

**Journée SITG 2013 : Investir le territoire numérique
15 Octobre 2013 – Fédération des Entreprises Romandes Genève**

Table ronde - Données publiques et économie : quelles plus-values ?

11h00 à 12h15

Modérateur : Daniel LOEFFLER, Promotion économique, Etat de Genève

Intervenants : Jean-Marc COMMENT – Archives fédérales
Christian FRISCH – Data Publica (France)
Thomas GAUTHIER – HES-SO Genève
David JONGLEZ – Esri France
Pascal BOYEAU – Here (France)

Introduction

M. Daniel LOEFFLER, Directeur du Service de la promotion économique de Genève, souhaite à tous la bienvenue à l'atelier intitulé « Données publiques et économie : quelles plus-values ? ». Il fait un bref rappel des caractéristiques de l'Open Data, notamment au niveau du service public.

Source d'une transparence accrue et d'un partage démocratique des données, l'Open Data va fournir, aussi bien aux citoyens qu'aux entreprises, des données pouvant être transformées en informations à forte valeur ajoutée. Cette table ronde a pour objectif de présenter quelques applications concrètes de l'Open Data, qui représente une sorte de plate-forme pour développer de nouvelles opportunités d'affaires et dynamiser l'économie.

Le SITG met à disposition des données sur plus de 650 couches d'information différentes, notamment sur les zones d'aménagement, les logements et les réseaux routiers. Les slides des présentations de la table ronde seront disponibles sur le site Internet de la manifestation.

M. Daniel LOEFFLER cède la parole au premier intervenant, M. Jean-Marc COMMENT, pour sa présentation sur le travail effectué dans le cadre du projet Open Government Data à la Confédération (OGD@Confédération), un portail pilote suisse visant à donner accès, de manière centralisée, à des données ouvertes de l'administration.

Présentation de M. Jean-Marc COMMENT, gestionnaire de l'intégration informatique aux Archives fédérales suisses et chef de projet du portail OGD@Confédération

Lancé le 16 septembre 2013 à Genève lors de l'Open Data Conference, le portail pilote suisse OGD@Confédération réunit actuellement l'Office fédéral de la statistique, swisstopo, MétéoSuisse, la Bibliothèque nationale et le canton de Zurich. Il a pour ambition de rassembler un maximum d'entités sur l'ensemble du territoire helvétique, non seulement à tous les niveaux administratifs, que ce soit la Confédération, les cantons ou les villes, mais également toutes les organisations privées, de manière à disposer d'un nombre toujours croissant de données.

Le projet OGD@Confédération vise à déterminer si l'Open Data peut augmenter les capacités de transparence d'une société et apporter un certain bien-être du point de vue économique. Thomas Jefferson, ancien président des États-Unis, a déclaré : « L'information est la monnaie de la démocratie. » L'information est une valeur, un soutien à l'innovation et aux consommateurs. Une bonne information permet en outre une économie de temps. Ainsi, les données ouvertes doivent être retravaillées et transformées en informations utiles à différents niveaux, pour l'économie et le citoyen.

Selon une étude de Deloitte faite en Grande-Bretagne, l'un des pays européens les plus avancés dans le domaine de l'Open Data, sur les 2,7 millions de téléchargements de données sur les trois sites principaux - l'Open Data du gouvernement anglais, l'office anglais de la statistique et l'open data de la ville de Londres -, les données les plus téléchargées concernent avant tout l'économie, puis, en deuxième position la démographie. Il a d'ailleurs été constaté une erreur dans la base de données du système de santé britannique, selon laquelle 20 000 hommes auraient été « enceints » en Grande-Bretagne. Il a également été remarqué que les applications pour Smartphone qui concernent les crimes intéressent beaucoup les Britanniques, par exemple les applications qui offrent des informations sur le lieu des crimes et l'intervention de la police.

Les transports viennent en troisième position des données les plus téléchargées en Grande-Bretagne. Dans le pays, 5 000 emplois informatiques auraient été créés dans le domaine des transports grâce à ces données publiques et aux applications développées à partir de celles-ci. À Londres par exemple, il existe un grand nombre d'applications, gratuites ou très bon marché, qui offrent une quantité remarquable de prestations. Cela donne une idée du modèle économique en développement dans le domaine de l'utilisation de ces données.

Au niveau de la Confédération, une étude a été menée par Mme Adélaïde BÜRGI-SCHMELZ, ancienne directrice de l'Office fédéral de la statistique et directrice des statistiques au FMI, pour déterminer ce que la mise à disposition à grande échelle de données publiques, principalement de la Confédération, pouvait rapporter à l'économie. Étant donné qu'il n'existe actuellement aucune méthode économique totalement fiable dans ce domaine, plusieurs méthodes ont été employées.

Selon les calculs réalisés avec la première méthode, basée sur des études européennes, un profit entre 1,2 et 6,5 milliards CHF serait dégagé pour l'économie suisse. La deuxième méthode a pris en compte le temps économisé si les travailleurs en Suisse disposaient de meilleures informations : la mise à disposition des données permettrait une heure d'économie par mois pour 4 millions de personnes travaillant à un coût horaire moyen de 26,70 CHF, ce qui correspond à 1,3 milliard CHF de profits supplémentaires en moyenne. Selon les résultats obtenus au moyen de diverses techniques de multiplicateurs économiques, le profit serait compris entre 0,6 et 1,6 milliard CHF.

Selon une enquête réalisée ensuite auprès de l'administration fédérale (l'étude pourrait ultérieurement être poussée au niveau des cantons et des communes), les coûts engendrés par la mise à disposition de ces données – par exemple la perte des recettes de Swisstopo qui vend actuellement ces données et mise en format plus pratique pour les programmeurs - se monteraient à environ 41 millions CHF sur une année.

D'un autre côté, la Confédération bénéficierait du développement de l'économie suisse engendré par la mise à disposition des données, avec des rentrées via la TVA et l'impôt fédéral direct (fourchette comprise entre 44 et 61 millions CHF). L'exercice serait certainement neutre pour la Confédération.

En plus de sa valeur économique, l'Open Data favoriserait la création d'emplois. En effet, selon des sources provenant de l'Office fédéral de la statistique, il est estimé que l'ouverture des données créerait dans le secteur informatique suisse entre 5 000 et 7 000 emplois, estimation faite en fonction de la part de l'informatique dans les emplois.

Le rapport conclut que la mise à disposition doit être poursuivie à condition que les pertes pour les administrations soient compensées par de nouvelles recettes. En septembre 2013, le Conseil fédéral a décidé de continuer à développer une stratégie Open Data pour la Suisse.

M. Daniel LOEFFLER remercie M. COMMENT pour sa présentation qui permet de constater que les produits économiques de l'Open Data sont importants, raison pour laquelle Genève s'y intéresse.

Il passe ensuite la parole à l'intervenant suivant, M. Christian FRISCH, de Data Publica (France), diplômé de l'École Polytechnique, titulaire d'un MBA de l'INSEAD. M. FRISCH a exercé chez iPIN, (San Francisco), Dell, Unilog Management et IBM Global Services.

* * * * *

Présentation de M. Christian FRISCH, co-fondateur en 2011 de Data Publica, société spécialisée dans la production de flux, de tableaux de bord et de visualisations personnalisées

La raison d'être de Data Publica, éditeur de jeux de données, est de valoriser l'Open Data et les données disponibles sur la Toile de manière générale. Plus concrètement, Data Publica identifie les sources de jeux de données spécifiés par le client, aussi bien de grandes entreprises comme Renault que de petites entreprises de quelques personnes, et rassemble les données des portails Web, les données internes de l'entreprise, les éventuelles données vendues par des éditeurs externes, ainsi que les données des réseaux sociaux. Puis, Data Publica fabrique des mécanismes de construction de jeux de données par toute une série de techniques d'assemblage et d'extraction d'informations intéressantes et livre au client ces jeux de données soit sous forme de données techniques qui seront introduites dans la base de données de l'entreprise pour que celle-ci puisse elle-même les valoriser ou créer de nouveaux services, soit sous forme d'une jolie visualisation Web présentant la valeur de ces données.

Voici trois cas concrets où l'Open Data rapporte une valeur économique importante pour les entreprises.

1) Tableaux de bord des territoires

Traditionnellement, les entreprises exploitent beaucoup de données externes - des données géographiques, économiques, de recensement, etc. -, la plupart du temps regroupées dans des tableaux Excel, et sur cette base établissent des graphiques et des rapports. Or, ceux-ci deviennent rapidement obsolètes puisque la plupart des données peuvent avoir changé.

Chez Data Publica, le processus est industrialisé. Les données sont prélevées sur les portails de façon régulière, tous les jours, et restructurées dans des bases de données. Les clients peuvent ainsi disposer de données toujours à jour sous forme de tableaux de bord. La société Data Publica possède beaucoup de données, qu'elle peut croiser à partir de sources multiples en respectant un certain nombre de contraintes, en termes de licences notamment.

Par exemple, Data Publica a créé pour l'association des notaires de France une application qui fournit toute une série d'informations, notamment des cartes, des prix, les villes principales,

les différents secteurs notariaux et environnements, des informations sur le parc provenant du ministère du développement durable, ainsi que des informations sur toutes les transactions immobilières des notaires en interne. Grâce à cette application, les notaires peuvent entre autres connaître les catégories socioprofessionnelles, l'âge et le sexe des acheteurs d'une zone donnée, ce qui leur permet d'adapter leur stratégie de communication.

2) Jeux de données sur les marchés publics

Data Publica a identifié un millier de sources Web qui référencent les marchés publics. En France, les marchés publics sont référencés par le Bulletin officiel d'annonces des marchés publics (BOAMP), dont l'équivalent européen est le Journal officiel de l'Union européenne (JOUE), mais chaque mairie peut également émettre des marchés publics et les publier sur son propre site Web. La société identifie à peu près 2 000 marchés par jour, au total environ 45 000 marchés actifs en France.

Data Publica examine les contenus de ces sites Web, ainsi que les documents associés aux marchés publics - très souvent des documents Word très mal formatés -, puis extrait et duplique les informations importantes, notamment le nom des contacts, les montants, les domaines d'activités et les types de marché, pour les fournir au client.

La société a référencé toutes les entités en France capables d'émettre des marchés publics - 170 000 ont été identifiés -, entre autres les mairies, les centres de recherche et les universités. En croisant tous les marchés avec ces entités publiques, Data Publica dispose d'une vision du marché public extrêmement précise et peut enrichir les marchés publics émis avec des informations sur les entités publiques, par exemple des contacts, et inversement enrichir les marchés publics avec des informations sur les entités publiques.

Grâce à cet outil, un client de Data Publica, une filiale du grand groupe médias français West France, a pu développer un moteur de recherches qui permet par exemple à un couvreur de déterminer quels sont les marchés publics accessibles dans sa région et d'obtenir toutes les informations associées, quand il en a besoin. Ainsi, les petites PME ont désormais accès à ce pan de l'économie de manière très simple, ce qui n'était pas le cas avant. Cet outil permet également d'avoir un observatoire très précis des marchés publics et une fiche d'identité sur chaque entité publique avec notamment les achats qu'elle fait.

3) Données sur les entreprises

Si les entreprises utilisent beaucoup les données comportementales, il est également intéressant d'exploiter les données des entreprises elles-mêmes. Dans tous les pays, il existe un registre public d'enregistrement des sociétés (dont l'accès est gratuit ou payant et la qualité bonne ou mauvaise selon les pays), qui contiennent généralement un numéro d'identification, un code d'activité de la société, souvent obsolète car l'activité a pu évoluer depuis sa création, ainsi que quelques informations sur ses dirigeants, habituellement les fondateurs, et éventuellement quelques informations financières.

Data Publica recherche les codes des entreprises, exploite toutes les informations que ces dernières communiquent, notamment les contacts, les technologies employées et l'utilisation ou non d'un moteur de e-commerce, et consulte les sites Web de manière systématique, afin de catégoriser les entreprises de manière beaucoup plus précise que les codes de style NAF (nomenclature d'activités française). En classant toutes les informations recueillies, Data

Publica détermine des groupes d'activités, d'entreprises, etc. et permet aux entreprises de mieux cibler leur prospect.

Par exemple, Data Publica a conçu pour l'annuaire d'entreprises Kompass une application qui analyse un très grand nombre d'entreprises de trois zones : la France avec toutes ses régions ; la France, l'Allemagne et l'Angleterre ; et l'ensemble de l'Europe économique. Cette application permet de connaître la répartition des entreprises par domaine d'activité et les lieux de fabrication de produits donnés. Data Publica a évalué environ 100 000 sites Web d'entreprises à travers l'Europe et a vérifié la présence de ces entreprises sur les réseaux sociaux tels que Twitter et Facebook, offrant une vue macroéconomique permettant une prospection ciblée.

Data Publica exploite l'Open Data au sens large, c'est-à-dire toutes les données disponibles aujourd'hui gratuitement, pour que les entreprises puissent avoir une connaissance accrue de leurs zones de chalandise ou du contexte socio-économique général de leurs zones d'activité, de manière à pouvoir adapter leur stratégie et éventuellement développer de nouvelles activités dans de nouveaux secteurs.

M. Daniel LOEFFLER remercie M. Christian FRISCH pour sa présentation très complète et passe la parole à M. Thomas GAUTHIER, Docteur en médecine expérimentale de l'Imperial College (Londres), diplômé du MIT et de l'École Supérieure de Physique et Chimie Industrielles (Paris), Ingénieur, directeur de recherche puis stratéliste chez Philips.

* * * * *

Présentation sur l'Open Data et les stratégies d'accès au marché de M. Thomas GAUTHIER, actuellement professeur de stratégie à la Haute École de Gestion (Genève)

Avec son équipe, à l'origine constituée d'un noyau d'ingénieurs logiciels et de mathématiciens appliqués, rejoints depuis par des designers apportant une réelle valeur ajoutée, Thomas GAUTHIER accompagne des sociétés de la région, du pays, de l'Europe et d'ailleurs, principalement pharmaceutiques et biotechnologiques, dans leur stratégie d'accès au marché en les aidant à valoriser des données ouvertes très spécialisées ou des données internes dont elles disposent.

Le secteur de la santé représente une part croissante dans l'économie, notamment en Suisse où environ 35 % des exportations sont liées de près ou de loin aux sociétés et aux industries des sciences de la vie, notamment les industries chimiques et pharmaceutiques. Les sociétés du secteur sont incitées à être plus analytiques dans leur stratégie d'accès au marché en raison principalement du nombre toujours croissant de parties prenantes, parmi lesquelles les patients, qui ont de plus en plus leur mot à dire sur les médicaments et les thérapies, les législateurs, les organismes payeurs, puisque les consommateurs ne sont pas forcément les payeurs, et enfin les leaders d'opinion, dont l'influence est considérable puisque le marketing direct des sociétés pharmaceutiques et biotechnologiques auprès des prescripteurs, des hôpitaux et des médecins n'est pas toujours efficace.

Pour identifier les leaders d'opinion et définir la topologie de leurs réseaux, ainsi que les types de liens qui les unissent, diverses données sont disponibles :

- 22 millions de résumés d'articles scientifiques de la littérature biomédicale mise à disposition par la bibliothèque nationale américaine de médecine, malgré quelques restrictions actuellement en raison du « shutdown » américain.

- 18 millions de documents légaux de l'OMPI, organisation basée à Genève, dans lesquels est décrite en détail la propriété intellectuelle globale.
- 140 000 essais cliniques dans plus 180 pays à travers le monde, recensés par ClinicalTrials. Un travail a été fait par les autorités américaines pour permettre que des tierces parties accèdent facilement à ces données.
- les réseaux sociaux. Il faut savoir que les médecins et les leaders d'opinion s'expriment de plus en plus sur Twitter, notamment au sujet de nouvelles thérapies avant la publication d'articles dans des journaux.

M. Thomas GAUTHIER présente l'étude de cas qui a occupé son équipe jour et nuit au cours des 3 derniers mois : une entreprise de petite taille du secteur pharmaceutique est prête à lancer une campagne marketing pour un nouveau traitement thérapeutique pour les patients souffrant d'une addiction alcoolique. Or, cette entreprise dispose d'une ressource terrain limitée et ne peut donc pas couvrir l'intégralité du territoire qu'elle souhaite aborder commercialement. En revanche, elle bénéficie d'un bon ancrage à Paris et de liens très forts avec les leaders d'opinion parisiens de l'addictologie et de la psychiatrie appliquée au traitement des addictions.

Pour son développement commercial, cette petite entreprise souhaite s'implanter en Bretagne, car des études socio-économiques y ont révélé une réelle opportunité de marché, mais ne sait pas quelle serait la meilleure option entre Nantes et Rennes.

En mobilisant les informations mentionnées plus haut, M. Thomas GAUTHIER et son équipe ont établi une carte de France, sur laquelle figurent des bulles de grosseurs différentes selon l'influence des leaders d'opinion dans chaque ville, l'influence étant une mesure composite intégrant plusieurs paramètres liés aux données disponibles, notamment le volume de publications, l'importance de ces publications, l'impact de celles-ci dans la communauté scientifique en calculant par exemple le nombre de fois qu'un article a été cité par d'autres chercheurs.

Cette première analyse a permis de déterminer avec précision quatre leaders d'opinion français, les plus importants à Paris dans le domaine de l'addictologie.

M. Thomas GAUTHIER présente un graphique permettant de comprendre comment les Bretons se désenclavent et collaborent avec des médecins chercheurs en dehors de leur région : chaque bulle représente un médecin chercheur, plus la bulle est grosse plus le médecin chercheur en question est influent. Les points bleu foncé sont des médecins chercheurs de Bretagne, les points bleu ciel sont les médecins chercheurs qui appartiennent au premier cercle de collaborateurs, les points blancs sont les médecins chercheurs qui appartiennent au deuxième cercle de collaborateurs. On remarque que certains groupes de leaders d'opinion sont autocentrés, c'est le cas par exemple du département de psychiatrie du CHU de Brest, dont les médecins chercheurs n'ont pas forcément d'appétence pour les collaborations hors de Bretagne.

Sur le graphique de la topologie des leaders d'opinion à Nantes, aucun des quatre leaders d'opinion parisiens cités précédemment n'apparaît, ni au premier, ni au second degré de collaboration. Ce qui révèle que, quand bien même l'entreprise est bien implantée à Paris, maîtrise parfaitement bien ses réseaux et ses leaders d'opinion à Paris, sa capacité à mobiliser ce réseau parisien pour s'implanter sur le marché nantais risque d'être limitée.

En revanche, sur la base du graphique de la topologie des leaders d'opinion de Rennes, on peut constater que l'influence des quatre leaders d'opinion parisiens y est importante. Cela

indique que vis-à-vis du marché rennais la société biopharmaceutique en question sera certainement en mesure de mobiliser sa bonne image et sa collaboration déjà en place avec les leaders d'opinion parisiens pour pénétrer le marché de manière plus efficace.

Ainsi, cette petite entreprise peut faire son choix stratégique très important, non seulement sur l'intuition, mais également grâce à des informations sur l'influence des leaders d'opinion, définie à partir de données totalement ouvertes.

Ce type d'outil pourrait également avoir une application, non économique mais qui mérite d'être mentionnée, dans l'analyse forensique des conflits d'intérêt, nombreux dans les industries pharmaceutique et biotechnologique.

M. Daniel LOEFFLER remercie M. Thomas GAUTHIER pour sa présentation très complète et donne la parole à M. David JONGLEZ, Ingénieur INSA Lyon, co-fondateur de Campocamp (2001-2009), directeur du centre d'excellence GIS Capgemini (2010-2013), responsable du business development chez Esri France. Membre des groupes OGC / OpenData (Afigéo) et Etalab, M. Jonglez travaille au montage d'OpenGeoData.fr pour Esri France.

* * * * *

Présentation sur la manière de faciliter l'accès aux données de M. David JONGLEZ, responsable du business development chez Esri France

La stratégie d'Esri a connu ce que Jack Dangermond en mars 2013 a qualifié de *turning point*, un « changement de direction », ou « double direction », c'est-à-dire qu'en parallèle de son marché historique, Esri développe aujourd'hui une stratégie de plates-formes pour gagner de nouveaux marchés en diffusant plus largement et en rendant plus accessibles et plus facilement appréhendables les technologies géospatiales, qui jusqu'alors étaient essentiellement utilisées par les spécialistes. Dans son ouvrage intitulé *The Age of the Platform* (2011), Phil Simon compare les stratégies et les modèles économiques des grands éditeurs des années 60-90 à ceux des années 2000-2010. La première période se caractérise par des écosystèmes relativement fermés, ainsi que par des *partnerships* assez étroits, directifs et concentrés. De nos jours au contraire, avec l'apparition de stratégies de plates-formes, une ouverture maximale est adoptée non seulement sur le plan technologique mais également sur le plan partenarial pour offrir un maximum de valeur aux écosystèmes multiples et aux utilisateurs.

Les plates-formes, ce sont des services accessibles en ligne via des API, du contenu et une large communauté d'utilisateurs (la conférence annuelle d'Esri France a réuni sur place près de 2 500 personnes (plus de 15 000 aux US pour la conférence mondiale annuelle) sur deux jours pour collaborer, échanger et présenter un certain nombre de choses à la communauté). Les stratégies de plates-formes représentent un enjeu important pour Esri, qui cherche à amener les développeurs à exploiter cette technologie dans des usages non géocentriques mais qui vont enrichir leurs apps ou leurs usages à travers une dimension géospatiale qui était inexistante ou difficilement appréhendable jusqu'alors. Cette approche d'Esri s'inscrit dans la volonté de rendre accessibles ces usages, via des apps ou des contenus. Il existe sur le marché de grands ténors de cette stratégie de plates-formes, notamment Facebook dans le domaine du social, Amazon dans le e-commerce, etc. Esri entend être, à travers *arcgis.com*, la plate-forme géo mondiale.

Le contenu est bien sûr l'essence des applications. Esri adopte une stratégie Open Data essentiellement indirecte dans la mesure où elle ne génère pas de revenus directs avec les données ouvertes mais vise à rediffuser et à rendre accessible plus facilement l'Open Data local, national ou global auprès de ses utilisateurs, afin de leur permettre d'être créatifs et de développer un maximum d'usages à travers une plate-forme technologique Web, arcgis.com. Ce portail comprend une partie interactive et collaborative, enrichie de nombreux contenus par l'écosystème de partenaires et d'utilisateurs.

Pour Esri, l'Open Data est à la fois une réelle opportunité et d'une certaine façon une menace car la très probable libération de grands jeux de données géographiques français se traduira à moyen terme par une accessibilité plus simple et plus ouverte des données de l'IGN français, ce qui impactera l'activité commerciale d'Esri sur sa vente de contenus, l'obligeant ainsi à adapter son activité. Par exemple, pour la base de données FranceRaster produit et coédité par Esri France et l'IGN, la société rasterise et distribue les fonds de carte rasters générés à partir des données IGN. L'activité de contenus d'Esri France représente 12 % de son chiffre d'affaires en France, soit environ 3 millions d'euros, ce qui fait vivre une équipe de 15 personnes.

Aujourd'hui, la stratégie Open Data d'Esri France comprend 3 axes :

- Rendre accessible les données Open Data qui manquent à l'heure actuelle encore de visibilité. En effet, même s'il existe des portails qui cataloguent les données ouvertes en France, notamment celui de l'État, il est encore problématique d'identifier, de qualifier, voire d'harmoniser ces jeux de données libérés de manière relativement éparse sur le territoire, aux échelles locale, départementale, régionale et nationale. Il y a environ un an, Esri a créé un portail expérimental qui sera industrialisé avant fin 2013. Sur ce portail basé sur la technologie Portal for ArcGIS, les jeux de données sont qualifiés, enrichis, et formatés, avant d'être rendus directement accessibles via l'ensemble des outils de la plate-forme auprès de tous ceux qui souhaitent les utiliser, conformément à la licence de données ouvertes sous-jacente à chacun de ces jeux de données.
- Se rapprocher de la communauté OpenStreetMap, extrêmement active dans le domaine de l'Open Data, avec laquelle Esri collabore depuis des années au niveau global et souhaite créer des synergies. En France, Esri cherche à déterminer à quel niveau les demandes de la donnée OpenStreetMap et des clients de la société se chevauchent afin d'identifier les besoins réels en termes d'usages. Avec le programme communautaire *acrOpole*, spécifique à Esri France, des services, des usages, des applications, des templates et de la donnée, spécifiques à des usages verticaux, sont créés en collaboration avec les grandes collectivités françaises. Un des premiers usages qui a récemment été créés avec OpenStreetMap concerne l'accessibilité pour les personnes à mobilité réduite. Il faut savoir qu'en France, le contexte législatif est extrêmement contraignant dans ce domaine pour les collectivités, qui se doivent de géoréférencer l'ensemble des objets géographiques liés à l'accessibilité. Esri a décliné un certain nombre de verticaux qui lui semblaient intéressants de développer entre OpenStreetMap et les collectivités en fonction des besoins de ces deux mondes. OpenStreetMap, ce n'est pas seulement de la donnée Open Data, mais une plate-forme, des API d'accès aux données, du contenu, des utilisateurs - environ 1 350 000 -, des développeurs, une marketplace basique et des outils utilisables. Dans le cadre de sa stratégie de plates-formes, Esri souhaite collaborer avec d'autres plates-formes, venir s'en enrichir et créer des synergies bijectives avec celles-ci.

- Collaborer avec les développeurs Open Data. Pour Esri, l'Open Data représente un nuage de plates-formes ou une seule et grande plate-forme, une sorte de méta-plate-forme mondiale avec laquelle l'entreprise souhaite collaborer. À travers le programme start-up et les concours Open Data, Esri entend apporter un maximum de valeur aux usagers des plates-formes pour leur permettre de développer de nouveaux usages sur la base de contenus ouverts dans le cadre des mouvements Open Data.

Esri a instancié un chantier au niveau de la production de données ouvertes mais s'attend à devoir faire face à certaines restrictions. Par exemple, l'utilisation d'OpenStreetMap, sous licences ouvertes par nature, nécessite un partage à l'identique (share alike). Ainsi, un certain nombre de problématiques existent mais celles-ci peuvent être surmontées en entamant une vraie réflexion sur les modèles économiques sous-jacents.

M. Daniel LOEFFLER remercie M. David JONGLEZ pour sa présentation et rappelle que les slides seront disponibles sur le site Internet. Il cède ensuite la parole à M. Pascal BOYEAU, Here (France) Ingénieur « Arts & Métiers » ICAM, Spécialiste des applications de haute technologie : consultant en CAO Mécanique et Gestion de données (Hewlett Packard) ; stations graphiques pour Réalité Virtuelle, Digital Média et CAO (Intergraph) ; gestion de données et Workflow (Motiva)

* * * * *

Présentation sur la création de données fixes et dynamiques de M. Pascal BOYEAU, spécialiste des solutions cartographiques / géolocalisation en Europe du Sud (HERE)

Navteq et son concurrent Tele Atlas sont les deux grandes sociétés de la cartographie numérique. La société Navteq a été fondée en 1985 pour travailler sur les premiers systèmes de navigation automobile (à l'époque, les cartes en papier étaient scannées). L'IGN, sous-traitant de Navteq en France, a créé les premières cartes vectorielles. En 2008, NOKIA a racheté Navteq, une des raisons étant que la cartographie est de plus en plus présente sur les téléphones portables. Entre deux, Navteq a racheté une société active dans le domaine du *digital mapping* qui fait notamment en France des photos de bâtiments pour les pages jaunes. Cette société fournit des voitures extrêmement automatisées avec caméras et e-dar qui permettent de cartographier les rues en 3D. Navteq a aujourd'hui été rebaptisé HERE.

HERE crée, achète - quand elles existent - et teste des données aussi bien statiques que dynamiques, notamment des rues, des adresses fixes et ponctuelles, des autoroutes, de l'info trafic, des bouchons, le prix du carburant et le prix des transports publics. La mobilité et les transports publics représentent aujourd'hui 50 % du travail des cartographes avec entre autres les informations concernant les bus, les tramways, les RER et les trains, la position des Vélib', etc. Ces cartographes passent beaucoup de temps à négocier avec les villes et les collectivités locales pour obtenir les données relatives à la mobilité.

HERE vend ses données à des sociétés comme Esri, qui disposent de leur propre plate-forme. Depuis plusieurs années, poussé par la concurrence de Google, HERE conçoit également des plates-formes qui permettent aux clients de développer en utilisant le langage de programmation JavaScript et sans rien devoir installer chez eux des applications, par exemple des cartes, le positionnement à l'intérieur d'un bâtiment, du géocodage, du calcul d'itinéraire personnalisé, du guidage et de l'info trafic. Cela permet à la société HERE d'adresser les marchés de systèmes SID et de systèmes d'exploitation pour appareils mobiles comme Android, iOS, Windows Phone, etc.

La société HERE occupe 90 % du marché mondial des cartes numériques de navigation automobile (9 voitures sur 10 sont dotées de cartes Navteq ou HERE). Lors du dernier Salon de l'automobile de Francfort, HERE a présenté sa plate-forme automobile, la seule solution d'aide à la conduite complète disponible sur le marché mondial. À l'avenir, la voiture sera de plus en plus « connectée » et pour cela, des plates-formes, semblables à celles utilisées pour les villes, doivent être développées. Sur les quatre marchés qu'adresse HERE - automobile, mobile, Web et entreprises/solutions professionnelles - les trois premiers représentent 95 % de son activité, ce qui montre que les données sont avant tout utilisées par le grand public.

Voici quelques types de données extrêmement prisés par les fournisseurs du privé : la position des feux, des ronds-points, des panneaux danger ; la transcription phonétique de tous les noms de rue (même les plus difficiles à prononcer doivent être codés de façon phonétique pour les systèmes de reconnaissance vocale) ; la hauteur des ponts ; les voies dans les parcs qui permettent aux piétons de passer et de prendre des raccourcis ; les données « indoor », c'est-à-dire de l'intérieur des bâtiments, notamment les gares, les aéroports et les centres commerciaux ; les données de navigation avec les points d'intérêt tels que les restaurants et les hôtels ; le prix des carburants ; le nombre de places disponibles dans les parkings ; les données visuelles (bâtiments 3D, symboles qui permettent par exemple de reconnaître le musée du Louvre sur une carte, etc.) ; les données ADA (Advanced Driver Assistance System), très importantes aujourd'hui pour HERE, utilisées dans les solutions professionnelles pour choisir les itinéraires qui consomment le moins d'énergie pour les véhicules électriques (celles-ci peuvent notamment adapter leur vitesse en fonction du relief grâce aux données sur les pentes) ; les données trafic (bouchons soit historiques, c'est-à-dire quart d'heure par quart d'heure, soit statistiques permettant de savoir s'il y a en général des bouchons à un endroit donné, à une heure donnée) ; les adresses notamment ponctuelles et les polygones de codes postaux.

Société internationale, HERE emploie environ 7 000 personnes, dans 80 pays, dont la Suisse avec une équipe de 7 cartographes issus de l'entreprise EGT, fondée en 1996. Ces cartographes collectent des données et négocient des contrats avec les collectivités locales. La société dispose par ailleurs de véhicules True Car, qui ont commencé à sillonner la Suisse pour récupérer des images 3D grâce à des systèmes de caméra vidéo extrêmement sophistiqués (e-dar 600 000 ou 1 300 000 points en fonction des technologies). Cette équipe suisse de cartographes/chefs de projet est aidée par des équipes de sous-traitants notamment en Inde et au Mexique. Des équipes similaires sont présentes dans le monde entier.

Sur les cartes, chaque rue comporte environ 200 attributs tels que les sens interdits, les bandes continues et autres informations non-disponibles sur les images satellite. Pour la Suisse : environ 127 000 km de route, 107 millions de points d'intérêt, 1,6 million d'adresses ponctuelles - point important de HERE pour pouvoir effectuer un géocodage d'excellente qualité. Avec le système, les utilisateurs peuvent en outre calculer un itinéraire par exemple entre Genève et Brest en empruntant les divers transports publics (pour cela, HERE doit fournir des données sans couture au niveau des pays, avec toutes les données transports. Par ailleurs, l'outil Natural Guidance est de plus en plus utilisé. Il vous donne des indications de navigation plus précises, par exemple « tournez à gauche à l'église » au lieu de « tourner à gauche dans 100 mètres ».

HERE dispose d'un système d'assurance qualité extrêmement poussé et n'accepte ni n'achète aucune donnée, que ce soit de l'Open Data, des journaux, des sources Internet, des contacts directs avec les mairies et les collectivités locales ou des images aériennes, avant de l'avoir testée et à condition que la qualité soit jugée supérieure à 97-98 %.

L'usage commercial de certaines données Open Data n'étant pas autorisée, HERE utilise uniquement celles que l'entreprise a le droit de vendre et de modifier et ce, ad vitam aeternam. Certaines données de HERE sont proposées gratuitement et en France, les cartographes de la société échangent des données gratuites des collectivités locales contre par exemple le droit de celles-ci d'utiliser l'affichage info trafic de l'entreprise.

SESSION DE QUESTIONS-RÉPONSES

Est-ce que Data Publica utilise opencorporates.com ?

M. FRISCH : Étant donné que ce site ne donne pas d'information sur les entreprises françaises, nous l'utilisons uniquement pour les entreprises anglaises. Nous nous en sommes servis notamment pour un client qui souhaitait connaître parmi 10 000 entreprises lesquelles exportaient et vers quels pays.

Est-ce que vous arrivez à recueillir également des données sur les petites entreprises ?

M. FRISCH : Nous nous intéressons principalement aux entreprises qui disposent d'un site Web. Parmi les 8 millions environ d'entreprises enregistrées en France, 1 million d'entre elles comptent plus de 2 personnes et ont été actives au cours des trois dernières années. Parmi ce million d'entreprises, environ 47 % possèdent un site Web.

Le manque de stabilité de la mise à disposition des données, comme dans le cas du shutdown américain actuel, ne met-il pas en péril les entreprises qui s'appuient sur ces données ? Est-ce que ces questions sont abordées avec les administrations ?

M. GAUTHIER : Même pendant le shutdown gouvernement fédéral américain, les données du site PubMed sont toujours disponibles - les administrations américaines qui mettent ces données publiques à disposition garantissent toujours un accès - mais ne sont pas mises à jour. En termes de risques des sociétés à but commercial qui basent une partie de leurs prestations sur ces données, il semblerait que les administrations publiques américaines aient pris des mesures pour ne pas les pénaliser en cas de shutdown.

M. COMMENT : La pérennité des données est l'un des trois grands problèmes concernant l'Open Data, les deux autres étant les licences permettant à une entreprise de faire du commerce et les formats des données. En ce qui nous concerne, nous utilisons des programmes qui parcourent les sites et vérifient si les données sont toujours disponibles. En cas d'absence de données, les fournisseurs en sont informés. J'estime que ce sont ces derniers qui doivent garantir la pérennité des données qu'ils proposent au public.

M. JONGLEZ : Il ne faut pas que la question de la pérennité soit un frein à l'ouverture des données. Cette question concerne également les éditeurs privés, car lorsque des entreprises ferment, des données disparaissent. Il est préférable de disposer de données, même si la pérennité de leur disponibilité n'est pas garantie, que de pas en disposer du tout. Dans la situation actuelle, les données sont encore insuffisantes.

M. GAUTHIER : Contrairement au site PubMed, data.gov s'est arrêté à cause du shutdown. Cela a relancé un vieux débat concernant le format de mise à disposition. Le gouvernement américain a beaucoup investi dans les API, qui sont des flux, et comme tous les flux, quand le robinet est coupé, plus rien ne coule. À l'inverse, les fichiers disponibles en téléchargement restent accessibles en cas de shutdown. D'ailleurs, la Sunlight Foundation a récemment critiqué cette stratégie d'API qu'elle considère non pérenne. Je pense que le shutdown offre une belle opportunité de réellement comprendre les conséquences de ce genre de situations.

M. JONGLEZ : Je suis convaincu que lorsque ce genre de cas se présente, l'intelligence collective peut apporter des solutions. Dès lors qu'un problème est survenu, il n'est plus un problème car il est solutionné. Les communautés réagissent très rapidement, en quelques jours. De plus, des associations peuvent s'organiser pour pallier aux défaillances potentielles de serveurs ou d'institutions.

Quelles sont les exigences de la société HERE en termes de mise à jour des données ?

M. BOYEAU : Nous avons 40 000 sources au niveau mondial, c'est-à-dire 40 000 contrats. Pour chacun d'entre eux, les mises à jour sont systématiquement vérifiées. Nous effectuons les mises à jour tous les trois mois dans le monde entier. Sur notre plate-forme en ligne, les mises à jour sont désormais mensuelles. D'ici quelques mois, elles seront hebdomadaires. La carte du système de navigation d'une voiture, même neuve, date d'au moins un an et demi et devient donc rapidement obsolète. Avec les véhicules connectés, les mises à jour sont effectuées chaque semaine. HERE considère la mise à jour comme essentielle et la société signe rarement un contrat si elle ne dispose pas de mise à jour. L'entreprise préfère dépenser des fonds pour avoir des mises à jour, plutôt que de prendre des données gratuites sans mises à jour. Dans certains cas spéciaux, HERE finance l'entité en échange de mises à jour.

Comment la Confédération peut-elle s'assurer que les profits potentiels de l'ouverture des données reviennent bien aux producteurs des cartes ?

M. COMMENT : Swisstopo gagne de l'argent en vendant les données qu'il produit. Si le Conseil fédéral déclare la gratuité de ces données, il faudra que le budget fédéral rémunère les administrations qui offrent ce service. J'ai constaté qu'une des peurs de l'administration à l'ouverture des données est que celle-ci entraîne une perte de compétences et d'informations. Il faut néanmoins savoir que la majorité des offices fédéraux, notamment les Archives fédérales, la Bibliothèque nationale et l'Office fédéral de la statistique, ne gagnent pas d'argent avec leurs données. Je considère personnellement cela normal dans la mesure où ces données sont créées grâce aux impôts. De plus, par rapport au budget de la Confédération, 40 millions ne représente pas une somme très importante.

En conclusion, M. LOEFFLER espère que cet atelier a donné envie de développer de nouvelles idées et de créer de nouvelles opportunités d'affaires, car nous ne sommes pas là pour deviner le futur mais pour le faire.