

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

Table ronde - Données publiques : quelles ouvertures ?

14h30-15h50

Modérateur M. Giorgio PAULETTO (Observatoire technologique, Etat de Genève)

Invités : M. Simon CHIGNARD  
M. Antoine LOGEAN  
M. Antoine STROH

\*\*\*\*\*

**Introduction**

M. PAULETTO introduit le sujet en posant plusieurs questions :

- *Qu'arrive t-il une fois que les données sont ouvertes ?*
- *Quelle utilisation peut-on en faire ?*
- *De quelle façon sont-elles réappropriées par la société ? Avec quelles interfaces ?*
- *Comment crée-t-on de la valeur à partir de ces données ?*

M. PAULETTO présente ensuite les intervenants :

Simon Chignard est l'auteur de l'ouvrage *L'open data, comprendre l'ouverture des données publiques* (éditions Fyp, 2012). Il est consultant indépendant, a participé activement depuis 2010 à l'ouverture des données sur le territoire de Rennes. Depuis cette année, il est également rattaché au réseau d'experts Etalab, le Service du Premier Ministre chargé de l'ouverture des données publiques et du développement de la plateforme française Open Data ([www.data.gouv.fr](http://www.data.gouv.fr)).

Antoine Logean travaille actuellement pour SwissRe en tant qu'expert en modélisation et en traitement de données (*data scientist*), il est docteur en bioinformatique et spécialiste du web sémantique. Il fait partie de l'association Opendata.ch dans le cadre de laquelle il explique, favorise et essaie de faire évoluer les administrations publiques et le secteur privé dans la mise en œuvre de l'ouverture des données publiques en Suisse.

Antoine Stroh est diplômé de l'Ecole Centrale de Lille et travaille aujourd'hui pour les Transports publics genevois (TPG). Il a participé depuis une dizaine d'années à des applications informatiques dans le domaine du transport. Au sein des TPG, il a été pendant plusieurs années chef de projet en charge du Système d'aide à l'exploitation et à l'information voyageurs (SAEIV). Depuis plus récemment, il est responsable des données de production et en charge du programme Open data des TPG.

## Présentations des intervenants

M. CHIGNARD répond d'abord à la première question par M. Pauletto en disant que la plupart du temps, une fois que les données sont ouvertes, il ne se passe rien de façon spontanée, il est donc nécessaire d'encourager la réutilisation. A l'aide du 1<sup>er</sup> schéma (que l'on peut retrouver sur son blog [www.donneesouvertes.info](http://www.donneesouvertes.info)), il présente la première catégorie d'actions d'animation : le hackathon, un marathon du développement où des développeurs sont placés dans une même salle pendant un certain temps afin de faire émerger des idées sur les données ouvertes. L'objectif est de réaliser des applications mobiles, des services et des médiations. La deuxième catégorie consiste en des appels à idées et à projets afin d'imaginer des réutilisations.

M. Chignard donne l'exemple de la filiale Ile-de-France de la SNCF, Transilien, qui a mis en place un concours d'idées sur Internet et a récolté 2'000 idées de services et d'applications. Parmi elles, l'idée d'application « C'est pas MaFaute » - qui n'a pas été retenue par la SNCF - repérait automatiquement la localisation de quelqu'un se trouvant dans un RER en retard et envoyait un bulletin de retard à son employeur ou à son université. Une autre application, Tranquilien, retenue elle par la SNCF, permet d'avoir une idée du taux de remplissage des rames de métro. Le véritable enjeu est de former un public plus large sur tous types de données, qu'elles soient collaboratives ou personnelles, pour faciliter leur appropriation et comprendre la culture de la donnée.

M. Chignard travaille actuellement sur le concept d'*Infolab*, un lieu où l'on réaliserait des projets à partir de données. *Infolab* est une reprise des Fab Labs, des laboratoires de fabrication avec des imprimantes 3D permettant de réaliser des projets. A l'aide du 2<sup>ème</sup> schéma, il explique que ces actions d'animation peuvent correspondre à des temps et à des publics différents, en fonction des réponses données à 5 questions :

1. Pourquoi ? Pour imaginer, réaliser, former ou sensibiliser ?
2. Comment ? Par des hackathons, des concours, des appels à projet ou des *Info Labs* ?
3. Pour qui ? Pour des développeurs, des journalistes, des acteurs économiques, des étudiants, le grand public ?
4. Quand ? En amont du lancement du portail par exemple.
5. Par qui ? L'animation, celui qui ouvre les données, des tiers, des cantines numériques, des collectifs Open data, des agences numériques ?

Après avoir répondu à ces cinq questions, M. Chignard estime qu'il est possible d'imaginer un millier de dispositifs d'animation. Le message clé qu'il veut faire passer est qu'il n'existe pas de réutilisation sans animation. Aujourd'hui, il existe une mauvaise visibilité de la réutilisation des données effectuée. En effet, il est possible de voir qui a participé à un concours de développement d'applications, mais pas qui a réalisé des applications sans avoir participé à un concours, ni qui a fait une visualisation sur un site web, ni quelles entreprises réutilisent ces données. La RATP (Régie autonome des transports parisiens) a publié les statistiques du nombre de voyageurs entrant dans les stations de métro, et le quotidien *20 minutes* a réutilisé ces statistiques afin de positionner des distributeurs de journaux aux sorties de ces stations.

Selon M. Chignard, si le journal n'avait pas relaté cette réutilisation, elle serait restée inconnue, et cela est lié aux deux principes fondateurs de l'Open data : « *Don't ask, Don't tell* », autrement dit, il n'y a pas de contrôle à priori des données et leur utilisation n'est pas divulguée. Idéalement, il faudrait garder le principe « *Don't ask* », en mettant simplement la donnée à disposition, et passer au « *Please tell* » afin d'avoir connaissance de l'utilisation des données faite par les différents acteurs.

\* \* \* \* \*

M. LOGEAN présente l'association suisse Opendata.ch dont il fait partie et où chacun a des compétences spécifiques : s'occuper de la partie juridique, médiatique, ou encore coder de l'Open data. A l'origine, c'est un projet qui vient de Zurich et de Berne. Giorgio Pauletto et lui relient la partie suisse-allemande à la partie francophone, notamment par des traductions. L'association comporte 78 membres ainsi que des entreprises affiliées, et depuis 2 ans, elle représente en Suisse l'*Open Knowledge Foundation*, une fondation qui vise à promouvoir l'Open data à une échelle internationale. Leur but est de rendre les données publiques accessibles et réutilisables et par ce biais favoriser l'innovation, la transparence et l'efficacité. Pour cela, l'association agit à 3 niveaux différents : une Conférence annuelle regroupe les acteurs de l'Open data en Suisse et s'adresse aux politiques et aux entreprises. 3 conférences ont été données depuis 2011. Elles se déroulent en parallèle en Suisse romande et en Suisse allemande. L'association s'adresse également à la communauté avec des hackathons bisannuels qui consistent à placer dans un même espace un panel de personnes avec des compétences variées afin de faire un *brainstorming* autour d'une thématique comme la mobilité ou la santé. Les personnes échangent leurs idées et sont ensuite regroupées en équipes autour de ces différentes idées.

Le résultat de ces hackathons est impressionnant. Bien que les développeurs soient très productifs, il est nécessaire de leur donner des consignes, alors qu'une personne qui travaille dans un Office de Tourisme par exemple aura spontanément des idées pour développer ses services à partir de sites comme Twitter ou Flickr. Enfin, l'association fait du lobbying en aidant les politiques à trouver la façon dont ils vont ouvrir les données. Plusieurs interventions ont déjà été faites au Parlement et le portail [opendata.admin.ch/](http://opendata.admin.ch/) existe en partie grâce à l'association.

M. Logean résume les actions passées par une chronologie : en 2011 a eu lieu la première Conférence à Berne, puis les premiers hackathons. Ensuite, l'association a été fondée et une deuxième conférence a eu lieu à Zurich. Suite à cela, la Ville de Zurich a lancé son portail Open data. Enfin, une Conférence internationale a eu lieu à Genève avec des personnes venant des Etats-Unis et d'Angleterre. Toutes les informations se trouvent sur [fr.opendata.ch](http://fr.opendata.ch), qui est également le site utilisé pour l'organisation des hackathons.

\* \* \* \* \*

M. STROH explique avoir commencé à travailler en 2006 pour l'entreprise publique TPG sur le projet SAEIV, dont la mission principale est de réguler les véhicules depuis un centre de régulation, comparable à une tour de contrôle. Il s'est aperçu en travaillant sur ce système qu'une base de données stockait en temps réel la position

des véhicules, mais qu'aucune utilisation n'en était faite. En 2009, la thématique de l'information voyageurs est devenue une vraie mission d'entreprise pour les TPG. Il a donc été consulté le service marketing en leur proposant d'utiliser cette base de données, et à partir de cela, une application iPhone et un site web ont été créés en juin 2011. Une application Android est également sortie plus tard. A l'époque, la question de l'Open data n'était pas encore évoquée. M. Stroh a seulement commencé à s'y intéresser après avoir consulté le site de Simon Chignard ainsi qu'une vidéo sur le potentiel d'utilisation des données. Suite à cela, il a lancé un projet, a consulté les Directions Informatique, Marketing, Communication et Juridique puis a confié un mandat à Simon Chignard afin de convaincre définitivement ces interlocuteurs. La mise en conformité avec la LIPAD en ce qui concerne l'aspect transparence était également un levier du lancement projet. Le projet a été validé en octobre 2012 et a obtenu un budget.

En novembre 2012, suite à une demande de Lift, les TPG ont participé au concours mondial « Urban Data Challenge » sur la mobilité avec les villes de San Francisco et de Zurich ([www.urbandatachallenge.org](http://www.urbandatachallenge.org)). Dans le cadre de ce concours, les tpg ont ouvert plus d'1 million de données sur les horaires de passages des véhicules et les montées de passagers pour la première semaine d'octobre 2012. En septembre 2013, les données des prochains départs en temps réel ont été ouvertes et en 3 semaines. Les TPG ont eu 100 demandes de données (clés d'API). Ils sont les premiers transporteurs suisses à avoir ouvert des données en temps réel et ont eu beaucoup de reprises médiatiques. Avec l'ouverture des données de perturbations tpg (accidents, charge de trafic automobile), les passagers peuvent désormais connaître les causes de retard des véhicules, et il s'est avéré que dans 50 % des cas, ce n'est pas du fait des TPG. Les données géographiques des 2'500 arrêts ont également été ouvertes. Au niveau marketing, cela a permis à l'entreprise de se rapprocher de ses clients et de contribuer à une meilleure image en plus des services clientèles existants telles que le service client, le site web, les réclamations, Twitter, Facebook. Le site Open data est [data.tpg.ch](http://data.tpg.ch), des ateliers ont été organisés et un concours lancé. La qualité de ces données peut également être améliorée grâce aux commentaires des utilisateurs sur le forum open data. Enfin, l'Open data a permis des échanges avec des entreprises de transport à Zurich et à San Francisco qui inspirent actuellement les TPG pour améliorer, par les données, le suivi de la performance de leurs lignes.

## Questions

*M. Boris Engelson* fait remarquer que s'il demandait aux TPG le classement comparé d'encombrement des véhicules par canton, il est probable que l'entreprise n'ait pas cette donnée ou ne veule pas la donner, car à Genève, les véhicules sont beaucoup plus encombrés qu'à Bâle ou à Zurich par exemple. Pour faire ce type de classement, l'entreprise doit en avoir envie et accepter que cela implique une baisse de tarifs. Il ne faut pas simplement ouvrir la donnée mais se poser les bonnes questions. Il aimerait bien consulter les données sur les appels d'offres, car il est toujours surpris de voir les frais d'investissement et le résultat souvent décevant du matériel. Il relève également qu'il n'y a jamais l'échelle sur les plans des TPG, et même le Directeur général en Conférence de presse n'a pas pu la lui donner. Pour lui, l'Open data est une imposture.

M. STROH répond en disant que concernant l'encombrement des véhicules et les données en temps réel du passage des véhicules de la première semaine d'octobre 2012 sont consultables en ligne dans le cadre des données ouvertes pour l' « Urban Data Challenge ».

M. LOGEAN ajoute qu'il ne peut qu'être d'accord avec M. Engelson. Ce n'est pas la première fois que l'on fait remarquer que les vraies données intéressantes ne sont pas ouvertes, mais il est d'avis qu'il faut bien commencer quelque part, et que le simple fait d'en parler aujourd'hui est déjà un progrès. L'ouverture des données ne peut pas se faire d'un coup, il faut laisser aux administrations le temps de se structurer. Il encourage les personnes qui font ce genre de remarques, mais explique qu'eux doivent être politiquement corrects et conciliants. Suite à l'exemple donné par M. Engelson de l'interdiction des données ethniques sur la criminalité en France, il répond qu'en effet, ils ne peuvent pas ouvrir ce type de données délicates en premier, au risque de tuer le mouvement en œuvre.

M. CHIGNARD ne pense pas que l'Open data soit l'outil de la transparence, mais plutôt le fruit de la culture politique anglo-saxonne. Que ce soit les entreprises, la collectivité ou l'Etat, aujourd'hui, chacun choisit ce qu'il publie. Il donne l'exemple de Data Publica qui a publié les données de la RATP sur le nombre d'accidents dans le métro parisien et qui ne sont pas des données que l'on va retrouver sur le portail de la RATP, et celui de l'association Regards Citoyens qui apporte une vision complémentaire par le site [nosdéputés.fr](http://nosdéputés.fr), sur lequel on ne trouve pas les mêmes informations que sur le site officiel de l'Assemblée Nationale.

M. Jonglez fait remarquer que si l'on ouvre les données, il faut une adaptation des organisations et des institutions publiques afin de fluidifier, rechercher les sources des données et les transformations qui ont été faites. Pour arriver à une logique d'Open data industriel, l'accent doit être mis sur la transparence du gouvernement.

M. CHIGNARD complète en donnant l'exemple du site de la Bibliothèque Nationale de France ([data.bnf.fr](http://data.bnf.fr)) qui a publié des données sur les différents ouvrages. Des personnes les ont réutilisées afin de voir quel type de livre est traduit dans quelle langue cible. Le Ministère de la culture et le Ministère des affaires étrangères avaient déjà évoqué ce projet là, et pour communiquer entre deux ministères, une convention est nécessaire. L'Open data a permis d'abaisser cette barrière conventionnelle et de travailler plus facilement, en passant d'un mode de convention de gré à gré à un mode extrêmement simple d'utilisation des données.

M. LOGEAN ajoute qu'ouvrir les données oblige à faire un état des lieux, qui est un travail fastidieux mais important. Cet inventaire permet de savoir quelles données il est possible d'ouvrir et lesquelles sont privées ou personnelles. Pour l'administration, ouvrir les données a un impact positif mais demande un investissement, car plus les données sont soignées, plus leur réutilisation en interne sera facile. La manière la plus simple de publier les données et de faire du Web service est d'utiliser le Web (http), ou les formats CSV, XML, RDF. Les fichiers Excel sont plus difficiles d'utilisation. Par exemple, plutôt que de passer par des systèmes propriétaires compliqués, le Département de la sécurité pourra utiliser une interface très simple pour avoir accès aux données d'un autre département.

M. STROH ajoute que la mission et les objectifs d'une organisation privée ou publique sont donnés par son autorité de tutelle. Pour que les TPG puissent ouvrir plus de données, il faudrait que la mission d'ouverture des données soit intégrée dans leur contrat de prestation.

*M. Pierre-Yves Gillieron* demande si les services de l'Etat pourraient à terme être affaiblis à cause de partenariats privés. Il donne l'exemple du SITG qui a des données gratuites de très bonne qualité : si le politique ouvre ces données, il y a un risque qu'à terme ces mêmes données deviennent payantes.

M. Pascal OEHRLI, directeur du centre de compétence du SITG, répond que le principe premier du SITG n'est pas de distribuer les données mais de les partager entre entités publiques pour leurs propres besoins. Il ne pense donc pas que des privés pourraient concurrencer cette mission, le SITG n'ayant pas vocation prioritaire à s'adresser au grand public. Si cela était le cas, ces privés le feraient pour le grand public avec plus de moyens et plus de valeur ajoutée pour les citoyens que pourrait le faire le SITG, ce qui est positif.

M. LOGEAN donne l'exemple de [swissmediatools.ch](http://swissmediatools.ch), un site actualisé en temps réel qui utilise les API des TPG et qui propose de partager les données entre entités publiques. Pour cela, il est possible payer CHF 29.- par an sans publicité avec des services complémentaires, ou profiter du site gratuitement mais sans fonctionnalités. Cependant, il doute qu'avec un portail en temps réel, ce site arrive à valoriser CHF 29.- par an.

M. CHIGNARD renvoie M. Gillieron vers une étude de l'Agence Spatiale Européenne qui figure sur son blog et qui avait analysé ce phénomène sur 15 ans, ainsi que vers un rapport de la Cour des comptes mandatée par Jean-Marc Ayrault (dit « rapport Trojette »). 35 redevances en place ont été analysées, et cela va d'organismes qui font des bénéfices de dizaines de milliers d'euros ou des organismes qui gagnent quelques millions d'euros. La plupart des modèles de tarification sont aujourd'hui favorables aux plus gros consommateurs de données et défavorables pour ceux qui entrent sur le marché. La recommandation de la Cour des comptes vise à changer de modèle de tarification afin que commencer à utiliser la donnée soit moins cher et que les plus gros utilisateurs de données contribuent encore plus fortement à l'ouverture des données qu'ils ne le font aujourd'hui.

*M. Olivier Travaglini* demande comment fonctionne l'association Opendata.ch et quel est son rôle par rapport aux personnes qui produisent déjà de la donnée. Notamment, il demande si le rôle de l'association est d'expliquer que mettre en place la donnée est une chose, mais qu'il faudrait le faire sous un format plus facilement utilisable que celles actuellement en ligne. Il donne l'exemple de l'Office fédéral de la statistique (ci-après OFS), qui permet de télécharger des fichiers Excel qui seraient tout à fait utilisables en théorie, mais qui ne le sont pas actuellement car il faut passer ¼ d'heure à rendre ces fichiers bruts. Il donne également l'exemple des polices cantonales romandes qui ont uniformisé leurs rapports statistiques mais sous forme de PDF.

M. LOGEAN répond que le portail [opendata.admin.ch/](http://opendata.admin.ch/), [map.geo.admin.ch/](http://map.geo.admin.ch/), l'OFS, MétéoSuisse et les Archives ont mis leurs données en ligne et ont immédiatement



reçu des mails sur le fait que ces fichiers n'étaient pas utilisables, sans que l'association intervienne. Avoir des formats plus faciles d'utilisation qu'Excel prend du temps et le processus a déjà commencé. Mais il est vrai qu'en théorie, inciter à publier des données publiques dans un format réutilisable fait partie de leurs buts.

M. PAULETTO précise qu'il existe des niveaux définis par des spécialistes du web sémantique. Le Niveau 0 correspond à avoir des données ouvertes ou du moins déclarées comme telles. Au Niveau 1, le format est de plus en plus ouvert, jusqu'à des API, qui permettent de faire de l'information mise à jour en temps réel et du *Linked Data*.

M. CHIGNARD complète en montrant un fichier de [data.gouv.fr/](http://data.gouv.fr/) sur la fréquentation des musées de 2006 à 2010. Il faudrait passer 2 heures afin de rendre ce fichier plus brut qu'il ne l'est, car il y a des codes couleurs, des sous-totaux, des numéros de référence qui le rend très difficile d'utilisation. Si ce type de fichiers existe, c'est parce que la philosophie de départ était de publier les données peu importe la forme.

*M. François Voisard* demande quels sont les risques que les utilisateurs déduisent de l'ouverture des données un droit de les obtenir à long terme ou que les grandes entreprises les réutilisent pour du *Big data*, les traitent et en tirent des conclusions.

M. CHIGNARD répond que dans les licences Open data, il y a toujours des clauses de désengagement de responsabilité, mais il est vrai qu'il n'est pas possible de se désengager de la responsabilité de l'usage de certaines données et de garder le contrôle de certaines autres données.

*M. PAULETTO clôt le débat.*