

Open Data : une mine d'or brut pour les professionnels de l'information

CAROLE TISSERAND-BARTHOLE

L'Open Data fait régulièrement la Une de l'actualité française depuis 2010 avec l'ouverture des premières données publiques dans notre pays. Nous avons eu plusieurs fois l'occasion de publier des articles dans Bases sur l'Open Data mais il s'agissait surtout de lister les nouvelles initiatives en France et ailleurs dans le monde. Nous avons décidé qu'il était temps de nous lancer dans un article de fond sur les apports de l'Open Data pour les professionnels de l'information.

Cet article reviendra donc sur le mouvement de l'Open Data en France mais également dans le reste du monde (dates-clés, concepts, etc.), nous nous intéresserons ensuite aux types de données proposées (quels organismes en proposent, est-ce du contenu exclusif et non disponible ailleurs, quel apport comparé à des données payantes, etc.) et nous effectuerons ensuite un focus sur la France avec un panorama des sources aujourd'hui disponibles.

Nous illustrerons tout cela avec une méthodologie de recherche pour retrouver des données sur les gares de voyageurs et de fret en France et dans différents pays européens. Enfin, nous

terminerons sur les actualités des derniers mois.

Open Data : Fondamentaux

L'Open Data, c'est avant tout l'idée qu'un certain nombre de données numériques (textuelles ou non textuelles comme des cartes, des formules chimiques, des statistiques, etc.) produites par des organismes publics mais également privés devraient être librement accessibles et réutilisables sans condition par tout un chacun. L'Open Data s'inscrit dans le mouvement dit de « l'Open » - Open Access, Open Source, etc. - mais il dispose de ses propres particularités.

L'Open Access par exemple se concentre exclusivement sur la mise à disposition de manière libre et gratuite de publications académiques tandis que l'Open Source concerne uniquement le type de licence sous laquelle est distribué un logiciel informatique. L'idée de diffusion libre et gratuite est en filigrane derrière toutes ces initiatives mais c'est le type de supports/documents/données / domaines d'application qui les distinguent.

Depuis les débuts de l'Open Data, de nombreux arguments en faveur ou contre le mouvement sont avancés par deux camps opposés.

Pour certains, les données appartiennent par nature à l'être humain, de l'argent public a souvent été utilisé pour les produire ou le fait de rendre plus facilement accessibles des données accélère le processus de recherche et d'innovation et donc justifie le mouvement de l'Open Data.

D'autres en revanche ont le sentiment que l'Open Data fait une concurrence déloyale au secteur privé, que la mise en forme et la mise à disposition de ces données représentent un coût humain non négligeable pour les organismes qui les diffusent ou encore qu'il est préférable de payer une somme minimum pour l'obtention de ces données mais d'avoir l'assurance d'obtenir des données vérifiées, fréquemment mises à jour et dans un format exploitable.

Nous ne tenterons pas de départager les deux camps mais toujours est-il que le mouvement est maintenant bien en route et en perpétuelle évolution et ce, dans la plupart des pays du monde (à des degrés très divers néanmoins). En Europe notamment, le mouvement des données ouvertes pour le secteur public est encadré par la directive 2003/98/CE et la directive INSPIRE de 2008 qui imposent aux collectivités de publier certaines de leurs données. Seul problème, notamment en

France et en Europe (mais pas aux Etats-Unis), les collectivités et organismes publics doivent publier une partie de leurs données mais sans aucune contrainte quant au format ou au support de publications. Ce qui conduit parfois à la publication de données complètement inexploitable.

Il serait dommage de ne pas en tirer parti que ce soit pour une veille ou pour une recherche d'information. Parmi les thématiques fréquemment abordées, il est souvent question de la réutilisation commerciale des données ou encore de l'anonymisation des données et du respect de la vie privée. Ces sujets sont certes très intéressants mais ce n'est pas l'angle sous lequel nous avons choisi d'aborder la question. L'idée étant pour nous de regarder l'Open Data comme une nouvelle source, un nouveau canal d'information au même titre que des bases de données professionnelles.

Quels types de données et qui les produit ?

Sur le papier, l'Open Data pourrait être présent dans n'importe quel domaine (santé, transport, sciences, économie, etc.) et les données pourraient provenir d'organismes publics (on appelle cela des données publiques) mais également d'entreprises privées. Dans la