parallèle, un travail d'intégration de ces données dans le progiciel Pégase est en cours.

Concernant les données du réseau Transbeauce, aucun droit particulier n'est indiqué « textuellement » dans l'actuelle DSP. Dans le cadre de la convention de mise en œuvre et d'exploitation du SIM régional, il a été indiqué que « tout membre signataire peut demander ponctuellement ou de manière récurrente l'extraction de données depuis la base régionale. Il doit cependant avoir l'accord des 2/3 des parties signataires concernées par les données échangées. Ces extractions auront pour seule finalité l'amélioration des dessertes et des correspondances ».

Les données peuvent être fournies au format XML Trident via un export du logiciel métier transporteur (logiciel servant par ailleurs dans le cadre d'un calculateur d'itinéraire monomodal départemental). Cet export est ensuite transmis au prestataire chargé de l'exploitation du SIM.

Le CG28 n'a pas encore établi sa stratégie de diffusion des données.

Conseil Régional Rhône-Alpes : le CR-RA est Maître d'Ouvrage du SIM multitud.org ; les données comprennent la BD horaires centralisée avec la position des arrêts (mais pas les tracés des itinéraires).

Les données brutes appartiennent à la SNCF, la Région n'a pas le droit de les transmettre à des tiers telles quelles. En revanche, une fois intégrées dans le SIM Multitud' ces données appartiennent à la Région (ou aux partenaires) et peuvent être utilisées selon les accords des partenaires (par ex. pour réalisation de cartes représentant les temps d'accès aux principales gares par l'agence d'urbanisme de Lyon), mis à disposition soit au format XML/trident, soit via des web services.

Il n'y a pas encore de stratégie pour la diffusion des données, les AO sont déjà régulièrement sollicités pour transmettre ce type de données, mais pas organisés pour le faire.

Les 12 partenaires du projet sont liés en effet par une convention qui stipule que la Région ne pourra pas retransmettre ces données à des tiers sans leur accord. De façon très pragmatique, la Région Rhône-Alpes réceptionne les

demandes/propositions d'utilisation de la BD dont elle fait état auprès des partenaires pour qu'ils valident (ou non) la transmission de la base de données au format Trident à des tiers.

Il est prévu de travailler à définir les conditions de diffusion et d'utilisation des données de Multitud avec le service juridique de la Région. C'est un des souhaits des partenaires également ; il y a un besoin de clarifier la manière de transmettre des données et sous quelles conditions. L'objectif pour la Région est de valoriser les données de Multitud' (par leur diffusion) dans des projets jugés intéressants (par exemple, pour la mise en œuvre de PDE/PDA, le calcul de temps d'accès aux gares, etc.) en s'assurant que ces données ne sont pas récupérées par un tiers qui proposeraient des services d'information identiques à Multitud'. Pour les partenaires de la Région ; il s'agit de pouvoir leur garantir l'usage qui serait fait des données qui seront diffusées, s'assurer que les organismes (privés notamment) à qui l'on transmettrait les données ne les réutiliseraient pas à des fins commerciales, ou pas du moins sans une juste rétribution pour le travail et l'investissement fournis par les partenaires.

Conseil Régional Pays de la Loire

Le SIM Destineo.fr centralise les données des réseaux TC de la région Pays de Loire, y compris les arrêts géo-référencés qui permettent des recherches d'itinéraire d'adresse à adresse. Le serveur internet peut également être interrogé en web service (c'est-à-dire, de manière programmatique depuis l'internet¹). La convention partenariale précise que "les données alimentant le site restent la propriété de chaque partie signataire", et que "la Région se porte garante d'une utilisation conforme par le prestataire des données des parties signataires au fin prévues dans le cadre du marché". L'annexe relative à l'usage des web services précise que « le partenaire souhaitant utiliser les webservices doit préalablement décrire à la Région la finalité d'utilisation du web service, et respecter cette finalité. Tout changement dans la finalité d'utilisation doit préalablement faire l'objet d'une autorisation écrite de la part de la Région, après avoir reçu une demande motivée de la part du partenaire."

Le Conseil Régional ne dispose pas en propre des données ; la SNCF gère l'essentiel des données et les transmet au prestataire qui gère Destineo. La convention liant la Région et la SNCF ne précise pas les questions de propriété de données.

Les données sont accessibles via des web services (pas directement sous forme de fichiers). Les conditions de diffusion sont définies dans une charte d'usage, et dépendent de l'usage envisagé. Le demandeur doit adresser au Conseil Régional un courrier précisant les utilisations des données qu'il prévoit. La charte prévoit que l'accès puisse être payant pour certains cas d'utilisation, mais les prix ne sont pas encore définis (mais devront l'être en 2010).

Sur ces question, la Région s'appuie sur son service juridique ainsi que sur un cabinet d'avocats. La Région a également saisi la CADA, pour savoir quelles étaient les obligations pour la Région, en tant que maître d'ouvrage de Destineo, de donner accès aux données d'offre TC, mais la CADA s'est déclarée incompétente (cf. référence web citée en annexe).

Conseil Régional Basse-Normandie : la Région est propriétaire des données d'offre TER et de la BD constituée pour le fonctionnement du SIM (commentjyvais.fr) ; une charte a été signée avec les AOT pour la mise à disposition de leurs données. l'ensemble des données (lignes, arrêts, horaires) est diffusable sous forme de fichiers mais la stratégie de diffusion n'est pas encore établie. Le Conseil Régional souhaite par ailleurs mieux diffuser l'information vers les mobiles et en gare. Plus largement, la Région s'intéresse au covoiturage, au vélo, aux informations tarifaires, au temps réel et donc aux données correspondantes et à leur diffusion éventuelle. L'objectif est également s'assurer une interopérabilité avec les SIM des territoires voisins et avec l'information touristique et culturelle.

Le Syndicat des Transports d'Ile-de-France (STIF) a signé en 2009 avec la Direction Régionale de l'Equipement de l'ile-de-France DREIF une convention de mise à disposition des données de l'offre TC francilienne, nécessaire pour les études conduites par ce service de l'Etat. Une réflexion est en cours sur la diffusion des données, en cohérence avec le Schéma Directeur de l'Information Voyageurs.

Syndicat Mixte des Transport Collectifs de l'Oise (SMTCO)

Dans le cadre de conventions d'échanges de données, les données (horaires et localisation des arrêts, ainsi que les tracés des lignes départementales du CG60) sont mises à disposition du SMTCO par les AOT, en vue d'alimenter SISMO, le Système Intégré de Services à la Mobilité dans l'Oise. Le Calculateur multimodal est en cours de réalisation à l'échelle de

¹ Une API (Application Programming Interface) est un document qui spécifie la manière dont on peut interroger un programme informatique, et l'implémentation de cette API permet que lorsqu'on l'interroge, le programme réponde selon ces spécifications. Une API en web service signifie que la spécification d'interface utilise les protocoles du web (HTTP, etc.). L'intérêt pratique que les fonctionnalités du SIM (horaires, calculs d'itinéraires, etc.) soient accessibles via une API web est que le SIM peut alors être interrogé par des programmes, et que les fonctionnalités du SIM peuvent ainsi être intégrées dans de nouveaux services ou applications.

tous les réseaux de transports du département de l'Oise, y compris les lignes TER de car régional. De manière plus large, le projet SISMO intègre des fonctionnalités de suivi de l'état du réseau routier dans le département, de suivi des véhicules de TC en temps réel (SAEIV), de billettique intermodale, et une centrale de covoiturage.

Les informations (fiches horaires, tarifs,...) sont mises à disposition gratuitement au public par un site internet www.oise-mobilite.fr et par la centrale téléphonique d'information voyageurs N° AZUR 0 810 60 00 60. Les partenaires ont accès aux statistiques sur le nombre de consultations et d'appel sur la centrale d'information voyageurs. En revanche, les données sources formatées sont confidentielles et sont la propriété du SMTCO. Le Syndicat Mixte envisage de faire appel à un conseil juridique en vue de préciser les conditions de diffusion de ces données, et est intéressé par d'éventuelles licences-types et outils qui faciliteraient cette diffusion.

Au-delà de la diffusion des données théoriques, le SMTCO souhaite également progresser sur les échanges de données entre systèmes régionaux (en particulier afin d'assurer l'interopérabilité de l'information et de la billettique avec l'Ile-de-France, et avec le réseau ferré), ainsi que sur les échanges de données concernant les perturbations sur les réseaux TC.

A noter que certaines collectivités soutiennent les logiciels libres, mais également les "données libres" (diffusées selon des licences de type [Creative Commons](https://creativecommons.org/)). Par exemple, l'agglomération de Brest Métropole Océane a fourni à OpenStreetMap (le principal site communautaire de données cartographiques libres) un premier jeu de données comprenant le cadastre vectoriel, et le plan de ville pour Brest et les 7 communes constituant l'agglomération BMO, ainsi que 1300 points d'intérêt (mais pas pour l'instant de données relatives à l'offre TC).

3.2 en résumé...

En pratique aujourd'hui à notre connaissance, aucune AO ne met à disposition ses données horaires de manière ouverte selon les termes d'une licence d'utilisation, ni a fortiori de manière libre sur l'internet ; les données sont fournies au cas par cas, en fonction des demandes.

Plusieurs raisons à cela :

- les données ne sont pas encore disponibles pour certains réseaux, pas encore validées, ou pas géo-référencées par exemple ;
- certaines AO ne sont pas explicitement propriétaires des données, ces dernières étant gérées par les transporteurs. En particulier, les Régions ne sont pas propriétaires des données SNCF ;
- il existe une réelle réticence à fournir les données à des opérateurs de service d'information (a fortiori des services commerciaux), d'une part parce qu'on considère qu'une fois qu'un SIM est mis en place et le site web correspondant ouvert au public, il n'y a pas besoin d'autres services d'information, d'autre part par crainte que les données diffusées par ces opérateurs ne soient pas à jour et donc fournissent des informations erronées aux usagers ;
- certains élus sont assez sensibles aux questions de « souveraineté » sur les données et les informations ; l'idée de mutualiser pour créer de nouveaux services ou améliorer le service aux usagers donne l'impression à certains d'être dépossédés de leurs prérogatives ; c'est vrai sur les questions d'intermodalité en général, et même pour la mise à disposition des données horaires d'offre théorique ;
- en général, la plupart des AO sont disposées sur le principe à fournir les données dont elles disposent (au moins d'offre théorique) aux différents organismes et services (en général publics ou associatifs) qui les sollicitent ; néanmoins en pratique cela demande parfois beaucoup d'efforts pour obtenir lesdites données
- certaines AO ne sont pas du tout sollicitées pour fournir des données ;
- bien qu'il y ait un consensus pour considérer que les données d'offre TC soient des données publiques (au sens de la loi), il n'existe pas d'avis juridique faisant autorité qui indique clairement quelles sont les conséquences pratiques du caractère public des données, le cas échéant ;
- l'élaboration d'une licence d'une utilisation des données demande un effort important et n'est pas prioritaire ; néanmoins la plupart seraient preneurs de solutions et retours d'expériences pour avancer sur ce point.

Il ressort que chaque AO devrait déjà s'assurer qu'elle dispose des données sur les réseaux qu'elle gère, de ses droits patrimoniaux sur ces données, de définir les modalités de mise à disposition des données. Trois modalités de diffusion sont notamment possibles (et sans doute d'autre):

- données brutes (fichiers XML type Trident/Neptune) en limitant éventuellement la diffusion à des utilisations non commerciales (études) et interdisant la rediffusion (si c'est possible juridiquement, ce qui n'est pas évident!), voire en diffusant les données de manière expérimentale pour voir quelles utilisations pourraient être faites (cf. expériences de ce type en Angleterre) ;
- via des interfaces logicielles (API, services web), avec une qualité de service supérieure mais éventuellement de

manière payante ;

- soit en fournissant un site web en 'marque blanche' de manière à ce qu'ils soient intégrés dans des sites partenaires, typiquement des pôles de déplacements tels que des zones commerciales (ce qu'ont fait les anglais pour Transport Direct).

Il faudrait envisager des conditions de mise à disposition différenciées selon l'usage, selon que les données sont réutilisées en vue de faire un service d'information ou des études, à des fins commerciales ou pas, avec droit de rediffusion, avec quelles exigences différenciées de qualité de service pour la mise à disposition : données à jour, temps de réponse, formats disponibles, etc. Pour une utilisation en vue de faire des études sans objectif commercial ni rediffusion, une simple fourniture des données existantes 'brutes', sous forme de fichiers XML ou autre, serait acceptable pour la plupart des utilisateurs.

3.3 Cas de la SNCF

Les données SNCF constituent un cas particulier important de données TC, et sont indispensables dans tous les SIM. La SNCF, via sa direction TER, a mis en place un Guichet Unique de données. Elle fournit sur demande des collectivités ses données horaires théoriques au format XML/Trident. Cette prestation est payante dans le cadre d'une convention signée avec l'AO demandeuse, ou d'un avenant à la convention d'exploitation régionale : en 2009, 30 k€ / an pour une Région, 8 k€/an pour un Département, ces tarifs correspondant aux coûts techniques liés à la prestation de mise à disposition qui inclut la fourniture au format Trident des données horaires concernant l'ensemble des circulations qui desservent un territoire géographique donné, leur mise à jour hebdomadaire et l'accès à une cellule d'assistance utilisateurs ouverte aux AO et à leurs prestataires. Les données peuvent ensuite être diffusées dans le cadre du site internet d'information multimodale, mais pas être rediffusées ni revendues, a fortiori.

Les données qui couvrent l'offre des 3 transporteurs (Voyages Grandes Lignes, TER, Transilien) sont fournies au format XML/Trident. Dans la mesure où ce format d'échange a vocation à devenir une norme, la SNCF souhaite pouvoir diffuser systématiquement selon ce format, y compris pour des demandes émanant de services études. Un des points très importants pour exploiter les données fournies est que dans le monde ferroviaire (en France en dehors de l'Ile-de-France en tout cas), la notion de ligne telle qu'elle existe dans les TC et dans Transmodel n'a pas cours : l'offre est organisée en trajets origine/destination et trains, qui seraient l'équivalent des courses TC. La notion de ligne telle que l'on retrouve sur les fiches horaires a été créée pour rendre plus lisible l'offre SNCF pour les voyageurs. Par conséquent, les fichiers fournis regroupent des trains desservant à peu près les mêmes gares, un peu comme dans les fiches horaires papier ; la notion de ligne est spécifiquement construite pour l'export et cela nécessite un travail pour l'utiliser correctement ensuite dans le SIM et son calculateur d'itinéraires.

Comme les AO, la SNCF est très soucieuse du fait que les données d'information horaire concernant les trains et les cars qui pourraient être fournis par d'autres services que les siens soient les bonnes. Pour l'instant, le pôle information voyageurs TER traite uniquement les demandes des collectivités AO dans le cadre de projets SIM (tels que Destineo, Itinisère, Multitud, etc.), mais l'idée est de mettre en place un guichet pour d'autres acteurs (opérateurs de services d'information, employeurs dans le cadre de PDE, etc.). La SNCF réfléchit donc actuellement aux conditions d'utilisation des données, ce qui pourrait se concrétiser sous la forme d'une ou plusieurs licences et du service correspondant. Les Régions elles-mêmes sont sans doute directement sollicitées, notamment par des services qui ont besoin des données en vue de faire des études.

Par ailleurs, il est intéressant de savoir que la SNCF commercialise tous les 6 mois un CD-ROM 'RIHO' des horaires de train en Europe pour un prix de 8 euros (12.50 euros en cas d'envoi à domicile), et que le CETE Nord-Picardie a développé un logiciel pour lire ces données. Comme les horaires sont des données publiques, rien n'empêche en principe de les utiliser dans des études à condition qu'elles ne soient pas rediffusées. Les données sont évidemment de moins bonne qualité que celles mises à disposition des SIM car elles ne disposent pas des mises à jour hebdomadaires qui prennent en compte les évolutions des circulations liées notamment aux travaux sur le réseau ferré tant régional que grandes lignes.

Au niveau européen, les exploitants ferroviaires ont mis en place une base de données MERITS contenant l'ensemble des horaires en Europe. Cette base permet à la SNCF d'informer ses clients sur une partie des horaires des trains étrangers, ou par exemple à la DB et aux CFF de fournir des informations horaires sur les trains français. Cette base actualisée mensuellement ne dispose pas de l'exhaustivité des modifications des circulations.

4 Perspectives

Par rapport aux attentes exprimées par les AO, on peut proposer quelques pistes pour progresser :

- faire élaborer une ou plusieurs licences types (adaptées aux différentes stratégies de diffusion possibles pour une AO et compatibles avec le droit en vigueur), ou rediffuser celles qui seraient élaborées par les AO les plus avancées ; la licence-type élaborée par l'APIE distingue la mise à disposition avec livraison unique des données, ou avec mise à jour régulière, ce qui pourrait correspondre aux deux principaux cas d'utilisation des données TC : étude, ou information usagers ;
- lorsque les données sont disponibles, les mettre à disposition des organisations qui les demandent en vue de faire des études, au format existant, 'brutes', sans garantir leur qualité : cela suffira à satisfaire la plupart des demandes de ce type, les utilisateurs des services d'étude se débrouillant ensuite pour les convertir au format de leurs outils ;
- pour les AO intéressées, mettre en place des expérimentations similaires à celles qui ont été conduites aux USA ou au Royaume-Uni. Ceci pendant quelques mois ou 1 an, en vue de tester l'intérêt d'une mise à disposition libre et gratuite des données, puis évaluer la pertinence des services qui pourront avoir été développés à partir de ces données par la 'communauté'.

Les SIM qui sont mis en place sur la plupart des territoires permettront de fournir ce service de mise à disposition des données pour le compte de leurs partenaires, néanmoins les AO peuvent aussi fournir en direct leurs données, cela pouvant simplifier la mise à disposition à court terme.

Il semble essentiel de clarifier dans les mois qui viennent les conditions d'utilisation des données d'offre théorique TC. Il est clair que la question se pose aussi pour d'autres données (perturbation, horaires TC temps réel, vélos en libre-service, information routière, etc.), et que certaines collectivités, dans le cadre de la réflexion sur la mise à disposition des données d'offre théorique TC, voudront réfléchir de manière plus large à ces questions, ce qui pourra peut-être retarder le processus. Sans parler des questions de protection des données individuelles, qui concernent indirectement les services d'information aux usagers (cf. la récente proposition de loi² visant à compléter la loi Informatique et Libertés).

En pratique, chaque AO devrait, dans l'ordre :

0. vérifier que les données existent, leur nature (comprennent-elles la position des arrêts, le tracé des lignes ?)
1. vérifier qu'elle possède les droits patrimoniaux (si nécessaire ? Selon DSP, marché, convention transport)
2. prendre en compte les obligations réglementaires pour définir sa politique de diffusion (en tant que collectivité, puis ensuite dans le cadre de partenaire d'un SIM)
3. ensuite, il s'agira de :
 - produire une ou des licence(s) d'utilisation
 - s'appuyer sur un SIM pour la diffusion (mutualiser les coûts avec les partenaires)
 - commencer par fournir des fichiers bruts, sans garantie, cela répondra déjà à certains besoins
 - envisager une fourniture plus industrielle, avec des mises à jour fréquentes (web services ?), éventuellement payante (selon volume)
 - selon sa stratégie de diffusion, expérimenter une diffusion plus large et libre sur l'internet

Au niveau national, l'animation de la réflexion sur la mutualisation des données transport devrait tout naturellement revenir à l'Agence Française de l'Information Multimodale et de la Billettique, qui devrait être créée en 2009 par le ministère du développement durable. Néanmoins, les questions transversales sur les données publiques notamment dépassent le cadre et pourraient être traitées par le secrétariat à l'économie numérique, par exemple, et des agences telles que l'APIE, la CADA, l'ARAF, selon leur domaine de compétences.

² <http://www.senat.fr/leg/pp109-093.html>

5 Annexes

5.1 Questionnaire

Utilisation des données issues des Systèmes d'Information Multimodale Questionnaire aux Autorités Organisatrices

CETE Méditerranée, 5 août 2009

5.1.1 – *contexte*

Les données disponibles dans les centrales d'information multimodale peuvent avoir beaucoup d'autres utilisations, que le site web de recherche d'itinéraire et d'info horaires généralement mis en œuvre dans le cadre des projets de SIM, que tout le monde gagnerait à valoriser.

Ce questionnaire s'inscrit dans le cadre d'une étude confiée au CETE Méditerranée par le CERTU, et soutenue par le ministère du développement durable en préparation de la création de l'AFIMB.

L'objectif est de produire un document qui recense les pratiques en matière de diffusion des données d'offre théorique TC, et les attentes des parties prenantes (en premier lieu des AO maîtres d'ouvrage de SIM), traduites si possible sous forme de cahier des charges qui pourrait être soumis à des juristes en vue de produire des outils et des solutions (licences ou autres).

5.1.2 - *organisme*

organisme :

porteur du projet de SIM (le cas échéant si l'organisme participe à un SIM) :

rôle :

vos nom et vos coordonnées (e-mail, téléphone) :

souhaitez que vos réponses à ce questionnaire soient

- citées dans le rapport du CETE

- regroupées anonymement dans l'état des lieux et l'expression des besoins

5.1.3 - *pratiques actuelles*

quelles sont les données disponibles ?

- base de données horaires (offre théorique du réseau TC) :

- concernant le réseau TC que vous gérez :

- mutualisées au niveau d'une centrale d'information

- données géo-référencées

- localisation des points d'arrêt :

- tracé des lignes :

- quels sont vos droits patrimoniaux sur ces données (adaptation, diffusion, duplication...) :

comment les données peuvent-elle être diffusées

- fichiers texte
- au format Trident
- autre format
- couche SIG
- web service (préciser dans son mode d'utilisation)

- de manière plus large, quelles sont vos pratiques, vos projets et votre politique concernant les autres données liées aux transport et à la mobilité : référentiel routier, stationnement, trafic, TC temps réel, covoiturage, traces, billettiques, résultats d'enquêtes, etc.

Avez-vous une stratégie de diffusion des données ? Oui / Pas encore / Non

Si oui, est-elle concrétisée sous forme de

convention :

licence :

organisation :

outils informatiques :

Les conditions de diffusion dépend-elle :

du type de données (TC, autres) ?

de l'utilisation commerciale ou non, de la rediffusion des données, de la fréquence des requêtes ? Est-elle gratuite ?

En pratique, comment cette diffusion est-elle mise en oeuvre ?

Nombre et nature des demandes...

vos compléments et remarques :

5.1.4 - attentes

Envisagez-vous de travailler sur cette stratégie de diffusion et sa mise en oeuvre ?

A quelle échéance ?

Envisagez-vous de faire appel à des prestataires extérieurs (juristes notamment) ?

Quelles sont les principales attentes de votre organisme ?

Quelles sont selon vous les principales attentes de vos partenaires dans le cadre du SIM auquel vous participez ?

Etes-vous intéressé(e) par des outils qui seraient produits au niveau national :

convention-type, licence-type

outils informatiques :

Seriez-vous prêt à contribuer à cette production ?

Oui / Non / ça dépend

5.1.5 – autres points

Avez-vous d'autres éléments à suggérer qui n'ont pas été abordés dans les questions précédentes ?

vos compléments et remarques :

5.2 Références

5.2.1 études et documents

- un guide « échanger des données localisées » produit par le CERTU

http://www.certu.fr/catalogue/p1783/ECHANGER_DES_DONNEES_LOCALISEES/product_info.html

- des documents de travail de la PREDIM dont le plus récent est sur le web :

<http://www.predim.org/spip.php?article2705>

Documents administratifs : droit d'accès et réutilisation, CADA, La documentation française, 2008, 210 Pages.

Avis de la CADA aux questions posées par la Région Pays-de-Loire :

http://www.cada.fr/fr/conseils/resultat.cfm?fiche_choisi=20090221

Mutualisation des données transport; étude CETE Méditerranée pour CERTU, février 2008:

http://www.cete-mediterranee.fr/tt13/www/article.php3?id_article=138

Participation des usagers à l'information déplacements; étude CETE Méditerranée, 2009:

http://www.cete-mediterranee.fr/tt13/www/article.php3?id_article=195

5.2.2 web

Documents de travail de la PREDIM dont le plus récent est :

<http://www.predim.org/spip.php?article2705>

journée PREDIM : http://www.predim.org/IMG/pdf/actes__journeedu260608.pdf

https://www.apiefrance.com/sections/acces_thematique/reutilisation-des-informations-publiques/la_reutilisation_des/

https://www.apiefrance.com/sections/acces_thematique/reutilisation-des-informations-publiques/licences_types_de_re7750/

<http://fr.creativecommons.org/>

<http://www.villes2.fr>

www.openstreetmap.org

www.showusabetterway.com

<http://news.bbc.co.uk/2/hi/technology/8241029.stm>

<http://www.telegraph.co.uk/news/uknews/road-and-rail-transport/6560308/Downing-Street-plan-to-put-transport-information-at-passengers-fingertips.html>

<http://www.internetactu.net/2009/09/15/critiques-du-web%C2%B2-14-quelles-donnees-libere-t-on/>

<http://www.thetransitwire.com/2009/08/25/who-owns-transit-data/>

<http://www.eot.state.ma.us/developers/>

<http://sf.streetsblog.org/2009/07/14/bart-a-national-leader-in-real-time-data-transparency-and-development/>

<http://www.trimet.org/apps/index.htm>

www.citygoround.org

<http://www.opentransitdata.org/>

<http://data.gov>

<http://transid.blogspot.com/2009/11/mise-disposition-des-donnees-transports.html>

<http://transid.blogspot.com/2009/10/un-concours-dapps-municipales-san.html>

<http://www.streetsblog.org/2009/09/23/the-case-for-open-mta-data-transparency-savings-and-easier-riding/>

http://ec.europa.eu/information_society/policy/psi/index_en.htm

5.3 personnes interrogées

Aurélie Ginguéné, Région Pays-de-Loire
 Jean-Louis Graindorge, URBA 2000
 Michel Girard, CG38
 Cédric Cherfils, CG28
 Xavier Beauvallet, CG37
 Laurence Pariès, CG33
 Monica Grassi, CG54
 Nicolas Bousselin, CR Haute-Normandie
 Claire Coulibaly, CR Basse-Normandie
 Florence Dunod, CR Rhône-Alpes
 Olivier Clarimon, Tisseo
 Jean Chaussade, Sytral
 Nicolas Maffre, Syndicat Mixte Transport Bouches-du-Rhône (Iepilote)
 Véronique Picard, Vincent Bui, Syndicat Mixte des Transports Collectifs de l'Oise (SMTCO)
 Michel Briand, Brest Métropole Océane
 Nathalie Nilson, SNCF / Proximités
 Yannick Denis, CERTU
 Roger Lambert, MEDDM/Mission Transport Intelligent

5.4 Glossaire

SIGLE	DEFINITION
AFIMBA	Agence Française pour l'Information Multimodale et la Billettique
AOs	Autorités Organisatrices
AOTU	Autorités organisatrices de transports urbains
APIE	Agence du Patrimoine Immatériel de l'état
ARAF	Autorité de Régulation des Activités Ferroviaires
CERTU	Centre d'études sur les réseaux, les transports, l'urbanisme et les constructions publiques
CADA	Commission d'Accès aux Documents Administratifs
CETE	Centre d'Etudes Techniques de l'Equipement
CG	Conseil Général
CR	Conseil Régional
FING	Fondation Internet Nouvelle Génération
IGN	Institut Géographique National
KML	Keyhole Markup Language
OGC	Open Geospatial Consortium
OSM	Open Street Map
PASSIM	Annuaire des Services d'Information Multimodal
PDE / PDA	Plan de Déplacements d'Entreprise / Administration
PREDIM	Plate-forme de recherche et d'expérimentation pour le développement de l'information multimodale
PREDIT	Programme national de recherche et d'innovation dans les transports
PTU	Périmètre de Transports Urbains
SAAS	Software as a Service
SAEI	Système d'aide à l'exploitation et à l'information
SDIV	Schéma Directeur de l'Information Voyageurs
SIG	Système d'Information Géographique
SIM	Système d'Information Multimodale
SRU	Loi relative à la Solidarité et au Renouvellement Urbain
STIF	Syndicat des Transports de l'Ile-de-France
TAD	Transport A la Demande
TC	Transports Collectifs
VLS	Vélo en Libre Service
VRU	Voie Rapide Urbaine
XML	eXtensible Markup Language