

CRITIQUES DU WEB² (1/4) : QUELLES DONNÉES LIBÈRE-T-ON ?

LE 16 SEPTEMBRE 2009 HUBERT GUILLAUD

Oui, les données valent de l'or. C'est aussi comme cela qu'on peut lire le Web à la puissance deux (Web²) de Tim O'Reilly et John Battelle. Mais si c'est effectivement le cas, si nous sommes tous assis sur un capital dont on mesure mal l'exploitation qui peut en être faite, pourquoi faudrait-il les libérer ? [...]

Oui, les données valent de l'or. C'est aussi comme cela qu'on peut lire le Web à la puissance deux (Web²) de Tim O'Reilly et John Battelle. Mais si c'est effectivement le cas, si nous sommes tous assis sur un capital dont on mesure mal l'exploitation qui peut en être faite, pourquoi faudrait-il les libérer ? Que gagne-t-on à les partager ?

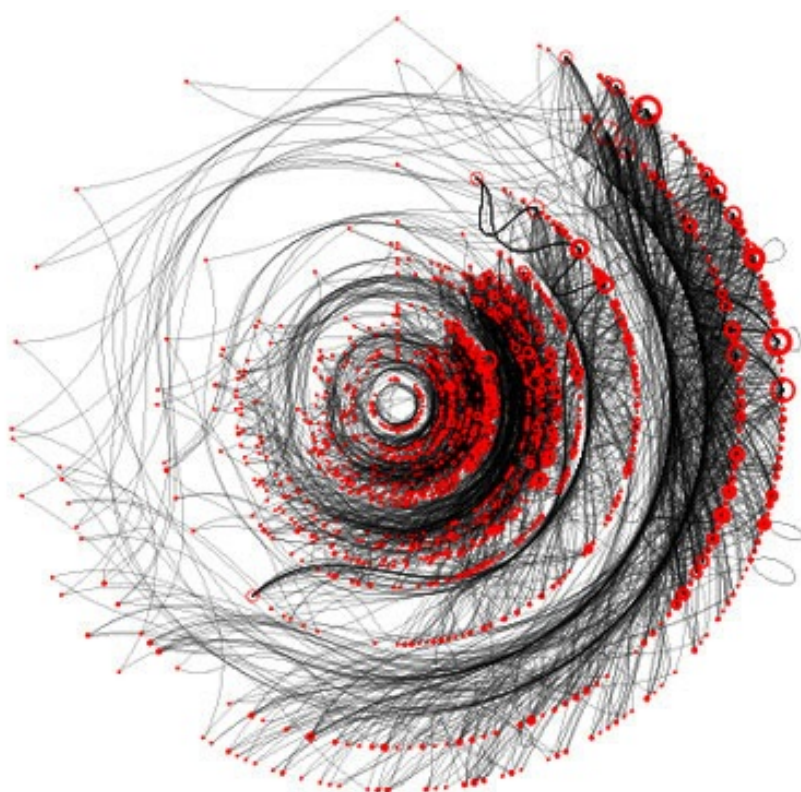


Image : Illustration de Boris Müller pour le festival de littérature allemand de 2006 montrant les relations entre les lettres dans un poème, mis en avant par l'australien Andrew Vande Moere, l'auteur du cultissime blog **infosthetics** sur l'esthétisme de la visualisation de données, pour **un article de PingMag** sur la beauté de cette nouvelle discipline de représentation de l'information.

Un des présupposés de l'article d'O'Reilly et Battelle est que les données que produisent les dispositifs sociotechniques, depuis les bases de données commerciales jusqu'aux réseaux de capteurs, en passant par les logs de sites web et les documents publics, "s'ouvriront" pour permettre à toutes sortes d'acteurs de les exploiter de manières nouvelles : les agréger, les croiser, les analyser, en extraire d'autres significations ou d'autres décisions...

Or l'ouverture des données ne va pas de soi pour les organisations. Et ce ne sont pas seulement des questions de droits ou de vie privée (centrales), mais aussi des questions de conception (politique) et de valeur (économique) des données elles-mêmes. Si les données sont le capital immatériel de demain comme l'expliquent Tim O'Reilly et John Battelle, **ou comme l'ont déjà dit Jean-Pierre Jouyet et Levy avant eux**, il sera difficile de dire qu'il faut les libérer, c'est-à-dire les ouvrir, les rendre accessible facilement – voir gratuitement. Surtout si demain, leur agrégation, leur compilation, leur croisement appellent à créer de nouvelles données, de nouveaux services et de nouvelles valeurs.

Indéniablement, la question de leur ouverture deviendra à l'évidence un sujet conflictuel.

“Libérer” une donnée, c’est la mettre en circulation et bien souvent, en perdre le contrôle, y compris économique. C’est aussi permettre à d’autres de créer à partir d’elle de nouveaux services, de nouvelles valeurs. En contrepartie de quoi ? Dans quelles conditions ? On peut bien sûr montrer que bien souvent, une donnée qui circule et que beaucoup de gens exploitent intelligemment produit plus de bien-être collectif qu’une donnée soigneusement enfermée dans son silo, mais la plupart des acteurs n’intègrent pas (encore ?) leur contribution au bien-être collectif dans leurs comptes...

Certes, on peut souhaiter qu’elles n’appartiennent plus seulement à ceux qui les produisent ou les collectent, puisqu’elles sont pour l’essentiel composées de nos identités et de nos activités. Mais n’est-ce pas un vain rêve ? Un miroir aux alouettes dont les reflets nous empêchent de regarder le problème plus en profondeur ?

Il y a au moins deux limites intrinsèques à la vision de Battelle et O’Reilly :

> Il n’existe pas de données brutes

Toute donnée est construite par le système qui la produit. Un capteur est installé à un endroit précis ; il mesure certains paramètres et pas d’autres, avec une certaine marge d’erreur, à une certaine fréquence. Un document est rédigé dans une langue donnée et avec un certain langage, administratif, scientifique, commercial, littéraire... Deux entreprises faisant le même métier collecteront des données différentes ; elles définiront des segments différents : par exemple, on pourra classer un âge dans les segments de 5 ans (15-20 ans, 20-25...) ou beaucoup plus larges (“jeune”, “actifs”, “vieux”...). Toutes les données auront du mal à discuter entre elles, quels que soient les progrès de la science des algorithmes ou de la sémantique

> On ne libère jamais tout

Quelles sont-elles ces données qu’on appelle à libérer ? Desquelles parle-t-on ? Desquelles ne parle-t-on pas ? Amazon par exemple libère certaines données via **ses interfaces de programmation** (API), oui. Celles qui permettent à d’autres services de construire des services depuis ses données à son propre profit... Amazon ne libère pas les données de ses clients. N’importe qui n’a pas accès aux livres que vos Kindle ont achetés. Nous n’avons pas non plus accès aux chiffres de ventes des produits que propose le commerçant. Pas plus que nous n’avons accès aux données de visites (les *logs*) de son site marchand. Où sont les données des plaintes ou des demandes de retrait de titres qu’il reçoit et qui justifie **certaines de ses reclassements de titres par exemple ?**

S’il y a de plus en plus de lois et/ou de décisions politiques pour libérer les données publiques – et encore, les exemples britanniques et américains demeurent pour l’instant isolés –, les entreprises, elles, n’ont pas cette contrainte et ne libèreront que ce qui leur semble soit non stratégique, soit plus probablement utile à l’émergence d’écosystèmes de services qui leur profiteront en dernier ressort.

La libération des données qu’appellent O’Reilly et Battelle ressemble certes à un horizon, une nouvelle frontière, mais elle n’a rien de magique en soi. Le plus dur assurément sera de créer les conditions de la libération des données pour ne pas que ces promesses ne se transforment en un vaste *Data-entertainment*, un “grand spectacle des données”, une industrie du divertissement autour des données, comme le propose Facebook.

—

> Article initialement publié sur Internetactu, sur lequel vous pouvez retrouver la suite de cette critique du Web²

1 ping

Les tweets qui mentionnent Critiques du Web² (1/4) : Quelles données libère-t-on ? | Owni.fr -- Topsy.com le 17 septembre 2009 - 10:06

[...] Ce billet était mentionné sur Twitter par Nicolas Voisin, Owni et Veronique Rabuteau. Veronique Rabuteau a dit: Fwd & RT @PierreTran: Critiques du Web² (1/4) : Quelles données libère-t-on ? <http://bit.ly/xSZaP> (via @Owni) (par... <http://ff.im/-8fZHJ> [...])

[WORDPRESS HASHCASH] The comment's server IP (208.74.66.43) doesn't match the comment's URL host IP (74.112.128.10) and so is spam.