

MIEUX EXPLOITER LES DONNÉES NUMÉRIQUES

Les avancées technologiques de l'informatique et de l'Internet ont permis le développement des datasciences. Ainsi, traiter et exploiter des données numériques produites par l'Etat, par les services municipaux et par les citoyens, permet d'assister les décideurs publics dans leurs prises de décision, de faire des économies, d'orienter l'action publique et de la mesurer ou de la corriger en temps réel.

DE LA DONNÉE BRUTE A L'ANALYSE PREDICTIVE

Depuis des dizaines d'années, l'informatisation et l'utilisation massive des logiciels a entraîné une augmentation de la quantité et de la diversité des données numériques. Plus récemment, leurs coûts de stockage (dans des data centers) et de circulation (via Internet) sont en constante diminution ce qui permet un accroissement exponentiel des échanges, des croisements, des traitements, etc.

Ainsi, Facebook utilise les données des utilisateurs pour prédire les amis de l'utilisateur, LinkedIn pour prédire les contacts professionnels, Netflix pour prédire les films que ses abonnés apprécient et Amazon pour prédire les produits que l'internaute est susceptible d'acheter.

Les stratégies fondées sur la donnée, dont l'analyse prédictive, recèlent un immense potentiel d'efficacité, et de maîtrise des dépenses. Grandes productrices de données, les collectivités territoriales auraient tout intérêt à intégrer ces pratiques.



REUTILISER DES DONNÉES EN INTERNE

En premier lieu, elles réutilisent très fréquemment leurs propres données. Plutôt que de les laisser « enfermées » dans les logiciels métiers, il serait plus judicieux de les extraire et de les publier sous une forme réutilisable (feuilles Excel, fichiers textes ou bases de données ouvertes) et facilement accessibles aux agents.

Cette possibilité d'usage interne de la donnée représente une source de réduction de coûts et facilite la transversalité entre des services souvent organisés en silos. Ce décloisonnement entre les services de la collectivité, une fois bien maîtrisé, peut faire l'objet d'une ouverture à d'autres acteurs externes (DSP, autres collectivités ou citoyens).

Ce travail interne à la collectivité permet de contrôler la qualité des données et la constitution de bases de données plus complètes, actualisées, mieux qualifiées. De fait, ces services qui se réunissent pour cette action ont une vision plus globale du travail de leurs collègues, des difficultés qu'ils peuvent rencontrer.

PARTAGER EN TEMPS REEL UNE INFORMATION

Partager une information en temps réel sur l'état d'un système permet à ses acteurs de modifier leurs comportements, de constater l'effet de cette modification de comportement et de l'ajuster.

Par exemple, des solutions numériques domotiques permettent de suivre en temps réel la consommation d'électricité de son logement depuis un écran



et de visualiser, en éteignant un équipement électrique, l'effet produit sur sa facture.

De la même manière, donner à des automobilistes une information prédictive une heure à l'avance sur des embouteillages attendus d'un axe routier permet de les contourner et, in fine d'en diminuer l'intensité.

A l'heure où la complexité des problèmes urbains croît inversement aux moyens financiers disponibles, les collectivités peuvent s'ouvrir à d'autres pratiques comme l'exploitation et la circulation des données pour trouver des solutions.

BONNES PRATIQUES DES DATASCIENCES

- Négocier les achats d'électricité en anticipant et en contrôlant les pics de consommation ;
- Mieux cibler les contrôles dans les réseaux de transport ;
- Optimiser les temps d'attente aux guichets ;
- Optimiser les feux de circulation pour désengorger et dépolluer les centres-villes.

RÉGULER LES DONNÉES : AIDER LES PETITES ENTREPRISES DANS UNE ÉCONOMIE MONDIALISÉE

LA RÉGULATION EN NEGOCIANT DES DONNÉES CONTRE L'AUTORISATION D'EXERCER

La plateforme Airbnb, née à San Francisco, occasionne trop d'effets négatifs puisque, au-delà des négociations pour collecter directement la taxe de séjour, le site fait grimper les loyers dans certains quartiers.

La mairie de San Francisco a récemment mis en place un bureau dédié à la location de très courte durée. Son objet est de simplifier les démarches administratives et « coordonner les poursuites contre ceux qui ignorent ou violent les réglementations ».

Son maire, M. Lee, souligne d'ailleurs que 15 procédures viennent d'être lancées contre des propriétaires soupçonnés d'avoir transformé leurs immeubles en hôtels illégaux.

Mais, pour contrôler cela, il faudrait pouvoir accéder aux données de l'entreprise, ce qu'Airbnb refusait jusqu'à présent. Le maire entend faire voter d'autres mesures encore plus restrictives à l'encontre de la startup avec l'aide du collectif citoyen *ShareBetter San Francisco*.

Ici, les données sont un instrument de négociation dans le rapport de force qui s'établit entre ces plateformes et les territoires sur lesquels elles opèrent.

<http://goo.gl/BVdELX>

<http://goo.gl/DVlczq>



LA RÉGULATION EN CRÉANT DES PLATEFORMES NUMÉRIQUES OUVERTES

UBER a considérablement déstabilisé les taxis en permettant, grâce à une application et une plateforme de données, la maraude électronique aux VTC. Or, l'action publique, en utilisant les mêmes armes d'UBER, peut tenter de rétablir un équilibre dans ce secteur.

En effet, en France, l'article 1er de la loi n° 2014-1104 du 1er octobre 2014 relative aux taxis et aux VTC prévoit la mise en place d'un registre géolocalisé des taxis.

Ainsi, le Secrétariat Général à la Modernisation de l'Action Publique (SGMAP) et le Ministère de l'Intérieur ont développé une plateforme « le.taxi » qui exploite ce registre et qui est ouverte de façon transparente aux applications des sociétés de taxi.

Ainsi, il va devenir possible de réserver son taxi immédiatement sur son smartphone, indépendamment de son appartenance à telle ou telle centrale de réservation. Charge aux Collectivités Territoriales d'en faire la promotion localement comme à Montpellier où « le.taxi » est en expérimentation.

Les pouvoirs publics jouent leur rôle de régulation avec les outils numériques en organisant la circulation d'une donnée (la géolocalisation des véhicules) et en posant des règles fortes (neutralité et gratuité de la plateforme).

<http://le.taxi>

ANALYSE DU BÂTI POUR ANTICIPER LES RISQUES

La Ville de New-York est une des premières à avoir mis en place une équipe de spécialistes de la donnée. Ainsi, ces derniers ont identifié des immeubles à risque d'incendie pour aider les pompiers dans leur action de prévention avec plus de 60 facteurs tels que le niveau de pauvreté du quartier, l'ancienneté de l'immeuble, l'état de son réseau électrique, la présence et la localisation des extincteurs, ou encore la présence ou non d'un ascenseur.

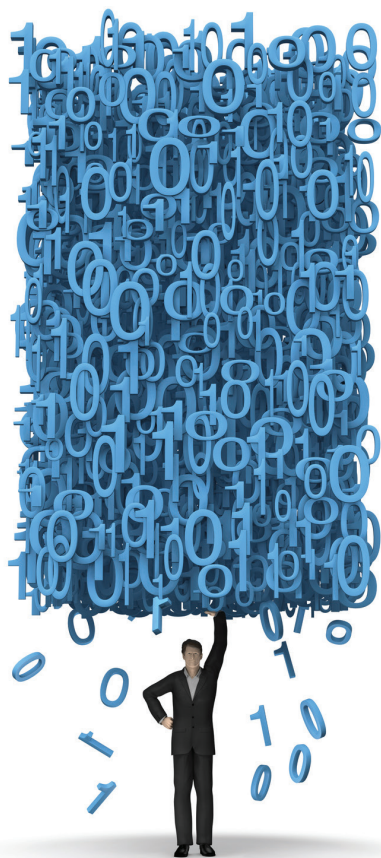
Les pompiers sont ainsi passés de 10 % de contrôles positifs à 78 % de contrôles positifs.

<http://goo.gl/uwY2Uj>



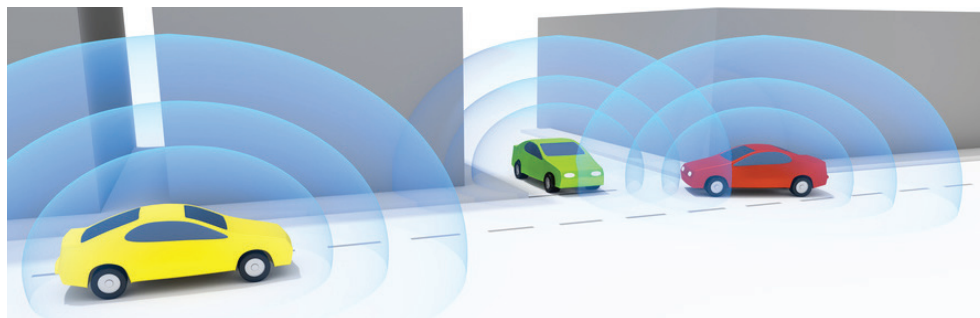
LUTTER CONTRE LE VOL DE VOITURE À L'AIDE DES DATASCIENCES

Plusieurs services de l'Etat collaborent pour mettre en œuvre un modèle prédictif des vols de véhicules pour optimiser les patrouilles des policiers et des gendarmes dans le département de l'Oise, un des plus touchés par ce délit. A l'aide de plus de 600 variables géographiques et socio-économiques ainsi que d'autres indicateurs tels que la météo et l'occurrence de vols les jours précédents et dans les quartiers voisins, un modèle a été mis au point dont les résultats sont prometteurs.



Allouer quotidiennement des patrouilles dans les 10 % de quartiers les plus risqués (probabilités estimées par le modèle) permettrait de couvrir plus de la moitié des vols des 5 derniers mois de 2014. Une collaboration directe avec les services de la sécurité intérieure présents sur le terrain est maintenant envisagée afin de construire un outil sur-mesure et adapté aux problématiques métiers.

<https://goo.gl/UrQVye>



AMÉLIORER L'OFFRE DE TRANSPORTS DANS UNE VILLE

ORANGE

Dans le cadre du programme Data4Development, l'opérateur Orange a mis à disposition de la communauté scientifique des données rendues anonymes et en particulier la localisation des utilisateurs de téléphone mobile en Côte d'Ivoire et au Sénégal, utilisées notamment pour en déduire les flux origine-destination au sein de la ville. Elles sont alors converties en parcours au niveau du réseau de transport existant.

Une équipe d'IBM a ainsi permis d'améliorer le réseau de transport de la ville d'Abidjan, de manière à augmenter le nombre de lignes et la satisfaction des usagers, à la fois en termes de parcours et de temps d'attente.

Obtenir un tel résultat (10 % de temps gagné pour chaque citoyen et 65 points d'amélioration) par des moyens traditionnels d'enquêtes et d'études sur le terrain aurait coûté bien plus cher à cette collectivité.

<https://goo.gl/M0eQX2>

WAZE

Autre cas, mettre en place des partenariats public/privé sur les données d'un territoire. Ainsi, Waze France cherche à mettre en œuvre des accords de partages de données avec les collectivités locales, comme elle le fait déjà à l'étranger. Cet accord donnant-donnant permet à l'entreprise d'améliorer son service localement et à la collectivité d'obtenir une connaissance en temps réel de son territoire.

<http://goo.gl/hMXj5T>

PRÉDIRE LES EFFETS MICROÉCONOMIQUES D'UNE RÉFORME FISCALE

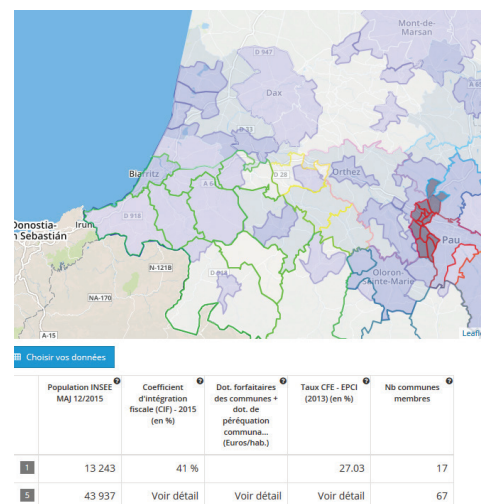
EPCI-FUSION est un outil de simulation gratuit pour accompagner les communautés dans la mise en œuvre de la loi NOTRe pour la fusion des EPCI.

Il intègre l'ensemble des données publiques pertinentes disponibles en open data dans une interface cartographique à l'échelle nationale. Il donne la possibilité de simuler l'évolution des périmètres proposés.

E-FUSION permet aux acteurs de la réforme de s'approprier les schémas départementaux et simuler une évolution des périmètres proposés en s'appuyant sur une interface cartographique et des calculs sur l'ensemble des données démographiques, physiques, et financières, dotations, indicateurs de richesses et de charges, etc.

Elle rend également possible le calcul de ces éléments dans un périmètre donné, de comparer différents scénarii, de les mémoriser et de les exporter. L'outil est accessible à tous, la sauvegarde des simulations est réservée aux abonnés de l'ADCF.

<http://epci-fusion.fr>



UNE ACTION PUBLIQUE PERSONNALISEE, PREDICTIVE, PREVENTIVE, PARTICIPATIVE MAIS GARANTE DES VALEURS DE SERVICES PUBLICS

La place croissante que prennent les données et les algorithmes est un phénomène très récent. D'un point de vue éthique, nous ne disposons pas encore du recul nécessaire et d'une maturité des usages qui permettraient de garantir que ces outils ne vont pas mettre à mal les valeurs de services publics (universalité d'accès, respect de la vie privée, respect de l'intimité, libre arbitre etc.).

Maîtriser pour réguler intelligemment

En 1978, le législateur avait eu un temps d'avance puisque la loi « Informatique et libertés » organise toujours, 40 ans après, les principales sécurités permettant de protéger la vie privée, créant un cadre d'analyse et de régulation.

Aujourd'hui, des premières régulations apparaissent puisque l'article 2 de la nouvelle loi pour la République Numérique ouvre le droit pour un individu de connaître le fonctionnement d'un algorithme qui a été utilisé pour prendre une décision l'affectant. Un des tous premiers exemple de cette situation est l'utilisation de l'algorithme d'Admission Post-Bac (APB) - le logiciel de répartition des étudiants

<http://goo.gl/VUYZfi>

S'autoriser à expérimenter

Se priver de ces nouvelles capacités d'action offertes par l'utilisation des données serait toutefois dommageable au moment où les moyens financiers diminuent et où les problèmes se complexifient. En attendant que se constituent des points de transparence et des contre-pouvoirs indépendants, les col-

lectivités peuvent commencer à expérimenter.

Nombre de services publics peuvent être améliorés à partir de données anonymisées comme les transports, la sécurité des biens et des personnes, les pratiques culturelles (fréquentations des spectacles, emprunts de documents à la bibliothèque, etc.).

Par ailleurs, il est intéressant de saisir toutes les opportunités possibles pour garder une capacité de négociation et de régulation avec ces entreprises qui bousculent des modèles économiques engendrant des troubles sociaux et sociétaux.

<http://goo.gl/pDLsaj>

LIMITES DES DONNÉES

Les données n'existent pas à l'état naturel. Elles sont constituées et recueillies par des moyens techniques ou humains avec un objectif déjà prédéfini qui comporte donc un biais. Ainsi, avant d'en faire une quelconque analyse, il faut déjà bien identifier le processus qui a conduit à leur création.

D'autre part, les données ne représentent jamais l'intégralité d'un phénomène car en plus des biais induits par leur constitution, elles peuvent nous amener à confondre causalité et corrélation et seul le bon sens nous permet d'y voir clair.

L'exemple type de cette limite peut s'illustrer lorsqu'on recherche l'explication à la montée parallèle des ventes de lunettes de soleil et de crèmes glacées. Aucune des deux n'explique l'autre, elles sont juste corrélées car c'est l'ensoleillement qui en est la cause.

Le problème est facile à résoudre lorsqu'on dispose de ces quelques paramètres mais si on manipule plusieurs dizaines de données différentes, il devient plus difficile de repérer ce qui relève de la causalité et de la corrélation.

3 ALGORITHMES CÉLÈBRES

GOOGLE

En perpétuelle évolution, l'algorithme du moteur de recherche détermine en fonction de plus de 200 critères l'ordre de la liste des sites répondant aux requêtes de chaque internaute.

NETFLIX

La plate-forme de vidéo à la demande fait tourner des algorithmes qui analysent le comportement de millions d'utilisateurs. La décision de produire la série House of Cards est issue de ce brainstorming logiciel.

WATSON

Le superordinateur d'IBM, vainqueur en 2011 du jeu Jeopardy!, intervient aussi dans le secteur de la santé, sur le cancer et le diabète, en partenariat avec des hôpitaux américains. Il assimile les données médicales d'un patient et l'ensemble de la documentation scientifique sur sa pathologie pour délivrer des options de traitement.

