

CHRISTMAS
EDITION

\$
FREE
\$



OPEN DATA

PDF EDITION

Pulp
OWNI
Digest



2010 et l'OPENDATA

LIBÉREZ LES DONNÉES !

«Nous ouvrons les gouvernements» : si le slogan de WikiLeaks semble avoir été entendu, au moins en partie, par les gouvernements anglo-saxons comme en témoignent les initiatives des Etats-Unis, de la Grande-Bretagne, de l'Australie et du Canada, la majeure partie des pays restent à la marge.

Ainsi, les pays européens peinent à suivre le mouvement, malgré la mise en place de la directive européenne [INSPIRE](#) en 2007, et la France ne déroge pas à cette règle. Les politiques nationales, conservatrices en la matière, ont peu évolué depuis la loi du 17 juillet 1978 et restent aujourd'hui contradictoires et inégales selon les secteurs, orientées parfois vers un mouvement de fermeture, comme en témoigne la récente loi Loppsi2.

Les initiatives les plus poussées se développent au niveau local, Bretagne en tête, et le mouvement de l'open data est d'abord porté par les citoyens, les journalistes, les universitaires et les communautés open source.

Que les gouvernements le veuillent ou non, le mouvement de libération est lancé et c'est à nous de les encourager à aller vers plus de transparence et d'ouverture pour le bénéfice de nos démocraties.

Libérez les données !

DONNÉES PUBLIQUES EN FRANCE: GIRONDINS VS MONTAGNARDS ?

PAR LIBERTIC LE 24 FÉVRIER 2010



La proposition n°22 du rapport Riester suggère la création d'une plateforme d'innovation de services ci-nommée « Etat Lab » qui permettrait aux acteurs tiers de développer des services innovants à partir de données publiques.

La semaine dernière, [@nicolasvoisin](#) indiquait sur Twitter la création prochaine d'un **état lab** en France. Je n'avais pas réussi à trouver la source de l'information mais c'est chose faite grâce à [@epsiplatform](#), il s'agit de la page 38 du [rapport Riester](#) remis la semaine dernière à [Eric Woerth](#) et [Nathalie Kosciusko-Morizet](#).

Etat lab

La proposition n°22 du rapport Riester suggère la création d'une plateforme d'innovation de services ci-nommée « **Etat Lab** » qui permettrait aux acteurs tiers de développer des services innovants à partir de données publiques*.

Cette plateforme prendrait appui sur le futur portail national de l'**APIE (Agence du Patrimoine Immatériel de l'Etat)** de mise à disposition des données publiques et serait initialisée fin 2010. Elle permettrait d'intégrer les bases de données et API, ce qui rappelle trait pour trait le principe du [datagov](#) aux Etats-Unis (notamment).

Ce que le rapport ne dit pas :

Quelles données libère-t-on ?

Sous quel format ?

Quels seront les types de licences pour l'usage de ces données ?

Les formats seront-ils user-friendly (usage citoyen) et/ou mashup friendly (developpeurs only) ?

Comment interagiront les plateformes de différents niveaux (locales,

régionales, nationales) ?

Qui finance ces plateformes ?

A partir de quelle phase privatise-t-on le service ?

S'il convient de garder ces questions en tête, elles ne devraient pourtant pas servir d'alibi à l'inaction. Pour **Hugues Aubin**, chargé de mission TIC à Rennes, plutôt que s'attarder sur la normalisation des données ([pointée du doigt à la FING lors de la conférence du 11 février sur l'ouverture des données](#)), partons des bases existantes pour avancer.

Et justement de l'avancement il y en a avec la plateforme [Data Publica](#) dont la fonction correspond tout à fait à la définition de l'Etat lab.

Data Publica va mettre en place et opérer une place de marché sur laquelle ceux qui possèdent des données viendront publier leurs méta-données ou leurs API ainsi que la ou les licences sous lesquelles elles peuvent être utilisées, et ceux qui développent des applications viendront prendre ces méta-données pour développer des applications au modèles économiques compatible avec la licence d'utilisation.

Data Publica fournira aux éditeurs de données une palette de licences parmi lesquelles ils pourront choisir et des outils pour publier leurs données et méta données. Data Publica fournira aux développeurs un annuaire des API et meta données et l'accès aux données.

Cette place de marché contribuera au développement des applications valorisant les données publiques ou privées. Le projet est financé partiellement dans le cadre de l'appel à projet « Web 2.0 innovant ». Il est développé par un consortium de trois entreprises : Araok, Nexedi et Talend.

Au fur et à mesure des avancées dans la révolution de l'ouverture des données publiques, les acteurs tendent à se positionner derrière deux camps. **Data Publica rejoindrait les [Montagnards](#)** misant sur la

centralisation. Leur plateforme semble prendre la voie d'un service exclusif aux développeurs, avec catalogue de données brutes, approche assez administrative, données en partie payantes et peu ou pas d'outils intégrés pour la visualisation ce qui rend ces données innaccessibles au citoyen lambda.

Les plateformes locales ([Keolis et la ville de Rennes à suivre](#)) sont en train de se constituer de manière plus rapide- relativement parlant- et polymorphe. Elles pourraient prendre une approche plus ludique et aussi accessible au public en proposant des outils de visualisation, de personnalisation et de comparaison. **Ce seraient les Girondins** de l'ouverture des données, proposant un service décentralisé

Ces approches vont-elles s'entrechoquer ou au contraire devenir complémentaires ?

Dans un article comparant la plateforme américaine [Datagov](#) et la plateforme anglaise [datagov.uk](#), Flowing Data indique sa [préférence pour la version anglaise](#), plus accessible, elle est plus démocratique. Un catalogue centrale des données est pourtant indispensable aussi pourquoi ne pas préparer dès aujourd'hui la conciliation des deux versions ?

Data Publica pourrait être le centre de référencement national pour trouver une donnée et l'en extraire pour les développeurs. Les plateformes locales permettraient l'accessibilité aux données par les développeurs mais aussi les citoyens, ce qui serait d'autant plus judicieux qu'il s'agira de données locales fortement plebiscitée: transport, tourisme, etc. Pour s'y retrouver et obtenir un service homogène, il faudrait d'ores et déjà mettre en place des règles, notamment un intitulé-type pour les sites de bases de données: ville.donnee.fr par exemple.

Et si on changeait l'équation Girondins vs Montagnards pour:
Girondins + Montagnards ?

* « Orange » a fait parti du groupe de travail pour le projet Riester qui préconise l'ouverture des données publiques, données qui sont à la base de bons nombres d'applications smartphones.

> [Article initialement publié sur Libertic](#)

> **Illustration de page d'accueil par [suzannelong](#) sur [Flickr](#)**

>Image by [DarFin Oil Painting](#) via Flickr

LIBÉRONNS LES DONNÉES ! DE QUELQUES ASPECTS JURIDIQUES

PAR MICHELE BATTISTI LE 5 MARS 2010



A l'heure où l'on évoque la libération des données publiques [1] et le Web des données [2], ou encore le [data journalisme](#), il semblait utile d'approfondir [cette question que l'on avait abordée en 2006 pour les données de la recherche](#).



Les enjeux

Dans le domaine scientifique, il est essentiel de partager les informations. Pouvoir reconnaître la qualité de biens collectifs aux résultats de la recherche devient donc un impératif. Quant au Web de données, autre exemple retenu, il implique que l'on puisse relier et interroger de manière simple les données dispersées sur le web [3], mais aussi d'en disposer sans se heurter à des obstacles juridiques.

Or, pour accéder à des données et avoir le droit de les utiliser, on s'aperçoit qu'il faut bien souvent contacter les ayants droit et négocier des licences, ce qui s'avère généralement lent et coûteux.

Pour pallier les risques croissants de réappropriation des données, liés à l'expansion actuelle de la propriété intellectuelle, ont été développées parallèlement, depuis plusieurs années, des approches techniques et juridiques dites « ouvertes ».

Les données libres et ouvertes permettent ainsi de disposer d'un fonds commun de données dans lequel n'importe qui peut puiser et ce, malgré les différences importantes des contextes juridiques et institutionnels de leur création.

Quelques rappels

DÉFINITIONS

donnée : un fait, notion ou instruction représentée sous forme conventionnelle convenant à la communication, l'interprétation ou au traitement par des moyens humains ou automatiques (afnor)

donnée publique : donnée collectée ou produite dans le cadre de sa mission, par un service public, sur des fonds publics.

donnée libre : une donnée que l'on est libre d'utiliser, de modifier et de rediffuser

donnée ouverte : l'ouverture s'applique la notion d'interopérabilité, assurée par des standards.

licence libre : contrat par lequel un titulaire d'un droit de propriété intellectuelle concède à un tiers tout ou une partie de la jouissance d'un droit, en accordant au moins la possibilité de modifier, de rediffuser et de réutiliser l'œuvre dans des [œuvres dérivées](#). Ces libertés peuvent être soumises à conditions [\[4\]](#) ([Wikipédia](#)).

domaine public : des œuvres protégées par le droit d'auteur dont la durée des droits patrimoniaux a expiré (droit d'auteur) ou des choses communes qui n'appartiennent à personne mais dont l'usage est commun à tous (art. 714 Code civil).

LE CADRE JURIDIQUE

Le droit d'auteur. Les données « brutes » ne sont pas protégées par le droit d'auteur, mais leur mise en forme (sous forme de

graphiques, diagrammes, etc.) pourrait l'être, dès lors que celle-ci s'avère originale et que la présentation des données ne découle pas automatiquement de l'utilisation d'un procédé technique ou d'un logiciel. Si, par ailleurs, la structure et choix des données sont originaux, la base de données sera protégée par le droit d'auteur.

Le droit *sui generis* des bases de données. Si le producteur de la base de données peut en outre prouver avoir investi de manière substantielle pour constituer et gérer sa base, il peut s'opposer à toute extraction ou réutilisation substantielle de celle-ci.

Le droit de réutilisation des données publiques. Dans ce cadre, l'administration met à la disposition du public les données qu'elle a produites ou qu'elle détient [5].

D'autres mécanismes. L'accès aux données peut être protégé *a priori* par un contrat et *a posteriori* par une action en concurrence déloyale, pour parasitisme ou pour [enrichissement sans cause](#).

Libérer les données

Pourquoi adopter une licence ?

Même dans le cas où les données ne sont pas protégées par un droit (droit d'auteur, droit *sui generis* du producteur de base de données), il est recommandé de les diffuser accompagnées d'un contrat de licence.

Les utilisateurs ont, en effet, besoin de savoir d'où viennent les données (notion d'attribution), si celles-ci elles ont été modifiées (souci d'intégrité et de validation) et de connaître toutes les modalités de réutilisation autorisées.

Les recours à des contrats de licences reconnus dans le monde entier, même si celles-ci nécessitent une adaptation au droit national, permet d'organiser le partage des données en évitant, puisque les contrats sont la loi des parties, de se pencher sur le problème de la loi applicable et de se heurter à des concepts juridiques différents.

L'extrême diversité des licences

Dans son [rapport](#), le Conseil supérieur de la propriété littéraire et artistique (CSPLA) note qu'il existe des œuvres sous licences mixtes (contenant des éléments sous licences libres et d'autres sous contrats propriétaires) ou encore des licences doubles (permettant d'opter soit pour une licence libre qui impose certaines contraintes, soit pour une licence propriétaire).

A côté des « domaines historiques d'expansion » que sont les logiciels libres, des modèles « généralistes » ([licences Creative Commons](#) ou [Art Libre](#), par exemple), diverses communautés (informaticiens, chercheurs [6], musiciens, ...) ont développé des licences qui répondent à leurs priorités. En outre, afin de les améliorer, pour s'adapter à de nouvelles pratiques ou de nouvelles législations, ces licences ont pu évoluer dans le temps et se présenter sous de nouvelles versions numérotées [7]. Le mouvement ne s'est étendu qu'ensuite et plus récemment aux données [8].

Quelle licence choisir ?

Pour s'assurer que les données restent libres et bâtir un fonds commun de données, il est important de pouvoir contrôler leurs usages. Abandonner tous ses droits de manière anticipée, en choisissant une licence qui fait entrer son œuvre dans le domaine public (comme l'[Open Data Commons – Public Domain Dedication & Licence \(PPDL\)](#) ou la licence [CC Zéro](#)), permet certes de les utiliser pour créer des œuvres dérivées, mais ne répond pas obligatoirement à un tel objectif, puisqu'il n'oblige pas à partager les résultats obtenus à partir des données utilisées. N'importe qui, une entreprise par exemple, pourrait se réappropriier ces données, d'où le paradoxe qui consiste à créer un droit de propriété, qui ne devrait pas exister, pour empêcher cette action.

Si l'on entend bâtir un fonds commun de collaboration, il s'avère de ce fait prudent d'adopter une logique de [Copyleft](#) qui « force » le partage. Classées dans [le rapport du CSPLA](#) parmi les licences offrant une liberté pérenne, les licences *Copyleft*[\[9\]](#) ne « se limitent pas à l'octroi du droit d'utiliser, de modifier et de diffuser l'oeuvre : elles veillent à ce que l'usage de l'oeuvre copiée ou modifiée demeure libre. Dès lors que sont combinés les éléments de l'oeuvre mise à disposition, l'oeuvre seconde devra obéir au même régime », ce qui permet de s'assurer de l'existence d'un fonds commun d'oeuvres que l'on peut utiliser librement.

La licence *Creative Commons BY-SA*[\[10\]](#) qui oblige à citer la source et à partager les données obtenues à partir des données utilisées sous la même licence CC BY-SA, a été adoptée notamment par les concepteurs du projet *OpenStreetMap*, projet qui connaît un grand succès, qui vise à cartographier des villes à partir de données géographiques libres. Considérant toutefois qu'une licence CC vise à protéger des oeuvres protégées par le droit d'auteur, [les concepteurs d'OpenStreetMap songent à adopter la licence ODbL](#), adaptée au droit sui generis des bases de données

Le choix d'une licence n'est pas neutre

S'il convient, en préalable, de s'assurer que l'on est en mesure de céder tous les droits que la licence va conférer (coauteurs éventuels, droits d'auteur sur les oeuvres incorporées, ...), il convient aussi d'en apprécier l'impact avant d'opter pour l'une d'entre elles.

Kobé

Mes plus vifs remerciements à [Benjamin Jean](#), [Linagora](#)

Ce texte sera également publié [sur le site de l'ADBS](#)

Notes

[\[1\]](#) Les portails gouvernementaux [en Australie](#), [aux Etats-Unis](#), ou encore [au Royaume-Uni](#) proposent tous des données en libre accès. Mais ces portails seraient moins riches en données que le portail français, tel que préfiguré aujourd'hui. Le niveau « d'ouverture » des données publiques accessibles *via* le portail français

Opendata - OWNIXMAS 2010

d'accès aux données publiques annoncé par l'Agence du patrimoine immatériel de l'Etat (APIE), variera selon les administrations. Le [Chapitre français Creative Commons](#) a interrogé les candidats aux élections régionales pour les sensibiliser à l'utilité de libérer les données publiques régionales.

[2] Voir aussi : [L'avenir de l'information : le web de données](#), Hubert Guillaud, *InternetActu.net*, 4 décembre 2008

[3] Démontrant ainsi la nécessité de l'interopérabilité, une condition facilitée par le mouvement Libre.

[4] La licence libre GPL (*General Public License*), par exemple, impose que l'on redistribue le logiciel « *enrichi* » après l'avoir récupéré, étudié et amélioré.

[5] Le principe d'une réutilisation des informations publiques ne s'applique pas aux informations élaborées dans le cadre d'une mission de service public à caractère industriel et commercial, ni à celles dont les tiers détiennent des droits de propriété intellectuelle, ni à celles qui contiennent des données personnelles non anonymisées lorsque l'accord des personnes concernées n'a pas pu être obtenu.

[6] Des licences spécifiques ont été développées pour certaines disciplines, *BiOS Open source* dans le domaine des biotechnologies, par exemple.

[7] La version 3.0 des licences *Creative Commons*, par exemple, tient compte du droit *sui generis* qui peut être conféré à certains producteurs de bases de données.

[8] Les réflexions autour des *Science Commons* [12] et le succès d'OpenStreetMap [17] ont joué à cet égard un rôle majeur

[9] La licence CC BY SA ou la licence Art Libre, par exemple

[10] Licence CC qui oblige à mentionner la source (BY) qui autorise les modifications, mais qui impose que les oeuvres dérivées soient proposées au public avec les mêmes options *Creative Commons* que l'œuvre originale.

Références

1. [Rennes et Keolis: ils ont osé ! LiberTic](#), 1er mars 2010
2. [Vers des données régionales en libre accès](#), Michaël Szadkowski, *blog du Monde*, 10 février 2010
3. [Rennes passe à l'open source pour diffuser ses infos pratiques](#), Christophe Guillemin, *01Net*, 24 février 2010
4. [Atelier Communia « Les institutions de mémoire et le domaine public » – Barcelone 1 & 2 oct. 2009 \(Compte-rendu et impressions\)](#), Calimaq, *S.I.Lex*, 13 octobre 2009
5. [The legal status of raw data : a guide for research practice](#), Madeleine de Cock

Opendata - OWNIXMAS 2010

Buning, Allard Ringnalda, *CIER* and Tina van der Linden (Centre for Intellectual Property Law, Surf Foundation, July 2009

6. [La mise à disposition ouverte des œuvres de l'esprit](#), Valérie-Laure Bénabou et Joëlle Farchy, Conseil supérieur de la propriété littéraire et artistique juin 2007
 7. [CCo : une nouvelle licence Creative Commons pour « marquer » le domaine public en ligne](#), Calimaq, *S.I.Lex*, 17 mars 2009
 8. [Mon œuvre est dans le domaine public](#), Michèle Battisti, *Actualités du droit de l'information*, 16 mars 2009
 9. [\(Petit\) Guide à l'usage des licences libres](#), Benjamin Jean, Intervention lors de la matinée juridique, Syntec informatique du 14 mars 2008.
 10. [Comment puis-je mettre mes documents sous licence Creative Commons ?](#) Michèle Battisti, *Actualités du droit de l'information*, janvier 2008
 11. [Principes pour des données publiques ouvertes](#), Hubert Guillaud *Internet Actu*, 20 décembre 2007
 12. [Science Commons : une solution pour le partage des résultats scientifiques](#), Michèle Battisti avec la collaboration d'Arabelle Baudette, *Actualités du droit de l'information*, n° 67, mars 2006
 13. [Les contrats Creative Commons](#), Michèle Battisti, *Actualités du droit de l'information*, n° 52, novembre 2004
- Définitions. Guide
14. *Open Definition*. Un tableau des licences. [Sur le site Open Definition](#)
 15. [Guide to Open Data Licensing](#). Sur le site *Open Knowledge Foundation Wiki*
 16. Openstreetmap. [Sur le site Wikipédia](#)

Présentations ppt

17. [Le statut juridique de la donnée libre](#), Benjamin Jean, *Linagora*, 17 septembre 2008
18. [Data sharing : social and normative](#), Kaitlin Thaney, 25 octobre 2009
19. [Rights Statements on the Web of Data](#), Leigh Dodds, 25th October 2009

Présentation audiovisuelle

20. Le web des données. Emmanuelle Bermès, 5 à 7, ADBS (en ligne prochainement)

Billet initialement publié sur [Paralipomènes](#) Image [bionicteaching](#) sur Flickr

LA DATAVOLUTION EST-ELLE POSSIBLE?

PAR GILLES BABINET (TRAD. NICOLAS KAYSER-BRIL) LE 30 MAI 2010



Il semblerait qu'une grande partie de la population pense que les données sont loin d'être devenues accessibles à tous. Les données les plus intéressantes seraient cachées dans des *datacenters* très protégés. L'argument ultime est que même si les données étaient disponibles, rien n'indique qu'elles seraient utilisées par tous.

Ces arguments sont en partie vrais. Nous les avons résumés en listant les principaux nœuds restant à défaire avant que l'on assiste à une véritable appropriation des données par le grand public, et à leur utilisation par le plus grand nombre.

1. Trouver un standard commun

Les données sont encore stockées dans une foultitude de formats différents. Des essais ont déjà été faits pour standardiser les données dans des formats communs, mais le moins que l'on puisse dire est que ces efforts n'ont encore rien donné.

Résultat : quiconque veut utiliser des données doit d'abord les reformater pour son usage personnel. Ca fait partie du boulot que fait data.gov, entre autres. On a plusieurs raisons d'être optimistes. Non pas que les données vont devenir standardisées, mais plutôt qu'on dispose de plus en plus d'outils nous permettant de passer d'un format à l'autre automatiquement.



Sur un autre sujet, et contrairement aux idées reçues, les données sont largement disponibles. On peut avoir accès à tout types de données sur la circulation, la météo, la pollution, la population, les exportations/importations de marchandises, l'énergie etc. Sur ce sujet, lisez l'article [The Data Deluge](#) de The Economist.

2. Faciliter l'accès aux données

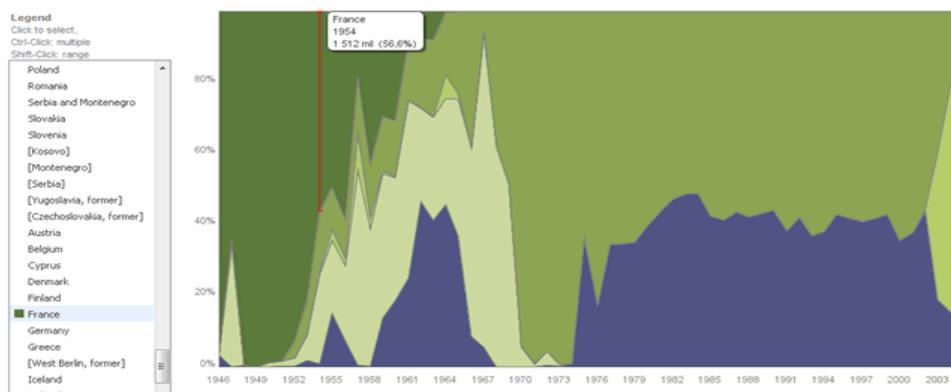
Bien qu'il y ait des quantités quasi-illimitées de données sur le web, les trouver n'est jamais facile. Il n'y a pas de véritable moteur de recherche ou de 'supermarché de la donnée' pour les indexer ou les entreposer. Quelques initiatives, la plus connue étant data.gov, tentent de résoudre le problème, mais on n'a toujours pas de Datapedia.

3. Rendre les données divertissantes

La donnée, c'est souvent ennuyeux. Rechercher ses amis sur Google ou Twitter reste beaucoup plus divertissant. Très peu de gens passent leurs weekends à jouer avec des séries de données sur Excel. Pour que chacun puisse jouer avec des données, on a besoin d'interfaces très faciles à utiliser, qui devront plaire tout autant à notre cerveau droit (l'émotionnel) qu'au gauche (le rationnel).

Ca veut dire que de nouveaux outils doivent faire leur apparition, et

qu'ils soient révolutionnaires par rapport à ce qui existe aujourd'hui. Surtout, ça signifie que l'ère des données sur tableur doit se finir, pour être remplacée par celles des données visualisées, ou, en d'autres termes, à une ère de représentation graphique.



Many Eyes, une première étape pour jouer avec les données. Ci dessus, une visualisation interactive de la répartition de l'aide US.

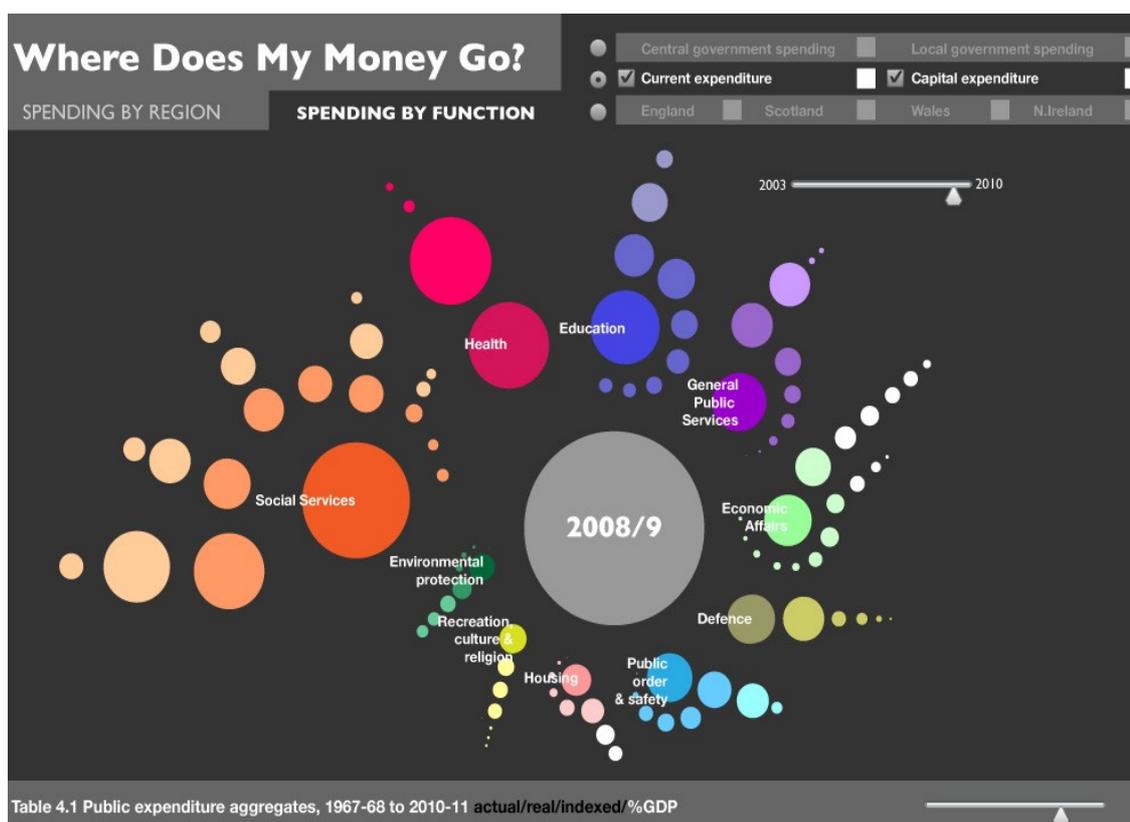
Surmonter ces trois obstacles est indispensable si l'on veut que les données soient au cœur de la prochaine révolution internet. Cette évolution nécessiterait aussi que les données soient plus acceptées culturellement parlant, c'est-à-dire que nous comprenions que ce n'est pas parce qu'une information est dans une base de données qu'elle est nécessairement difficile à comparer ou utiliser. Il serait urgent de s'attaquer à ce problème, et nous sommes confiant qu'il sera résolu très bientôt.

Billet initialement publié chez [Captaindash](#)

Photo CC [designbyfront](#) et [le o j](#)

L'EFFET PERVERS DE L'OPEN DATA PAYANT

PAR ROBIN BERJON LE 8 JUILLET 2010



Pour que les données jouent vraiment un rôle sur Internet et dans l'espace public au sens large, il reste trois problèmes à résoudre: le format des données, la facilité d'y accéder et la possibilité pour chacun de jouer avec.

La semaine dernière je suis allé à la “conférence débat” “[Les actifs immatériels publics, leviers de création de richesse et de modernisation de l'Etat](#)” de l'[APIE](#) (Agence du Patrimoine Immatériel de l'État). J'avoue que j'étais assez excité à l'idée d'aller voir de plus près ce qui se trame en matière d'Open Data en France. Je n'en ai été que plus déçu à la fois par la totale absence du débat annoncé, et par le peu de vision claire ou prometteuse fournie.

Il est difficile de tirer les grandes lignes de ce qui a été présenté dans la première partie, étant celle qui m'intéresse ici (la seconde étant sur les marques, principalement une publicité pour l'INPI, où l'on appris que les Français étaient champions du monde du nombre d'enregistrements de marque par tête, et que c'est là un indicateur fort du bon état de notre économie). La présentation n'était pas tant un débat qu'une accumulation de personnes parlant de leur propre petit bout de sujet, et ces îlots mêmes n'étaient abordés que très superficiellement.

Le choix des participants était d'entrée de jeu étrange: que faisait l'AFP — qui est indépendante et ne produit pas de données gouvernementales — autour de cette table alors que n'y étaient pas par exemple l'INSEE ou l'IGN ? Il semblerait que leur rôle était principalement d'abonder dans le sens de l'APIE en disant le plus de mal possible de la gratuité sur Internet.

Que l'AFP ne désire pas fournir ses contenus gratuitement, c'est là une chose parfaitement normale, voire louable. Elle est en effet sensée obéir à des impératifs commerciaux afin de s'assurer un

maximum d'indépendance. En outre, les problèmes de la presse par rapport à la gratuité sont bien connus, même si irrésolus, mais n'ont aucun rapport direct avec ceux de la gratuité des données gouvernementales.

L'obsession Google

La première chose qui se dégage cependant des positions des divers intervenants est une incroyable obsession vis à vis de Google. On croirait, à les entendre, que si ce n'était pour Google, la France ferait deux points de PIB annuel en plus, que les réunions de rédaction se feraient au champagne, et que d'un seul homme, d'une seule femme, l'économie française dans son intégralité, du textile aux visagistes et de la métallurgie aux boulangers, se réveillerait d'un pied radicalement innovant.

Je ne puis retranscrire ici le point auquel cette obsession m'a semblé manifeste — “Google” aurait été un mode de ponctuation oral de la langue française qu'il n'eût pas été cité plus souvent. Sur place, j'ai eu l'impression d'assister à un déraillement ferroviaire au ralenti. Je pensais à la campagne pour Paris de Françoise de Panafieu qui n'a parlé que de Bertrand Delanoë, ou à cette partie de la gauche qui ces dernières années ne discute de rien d'autre que de Sarkozy.

L'obsession est une stratégie de défaite. Je me fous, cher lecteur, du pays dont vient le plus gros ceci ou cela de telle ou telle industrie. Peu me chaut que certains membres de l'administration s'entichent d'une psychose inutile. Mais j'aimerais bien que là où j'habite, travaille, et entreprends on évite d'inhiber de par trop l'innovation au nom d'illusions protectionnistes. J'aimerais, surtout, pouvoir participer au renouvellement d'une confiance citoyenne en l'État — laquelle passe désormais par une réelle transparence de son fonctionnement.

L'Open Data payant: "un open bar où l'on paie ses coups à boire"

Et c'est là que le bât me blesse. Car le résultat de cette obsession sur Google est que l'APIE semble déterminée à faire payer pour les données gouvernementales. Oui, de l'Open Data payant. C'est un peu comme un open bar où l'on paie ses coups à boire. Non. C'est comme un open bar où l'on amène ses propres bières pour ensuite devoir les payer.

Quelle est la logique supputée de cette approche? Eh bien, comme Google n'est pas imposé en France sur la plupart de ses revenus, il faut lui faire payer pour accéder à nos données parce qu'ils pourraient les exploiter et gagner de l'argent avec. Si les facteurs en jeu se limitaient à ceux énoncés au paragraphe précédent, ça pourrait presque faire sens. Mais ça n'est pas le cas.

Lors de la conférence, tout le monde se targuait de ne surtout pas vouloir créer de barrière à l'entrée pour l'exploitation de ces données, en dépit de leur caractère payant. Une idée suggérée était de les rendre gratuites aux JEI (Jeunes Entreprises Innovantes) pendant deux ans, les faisant payer par la suite en fonction de leur chiffre. Ce genre de proposition a toujours l'air rassurant — ça tient en une phrase, ça semble énoncer une solution, ça cause jeunesse et innovation. Mais il suffit de détailler l'approche au-delà de l'idée en l'air pour y voir toutes les barrières. Premièrement, il faut être une entreprise. Ça élimine déjà tous les bricoleurs citoyens. Ensuite, il faut être une JEI. Ça élimine (à ma connaissance) tous les indépendants et les auto-entrepreneurs, toutes les PME de plus de huit ans qui pourraient vouloir s'intéresser à un nouveau domaine sans pour autant avoir beaucoup de ressources de R&D. Ça demande de passer le temps qu'il faut à obtenir le label, temps qui est une barrière en elle-même dans toute PME dont les capacités administratives sont limitées. Finalement, ça pose une barrière à des

sociétés étrangères qui pourraient faire bénéficier les citoyens français de diverses innovations en matière de transparence gouvernementale.

En bref, autant que je puisse le concevoir il est impossible de faire de l'Open Data payant sans créer une entrave à la participation.

Ceci étant, si ces barrières sont effectivement un problème pour les acteurs les plus petits, elles ne sont qu'un détail pour les plus gros. Remplir divers formulaires administratifs? On a du personnel pour ça. Payer? On a des sous pour ça. Suivant ce raisonnement, une seule conclusion s'impose: non seulement toute forme de contrôle d'accès administratif aux données gouvernementale présente une barrière pour les petits entrants, mais elle ne présente aucun problème aux gros existants.

Une politique d'Open Data payant non seulement musèle l'innovation locale, mais favorise donc directement la Némésis de l'APIE: Google. Son effet est exactement l'inverse de celui escompté.

Il faut donc mettre un terme à ces absurdités. Un véritable effort Open Data français permettrait le développement de projets locaux, avantagés précisément du fait de leur localité, et se rémunérerait tout simplement sur l'impôt prélevé sur une plus grande valeur produite. Je ne parle même pas des avantages citoyens qui découlent de ce type de projet, tant ils semblent avoir échappé à nos amis de l'APIE.

[Article initialement publié sur Berjon.com](#)

Illustration CC FlickrR par [jwyg](#)

L'OPEN DATA: UNE IDÉE DE GAUCHE?

PAR SIMON CHIGNARD LE 4 OCTOBRE 2010



L'ouverture des données publiques est un sujet d'actualité, en particulier à l'heure où des collectivités comme Rennes Métropole se lancent dans des projets concrets. Le sujet reste pourtant souvent traité sous l'angle technologique, juridique (licences) ou économique. On pourrait laisser croire que l'Open Data n'est pas une question politique, qu'elle ne relève d'aucune idéologie. Qu'en est-il véritablement ? **L'Open Data est-il plutôt une idée de gauche ou de droite ?** Peut-on placer cette idée sur l'échiquier politique ?

Pour comprendre la dimension politique de l'Open Data, il faut aller de l'autre côté de l'Atlantique. L'ouvrage collectif "[Open Gouvernement : collaboration, transparency and participation in practice](#)" regroupe les contributions des penseurs de ce mouvement, issu de la sphère Internet et technologies. Ils ont profondément inspiré la campagne présidentielle d'Obama.

Une contribution majeure « Government as a platform » est rédigée par Tim O'Reilly, l'éditeur-auteur-penseur américain à qui l'on doit le terme de Web 2.0.

Pour résumer son propos : l'Open Government – dont l'Open Data est l'un des piliers – ce n'est pas utiliser les techniques du Web 2.0 (participation, réseaux sociaux, ...) pour gérer les affaires publiques, c'est avant tout une démarche de ré-invention et de retour aux sources. Il propose une analogie avec le web des années post-bulle Internet, où le secteur paraît dévasté et incapable de se relever. A l'époque, la Silicon Valley opère un retour aux fondamentaux de l'Internet, à son « ADN primitif » c'est à dire un réseau d'égal à égal, avec une certaine symétrie entre « producteur » du contenu et « consommateur » (d'où aussi l'expression de Read/Write Web).



Ré-inventer l'action publique

Ce qui définit l'Open Government ce serait donc d'une part le constat d'une défaillance du mode de gouvernance (le fameux « Washington is broken » d'Obama) et d'autre part le nécessaire travail de [ré-invention de l'action publique](#). Or, poursuit O'Reilly, l'Etat trouve sa légitimité dans le fait qu'il y a des problèmes qui sont mieux traités au niveau collectif plutôt qu'individuel. Il cite d'ailleurs l'un des pères fondateurs de la nation américaine qui voulait faire de chaque citoyen américain un participant à l'action politique « tous les jours et pas uniquement le jour de l'élection ». (*Sur l'influence toujours présente de la pensée des pères fondateurs dans le débat américain, lire aussi l'ouvrage de la correspondante du journal Le Monde Corinne Lesnes « Aux sources de l'Amérique »*).

L'Open Government a pour programme d'appliquer aux affaires publiques les principes au coeur du mouvement Open Source (ouverture, collaboration de pair-à-pair vs. hiérarchie, ...). Le mouvement Open Government a profondément influencé la

campagne d'Obama que l'on présente d'ailleurs parfois comme le premier « We President ».

Alors, pour revenir à notre question initiale, l'Open Data, en France, est-il plutôt une idée de gauche ou de droite ?

De la gauche, l'idée d'Open Data reprend la notion de démocratie participative, de capacité du citoyen à participer à la décision publique, de notion de bien commun, de décentralisation. De la droite libérale, l'Open Data reprend la volonté de ne pas laisser à l'Etat (sous toutes ses formes) le monopole des questions et des « réponses » publiques. On retrouve aussi la croyance en la capacité du marché et du secteur privé à assumer certaines fonctions de l'État – et de quoi parle-t-on lorsqu'on donne la possibilité au secteur privé de développer des applications et services que la collectivité ne peut ou ne souhaite pas développer ?

L'Open Government donne des outils – au premier rang l'open data – pour mesurer l'efficacité de l'action publique. On connaît l'exemple du site See, Click, Fix qui permet de signaler un problème de voirie par exemple. Comme le soulignent les auteurs de la contribution « The Dark Side of Open Government » dans l'ouvrage précité, cette transparence et cette exigence de rendre des comptes (accountability) surexposent mécaniquement les défaillances et les limites de l'action publique, plutôt que les réussites ou le travail réalisé. Autre point-clé, l'open data concerne aujourd'hui essentiellement les données issues du secteur public. C'est donc sur le secteur public que la pression de la transparence est mise. Et quid de l'influence du secteur privé sur nos vies quotidiennes ? Les entreprises ne détiennent-elles pas des données dont le partage serait profitable à tous ? Les opérateurs de télécommunications par exemple, disposent de données anonymes précises sur la fréquentation des quartiers de la ville, sur le trafic sur

les grands axes routiers, ... Autant d'éléments pour mieux comprendre et améliorer la ville. On pourrait se poser les mêmes questions concernant les fournisseurs d'énergie, les entreprises de collecte des déchets, ...

Un Nicolas Hulot pour l'Open Data français ?

J'ai entendu récemment qu'il manquait à l'Open Data en France une figure politique majeure, quelqu'un qui s'empare du sujet et en devienne le porte-drapeau. Que cette « incarnation » soit indispensable à la prise de conscience politique, on peut en débattre. Le risque de voir la démarche associée à une couleur politique est-il moins fort que les gains à attendre ? L'exemple de la Grande-Bretagne est intéressant. Sous la pression du lobbying public – initié notamment par la campagne du Guardian – les principaux candidats aux dernières élections législatives se sont engagés en faveur de l'ouverture des données publiques.

Force est de constater pourtant que la notion de participation citoyenne n'est pas en ligne avec la volonté affichée au niveau national d'un pouvoir et d'un État fort et centralisateur.

L'Open Data est peut-être un objet politique non identifié au milieu des traditionnels clivages droite/gauche. Cela pose en fait la question du troisième acteur qui n'est ni l'État, ni le marché : le citoyen.

Le citoyen est-il jugé assez « compétent » pour prendre part à la décision et, in fine, à l'action publique ? L'Open Data pourra-t-il contribuer à l'émergence d'un nouveau rapport au politique ? En a-t-il l'ambition et les moyens ?

Article initialement publié sur le blog de l'[association Bug](#)

Illustration CC Flickr [opensourceway](#)

OPENDATA: DE LA MARGE À LA NORME

PAR NICOLASKAYSER-BRIL LE 5 OCTOBRE 2010



En politique, la mode était à la transparence et à [l'ouverture](#) à la fin des années 2000. Chaque candidat, dans une bonne partie des pays, a brandit les idéaux de responsabilisation et [d'évaluation](#) dans son programme.

Mais la mode change, en politique comme ailleurs, et nous devons maintenant faire en sorte que l'ouverture des données devienne la norme plutôt qu'une passade. Les intervenants du *[Personal Democracy Forum](#)* de Barcelone ont évoqué quelques pistes en ce sens.

Rentrer dans les têtes

John Wonderlich, directeur à la *[Sunlight Foundation](#)*, a mené le combat pour l'opendata aux Etats-Unis avec le succès que l'on sait, puisque l'administration Obama a mis en place – et alimente – le site [data.gov](#). Depuis 2006, il met les administrations locales et fédérales sous pression pour qu'elles libèrent leurs données. **Pourtant, aujourd'hui, il fait face à un 'reality check'**. C'est le moment où l'on va voir si l'ouverture des données à pénétré les mentalités.

Il explique comment, en préparation des élections de mi-mandat cette année, ses alliés de 2006 traînent des pieds pour libérer leurs données, maintenant qu'ils sont devenus des candidats sortants. Ils prennent peur et inventent des excuses pour ne pas que l'on puisse juger leur action à l'aune des données publiques.

La semaine dernière par exemple, un comité du Congrès lui a dit qu'ils ne pouvaient pas stocker de données sur plus de trois ans, alors qu'il ne s'agit que de quelques giga octets en plus. Soit un investissement d'une centaine de dollars, pour une institution disposant d'un budget de plusieurs milliards !

Cette mauvaise foi, qui ne surprendra pas un Européen, paraît déplacée aux Etats-Unis. Là bas, toujours selon John, le principal problème auquel le mouvement opendata a du faire face a été de convaincre les politiciens de s'engager pour publier des documents en sachant qu'ils n'auront pas de couverture médiatique. Ca paraît bien peu quand on sait qu'en Europe, le conseiller de la commissaire chargée de la société numérique ne connaît même pas les textes régissant l'ouverture des données dans sa propre organisation !



Mettre les journalistes de notre côté

Si les politiciens sont et resteront une cause perdue, le mouvement opendata doit trouver d'autres relais au sein de la société civile pour diffuser leur message et lui faire prendre racine. Une partie des ressources de la *Sunlight Foundation* sont ainsi dirigées vers une équipe de journalistes d'investigation chargés d'extraire les informations les plus intéressantes des documents publics.

Surtout, la *Sunlight Foundation* offre des formations à l'utilisation de données pour journalistes. Selon John, après ce cours, les journalistes rentrent chez eux et envoient plein de requêtes à leurs administrations locales pour obtenir des informations publiques, l'équivalent chez nous de requêtes CADA.

L'importance des journalistes a été également évoquée par **Marko Rakar**, un Croate conseiller politique le jour et blogueur politique la nuit. Il raconte son combat pour obtenir le budget de l'Etat, « le plus important document d'un pays, » selon lui. Le gouvernement croate s'est toujours refusé à communiquer les chiffres au format tableur. En effet, le gouvernement doit fournir au Parlement croate une et unique version de la proposition de budget. Le document officiel est imprimé avant d'être transmis. Pour Zagreb, seul ce document de référence – imprimé – peut être transmis aux journalistes, tous les autres n'étant que des versions de travail susceptibles de contenir des erreurs. Libre ensuite à qui veut de parcourir les 14 000 lignes budgétaires sur papier.

Marko ne s'est pas laissé démonter et s'est "procuré" une version Excel du budget. Son analyse a révélé plusieurs histoires juteuses, concernant notamment le parc automobile présidentiel et les salaires des fonctionnaires. Les médias se sont alors emparés de l'affaire, qui a, selon lui, été reprise par plus de 40 titres, y compris le journal télévisé du soir.

Convaincre la population

Le 4e pouvoir reste le soutien le plus évident de l'opendata, les données publiques contenant de nombreux sujets d'articles. Mais on sait bien que la presse périclité et que les journalistes n'ont pas tous les compétences nécessaires à l'analyse de données, quand ils ne sont pas tout simplement de mèche avec les administrations locales.

Pour pérenniser le mouvement de l'opendata, il faut que

chaque citoyen considère comme normal le droit d'accès aux documents administratifs. Que chaque fonctionnaire trouve normal qu'on puisse lui demander le montant de son salaire. Que les ministres ne soient pas offusqués lorsqu'on publie le prix de leur hôtel.

Sur ce dernier point, suite à un sondage exclusif et très peu scientifique, je suis en mesure d'affirmer que les Européens ne sont pas en retard. Alec Ross, conseiller d'Hillary Clinton, a jugé "normal" que les contribuables américains connaissent le montant de ses notes de frais, mais n'en a pas dit plus. Constantijn van Oranje-Nassau, de la Commission Européenne, a, lui, admis ne pas connaître le prix de sa chambre au Novotel, mais a affirmé payer les taxis de sa poche.

Au-delà des ministres, force est de constater que les citoyens Européens ne connaissent que très peu la législation sur l'ouverture des données. **Helen Darbshire**, directrice de l'ONG [Access Info](#), raconte qu'elle avait monté un stand, carrefour de l'Odéon à Paris, et distribué des flyers aux passants sur la loi du 17 juillet 1978. Pas un ne connaissait le texte !

L'apathie des citoyens et des contribuables reste le problème numéro un. Sans soutien populaire, les avancées effectuées sur l'opendata ne se concrétiseront jamais. **Alvaro Ortiz**, l'un des pionniers du mouvement en Espagne, le dit sans détour. Le sujet était attirant avant la crise, explique-t-il, mais la montée du chômage l'a relégué en bas de la liste des promesses politiques.

Dejan Milovac, qui dirige une [ONG luttant contre la corruption](#) au Monténégro, explique comment il a trouvé un soutien populaire pour combattre un projet de construction illégale sur le littoral. Un consortium russe a commencé à construire un projet de 200m€, sans bien sûr attendre l'approbation du permis de construire – le parrain local couvrait l'affaire.

Il a montré quels étaient les réseaux de pouvoir à l'œuvre dans

l'affaire, et comment l'argent public était détourné ou les taxes impayées. Pour lui, il est fondamental de faire comprendre aux citoyens qu'il existe bien une différence entre 1m€ et 3m€, par exemple, en expliquant ce qui peut être fait avec l'argent manquant. La mobilisation a été telle que Dejan a obtenu que les permis de construire soient désormais disponibles en ligne, afin que chacun puisse vérifier la légalité des projets de construction.

Crédits image CC Flickr par [opensourceway](#)

ENTRETIEN AVEC SIMON ROGERS, LE MONSIEUR DATA DU GUARDIAN

PAR ZIAD MAALOUF (ATELIER DES MÉDIAS) LE 14 NOVEMBRE 2010



Il y a quelques jours, L'Atelier des Médias de RFI a profité du [passage](#) par Paris de [Simon Rogers](#), "Monsieur data" au [Guardian](#) pour l'interroger sur son parcours et sur cette nouvelle tendance du journalisme : le *datajournalism* ou journalisme de données. Simon Rogers édite le [Data Blog](#) du Guardian. Il a participé pour son journal à l'exploitation des [fuites de Wikileaks sur l'Afghanistan et sur l'Irak](#). Malgré sa modestie et la simplicité avec laquelle il expose ses points de vue, c'est certainement une des principales personnalités et [un des principaux moteurs](#) du data-journalisme et de la visualisation de données dans le monde. Ci-dessous, une retranscription aussi fidèle que possible de cet entretien que vous pouvez également écouter en VO ou en VF dans les deux players sur [le site de l'Atelier](#) ou à télécharger ([VF](#) ou [VO](#)).

Ecoutez ici l'interview en VF:

[audio:http://owni.fr/files/2010/11/ADM154-1_Simon-ROGERS-VF1.mp3]

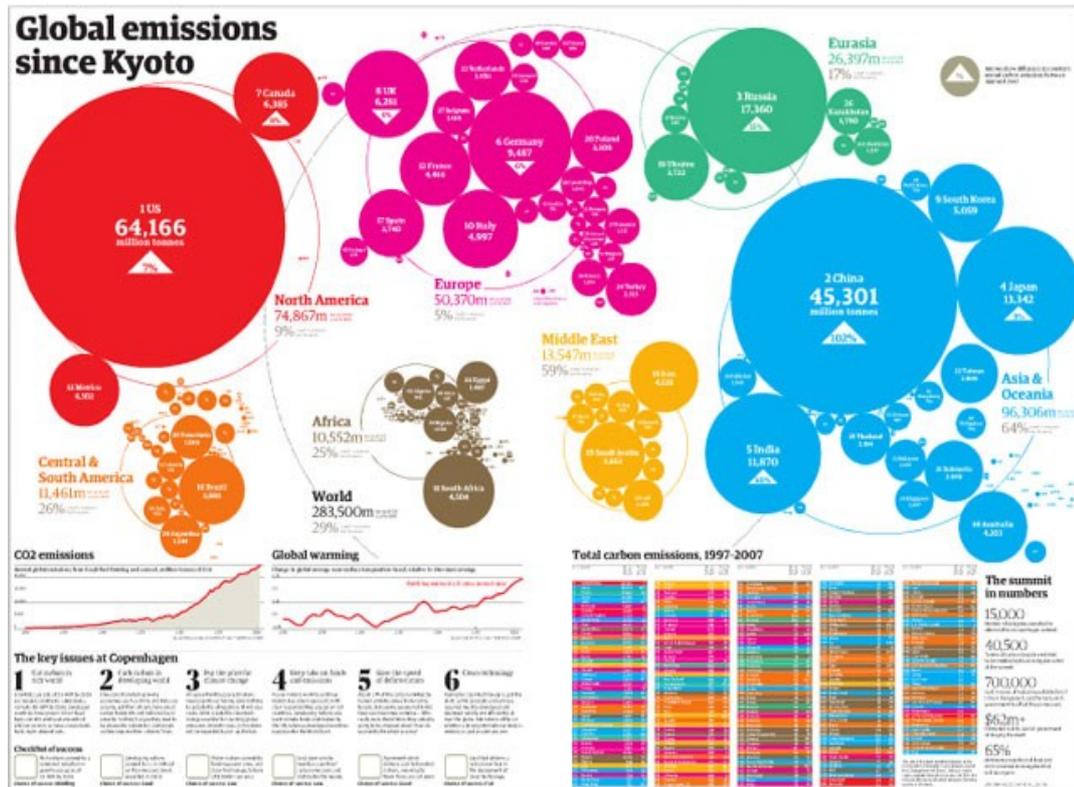
Atelier des Médias (ADM) : Pour commencer, est-ce que vous pouvez nous présenter votre blog et nous parler de votre parcours?

Ce que nous faisons c'est publier des ensembles de données brutes. On les met à disposition du monde et on demande à notre public de s'en servir, d'en faire quelque chose. Parallèlement, on dirige un service au sein du Guardian qui analyse et rend exploitable des données par nos journalistes, c'est ce qu'on appelle en fait du data-journalisme, du journalisme de données : transformer une masse de données en information, en histoire qu'on peut raconter à nos lecteurs. J'ai commencé à travailler au [Guardian le 10 septembre 2001](#), c'était mon premier jour ! Le lendemain, j'ai vu le monde devenir complètement fou. Il se passait tellement de choses en même temps qu'on a eu besoin de [produire des infographies pour les expliquer](#). Il y avait trop d'infos, de données, à gérer pour confier cela

uniquement à des graphistes. Je me suis donc retrouvé à bosser sur des graphiques avec des designers pour expliquer l'information avec des visuels. C'est comme cela que j'ai commencé à collecter des masses de données.

Par la suite, en mars 2009, le Guardian a lancé sa [plateforme ouverte \(open platform\)](#) destinées aux développeurs et aux technophiles. Elle leur permet [d'interroger les données](#) du Guardian et de réaliser [des applications](#) à partir de ces données. Par exemple, on peut fabriquer une application qui permet, quand on est dehors, d'entrer notre adresse et l'application nous dit quels sont les restaurants alentours avec les critiques du Guardian. C'est cela l'idée de la plateforme ouverte. Au moment où le journal a proposé ce service d'accès à ses données, on a pensé que ce serait bien [d'ouvrir parallèlement un data blog](#), un blog de données. On a publié quelques paquets de données pensant que cela intéresserait un petit nombre de développeurs. En fait cela a vite décollé, c'est devenu un blog qui reçoit deux fois plus de visites que le [département officiel des statistiques du Royaume Uni](#). Beaucoup de gens visitent notre site parce qu'ils cherchent des informations brutes sur, par exemple, [les émissions de carbone dans l'ensemble du pays](#) ou [quelles aides a reçu le Pakistan](#) en provenance de quel pays après les inondations. En fait, tout peut être envisagé comme des données.

On a par exemple publié tous [les mots utilisés par les Beatles dans leurs chansons](#). Combien de fois ont-ils employé le mot « love » (amour) ? 613 fois figurez-vous. Tout peut être transformé en données. Et ce qu'il y a de magique, c'est qu'une fois que vous avez transformé quelque chose en chiffres et en tableaux, les gens peuvent s'en servir pour faire des visualisations, des graphiques et des analyses.



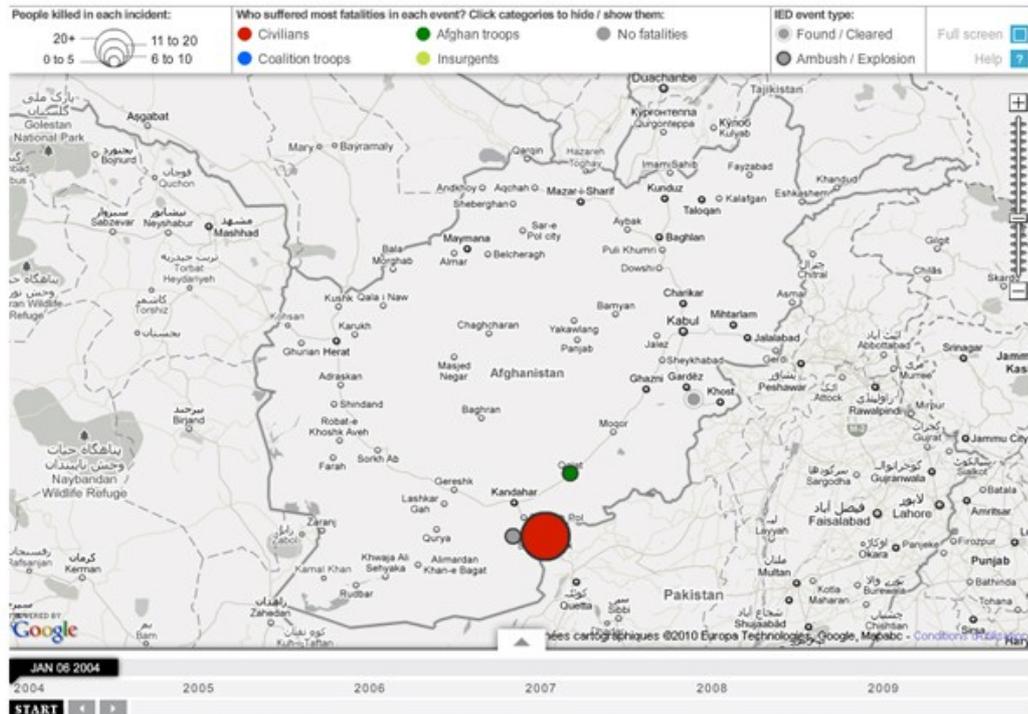
ADM : Quand on vous écoute, on a l'impression que c'est presque par chance ou par hasard que vous avez développé le journalisme de données dans votre journal, alors que c'est un phénomène qu'on voit prendre de l'ampleur dans beaucoup de médias. Comment cela s'est passé concrètement ?

En fait quand on a mis en ligne [le data blog en mars 2009](#), c'était une thématique encore très confidentielle. Après cela, le gouvernement américain a lancé [data.gov](#), et puis cela s'est étendu au monde entier. Depuis un an, de plus en plus de gouvernements publient leurs données. Ils ont maintenant une vraie pression qui les incite à rendre leurs données publiques dans des formats exploitables par tous. Plutôt que d'utiliser du .pdf, qui est le format traditionnellement utilisé par les gouvernements, on leur demande [des formats exploitables](#) comme le .csv ou .xls. En résumé, on a eu de la chance et puis il faut avouer qu'il y avait une tendance quand même. Les gens veulent de l'information brute.

Mon grand-père veut savoir [le prix du pétrole](#), où puis-je trouver cette information ? Et comment savoir si ce sont des données fiables ? Par exemple au Royaume Uni, le PIB est une donnée très importante, c'est publié tous les mois. [Si vous allez voir sur le site des statistiques officielles](#), vous allez trouver 9 ou 10 manières différentes de calculer le PIB. Comment savoir quelle est la bonne méthode ? Ça c'est une question qu'on s'est déjà posée au Guardian [à laquelle on a déjà trouvé des réponses](#) parce qu'on en a besoin, donc on partage notre savoir faire.

ADM : Comment vous voyez la relation du journalisme de données avec le journalisme en général ? Est-ce que cela doit être une partie du journalisme ou est-ce que cela doit être un peu partout ?

De mon point de vue, chaque sujet comporte une dimension « données » en lui. D'un point de vue technique, ce n'est pas si compliqué, ce doit être un outil de plus pour les journalistes. On ne doit pas envisager cela comme un machin insurmontable et flippant. C'est juste une autre manière d'accéder à l'information. En réalité, la plupart des choses que l'on fait sont très simples. On utilise des logiciels que tout le monde possède. On a tous un tableur sur notre ordinateur, [Excel](#), [OpenOffice](#) ou [quoi que ce soit](#) d'autre. C'est un outil aux capacités extraordinaires. Pas besoin d'être mathématicien ou statisticien. Tout ce qu'il faut faire c'est traiter cela comme toute autre information : il faut douter, interroger et se poser des questions de base.



Carte géolocalisée et chronologique des bombes artisanales en Afghanistan

ADM : Une autre partie du journalisme de données, c'est ce qu'on appelle la visualisation, la manière de montrer ces données. Quelle relation vous voyez entre les données et la visualisation ?

Ce qui est intéressant là dedans c'est que par le passé, on avait tendance à garder les données pour soi. Aujourd'hui, on les publie et par conséquent, notre rôle est d'aider les gens à les interpréter, à les analyser, à les montrer. Dans certains cas, un petit tableau suffit ou même la mise en avant d'un chiffre clé ou la comparaison entre un chiffre et un autre. Parfois ce simple travail est suffisant mais, dans d'autres cas, on veut vraiment montrer ces chiffres physiquement. Pour ce faire, le mariage de la visualisation et des données peut avoir un impact très fort. Par exemple, dans [le cas des fuites de wikileaks](#), on a [cartographié](#) toutes les bombes artisanales en Afghanistan de manière chronologique. Grâce à [cette visualisation](#), on comprend comment leur usage s'est multiplié et pourquoi c'est devenu un paramètre essentiel de la guerre. Seule la visualisation permet de montrer cela.

Maintenant, est-ce qu'il faut être un graphiste pour faire des visualisations ? Il y a plusieurs approches. Il y a des personnes qui sont avant tout journalistes et qui réalisent des visualisations comme [David McCandless](#) et son « [Information Is Beautiful](#) ». Il a une formation de journaliste, il n'est pas graphiste mais il a produit [de très belles visualisations](#) qui expliquent bien les choses. En revanche, il arrive qu'on ait vraiment besoin d'un graphiste pour réaliser une visualisation complexe.

Il existe, par ailleurs, des outils très simples, accessibles en ligne, qui permettent de réaliser des visuels : [Manyeyes](#), [Timetric](#) et même [Google propose des outils de visualisation](#) que chacun peut utiliser pour raconter une histoire.

ADM : Pour clore cet entretien Simon Rogers, je voudrais évoquer avec vous deux tendances assez fortes aujourd'hui. La première concerne l'Opendata, on parle aussi d'open government, c'est un mouvement qui incite les institutions et les gouvernements à ouvrir et rendre publiques le plus de données possibles. Qu'est-ce que vous pensez de cela ? Quelles sont vos attentes ?

C'est une évolution très intéressante. Il y a encore un an, on militait pour accéder aux données et aujourd'hui on croule sous les données que nous recevons. Par exemple, au Royaume Uni, le Trésor avait une base de données importante appelée [COINS](#) qui répertorie tout ce qui est dépensé par le gouvernement, dans le moindre détail, il y a des millions d'informations. Le nouveau gouvernement a rendu ces données publiques, [on avait beaucoup milité pour ça](#). Il y a un an, on voulait accéder à ces infos, aujourd'hui on les a et il s'agit maintenant d'aider les gens à s'en servir, à les analyser. Donc [ce qu'on a fait avec la base de données COINS](#), on a choisi le meilleur format, le .csv. On a choisi la meilleure période et les départements les plus faciles à comprendre et on a [rendu l'ensemble plus intelligible](#).

Pour moi, le rôle des journalistes va, de plus en plus, ressembler à ça.

des sociétés ?

Ce qu'il faut comprendre c'est que les gouvernants n'ont rien à perdre. Aux USA, au Royaume Uni, en Australie ou en Nouvelle-Zélande, le monde ne s'est pas écroulé parce que les données gouvernementales ont été rendues publiques ! Cela a tout simplement rendu les choses plus ouvertes et plus transparentes à une époque où on ne fait plus confiance aux politiciens, on ne fait plus confiance à la politique. Vous voulez qu'on vous fasse confiance ? Il faut être ouvert. Rendre ses données publiques c'est essentiel pour cela, il faut le faire dans un format pratique pour encourager les gens à s'investir.

Quand Data.gov.uk a été mis en ligne, ils ont impliqués plein de développeurs pour permettre aux gens d'accéder à cette masse d'infos, pour moi c'est le rôle du gouvernement de faire ça. Il y a beaucoup d'exemples aujourd'hui que [ceux qui militent pour l'ouverture des données](#) peuvent utiliser pour montrer que les pays ne se sont pas effondrés à cause de l'open data. L'open data ne peut que renforcer la foi des peuples dans leurs gouvernants.

Everything you need to know about today's cuts

Today is the big day: George Osborne's comprehensive spending review. This is your pull-out guide to what to look out for, the key numbers and how it will affect you

How far public spending changes
 The Chancellor has announced a 38.8% cut in public spending over the next three years. This is a percentage of GDP. Other countries may be able to reduce public spending by a similar amount.

What is the deficit?
 The deficit is the difference between what the nation has spent in a year and what it has taken in. The receipts have gone down with the recession and public spending has gone up - which is why government has had to borrow more money and widen the deficit.

The UK's deficit for 2009-10 was
£106,369,000,000

How much is being cut by 2014?
 In his emergency budget, Osborne said that total public spending will be cut by £49bn over the next three years. This is a total of £131bn over the next five years. This is a total of £131bn over the next five years.

How big is Britain's public sector?
 The public sector is the part of the economy that is owned and controlled by the state. It includes the NHS, the police, the fire service, and the education system.

How big is Britain's total debt?
 The total debt of the UK is £13,187.79 billion. This is the total amount of money that the government has borrowed from the public and abroad.

How do the UK's deficit and debt compare?
 The deficit is the amount of money that the government has borrowed in a year. Debt is the total amount of money that the government has borrowed over its lifetime.

Five things Osborne might say ... and what he really means

The chancellor's deficit bingo

The big questions

Follow the speech guardian.co.uk/politics/spending-review

ADM: Pour finir, je voudrais parler avec vous de [Wikileaks](#), ce site qui se sert de lanceurs d'alertes, de personnes qui divulguent des informations secrètes dans le monde et qui les diffuse. Vous avez été impliqués dans ce qu'on appelle les War logs en Afghanistan (et en Irak), qu'est ce que vous pensez de cela ?

C'est une nouvelle tendance du journalisme. Les journalistes vont devoir de plus en plus gérer des masses de données et trouver des infos et des histoires dedans. A l'avenir, un journaliste ne pourra plus dire « je ne m'occupe pas de tableurs... », cela fera vraiment partie de son boulot. Au Guardian, cela a vraiment changé notre manière de travailler. Aujourd'hui les données sont une partie intégrante de notre job. Maintenant, je pense que les infos de Wikileaks sont excellentes pour comprendre ce qui se passe. Si vous regardez les rapports divulgués sur l'Afghanistan, ils montrent à quel point cette guerre est difficile. Cela raconte notamment comment des gens tentent d'apporter de l'aide humanitaire dans un village et cela n'intéresse personne, parce que les villageois ont peur de la corruption etc...Il y a ce genre d'infos où on voit que les militaires se sentent responsables du bien être des populations locales. On finit en fait par se demander pourquoi l'armée ne publie pas ces rapports.

Il n'y a rien de sensationnel, ça ne change même pas la manière dont on perçoit cette guerre. Cela fait en tous cas de la guerre en Afghanistan [une des plus documentées de l'histoire](#), grâce à ces fuites. Est-ce que cela a nuit aux États-Unis ? Pas vraiment je pense. Le gouvernement américain nous a aidé à [retirer les noms cités](#) dans les rapports pour ne pas mettre les personnes en danger.

ADM : Et sur le futur du secret, des choses secrètes, est-ce que cela pose une vraie question pour vous ?

Cela va être intéressant de voir comment les autorités gèrent cela. Il y a eu des menaces mais rien de très concret. Et puis il va y avoir bientôt de nouvelles lois de transparence et de [protection du journalisme en Islande](#). Cela va devenir un abri pour les journalistes. Sur le même modèle que les paradis fiscaux, on aura des paradis journalistiques, des lieux protégés pour la liberté sur internet. Les gens iront là-bas pour publier des données sensibles.

En tous cas, Wikileaks a complètement changé le rapport que l'on a avec les documents que l'on produit, géographiquement notamment. Je pense en tous cas qu'à l'avenir ce type de fuites ne va faire que se multiplier.

>> Interview initialement publiée sur l'[Atelier des Media de RFI](#).

>> Crédit photo : [Jessica Chekroun](#) (la photo n'est pas en Creative Commons)

[MAJ] LA LOPPSI, CONTRE L'OPEN DATA?

PAR OLIVIER TESQUET LE 22 NOVEMBRE 2010



Bug législatif

Pour les profanes du langage législatif, la loi du 17 juillet 1978 est le texte qui permet à tout un chacun, qu'il soit journaliste ou citoyen, d'accéder aux données publiques. Or, [comme le relève Regards Citoyens](#), l'argument sécuritaire brandi par les défenseurs de la LOPPSI pourrait bien menacer cet acquis, en affaiblissant encore un peu plus la Commission d'accès aux documents administratifs ([CADA](#)). *“L'OpenData est-elle un produit dangereux pour le gouvernement?”*, s'interroge le collectif, avant d'évoquer *“un bug législatif [...] qui vient non seulement s'opposer au droit communautaire, mais réalise une grave confusion entre les données publiques accessibles accessibles via la CADA et les données personnelles régies par la CNIL”*.

En substance, le gouvernement pourrait ainsi être tenté de restreindre l'accès à ses documents sous prétexte de vouloir préserver des informations personnelles, telles que les plaques d'immatriculation, alors que, selon [Regards Citoyens](#), *“les lois existantes permettent déjà un niveau de contrôle”*. Dans ces conditions, pas étonnant qu'une telle discrimination soit perçue par les promoteurs de l'open data comme *“une atteinte à la liberté d'information”*.

Dans un registre plus large, quelques heures avant le nouvel examen du projet de loi, une trentaine d'organisations [se mobilisent](#) pour tenter de stopper sa mise en place. Rassemblées sous la bannière du collectif Liberté Égalité Justice, le Droit au logement, la Quadrature du Net ou la Ligue des droits de l'homme, mais aussi le Syndicat de la magistrature ou le Parti communiste, dénoncent un *“fatras sécuritaires”* et appellent à [signer un textesans](#) équivoque: *“LOPPSI: pas en notre nom!”*

La loi controversée sera débattue dans l'hémicycle jusqu'au 26

novembre.

—
Crédits photo: Flickr CC [powazny](#), [Muffet](#)

MAJ du 21 décembre 2010: Alors que la Loppsi est [étudiée en seconde lecture](#) à l'Assemblée Nationale depuis le 15 décembre, les quelques lignes pointées avec inquiétudes par les défenseurs de la libération des données ont visiblement été remarquées par les législateurs. Ainsi, le collectif [Regards Citoyens note](#) que le texte "épargne l'Open Data" en introduisant un amendement "qui rectifie le tir". Néanmoins, si la référence à la loi CADA de juillet 78 est supprimée, une faiblesse subsiste dans l'interprétation de la notion de "données personnelles": l'amalgame entre données privées et données publiques.

TIM BERNERS-LEE: "LES GOUVERNEMENTS DEVRAIENT ENCOURAGER L'OUVERTURE DES DONNÉES"

PAR CYRILLE CHAUSSON (LEMAGIT) LE 1 DÉCEMBRE 2010



A l'occasion d'une conférence annuelle du W3C qui s'est tenue le 2 novembre dernier à Lyon, la rédaction du MagIT a rencontré Tim Berners-Lee, le père du Web et un des patrons du consortium. Au programme, le Web sémantique, l'ouverture des données, HTML 5 et la fondation W3. Tim Berners-Lee livre aussi sa réponse à [l'article de Wired](#) ("Le Web est mort. Vive Internet").

Le Web sémantique évolue-t-il au rythme que vous espérez ?

Tim Berners-Lee : Je n'avais pas de prévisions à proprement dit. On ne peut pas prédire le futur et savoir véritablement comment les choses évoluent. Je pense en revanche que le Web sémantique s'inscrit dans une croissance exponentielle. Il y a quelques années, lorsque nous avons commencé à parler des Linked Data [un concept qui consiste à relier les données entre elles via le Web et non plus uniquement des pages - le principe même du Web sémantique, NDLR], les gens ont été quelque peu surpris car ils avaient une idée très simple de ce que cela pouvait être, illustrant du coup la nature essentielle du web sémantique. Il existe de très bons et nombreux outils autour du Web sémantique. Les Linked Data, qui demeurent très importantes pour les données publiques des gouvernements, sont une idée relativement simple qui génèrent beaucoup d'enthousiasme.

L'approche sémantique émerge également sur le desktop, comme [Nepomuk](#) (un projet de desktop sémantique qui fait une première apparition dans Mandriva 2010), peut être plus rapidement sur le Web. Pourquoi ?

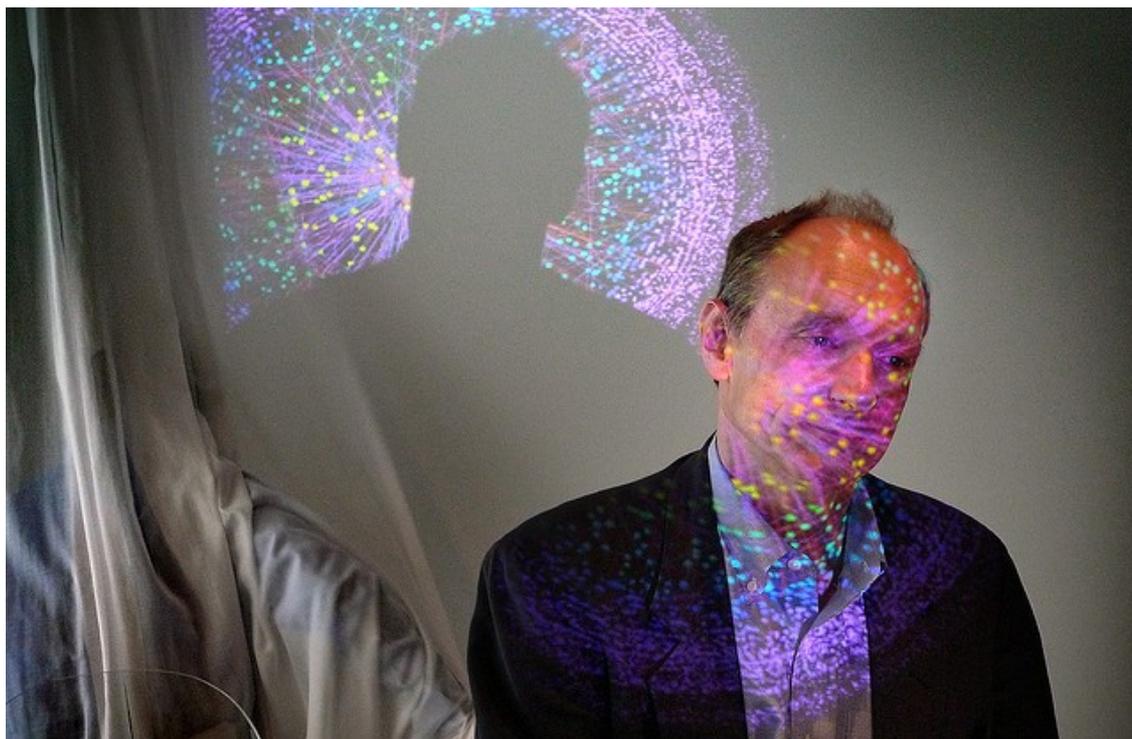
TBL : Ce qui est intéressant dans les Linked Data, c'est qu'il s'agit d'une technologie d'intégration très puissante, qui permet notamment d'intégrer tout forme de technologie : des bases de données, des fichiers XML, des agendas, voire des cartes de visites électroniques. Les bénéfices de cette approche? Avoir sous la main votre vie, tout votre environnement où toutes les données

importantes sont connectées entre elles. Avec Nepomuk, ils ont réalisé cette prouesse sur les données privées. Et les données privées sont une partie capitale du Web sémantique. J'utilise moi-même le principe des Linked Data à titre personnel, et il existe probablement d'autres personnes qui le font, mais on ne le sait pas. Et c'est la même chose en entreprise : les données personnelles de l'entreprise sont manipulées à l'intérieur du firewall. On ne sait donc pas que le Web sémantique est finalement utilisé. Les journalistes n'écrivent donc pas dessus. En revanche, ils s'intéressent aux Open Data Cloud, un *cloud* de données ouvertes pour les gouvernements par exemple. Il existe tout un spectre entre ce qui est public et ce qui est personnel. [...] Les entreprises peuvent copier ses données et les utiliser dans leur systeme d'information.

Quel est aujourd'hui le niveau de maturité des outils en place ?

TBL : Quand on considère le Web sémantique, développer de nouveaux outils est toujours fascinant, tout se connecte et s'auto-alimente. Développer les Open Linked Data est une de nos principales priorités, car il existe tellement de données éparpillées et que grâce à ce procédé, seule la valeur de la donnée compte. Et pas seulement pour les gouvernements, mais également pour la chaîne de valeur par exemple. Il existe une véritable valeur à publier ses données.

Toutefois, je pense qu'il existe encore un marché pour des outils plus puissants. Dans le domaine des interfaces utilisateurs, des outils qui permettent de naviguer dans ces données [...] et de mieux les présenter. Il existe une véritable valeur à ouvrir les données, mais une fois ces données connectées entre elles, comment l'utilisateur doit les explorer, comment lui faire passer l'information, pour ensuite produire une analyse pertinente, selon les besoins et les métiers. Les *spreadsheets* existent bien mais ne sont pas suffisamment performants pour donner du sens à un pool de données.



Tim Berners-Lee, prophète de l'open data?

Les gouvernements européens devraient-ils davantage exposer leurs données publiques - ce qui est très lié aux projets d'e-administration ?

TBL : Oui bien sûr, ils devraient encourager cette ouverture car tout le monde y est gagnant et cela apporte de meilleurs services aux citoyens. Les gouvernements sont parvenus à résoudre les problèmes de production des données, ils doivent aujourd'hui aller plus loin. [...] Donc oui je pousserais les gouvernements à développer les données publiques sur le Web. Mais ces données, liées aux fonctionnements du pays, ne sont qu'une partie de l'Open Data Cloud. Des données géographiques, sur les CDs, les artistes, les découvertes en matière de santé: il existe un grand nombre de type de données prêtes à être publiées sur le Web.

Le degré d'ouverture est-il aujourd'hui suffisant ?

TBL : En France, il y a le cadastre par exemple. La Commission européenne est également très intéressée par l'ouverture des données, pour notamment informer sur l'état des pays en cas de

voyage dans Union européenne. En 2009, aux Etats-Unis comme au Royaume Uni, des projets comme data.gov ont été un vrai succès, engageant une compétition transatlantique. [...] Ces projets sont certes globaux, mais les données locales sont également très importantes, regardez le cadastre en France. Cette classification n'est pas implantée dans tous les pays. Certains concepts sont nationaux, d'autres mondiaux. [...]

Quel conseil donnerez-vous aux développeurs Web concernant HTML 5 ? Attendez-vous que les spécifications soient finalisées pour les intégrer dans leurs développements, ou y passer dès maintenant ?

TBL : Ils doivent dès aujourd'hui s'intéresser à HTML 5. Il est vrai que les travaux sont encore en cours de standardisation, certaines spécifications sont plus stables que d'autres, mais un bon développeur écoute toujours où se portent les débats, où en sont les travaux du W3C, quels éléments sont susceptibles de changer. Il est important de comprendre l'étendue des possibilités de HTML 5. Il est également nécessaire de comprendre le calendrier d'implémentation des fonctions dans les différents navigateurs. Mais HTML 5 est conçu pour rester compatible avec l'existant. Donc voici mes conseils : apprenez le, utilisez le, expérimentez. Vous comprendrez également vers quelle direction se dirige le standard, et si vous le souhaitez, prenez part au débat du groupe de travail au W3C. Les travaux sont certes en cours, mais des éléments de HTML 5 sont aujourd'hui largement stables.

Comment se porte aujourd'hui votre fondation W3 ?

TBL : Aujourd'hui, 20% seulement de la population mondiale utilise le Web. Il en reste encore 80%. L'avons-nous développé uniquement pour nous ? A-t-il été développé par les pays développés pour les pays développés ? Nous avons donc décidé de monter cette fondation pour répondre à cette question, dont la portée est plutôt très large. Nous avons été très chanceux d'obtenir une première subvention sur cinq

ans de la part de la [Knight Foundation](#), pour démarrer. Nous avons commencé à regarder comment combler les fossés qui existent entre ceux qui ont accès au web et ceux qui ne l'ont pas. Et il existe des résultats intéressants. Dans les pays en développement, on mise en premier sur l'apport de l'eau, des soins médicaux puis on favorise les développements économiques. Mais Internet est considéré comme un luxe. Pourtant, il ne doit pas être pris de la sorte: il doit être vu comme un élément indispensable auquel les gens doivent avoir accès. Avec une connexion bas débit, ils peuvent trouver un emploi, obtenir de l'information sur leur environnement, l'utiliser pour partager des informations sur la santé. L'idée est que ces personnes entrent dans "l'Internet Society". La Web Foundation travaille en ce sens. Nous nous concentrons aujourd'hui sur l'Afrique, mais la portée est bien plus vaste. Et portera également sur l'accès à Internet dans les pays développés.[...]

LeMagIT : Le magazine [Wired a décrété la mort du Web](#), invoquant notamment le schéma des applications mobiles qui ne donnent plus accès au Web et aux pages HTML, mais aux seules applications qui sont elles-mêmes connectées à Internet. Quelle est votre position ?

http://www.youtube.com/watch?feature=player_embedded&v=FjqG7cZdSjY

[Article initialement publié sur LeMagIT](#), sous le titre "Tim Berners-Lee : 'les développeurs doivent s'intéresser dès aujourd'hui à HTML 5'"

Illustration CC FlickrR par [Dunechaser](#), [Documentally](#)

DONNÉES LIBÉRÉES, CHERCHEURS DÉBRIDÉS, SOCIÉTÉ IMPLIQUÉE

PAR ANTOINE BLANCHARD LE 8 DÉCEMBRE 2010



Jusqu'à très récemment, les données étaient les parents pauvres de la recherche scientifique, particulièrement en biologie. S'accumulant dans les laboratoires et les centres de séquençage du génome, isolées sur le disque dur des chercheurs, elles étaient invisibles et difficilement accessibles — enfouies sous la montagne d'articles scientifiques auxquels elles contribuent à donner naissance.

Et si les acteurs ont pris conscience de la nécessité de libérer leurs publications, notamment à travers l'[accès libre aux résultats de la recherche](#) (*open access*), ils avaient encore négligé leurs données.

Données libérées

Mais que sont exactement ces données ? Ce sont les résultats des expériences ou observations menées par les chercheurs, préludes à de nouvelles découvertes, à de nouvelles théories explicatives et à des publications scientifiques. Depuis la Révolution scientifique, elles étaient conservées consciencieusement afin de garantir la transparence inhérente à la recherche, mais rarement mises en commun. À ce titre, les publications scientifiques faisaient état des résultats de l'analyse de ces données sans jamais mettre les données elles-mêmes, brutes, à disposition. Ou quand elles le faisaient, elles étaient "emprisonnées" dans le format étroit et inutilisable du papier. Seuls des contacts informels entre chercheurs permettaient de s'échanger des données qu'Internet, désormais, permet de mettre à disposition de tous d'un seul clic.

Avec Internet, nous assistons en effet au développement de l'e-Science. Sous ce terme se cache un phénomène général d'appropriation par les chercheurs des technologies de l'information afin de [démultiplier les possibilités d'analyse, de stockage, de publication et de partage des données](#) mais aussi des articles, et autres résultats de la recherche ; sous ce nom se cache aussi des

programmes institutionnels de promotion de cette nouvelle ère de la science, comme en Allemagne ou en Grande-Bretagne ([National e-Science Center](#)). C'est ainsi que l'on voit grossir les rangs des bases de données accessibles sur Internet, des chercheurs qui stockent leurs articles en accès-libre sur le serveur de leur laboratoire, des logiciels libres (open source) qui permettent d'exploiter ou d'analyser les résultats d'une puce ADN ou d'une analyse phylogénétique.

Ce mouvement a véritablement décollé lorsque les généticiens du Consortium international de séquençage du génome humain se sont réunis aux Bermudes en février 1996 pour convenir, ensemble, du mode de publication de la séquence du génome humain. Ils se mirent d'accord sur la diffusion automatique (et si possible sous 24h) des séquences de plus de mille bases, la publication immédiate des séquences annotées, le tout en accès libre. Ces "principes des Bermudes" régissent depuis lors (avec quelques modifications) les projets de séquençage financés sur fonds publics et sont les fondements de la publication de données génomiques sur Internet.



Dans la perspective de l'e-Science, les données autant que les articles doivent être ouverts, libérés. C'est d'une importance croissante avec la collaboration d'équipes à travers le monde entier, dans des projets pharaoniques comme l'étude du virus de la grippe aviaire H5N1. D'autant que les données nécessaires sont rapidement volumineuses, comme l'ont appris à leurs dépens les chercheurs en génétique, neurosciences et plus largement des sciences de la complexité.

Parmi elles, la biologie intégrative cherche à corréliser les données issues de différents niveaux d'observation et donc à croiser des données tierces pour les mettre en perspective, les fouiller par des méthodes de fouille de données (data mining) etc. Cette nécessaire libération des données se joue à trois niveaux : l'accessibilité des données, qu'encouragent notamment les revues qui imposent aux auteurs de déposer leurs données — moléculaires, génétiques, écologiques... — dans les bases de données disponibles sur Internet, et qui s'opposent parfois à l'appropriation de droits de propriété intellectuelle ; l'inter-compatibilité des nombreuses bases de données et l'existence de métadonnées standardisées permettant de connaître leur origine, le contexte de leur obtention, etc., défendue par les bio-informaticiens ; et enfin, l'absence de frein légal à la réutilisation des données, par exemple en utilisant une licence libre de type [Science commons](#) comme le fait UniProt, la plus grande base de données sur les protéines.

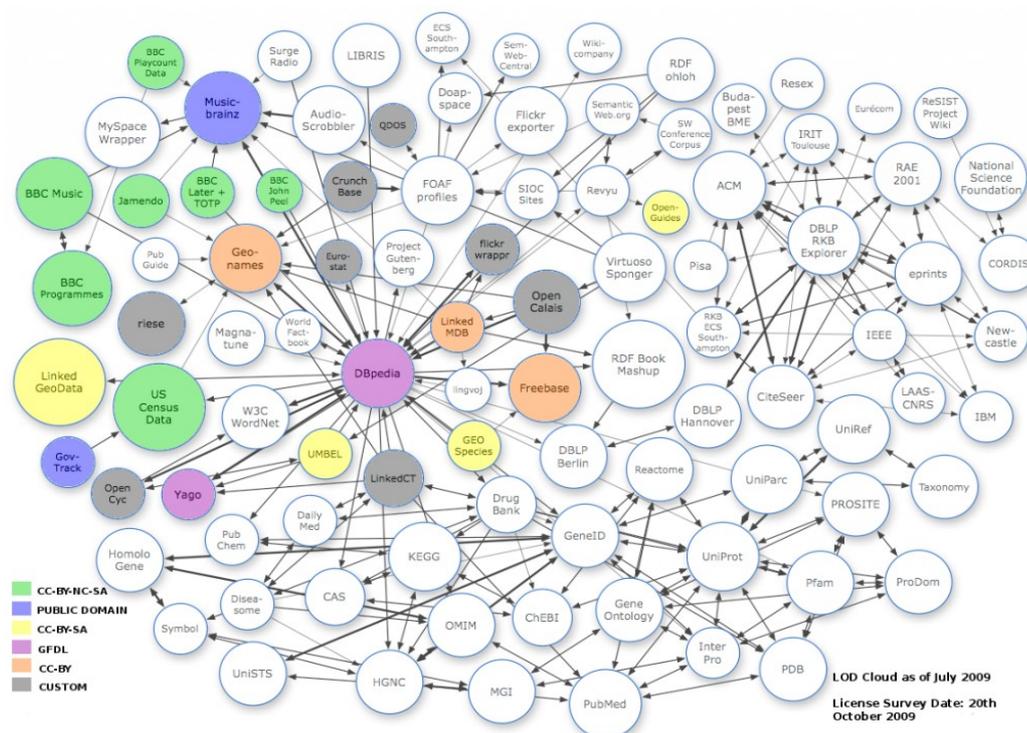
Chercheurs débridés

Les chercheurs, acteurs de ce mouvement, en sont aussi les premiers bénéficiaires et ont y vite vu l'occasion de débrider leur créativité, comme le montrent de nombreux exemples récents. En 2005, le lancement de la base de données de structures chimiques ouverte [PubChem](#) gérée par la National Library of Medicine

américaine, a augmenté la visibilité et l'accessibilité de ce type de données sur Internet. Une aubaine pour le chercheur qui dispose ainsi d'une offre plus grande, en terme de contenu et de moyen d'accéder à l'information qui l'intéresse. Ou encore, la mise à disposition du logiciel d'images satellitaires et données géographiques Google Earth — qui couvre toute la planète avec une excellente ergonomie et précision — a inspiré de très nombreuses utilisations originales. Le journaliste de Nature Declan Butler a par exemple construit une [représentation spatiale des foyers de grippe aviaire](#) (fichier .kml), en croisant la puissance de Google Earth avec les données de l'Organisation mondiale de la santé et de l'Organisation des Nations Unies pour l'alimentation et l'agriculture.

On nomme cette intégration automatisée de données complémentaires en provenance de plusieurs sources le "mash-up", sur lequel parient aussi les écologues avec leur projet [iSpecies](#) qui intègre, pour chaque espèce recherchée, des données bibliographiques issues de Google Scholar, des données iconographiques issues de Yahoo Images et des données taxinomiques du National Center for Biotechnology Information. Le mash-up autorise aussi l'intégration des disciplines et échelles biologiques pour mieux comprendre des systèmes biologiques complexes comme le cerveau ou la biodiversité ; ces deux thématiques font l'objet d'une politique d'incitation dans ce sens, sous l'impulsion d'un rapport d'un groupe de travail de l'OCDE. Dans le second cas, cela se concrétise notamment à travers l'initiative [Global Biodiversity Information Facility](#).

Opendata - OWNIXMAS 2010



L'ouverture des bases de données et la standardisation des normes de stockage et d'annotation permet d'envisager la fouille de données, utile pour faire parler des données volumineuses et en tirer des informations complètement nouvelles, par exemple sur les relations entre gènes et maladie, facteurs environnementaux et maladie ou autres. Sans compter que ces pratiques émergentes vont se développer en même temps qu'elles se banaliseront et que les outils existants s'y adapteront.

Enfin, l'existence de bases de données ouvertes est aussi un espoir pour les chercheurs des pays en voie de développement qui ne peuvent se payer l'accès aux bases de données privées et payantes. Ainsi, l'apparition du PubChem comme possible alternative gratuite à l'incalculable base Chemical Abstracts a été vécue comme un soulagement dans ces pays, et combattue avec vigueur par l'American Chemical Society qui gère Chemical Abstracts.

Société impliquée

De manière plus inattendue, l'ouverture des données issues de la recherche peut contribuer à impliquer de nouveau la société civile dans la science. D'abord parce que c'est une manière de lui retourner les résultats de la recherche qu'elle finance indirectement par ses impôts. Ensuite parce que c'est un moyen d'intéresser le public à la science, comme le suggère l'engouement incroyable autour du logiciel Google Earth et son appropriation par le jeu des mash-ups (voir des exemples sur le [Google Earth Blog](#) ou [Ogle Earth](#)). À ce titre, la [transposition par la France](#) (avec plus d'un an de retard) de la [directive européenne Inspire](#) qui prévoit l'ouverture des données géospatiales est révélatrice d'un changement de mentalités.

On peut y voir aussi un facteur d'implication des citoyens dans les processus de décision de la recherche et de l'avènement d'une démocratie scientifique. Dans un modèle de science entièrement ouverte comme l'expérimente le forum [Synaptic Leap](#), des chercheurs du privé et du public travaillent ensemble sur des médicaments contre certaines maladies tropicales négligées, notamment le paludisme et les schistosomoses ; tout le projet est mené en ligne, les données y sont publiées et les résultats sont sous licence libre. Et les citoyens de proposer des thèmes de recherche, [comme la chikungunya](#).

Ou encore, s'ils s'engagent dans des débats de plus en plus complexes scientifiquement, ils gagnent à avoir accès aux mêmes données que les chercheurs — à condition qu'ils aient la culture scientifique nécessaire à une bonne interprétation de ces données. [L'affaire du "climategate"](#) marquée par le piratage des serveurs d'un laboratoire de recherche sur le climat en Grande-Bretagne, a montré que la

science avait plus à perdre — notamment en terme de confiance — à cacher ses données qu'à les rendre publiques. Après avoir refusé à plusieurs reprises d'obtempérer à des [demandes relevant du Freedom of Information Act](#), ce laboratoire [a lancé une réflexion](#) sur l'ouverture de ses données, en partenariat avec le British Atmospheric Data Centre, habitué de la publication des données sur le climat.

Finalement, après l'accès libre aux articles scientifiques qui se dessine depuis le début des années 2000, l'accès libre aux données de la recherche est la deuxième étape de cette ouverture de la science en marche. Il ne fait aucun doute que ce mouvement va s'amplifier et s'imposer dans le futur, et il appartient aux chercheurs mais aussi aux citoyens de le soutenir. L'étape suivante sera l'accès libre aux technologies — à commencer par les biotechnologies —, mise en commun nécessaire pour sortir du piège du "tout-brevet". Cette étape s'ébauche déjà à travers quelques projets et devrait avoir *in fine* un impact considérable sur la science mais aussi l'économie et la société.

Pour en savoir plus

- Brooksbank, Cath et Quackenbush, John. [Data standards: a call to action](#). *OMICS: A Journal of Integrative Biology*. 10(2): 94-99, juin 2006
- Butler, Declan. Mashups mix data into global service. *Nature*. 439:6-7, 5 janvier 2006
- Commission européenne, [Riding the wave - How Europe can gain from the rising tide of scientific data](#), octobre 2010
- Cragin, Melissa. [Toward Integrative Science: Organizing Biodiversity and Neuroscience Data](#). *Bulletin of the American Society for Information Science and Technology*. 30 (1),

Opendata - OWNIXMAS 2010

octobre/novembre 2003

- OECD Megascience Forum Working Group on Biological Informatics. [Final Report of the OECD Megascience Forum Working Group on Biological Informatics](#), janvier 1999
- Rumble Jr, John et al. [Developing and Using Standards for Data and Information in Science and Technology](#). Congrès Ensuring Long-term Preservation and Adding Value to Scientific and Technical Data, The Royal Society (Edinburgh), 21-23 novembre 2005

>> [Article initialement publié sur le blog "Mon bouillon"](#)

>> Illustration CC [Unhindered by Talent](#), [ldodds](#) et [Jerrycharlotte](#)

>> Retrouvez tous les articles d'[OWNIsciences](#)



OPEN DATA
OPEN DATA
OPEN DATA

LIBERTIC
MICHELE BATTISTI
GILLES BABINET
(TRADUCTION MKB)
ROBIN BERJON
SIMON CHIGNARD
NICOLAS KAYSER-BRIL
ZIAD MARLOUF
(ATELIER DES MEDIAS)
OLIVIER TESQUET
CYRILLE CHAUSSON (LEMAGIT)
ANTOINE BLANCHARD

Pulp
OWNI
Digest

CHRISTMAS · EDITION · OWNI



WWW.OWNI.FR

CHRISTMAS · EDITION · OWNI · 12/12 / 2010