

OPEN DATA: LES DONNÉES LIBÉRÉES DOIVENT-ELLES GRATUITES?

Entretien avec Samuel Goëta

Samuel Goëta est doctorant à Télécom ParisTech, et écrit une thèse sur la sociologie de la production et de la libération de données publiques.

- *DataPublica étant une entreprise qui retraite les données, quel point de vue peut-on avoir à son égard par rapport à la gratuité des données ?*
- « DataPublica, beaucoup de gens s'en servent parce que, l'avantage, c'est qu'ils retraitent leurs données, ils mettent de bonnes méta-données, et on peut les retrouver plus facilement que sur data.gouv.fr, et donc ils apportent une valeur ajoutée. C'est un peu la vitrine dont ils se sont servi au début : mélanger données gratuites et données payantes. C'est un business model qui n'a pas trop marché ; donc ils ont réorienté leur business, proposent des tableaux de bords, des processus en interne dans les entreprises pour croiser les données, faire de la business-intelligence en croisant différents modèles. La vente des données est quelque chose qui ne marche pas bien, est sous optimum et n'est pas rentable en général. Au départ, ils pensaient qu'ils pouvaient vendre des données, et ils ont quand même eu cette intelligence de mélanger données gratuites et payantes. Il y a un travail pour les améliorer et les rendre accessibles, mais les données payantes, ils n'en ont pas vendu tant que ça.
A l'heure actuelle, il y a un questionnement sur : est-ce qu'on doit continuer à les vendre, est-ce que la valeur ajoutée d'un service est dans la vente des données ? De moins en moins, plus dans le retraitement, la donnée elle-même étant quelque chose qui a plus de valeur quand elle est publique et réutilisable par tout le monde. »
- *Où en est le gouvernement ? EtaLab étant surtout rempli des données de l'INSEE, il n'y a pas tant de données différentes que ça...*
- « En fait il y a plein, énormément, de données. Les données ont été publiées au lancement de data.gouv.fr, en pleine campagne électorale (Sarkozy et l'état sont transparents...) et du coup ils ont fait un coup de comm' de l'OpenData, ils ont joué à fond là-dessus, il fallait faire un coup et mettre des données d'un coup sur la plateforme, sans vraiment créer de routine chez

ceux qui publient les données (dans les administrations).

Aujourd'hui la publication de données n'est pas une obligation, pas une routine, pas une habitude. Il faut déjà qu'ils prennent conscience qu'ils détiennent un jeu de données, et les pousser à mettre en ligne ces données. Plein de choses demandent à être publiées, cela demande du travail (effacer les commentaires, etc...), ce qui demande alors de la formation (du gouvernement) ie un travail de longue durée. Dans la feuille de route d'EtaLab, il y a beaucoup d'actions internes. Aller dans les ministères, les inciter à publier leurs données, casser l'habitude de publier un rapport avec un tableau sans publier les données qui sont dans ce tableau.

La feuille de route propre au gouvernement (Henri Verdier) se base sur le développement de l'éco système de l'OpenData, encourager de manière large la réutilisation des données. Puis mettre en œuvre un chantier sur les données pas encore publiées : domaines de carences comme dans le domaine **de la santé** ! Revendication de certaine données retenues par la CAF. (cf. Fourmi Santé poursuivi pour avoir utilisé les données d'un site public). Données de **l'environnement**, peu fourni ! Dans les transports : les publications des données complexes sont freinées par la peur de Google qui pourrait créer un système de transit (les marges de Voyages SNCF se font actuellement sur la vente de chambre d'hôtels, etc.).

L'Application *MetroParis* paye chaque année la RATP pour avoir les horaires. DataPublica a publié des données trouvées on ne sait où sur le trafic par mois, alors que sur le site de l'OpenData RATP on n'y trouve que le trafic par an (quelque chose de déjà connu). **C'est donc quelque chose de très lent !** Velib' ne publie pas certaines données (trafic, les trajets effectués) de peur d'être attaqué et critiqué, ce qui remettrait en cause leur accord avec les mairies.

Quand la loi CADA a été faite en 78, l'important c'était d'accéder à l'information. C'était un mouvement qu'on appelait les Whistle Blower. Pendant la guerre du Vietnam, aux Etats-Unis, des gens ont récupéré des données publiques qui allaient critiquer, etc... Donc on accédait à l'information sans la réutiliser. Il y avait des services MiniTel qui se faisaient de l'argent avec des données publiques : par exemple les résultats du BAC se consultaient par MiniTel et étaient payants. **L'accès était gratuit, mais la réutilisation était payante. Aujourd'hui on veut une réutilisation gratuite.**

OpenDefinition.Org définit ce qu'est un contenu ouvert. Une donnée ouverte n'est pas nécessairement gratuite, mais si possible. Ce n'est pas efficient de faire payer les données, mais si on les publie elles doivent être à un coût marginal : juste le coût de publication, pas de bénéfices dessus. »