

Le web des données laisse-t-il une place au bien commun ?

Le web entre dans une troisième phase de son existence. Né tout d'abord comme un espace d'exposition déterritorialisé, donnant à voir des informations, le web est dans un second temps devenu un espace de coproduction, dans lequel tout un chacun est libre de s'exprimer, de partager, de commenter, de créer, de penser à voix haute. Aujourd'hui le web se transforme en un gigantesque réservoir de données numériques. C'est sur le web que s'accumulent et se croisent les savoirs de l'humanité — savoirs savants et savoirs profanes —, mais aussi ses créations, ses imaginaires, ses dialogues, ses émotions... dessinant ainsi les nouveaux contours de cette « société de la connaissance », en émergence.

UN RUISSELLEMENT DE DONNEES AUX ORIGINES MULTIPLES

Associer web et données peut sembler réducteur : comme si tous ces contenus qui font aujourd'hui la richesse du web pouvaient être réduits à des entités alignées, rangées, manipulables, des « bases de données ». Comme si nous avions oublié que le web est devenu un lieu d'interactions, de création de lien social et d'échanges. En réalité les deux approches ne sont pas antinomiques, bien au contraire. Parler de web des données, c'est faire un double constat : d'une part, celui de l'existence d'une masse en croissance exponentielle d'informations disponibles sur la toile sous des formes multiples (à titre indicatif, on estime que pour gérer ces informations, Google est passé de 40 000 serveurs en 2004 à 1 million en 2007, 2 millions en 2008 et on parle de 10 millions dans un immense nuage de serveurs en 2010 ; Facebook seul mobiliserait 30 000 serveurs en octobre 2009) ; et d'autre part, celui d'une difficulté corrélée à exploiter — chercher, trouver, croiser — ces données dans toute leur profondeur et leur richesse. Poser le web des données comme objet de réflexion, c'est se donner les moyens, techniques et politiques, de continuer à l'améliorer, à le faire gagner en créativité.

Autant qu'à la massification, cette irruption des données sur le devant de la scène du web est liée à un saut qualitatif : les sources de données qui viennent se jeter dans le web sont en train de se diversifier à vitesse accélérée. Données publiques, fournies par les gouvernements, les collectivités locales, etc. ; données scientifiques produites par les chercheurs ; données autoproduites par les usagers du web eux-mêmes ; données issues des activités du secteur privé ; métadonnées attachées et qualifiant les documents numériques ; enfin demain données générées par les machines et les objets du monde réel... la diversité croise la profusion.

Ces données numériques nous interpellent politiquement et économiquement : qui peut les utiliser ? À quelles conditions ? Pour quelles finalités ? Qui les contrôle ? Comment assurer leur croisement et leur réutilisation, source d'innovation renouvelée ?

HISTOIRE DE L'INTERNET ET DU WEB : UNE COEXISTENCE CREATIVE ENTRE PROPRIETE PRIVEE ET BIEN COMMUN.

Derrière toutes ces interrogations, un fil rouge : celui de l'équilibre entre économie de bien commun et économie marchande. Toute la brève et fulgurante histoire de l'internet et du web s'est construite sur cette coexistence équilibrée entre les deux formes d'organisation économique. L'internet, en tant qu'infrastructure qui sous tend le web, tout autant qu'ensemble de protocoles d'interconnexion, est né de la recherche publique. Il en a gardé un statut de bien commun grâce au principe essentiel de la « neutralité du net », principe qui garantit à tous, producteurs, créateurs ou utilisateurs, petits ou gros, un

droit équivalent pour emprunter dans les mêmes conditions ces « autoroutes de l'information » comme on les appelait autrefois. Les standards de l'internet, notamment le protocole TCP/IP qui en constitue la pierre angulaire, sont également le fruit partagé du travail d'une communauté, celle des ingénieurs réunis au sein de l'IETF¹, communauté de bénévoles, délégués par leurs entreprises ou leurs centres de recherche dans une démarche transparente. Quant au web lui-même, en tant que contenant, c'est-à-dire comme infrastructure applicative, avec toutes les technologies spécifiques qui lui sont associées, il est également le résultat d'un travail coopératif élaboré au sein du collectif ouvert, qu'est le W3C². Ce processus garantit l'interopérabilité entre les documents et les échanges d'informations.

Par ce processus coopératif d'élaboration des normes techniques privilégiant l'interopérabilité et l'ouverture, le web et l'internet constituent une véritable mutation dans la construction d'un bien commun mondial, et ce faisant, dans la gouvernance politique internationale. Jusqu'ici la construction de règles supranationales et la mutualisation de ressources étaient d'abord l'apanage des États nations, par exemple l'interopérabilité des télécommunications, le statut de l'Antarctique ou de celui de la haute mer. Dans le monde du numérique des communautés d'intérêt auto-constituées et autogérées, jugeant le rythme des négociations interétatiques trop lent au regard de celui de l'innovation, ont coproduit de la régulation et du bien commun. Voilà pourquoi les deux piliers de l'ère informationnelle que sont l'IETF et le W3C ne sont pas aujourd'hui dans le giron de l'UIT (Union internationale des Télécommunications, structure multilatérale dépendant des Nations Unies) — ce dernier aurait dû en constituer le berceau naturel, n'eût été son incapacité structurelle à accompagner ces révolutions.

Sur la base des normes ouvertes du web ont ensuite pu fleurir aussi bien des technologies propriétaires que des technologies libres, laissant à chaque acteur le choix des outils et de leur statut. Bien qu'organisé par des communautés autogérées, l'internet est pour les acteurs du secteur l'équivalent d'un domaine public, ouvert aussi bien aux initiatives privées et aux règles économiques afférentes, qu'à l'usage par toute la société, souvent dans des buts publics (éducation, gouvernement, etc.) ou non lucratifs (contre-pouvoirs associatifs, ONG, expression directe des individus, etc.). L'expérimentation sur le web peut s'appuyer sur des applications propriétaires (la base de données Access de Microsoft, le serveur IIS, le gestionnaire de contenu DocuShare de Xerox, ou K-Portal utilisé dans de nombreuses universités en France, et bien évidemment des outils de bureautique comme Microsoft Word ou DreamWeaver) ou sur les logiciels libres (MySQL pour la base de données, Apache, le serveur le plus utilisé, les gestionnaires de contenus comme SPIP ou Drupal, ou la bureautique libre avec OpenOffice). Mais dans tous les cas, l'infrastructure générale reposera sur des normes publiques, non propriétaires, exemptes de brevets et finalement sur des applications libres insoupçonnées par l'utilisateur (la gestion du cœur de réseau par le DNS Bind par exemple).

CONTENUS EN LIGNE : LES CREATEURS SE SONT DONNES LES MOYENS DE CHOISIR

Si l'on observe maintenant le web du point de vue de ses contenus, c'est-à-dire de la multitude de documents numériques qui en constitue la richesse, on retrouve la même coexistence entre contenus privés, c'est-à-dire soumis aux différents droits de propriété intellectuelle et contenus de bien commun. Par défaut les contenus (photos, films, musique, articles,...) placés en ligne par leurs auteurs et/ou leurs ayants droit sont protégés par le droit d'auteur en Europe, le copyright aux États-Unis. Mais rien n'interdit aujourd'hui à ces mêmes auteurs de leur assigner un statut juridique qui en facilite la circulation et la réutilisation. Licence Art Libre³, contrats Creative Commons⁴, licence GFDL⁵,... ces différents dispositifs

¹ IETF : *Internet Engineering Task Force* organise le processus des débats techniques de normalisation de l'internet et publie les documents correspondant ; <http://ietf.org>. Son fonctionnement communautaire est décrit dans « *The Tao of IETF : A Novice's Guide to the Internet Engineering Task Force* » ; <http://www.ietf.org/tao.html>

² W3C : *World Wide Web Consortium* organise le processus de normalisation des technologies du web et publie les « recommandations » ; <http://w3c.org>

³ <http://artlibre.org/>

juridiques apparus dans le courant des années 2000 insufflent dans le web une dose de bien commun, sans pour autant chercher à interdire ou à bloquer les contenus privés.

Là encore, la créativité juridique n'est pas née d'une négociation intergouvernementale mais a été portée par des communautés de créateurs, soucieux de ne pas se laisser enfermer dans un appareil juridique qui n'est adapté ni à la réalité des usages, ni à leur désir de voir leurs œuvres circuler largement au travers des réseaux. Dans ce cas, l'auto-saisine des communautés concernées s'explique par une divergence idéologique de fond avec les instances intergouvernementales théoriquement en charge de faire évoluer le droit de la propriété intellectuelle (Organisation Mondiale de la Propriété Intellectuelle (OMPI), Organisation Mondiale du Commerce (OMC) au travers des Accords sur les ADPIC...). Ces dernières instances ont abandonné depuis longtemps l'objectif initial de coexistence entre bien commun et propriété privée, au profit d'un renforcement récurrent des droits privatifs⁶.

DONNEES BRUTES ET BIEN COMMUN : RIEN N'EST JOUE

A côté de ces contenus relevant des différentes formes de droit d'auteur, on voit se développer à vive allure des contenus qui ne sont pas des œuvres originales, de la création, et dès lors dépendent d'autres formes de protection juridique. C'est ce qu'on appelle les données brutes : statistiques démographiques, contenus des annuaires, catalogues de magasins, horaires de transports, catalogues de bibliothèques, résultats d'études épidémiologiques, informations sur le contenu et l'origine d'un film ou d'un morceau de musique, résultats d'expériences scientifiques, etc. Toutes ces données sont en train de migrer vers le web, constituant *de facto* et sous nos yeux le « web des données ».

En cherchant à revisiter notre question de l'équilibre entre bien commun et propriété privée, on constate que sous des abords clairs — l'existence d'un cadre juridique qui régit les bases de données et leurs contenus⁷ —, la question varie considérablement d'une catégorie de données à l'autre.

LES DONNEES PUBLIQUES

Les choses semblent *a priori* simples lorsque l'on se penche sur les données d'origine publique. Données démographiques, flux migratoires, prélèvements obligatoires, statistiques d'activité des juridictions, fonds cartographiques... Ces données, recueillies dans le cadre de missions de service public, financées par l'argent public, semblent avoir une vocation naturelle à être publiées, diffusées, reproduites, utilisées, en un mot, à être libres. La réalité est un peu plus complexe. S'il existe de longue date des obligations de publications pour les administrations, celles-ci se limitaient jusqu'à peu à une publication au *Journal officiel* ou dans le journal d'une collectivité territoriale. Depuis quelques années est apparu un mouvement de fond, aussi bien en Europe qu'aux États-Unis, en faveur d'une libération des données publiques⁸. Ce mouvement part de volontés émanant des différents secteurs de la société civile. En France, un collectif a

⁴ <http://fr.creativecommons.org/>

⁵ <http://www.gnu.org/licenses/fdl-1.3.html>

⁶ Voir dans ce livre les articles de Xuan LI « *Propriété intellectuelle, normalisation, biens communs et responsabilité des pouvoirs publics* », et Gaelle Krikorian « *Accès à la santé ou renforcement des droits de propriété intellectuelle : quelle hiérarchie pour les normes internationales qui régissent nos vies* ».

⁷ En France ; la Loi n°98-536 du 1 juillet 1998 port ant transposition dans le code de la propriété intellectuelle de la directive 96/9/CE du Parlement européen et du Conseil, du 11 mars 1996, concernant la protection juridique des bases de données.
<http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000573438>

⁸ « Les enjeux publics, économiques et citoyens de l'ouverture des données : l'expérience britannique », Valérie Peugeot, In : *Enjeux politiques du document numérique, Actes de la troisième conférence Document numérique et société, Aix-en-Provence, 15-16 novembre 2010*, sous la direction de Évelyne Broudoux et Ghislaine Chartron, Ed. ADDBS, 2010, p. 277-298

été lancé pour « *la création d'un groupe de travail pour des données ouvertes et mises en commun* » en janvier 2008⁹. Au Royaume-Uni, c'est dès mars 2006 que deux journalistes du *Guardian*, sous le titre « Rendez nous les joyaux de la couronne », appelaient à une libération des données publiques¹⁰. Depuis différents collectifs se sont mis en place, dont « *Mash the state* » en Grande-Bretagne, qui invite plus particulièrement les collectivités locales à partager leurs données via de simples fils RSS¹¹. Mais c'est certainement le cri lancé au cours de la conférence TED de 2009 par Tim Berners-Lee, l'un des pères fondateurs du web et dirigeant du W3C, qui a eu le plus de retentissement : « *Free raw data !* »¹²

Les gouvernements semblent commencer à entendre le message. Aux États-Unis Barak Obama a demandé à Vivek Kundra, *Chief Information Officer*, de prendre le chantier à bras le corps. Ceci s'est traduit par le lancement en juin 2009 du site *data.gov* qui rassemble pour l'heure près de 600 bases de données produites par les services fédéraux aux États-Unis. Au Royaume-Uni, Tim Berners-Lee s'est vu confier en juin dernier une mission équivalente par le gouvernement britannique, dont on attend encore les fruits. Au Danemark, la plate-forme *Digitaliser.dk* pointe vers toutes les sources de données publiques ouvertes. En Finlande, un concours « *Apps for democracy Finland* » invite les développeurs à présenter les *mashups* qu'ils réalisent à partir des API et des données ouvertes du gouvernement. Ils peuvent également suggérer le nom d'autres bases de données qui mériteraient d'être rendues accessibles. En France, c'est l'Agence pour le Patrimoine Immatériel de l'État (l'APIE) qui est censée mettre un place un portail attendu pour 2010, mais dont nul ne sait encore s'il contiendra réellement des données ouvertes. Pour autant le chemin à parcourir pour disposer de données publiques véritablement libres et ouvertes reste long : format de données hétérogènes et donc difficiles ou impossibles à croiser, données incomplètes, souvent non mises à jour, sites bloqués... Le mouvement vers l'ouverture des données publiques est certes engagé, mais il souligne en creux les réticences de certaines entités publiques, qui ont du mal à effectuer la mutation culturelle de la sortie du secret et qui certainement, craignent, en partageant ces informations, de perdre une partie de pouvoir ou de légitimité.

DERRIERE LES DONNEES SCIENTIFIQUES, LE STATUT DE LA SCIENCE DANS LA SOCIETE

Une seconde grande source de données mérite notre attention : les résultats des travaux des équipes de recherche. S'inscrivant dans un mouvement plus large en faveur de la science ouverte *open access*¹³, un grand nombre de chercheurs multiplient les initiatives pour que les résultats de leurs travaux circulent : revues en libre accès à modèle économique alternatif, archives ouvertes et données ouvertes constituent les trois piliers de ce mouvement. Au fur et à mesure que la pression s'accroît sur les organismes de recherche publique pour qu'ils trouvent les moyens de leur autonomie financière, notamment par le biais de dépôt de brevets, la résistance s'organise chez les scientifiques qui considèrent que le fruit de leurs travaux doit, par nature, appartenir au bien commun.

Côté données, c'est dès 1995 que le CGDI (Comité sur les Données géophysiques et environnementales du Conseil national de la Recherche aux États-Unis) publie un document séminal intitulé « *de l'échange*

⁹ *Des biens publics vers les biens communs : création d'un groupe de travail pour des données ouvertes et mises en commun* ; <http://www.a-brest.net/article3782.html>

¹⁰ « Give us back our crown jewels », Charles Arthur and Michael Cross, *The Guardian*, Thursday 9 March 2006 ; <http://www.guardian.co.uk/technology/2006/mar/09/education.epublic>

¹¹ *Mash The State* : We are a grassroots campaign to encourage UK government and public sector organisations to make their data available to the general public. <http://www.mashthestate.org.uk/>

¹² « Libérez nos données brutes », Tim Berners-Lee on the next Web, *TED 2009*, février 2009. http://www.ted.com/index.php/talks/tim_berniers_lee_on_the_next_web.html

¹³ Voir dans cet ouvrage « *Connaissance, réseaux et citoyenneté : pourquoi le libre accès ?* » de Jean-Claude Guédon et « *De l'accès libre à la science ouverte* » de Philippe Aigrain.

complet et ouvert des données scientifiques »¹⁴ qui constate que « les programmes internationaux de recherche sur le changement global et la surveillance environnementale dépendent du principe d'un échange complet et ouvert des données (i.e. les données et les informations sont rendues disponibles sans restriction, sur une base non discriminatoire, pour un coût réduit à celui des besoins de reproduction et de distribution) » et appelle à un accès libre des données atmosphériques, océaniques, biosphériques. Depuis on ne compte plus les appels de ce type dans tous les secteurs scientifiques : chimie, recherche géospatiale, recherche sur le génome...

Le statut des données d'expériences scientifiques a des conséquences majeures sur la science elle-même, sur la capacité à vérifier la crédibilité des conclusions, et à participer à la critique collective, tant entre scientifiques que par les populations. Un exemple récent est offert par la polémique portant sur l'efficacité de l'antiviral Tamiflu menée par le prestigieux journal médical *BMJ* (*British Medical Journal*) en décembre 2009. Ce médicament était donné pour très efficace, notamment contre les gripes aviaires et A, suite à des publications médicales datant de 2005. Or, des éléments suspects provenant du Japon, où il est très utilisé, ont mis en doute ces publications. Un organisme de vérification interne au secteur médical, le Cochrane Collaboration a dès lors tenté d'obtenir du trust pharmaceutique Roche, qui commercialise le Tamiflu, les données brutes de ces expériences. Ce qui fut impossible, ou alors à des conditions qui auraient interdit toute revue critique. Comment juger dès lors de la fiabilité des informations produites par les entreprises pharmaceutiques ? Ce qui a conduit l'éditrice du *BMJ* à titrer son éditorial: « *We want raw data, now* »¹⁵. Dans le même ordre d'idée, les données fournies par Monsanto à la Communauté européenne pour faire valider son maïs transgénique MON863 ont été rendues publiques, suite à plusieurs années de bataille juridique menées par Greenpeace Allemagne. Ceci a permis au professeur Gilles-Éric Séralini et son équipe du CRII-GEN de relever de nombreuses incohérences dans les traitements statistiques... suffisamment pour remettre en cause la diffusion de cet OGM.¹⁶

DES DONNEES VOLONTAIREMENT AUTOPRODUITES ET PARTAGEES PAR LES UTILISATEURS

Tout comme les scientifiques, les utilisateurs du web sont de plus en plus nombreux à vouloir partager des données en tout genre, sans limite et sans restriction. Bref à vouloir placer leurs données en bien commun. Car, en dehors de leurs capacités créatives, les usagers sont de plus en plus dotés d'outils d'enregistrement, de géolocalisation ou de capteurs, et nombreux sont ceux qui souhaitent partager leurs captures d'information sur des plates-formes collectives. Les raisons en sont aussi nombreuses que les systèmes qui hébergent ces données. Geste de mémoire ou volonté esthétique pour les utilisateurs qui confient les sons de leur quotidien à Saveoursounds, plate-forme proposée par la BBC ? Cris du vendeur de poissons qui traverse les rues de Luanda, sons dans le terminal de bus de La Paz, sirènes de police la nuit à Los Angeles... un paysage mondial de nos environnements sonores se tisse, au fil des contributions¹⁷. Envie de sortir de la solitude face à la maladie, de contribuer à des avancées scientifiques pour les contributeurs

¹⁴ On the full and open exchange of scientific data, Committee on Geophysical and Environmental Data, National Research Council, Washington, D.C., 1995.
<http://www.nap.edu/readingroom/books/exch/exch.html>

¹⁵ We want raw data, now, Fiona Godlee, *BMJ*, 10 Decembre 2009, doi:10.1136/bmj.b5405

¹⁶ MON 863 : un maïs OGM autorisé impropre à la consommation - Mars 2007, communiqué du CRII-GEN ; http://www.criigen.org/index.php?option=com_content&task=view&id=122&Itemid=1
L'article de référence de ce travail scientifique : « New Analysis of a Rat Feeding Study with a Genetically Modified Maize Reveals Signs of Hepatorenal Toxicity », Gilles-Eric Séralini, Dominique Cellier and Joël Spiroux de Vendomois, *Archives of Environmental Contamination and Toxicology*, Volume 52, Number 4 / mai 2007, DOI 10.1007/s00244-006-0149-5 ;
<http://www.springerlink.com/content/02648wu132m07804/>

¹⁷ <http://www.bbcworldservice.com/saveoursounds>

de PatientsLikeMe¹⁸ qui confient en ligne des informations sur leur santé, leurs maladies, leurs évolutions, leurs souffrances ? Volonté de donner du sens à sa passion ou de pallier aux insuffisances des savoirs savants pour les contributeurs de TelaBotanica¹⁹, vaste banque de données en botanique constituée par les 11 000 botanistes francophones inscrits au réseau ? Volonté de donner à voir l'invisible et volonté de résistance politique pour les contributeurs de Disparitions.eu, qui racontent comment les sans-papiers sont enlevés à leur vie, leur ville, leur école ? Le témoignage brut, est au point d'équilibre entre le partage de données et la construction d'informations. Ce sont évidemment souvent des *documents* (textes, photographies,...). Mais la manière dont ils sont inscrits directement par les usagers (UGC — *User Generated Content*) modifie leur statut : notes ouvertes aux croisements (*mashups*) et à l'analyse statistique (moteurs de recherche), ils deviennent des *données* pour la production de synthèses ou de veille. Les informations « temps réel » distillées par Twitter sont un exemple frappant : les catastrophes sont souvent annoncées par l'analyse de ce service avant même que les secours ne soient opérationnels.

UNE OUVERTURE A GEOMETRIE VARIABLE

Pour ces trois premières catégories de données — d'origine publique, scientifique ou volontairement partagées par les utilisateurs —, selon les sites dans lesquels ces données sont versées, leur degré d'ouverture, et donc la qualité de bien commun, ne seront pas les mêmes.

Trois variables, technique, juridique et économique, permettent d'en évaluer la capacité à construire du bien commun.

Techniquement, il ne suffit pas de placer des données en ligne pour qu'elles soient réutilisables par des tiers. Encore faut-il qu'elles soient mises à disposition dans des formats qui soient facilement ré-exploitable, si possible des formats non propriétaires et qui surtout permettent de croiser différentes bases, donc des formats interopérables. Un site comme data.gov livre certes 600 bases de données, mais dans des formats hétéroclites et parfois propriétaires. Techniquement toujours, les possibilités d'usage ne sont pas les mêmes selon que les contenus des bases de données sont mis à disposition via une API (auquel cas les données restent stockées dans la base d'origine) ou en *dumps*²⁰ ce qui permet de « retravailler » les données avec tous les outils à disposition localement (logiciels de calcul, de statistique, de visualisation, etc.).

Juridiquement, certaines plates-formes vont accueillir les données sans clarifier les droits qui leurs sont attachés. Ce qui peut induire une incompréhension par les déposants sur ce que cela implique en termes d'usages par des tiers. Conscient de ce problème et lucide sur le fait que tous les scientifiques ne sont pas des juristes, le chapitre « Science commons » du mouvement Creative Commons a mis en place un protocole intitulé « *Science Commons Database Protocol* »²¹. Celui-ci, sorte de mécanisme de levée de droits, aide les scientifiques à placer leurs données dans un régime libre de droits en tous genres (brevet, droit d'auteur, droit *sui generis* des bases de données en Europe...), afin de faciliter la réutilisation et l'interopérabilité des données scientifiques.

Deux autres mécanismes, portées par deux groupes différents, permettent aux divers producteurs de données brutes de placer celles-ci dans le domaine public. L'*Open Knowledge Foundation* propose ainsi la PDDL (*Public Domain Dedication and Licence*) qui permet de placer des données dans le domaine public²²

¹⁸ <http://www.patientslikeme.com/>

¹⁹ *Tela Botanica : le réseau de la botanique francophone* ; <http://www.tela-botanica.org/>

²⁰ Possibilité de téléchargement de la base dans son intégralité, ou de fractions significatives suivant des requêtes.

²¹ *Protocol for Implementing Open Access Data* ; <http://sciencecommons.org/projects/publishing/open-access-data-protocol/>

²² <http://www.opendatacommons.org/licenses/pddl/>

et depuis juin 2009 une licence intitulée ODbL (*Open Database Licence*)²³, qui permet de placer des bases de données dans un régime intermédiaire, à savoir permettre leur partage, réutilisation, transformation par des tiers, sous trois conditions : mention de la paternité, partage à l'identique — c'est-à-dire que toute nouvelle base de données construite en utilisant la première doit être placée sous une licence ODbL — et le maintien de l'ouverture technique — c'est-à-dire le non usage de mesures techniques de protection de type DRM. De son côté, Creative Commons, ressentant le besoin de libérer des données de tout droit a proposé l'option CCo²⁴.

L'ENJEU DES LICENCES POUR PROTEGER LES BIENS COMMUNS

Cette clarification juridique devrait avoir entre autres mérites, de rendre irréversible le statut de bien commun indépendamment du devenir du consentement de ceux qui l'ont accordé. Plusieurs communautés productrices de données libres ont dans le passé fait les frais de l'absence de protection de ce type. L'aventure advenue à la base de données CDDB (*Compact Disc Data Base*) en est une illustration trop parfaite. Cette base en ligne contient les informations utiles sur chaque titre des CD du commerce, les métadonnées (auteur, compositeur, année, genre). C'est son existence qui permet aux logiciels de lecture ou de numérisation de disque compact (de type iTunes, Winamps ou Xmms) d'aller immédiatement chercher les métadonnées du disque qui est en train d'être écouté ou « ripé » et de les afficher ou les intégrer dans les enregistrements mp3. Créée en 1993 par Ti Kan, CDDB est passée en version en ligne en 1995. Dès sa création, les utilisateurs ont pu contribuer à nourrir cette base de données, d'abord en envoyant des mails à Ti Kan, puis en ligne en renseignant les morceaux des disques dont ils disposaient. Des milliers d'amateurs de musique ont ainsi pendant des années contribué à constituer ce formidable bien commun de la musique qu'était CDDB, dont l'accès était totalement libre. Vendue en 1998 à Escient, rebaptisée Gracenote en 2000, son accès devient progressivement soumis à des licences de plus en plus restrictives. En 2001, toutes les applications ne disposant pas d'une licence pour accéder à la nouvelle version de CDDB étaient bloquées. C'en était fini de ce bien commun de la musique, constitué collectivement par tous les usagers et finalement accaparé par une unique entreprise. En 2008, Gracenote était racheté par Sony, rejoignant le giron des *majors*. Heureusement, en réaction à ce détournement de bien commun, des internautes ont collecté au moment du changement le contenu de la version « coopérative » de CDDB (pré-commerciale) pour constituer Freedb, depuis associée à tous les logiciels libres de musique. L'enjeu d'une licence garantissant le maintien du bien commun est apparu ainsi très clairement. Freedb est sous licence GPL (*General Public Licence*). Un projet similaire, Musicbrainz²⁵, créé au même moment en résistance à la privatisation de CDDB, utilisant une formalisation basée sur RDF, et qui est en passe de devenir une référence plus complète, un « wikipédia de la musique enregistrée » a pour sa part choisi la licence *Creative Commons by-nc-sa*. Dans les deux cas, le choix d'un « partage à l'identique » est apparu comme la meilleure façon de protéger un travail coopératif, un bien commun.

Du point de vue économique, la question de la gratuité de la mise à disposition des données est bien entendu un critère essentiel. Ainsi en Europe une directive de 2003, transposée en droit français en 2005, impose d'ores et déjà aux pouvoirs publics de mettre à disposition leurs données, mais les autorise simultanément à en monnayer l'usage sous forme de redevance. C'est, entre autres, au nom de ces revenus réels ou supposés (le montant est aujourd'hui inconnu ; l'Agence du Patrimoine immatériel de l'État, l'APIE, annonce qu'elle publiera en 2009 une évaluation des actifs immatériels lié aux bases de données²⁶) que les administrations publiques opposent une résistance à la libération des données publiques. Inversement, aucune évaluation de la richesse économique induite par le lancement de services et les

²³ <http://www.opendatacommons.org/licenses/odbl/summary/>

²⁴ <http://creativecommons.org/publicdomain/zero/1.0/deed.fr>

²⁵ <http://musicbrainz.org/>

²⁶ https://www.apiefrance.com/sections/acces_thematique/comptabilisation-des-actifs-immateriels/recensement_du_patri/

usages sur bases de données publiques ouvertes n'a été menée²⁷.

DONNEES PERSONNELLES : LE BIEN COMMUN NE PEUT SE CONSTRUIRE EN SACRIFIANT LA VIE PRIVEE

À côté de ces données placées volontairement en statut de bien commun, les utilisateurs de services web nourrissent au fil de leurs butinages et vagabondages en ligne de nombreuses bases de données, propriétés des fournisseurs desdits services. Dès l'inscription, l'utilisateur se met à parler de lui-même : nom, pseudo, âge, genre, adresse, religion, statut amoureux... Et poursuit en donnant à voir ses goûts (les vidéos qu'il ou elle a visionnées, les livres achetés, les jeux en ligne joués, etc.), ses voyages (par exemple par l'application MyTravel), ses relations interpersonnelles à travers les réseaux sociaux, mais aussi son appartenance à des groupes d'intérêt plus ou moins formels (par exemple, adhésion au groupe de Médiapart sur Facebook) ou informels (groupe Facebook « mais non maman il n'y aura pas d'alcool à cette soirée »).

Toutes ces informations sont en principe des données d'ordre privé, protégées par des directives européennes de 1995, 2002 et 2006²⁸. Mais cette protection semble se fragiliser chaque jour un peu plus : nos données, des plus anodines aux plus intimes se mettent à circuler, à passer d'un service à l'autre, à être exploitées par différents acteurs du web, sans que nous en soyons toujours conscients. Les raisons de cette érosion sont multiples, dont la justification et la légitimité varient considérablement.

La première cause est à chercher du côté du modèle économique du web. La très large majorité des services accessibles sur le web n'appellent pas de contribution monétaire de la part des utilisateurs. Pour autant ils ne sont pas gratuits, puisque leur modèle économique repose sur la publicité et que le visiteur paye de son attention, de son acceptation à subir de la publicité, le droit d'utiliser le service. Alors que le secteur publicitaire est touché de plein fouet par la crise économique, seul le segment de la publicité en ligne poursuit sa croissance. Condition de son succès : une publicité toujours plus pertinente, plus adaptée, plus ciblée au regard de la personne qui y est confronté. Derrière ce ciblage, une seule solution : rassembler le plus d'informations et donc de données sur chaque internaute. D'où une industrie florissante du *datamining*, qui vient maintenant puiser à des sources inédites, comme les données de géolocalisation récupérées sur nos téléphones mobiles, ou l'usage intensif des cookies intersites déposés par les sites de bannières publicitaires.

La seconde raison à cet effritement de la protection des données personnelles, certainement la plus difficile à circonscrire, est liée à la transformation indubitable de la relation de nos contemporains aux notions de sphère privée/sphère publique. Alors que Serge Tisseron parle d'*extime* (vs. intime) on constate une évolution naturelle des passants du web à s'exposer, parler d'eux-mêmes, y compris dans ce qu'ils ont de plus intime, en particulier chez les plus jeunes²⁹. À cette impudeur s'ajoute un sentiment

²⁷ James Boyle, dans une colonne du *Financial Times* compare cette absence d'évaluation à une décision portant sur les médicaments qui se ferait sans données expérimentales. « A natural experiment », James Boyle, *Financial Times*, 22 novembre 2004 ;

<http://www.ft.com/cms/s/2/4cd4941e-3cab-11d9-bb7b-00000e2511c8.html>

Il montre par ailleurs l'impact de la libération des données en comparant les industries liées à la météo aux États-Unis, où les données publiques sont gratuites, et en Europe. « Public information wants to be free », James Boyle, *Financial Times*, 24 février 2005 ; <http://www.ft.com/cms/s/2/cd58c216-8663-11d9-8075-00000e2511c8.html>

²⁸ Directive 95/46/CE du Parlement européen et du Conseil, du 24 octobre 1995, relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données ; Directive 2002/58/CE du Parlement européen et du Conseil, du 12 juillet 2002, concernant le traitement des données à caractère personnel et la protection de la vie privée dans le secteur des communications électroniques (directive « *vie privée et communications électroniques* »). Cette dernière a été modifiée en 2006 par la directive 2006/24/CE.

²⁹ « Vie privée : le point de vue des "petits cons" », Jean-Marc Manach, *Internet Actu*, 4 janvier 2010 ; <http://www.internetactu.net/2010/01/04/vie-privee-le-point-de-vue-des-petits-cons/>

d'invulnérabilité. L'idée d'un mésusage de ces données, pour des finalités politiques, idéologiques, liberticides semblent à la plupart d'entre nous une inquiétude sans bien-fondé, voire légèrement paranoïaque, ou du moins qui ne concernerait que « ceux qui ont quelque chose à se reprocher »³⁰, oubliant que les critères de « ce qui peut être reproché » varient au gré des gouvernements, des époques, des lois... Qui sait par exemple que le gouvernement Bush des États-Unis d'Amérique a demandé et obtenu des opérateurs télécoms les traces des communications téléphoniques (les *call logs*) de milliers de personnes, au nom de la guerre contre le terrorisme ?

La troisième est liée au besoin de simplicité exprimé par les utilisateurs eux-mêmes. On constate depuis quelques années un phénomène de « fatigue » des utilisateurs du web, éprouvés par le besoin de remplir des formulaires longs et fastidieux, chaque fois qu'ils veulent découvrir un nouveau service web, ou exaspérés par l'impossibilité d'emporter avec eux leurs données, leurs contenus, lorsqu'ils veulent migrer d'un service à un autre. En réaction à cette fatigue, un mouvement a émergé au sein de communautés *geeks* pour promouvoir la portabilité des données, c'est-à-dire la faculté pour tout un chacun d'être réellement maître de ses informations et de pouvoir les transporter et réutiliser ses données personnelles d'un service web à un autre, ou bien dans plusieurs applications. Pour faciliter la circulation et l'échange aussi bien de données d'authentification (utilisées pour se connecter au service) que de réseau social (le carnet d'adresse des « amis » et les interactions entre eux) ou de contenus mis intentionnellement en ligne (photos, cadeaux, commentaires, articles, présentations), il est nécessaire de disposer de formats ouverts et interopérables. Ces standards ouverts, à l'image des microformats ou de RDFa, sont poussés soit par des organismes de normalisation établis comme l'IETF et le W3C, soit par des initiatives ascendantes portées par des communautés de développeurs, comme dataportability.org ou le projet DiSO³¹. Face à ces initiatives, les ténors de l'industrie du web ont eu une double réaction : la plupart ont affiché publiquement leur soutien à ces initiatives, voire en sont devenus membres ; simultanément ils ont organisé une contre-offensive à travers des standards de fait, appuyés sur une logique d'API utilisables par des tiers, mais assurant la conservation des données sur le site ayant le premier su capter l'utilisateur. Le domaine qui illustre certainement le mieux cet affrontement entre deux philosophies du web est celui des données d'authentification. Une communauté auto-organisée a développé OpenID, standard d'authentification, qui permet de s'enregistrer une fois pour toute chez un tiers de confiance puis d'utiliser cet identifiant à chaque fois que l'on se connecte sur un nouveau service web. Il s'agit d'un standard ouvert, qui ne dépend d'aucun acteur en particulier. Officiellement la plupart des grands acteurs soutiennent OpenID. Dans la pratique, il n'est pratiquement pas utilisé, car dans un même temps Facebook a poussé son kit d'authentification — Facebook Connect — qui a ce jour semble devenir le standard de fait et est utilisé par la plupart des grands médias sociaux en ligne : pour vous y connecter, vous autorisez le site tiers à aller chez Facebook prendre vos données d'identité. Pas de mystère derrière ce choix des acteurs : il s'agit de capter une partie des 350 millions d'utilisateurs de Facebook en leur simplifiant le passage d'un service à l'autre. Mais en échange, c'est Facebook qui continue à centraliser les données personnelles et qui peut suivre les activités de ses membres même sur les autres sites.

Au final, le paradoxe règne autour de la portabilité des données : ce mouvement, initié par des développeurs cherchant à rendre aux utilisateurs le contrôle de leurs données, permet en réalité à

³⁰ Ce point de vue est d'ailleurs largement promotionné par les responsables des grands sites effectuant du traçage des individus, à l'image d'Eric Schmidt de Google déclarant : « *If you have something that you don't want anyone to know, maybe you shouldn't be doing it in the first place.* » ; <http://gawker.com/5419271/google-ceo-secrets-are-for-filthy-people> ou encore Marc Zuckerberg, patron de Facebook répondant à une question sur la récente ouverture au public des profils de Facebook : « *We view it as our role in the system to constantly be innovating and be updating what our system is to reflect what the current social norms are* », sonnante ainsi la fin de l'idée de vie privée pour mieux justifier ses pratiques commerciales ; http://www.readwriteweb.com/archives/facebooks_zuckerberg_says_the_age_of_privacy_is_ov.php

³¹ <http://diso-project.org/>

quelques acteurs majeurs du web de consolider leur hégémonie et de devenir la porte d'entrée de tous les médias sociaux. Et ce faisant, de récolter encore et toujours plus d'information sur les cyber-faits et gestes des internautes.

Enfant grandit bien vite, le web garde dans ses gènes les inspirations libertariennes de bon nombre de ses pères fondateurs. À ce titre, l'idée de « libérer » les données semble posée comme un *a priori* positif, qu'il s'agisse de données publiques — qui semblent par nature avoir vocation à rejoindre le domaine public — ou de données privées — dont les propriétaires doivent pouvoir les faire circuler à volonté. Encore faut-il que cette libération de nos données ne soit pas prétexte à une érosion de nos libertés individuelles !

Ces tensions entre constitution d'un bien commun des données et libertés individuelles ne peut qu'aller en s'accroissant quand on assiste à l'entrée sur scène progressive d'une cinquième catégorie de données : celles produites par les objets, par *nos* objets.

UNE CATEGORIE EMERGENTE ET MIXTE : LES DONNEES EMISES PAR LES OBJETS

Nos machines, nos objets, les plus usuels, les plus banals, se mettent progressivement à s'exprimer et à converser « entre eux » via le web. Aujourd'hui nos téléphones mobiles envoient à toute application qui le demande (et que nous « autorisons »), nos coordonnées de géolocalisation. Ceci permet l'émergence de réseaux sociaux localisés, pour permettre à nos « amis » de savoir que nous sommes dans le quartier (à l'exemple de Aka Aki³²) ou de publier nos performances de coureur de fond pour comparer nos résultats avec ceux des autres sportifs empruntant les mêmes chemins (*Nokia Sport Tracker*³³). Demain ce sont les objets de la maison qui deviendront bavards à l'image de Watson, cet appareil qui mesure la consommation d'électricité du foyer et partage les résultats en ligne avec la communauté des personnes qui en sont équipées³⁴. Ce seront aussi les objets de l'espace urbain, comme des lampadaires équipés de capteurs permettant de mesurer la pollution sonore ou de l'air. Ou encore dans les campagnes avec ces instruments enterrés dans les sols, susceptibles d'en mesurer le taux d'hydratation et de surveiller par exemple la progression d'une sécheresse en temps réel. Déjà nos compteurs de gaz et d'électricité, nos voitures, nos biens de consommation sont équipés pour devenir communicants. Il s'agit pour l'heure d'une communication affectée, c'est-à-dire monodirectionnelle entre l'émetteur et le récepteur : le compteur informe notre fournisseur d'énergie de notre consommation, notre voiture prévient notre assureur de notre accident, l'étiquette RFID attachée au manteau prévient l'acheteur que le fournisseur a transféré la marchandise chez le transporteur etc. Mais rien n'interdit que demain ces informations, agrégées, ne puissent être partagées en ligne. Les consommateurs ne voudront-ils pas visualiser le trajet de leurs achats, depuis le producteur, l'artisan, jusqu'au magasin, pour en vérifier l'origine ? Les citoyens d'une ville n'exigeront-ils pas l'accès à l'information sur la consommation énergétique des immeubles publics de leur commune comme le revendiquent déjà les britanniques³⁵ ? Les personnes à la recherche d'un logement ne voudront-elles pas disposer simplement via une interface web des données de pollution sonore et atmosphérique des quartiers où elles cherchent à s'installer ? etc.

Quel sera le statut de ces données ? Qui donnera le droit de les rendre libres ? Avec quelle granularité d'agrégation ? Ainsi est-ce EDF seul qui pourra décider de cartographier les flux de consommation électrique des populations françaises ? Avec quel degré de précision : la ville dans son ensemble, le quartier ou la rue ; le village ou un périmètre rural de plusieurs kilomètres carrés ? Pour quels usages : vigilance environnementale ou surveillance des consommations de ses voisins ? Tout ceci reste largement à explorer, à construire, à penser, afin que ces océans de données que vont progressivement déverser les

³² <http://www.aka-aki.com/>

³³ <http://sportstracker.nokia.com/nts/main/index.do>

³⁴ <http://www.diykyoto.com/uk>

³⁵ <http://www.guardian.co.uk/environment/georgemonbiot/2009/sep/11/energy-efficiency-public-buildings>

machines ne constituent pas une faille de plus dans la protection des données personnelles, trouvent une réelle utilité sociale et au final méritent l'appellation de bien commun.

BIEN COMMUN ET LIBERTES PUBLIQUES

Dans cette tension entre bien commun et libertés publiques, le web révèle plus que jamais sa nature essentiellement politique. Pourtant certains sont tentés de balayer cette dimension d'un revers de main pour le réduire à une infrastructure technico-économique. Lorsque Mark Zuckerberg, le fondateur de Facebook, affirme que « *la norme sociale a évolué avec le temps* » et qu'il est donc normal que la plupart des informations qu'affichent les utilisateurs de son service soient publiques³⁶, il semble considérer que la technique façonne de façon unilatérale et irrépressible les comportements sociaux. À nous de démontrer qu'à l'inverse, le corps social peut aussi s'emparer de la technique et la réguler pour l'accorder à un projet de société dans lequel le souci du bien commun demeure le fil rouge.

³⁶ <http://www.ustream.tv/recorded/3848950>, entretien vidéo réalisé par *Techcrunch* ; <http://www.cnetfrance.fr/news/facebook-fin-confidentialite-39712122.htm>