Guides & Kits (Cats)

Il s'agit ici de feuilleter ce que certaine.s nomment le BABA (au rhum) : auto défense numérique, médias et informations et de s'outiller pour des animations débats sur les questions de critique citoyenne.
Tic, Tac & NTIC
Illustration : Sonia Soraya

- Guide Auto Défense Numérique
- Internet un monde meilleur ? La question de la régulation d'internet
- Le logiciel libre
- Les données ouvertes : un bien commun
- Un documentaire et un débat pour questionner la monétisation des données
- Guide d'autodéfense intellectuelle face à l'information
- Un documentaire pour questionner les données et leur €

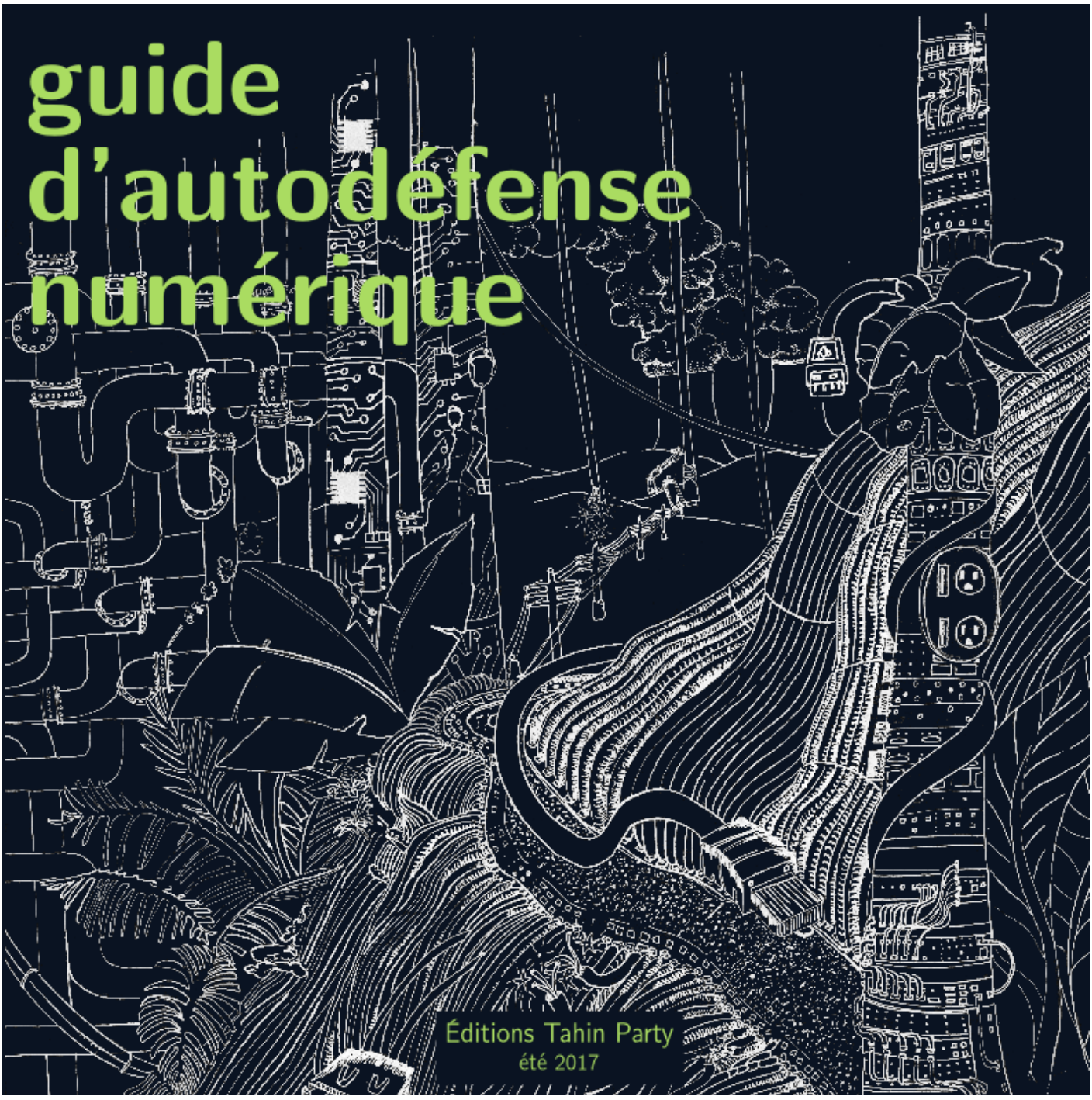
Guide Auto Défense Numérique

L’enjeu de ce guide est de fournir cartes, sextant et boussole à quiconque veut cheminer sur la route de la compréhension et pratiques d'auto défense numérique.

BOUM ici Tome 1 Hors Connexions

BOUM là Tome 2 En ligne Utiliser Internet

guide d'autodéfense numérique



Éditions Tahin Party
été 2017

Internet un monde meilleur ? La question de la régulation d'internet

Faut il réguler internet ?

Faut il réguler internet ?

Quelle question ! A l heure ou les marchandises, ~~les personnes~~ circulent .

Sources : Par la quadrature du net

Le modèle de l’économie de l’attention

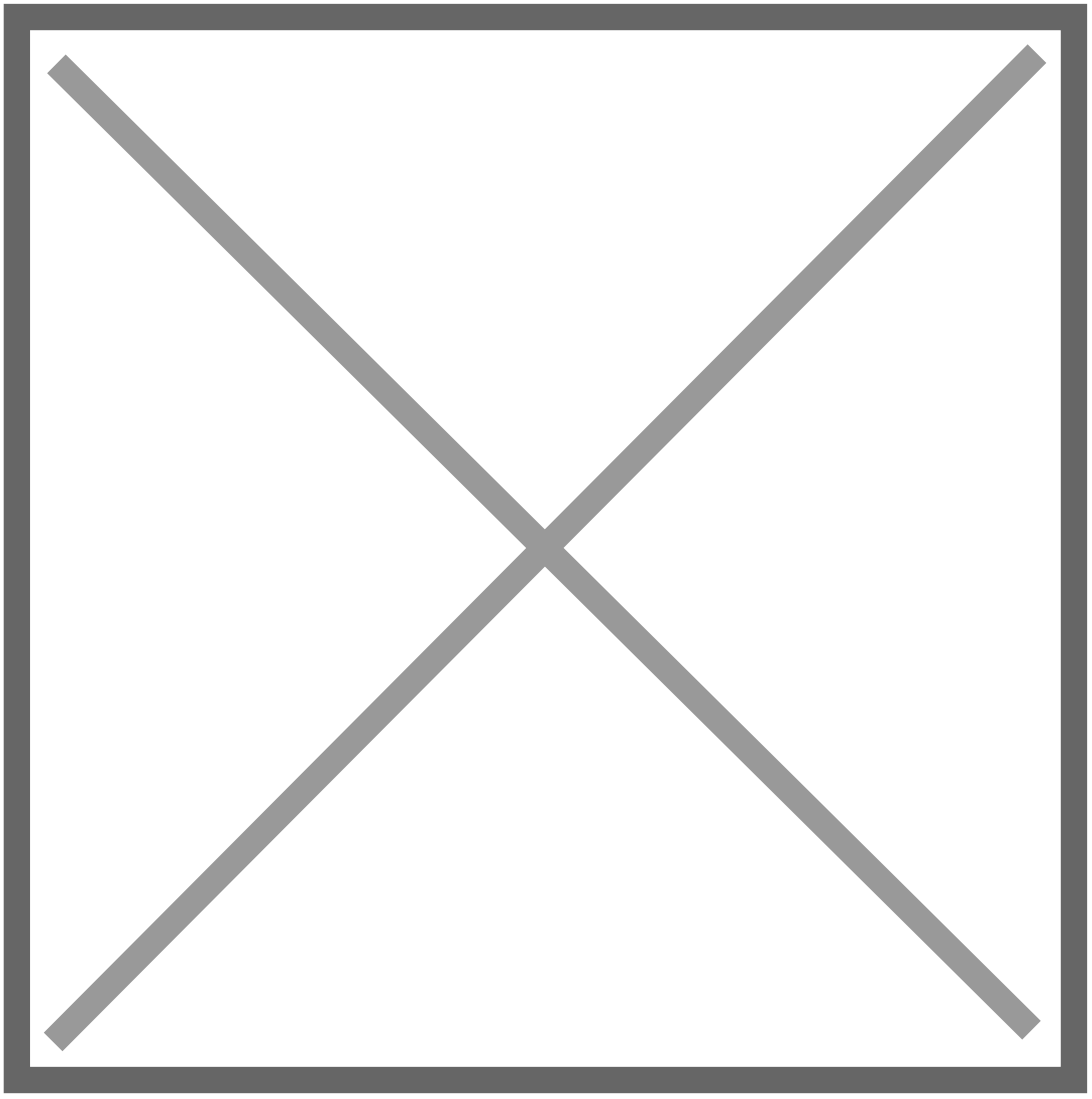
Si le rêve initial d’Internet était celui du partage, force est de constater qu’aujourd’hui les « Géants du Web » ont acquis un quasi-monopole sur de très nombreux aspects de nos activités en ligne : nos recherches sur le Net (Google Search possède par exemple 91 % des parts de marché en France), nos téléphones mobiles (le système Android est installé sur la plupart des portables, près de 8 sur 10 en France, et si l’on ajoute Apple et son système iOS, on couvre la quasi-totalité des portables), nos communications via les réseaux sociaux (qui passent en grande majorité par Facebook et Twitter) et les liens sociaux tissés par la même occasion…

Et ces entreprises n’ont pas développé ces activités dans un but philanthropique : elles sont là pour faire de l’argent, et leur modèle économique est essentiellement basé sur l’affichage publicitaire (cela représente par exemple 85 % du chiffre d’affaires de Google et même 98 % de celui de Facebook). Pour augmenter leur chiffre d’affaires et leurs bénéfices, ils font donc tout ce qu’ils peuvent pour mieux nous cibler et nous garder le plus longtemps possible sur leurs sites ou leurs applications. C’est ce qu’on appelle « l’économie de l’attention » : l’idée est de nous garder le plus longtemps possible « captif·ves », à la fois pour récupérer un maximum de données nous concernant, mais surtout pour bénéficier de plus de « temps de cerveau disponible » et mieux vendre leurs espaces publicitaires aux annonceurs.

Prenons l'exemple du réseau social Facebook. L'entreprise explique sans pudeur son fonctionnement : des personnes qui souhaitent diffuser un message (une publicité, un article, un événement, etc.) désignent à Facebook un public cible, selon certains critères sociaux, économiques ou de comportement, et paient l'entreprise pour qu'elle diffuse ce message à ce public, dans les meilleures conditions.

Afin de cibler correctement ces publics, Facebook analyse donc les contenus produits par ses utilisat·rices : contenus publics, messages privés, interactions diverses, contenus consultés, appareils utilisés pour se connecter, temps de visionnage d'un message, etc. Ces contenus ne se résument pas à ce qu'on publie consciemment, mais à l'ensemble nos activités, qui dévoilent nos caractéristiques sociales, économiques… Pour voir à quel point cette analyse peut s'avérer redoutable, on peut évoquer une étude conduite par l'université de Cambridge en 2013 : 58 000 personnes ont répondu à un test de personnalité, puis ce test a été recoupé à tous leurs « j'aime » laissés sur Facebook. En repartant de leurs seuls « j'aime », l'université a ensuite pu estimer leur couleur de peau (avec 95 % de certitude), leurs orientations politique (85 %) et sexuelle (80 %), leur confession religieuse (82 %), s'ils fumaient (73 %), buvaient (70 %) ou consommaient de la drogue (65 %).

Ces données récoltées et traitées (sans forcément que les utilisateur·rices en soient averti·es ou conscient·es) permettent ensuite à Facebook de proposer des messages publicitaires « adaptés », aux moments et aux formats les plus opportuns pour fonctionner sur les personnes ciblées et les influencer dans leurs choix. Mais cela ne s'arrête pas là, Facebook collabore aussi avec des tiers pour pister des personnes n'ayant même pas de compte chez eux, via les cookies (petits fichiers stockés sur nos appareils, initialement faits pour faciliter la navigation, mais qui sont susceptibles d'être utilisés pour traquer les personnes), les boutons « j'aime » sur des pages de sites Internet (si une personne voit un de ces boutons s'afficher sur la page qu'elle consulte, des données la concernant – telle que son adresse IP par exemple – sont envoyées à Facebook et ce même si elle n'a pas de compte auprès de ce service), le Facebook login (bouton qui permet de se connecter sur un site via son compte Facebook, et qui renvoie lui aussi un certain nombre d'informations à Facebook)…



Mais il est aussi nécessaire de capter le plus longtemps possible l'attention des utilisat·rices dans le cadre de cette « économie de l'attention ». Prenons cette fois l'exemple de Youtube, la plus grosse plateforme vidéo d'Internet, propriété de Google et l'un des sites les plus visités au monde. Youtube ne se contente pas seulement d'héberger des vidéos : il s'agit d'un véritable média social de contenus multimédias, qui met en relation des individus et régule ces relations : lorsqu'une vidéo est visionnée sur Youtube, dans 70 % des cas l'utilisateur·rice a été amené·e à cliquer sur cette vidéo via l'algorithme de recommandation de Youtube. Et le but de cet algorithme n'est pas de servir l'utilisat·rice mais de faire en sorte que l'on reste le plus longtemps possible sur la plateforme, devant les publicités. Cet algorithme ne se pose certes pas la question du contenu, mais en pratique l'ancien employé de Google Guillaume Chaslot [dans un entretien dans le numéro 5 de la revue Vraiment, paru le 18 avril 2018] constate que les contenus les plus mis en avant se trouvent être les contenus agressifs, diffamants, choquants ou complotistes. Il compare : « C'est comme une bagarre dans la rue, les gens s'arrêtent pour regarder ».

Youtube, désirant ne pas perdre une seconde de visionnage de ses utilisat·rices, ne prend pas le risque de leur recommander des contenus trop extravagants. L'ancien employé déclare qu'ils ont refusé plusieurs fois de modifier l'algorithme de façon à ce que celui-ci ouvre l'utilisat·rice à des contenus inhabituels. Dans ces conditions, le débat public est entièrement déformé, les discussions

les plus subtiles ou précises, jugées peu rentables, s'exposent à une censure par enterrement.

En tant qu'hébergeur de contenu, Youtube est tenu de retirer un contenu « manifestement illicite » si celui-ci lui a été notifié. Compte tenu de la quantité de contenus que brasse la plateforme, elle a décidé d'automatiser la censure des contenus « potentiellement illicites », portant atteinte au droit de certains auteurs, au moyen de son RobotCopyright appelé « ContentID » (système d'empreinte numérique qui permet de comparer tout nouveau contenu déposé sur la plateforme à une base de données préexistante, alimentée par les titulaires de droit d'auteur). Pour être reconnu auteur·e sur la plateforme, il faut répondre à des critères fixés par Youtube. Une fois qu'un contenu est protégé par ce droit attribué par Youtube (en pratique, il s'agit en majorité de contenus issus de grosses chaînes de télévision ou de grands labels), la plateforme se permet de démonétiser [1] ou supprimer les vidéos réutilisant le contenu protégé à la demande des titulaires de droit. Ce ContentID est un moyen de censure de plus qui démontre que Youtube ne souhaite pas permettre à chacun·e de s'exprimer (contrairement à son slogan « Broadcast yourself ») mais cherche simplement à administrer l'espace de débat public avec pour but de favoriser la centralisation et le contrôle de l'information. Et pour cause, cette censure et cet enfermement dans un espace de confort est le meilleur moyen d'emprisonner les utilisat·rices dans son écosystème au service de la publicité.

De la nécessité de réguler

Nous avons donc des entreprises qui récoltent nos données personnelles et en font la base de leur modèle économique, le plus souvent de manière illégale : en effet, depuis mai 2018, le RGPD prévoit que notre consentement à la collecte et au traitement de nos données personnelles n'est pas considéré comme valide s'il n'est pas librement donné, et notamment si l'on n'a pas d'autre choix que d'accepter pour pouvoir accéder au service (article 7, §4, et considérant 43 du RGPD, interprétés par le groupe de l'article 29 [2]). Or pour avoir un compte Facebook ou pour consulter des vidéos sur Youtube nous n'avons actuellement pas d'autre choix que d'accepter que nos données soient collectées par leurs outils de pistage/ciblage.

Et ces mêmes entreprises sont aussi celles qui mettent en avant des contenus polémiques, violents, racistes ou discriminatoires en partant de l'idée que ce sont ceux qui nous feront rester sur leur plateforme. Le rapport visant à « renforcer la lutte contre le racisme et l'antisémitisme sur Internet », commandé par le Premier ministre français et publié en septembre 2018, l'explique très bien. Il dénonce « un lien pervers entre propos haineux et impact publicitaire : les personnes tenant des propos choquants ou extrémistes sont celles qui « rapportent » le plus, car l'une d'entre elles peut en provoquer cinquante ou cent autres. Sous cet angle, l'intérêt des réseaux sociaux est d'en héberger le plus possible ». Plus généralement, le rapport regrette la « règle selon laquelle un propos choquant fera davantage de « buzz » qu'un propos consensuel, alimentant de façon plus sûre le modèle économique des plateformes ».

Face à ces dérives, quelles solutions s'offrent à nous ? Voici quelques propositions formulées et soumises au débat par La Quadrature du Net (voir la fiche « Faut-il réguler Internet ? (2/2) »)

Notes

[1] Nombreux sont les Youtubeur·ses qui vivent en partie des revenus publicitaires liés à leurs vidéos. Démonétiser une vidéo revient à supprimer les revenus publicitaires liés à cette vidéo et donc la rémunération de son auteur·rice, en cessant de mettre cette vidéo en avant ou en cessant de l'associer à des publicités.

[2] https://www.cnil.fr/fr/le-g29-groupe-des-cnil-europeennes et https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=623051 - référence en anglais

Le logiciel libre

Le logiciel libre comme fer de lance des biens communs numériques

Présenter le logiciel libre en quelques mots et sourires.

Source : https://www.ritimo.org/Le-logiciel-libre-comme-fer-de-lance-des-biens-communs-numeriques

Cet article présente les logiciels libres comme fer de lance des communs numériques. Pour reprendre les termes de Wikipédia (19 novembre 2013), « un logiciel libre est un logiciel dont l'utilisation, l'étude, la modification et la duplication en vue de sa diffusion sont permises, techniquement et légalement ». Les logiciels libres sont intéressants à analyser pour eux-mêmes, mais également pour la manière dont ils inspirent de nombreuses autres initiatives que nous étudions dans ce dossier. Toutefois, bien que le développement de certains logiciels libres reste aujourd'hui encore fortement basé sur l'approche « communautaire » et non-commerciale, d'autres sont désormais fortement investis par des entreprises commerciales comme IBM, Sun ou Google. En cela, les logiciels libres apparaissent comme un cas paradigmatique des communs numériques, car ils mettent de l'avant les tensions qui existent au niveau de la gouvernance des communs.

Brève histoire des logiciels libres

Les principes des logiciels libres ont été mis en place dans les années 1980. Au cœur de ces principes se retrouve l'accès au code source. Le code source est en quelque sorte la recette derrière le fonctionnement d'un programme ou d'un logiciel. Plus précisément, le code source est l'ensemble des instructions écrites dans un langage de programmation humainement lisible, spécifiant le fonctionnement d'un logiciel. L'idée derrière les logiciels lires n'est pas tant que ceux-ci soient gratuits, mais plutôt que l'accès à leur code source reste libre. La Fondation pour les logiciels libres les définit par quatre libertés fondamentales, que nous citons ici [1] :

- la liberté d'exécuter le programme, pour tous les usages (liberté 0) ;

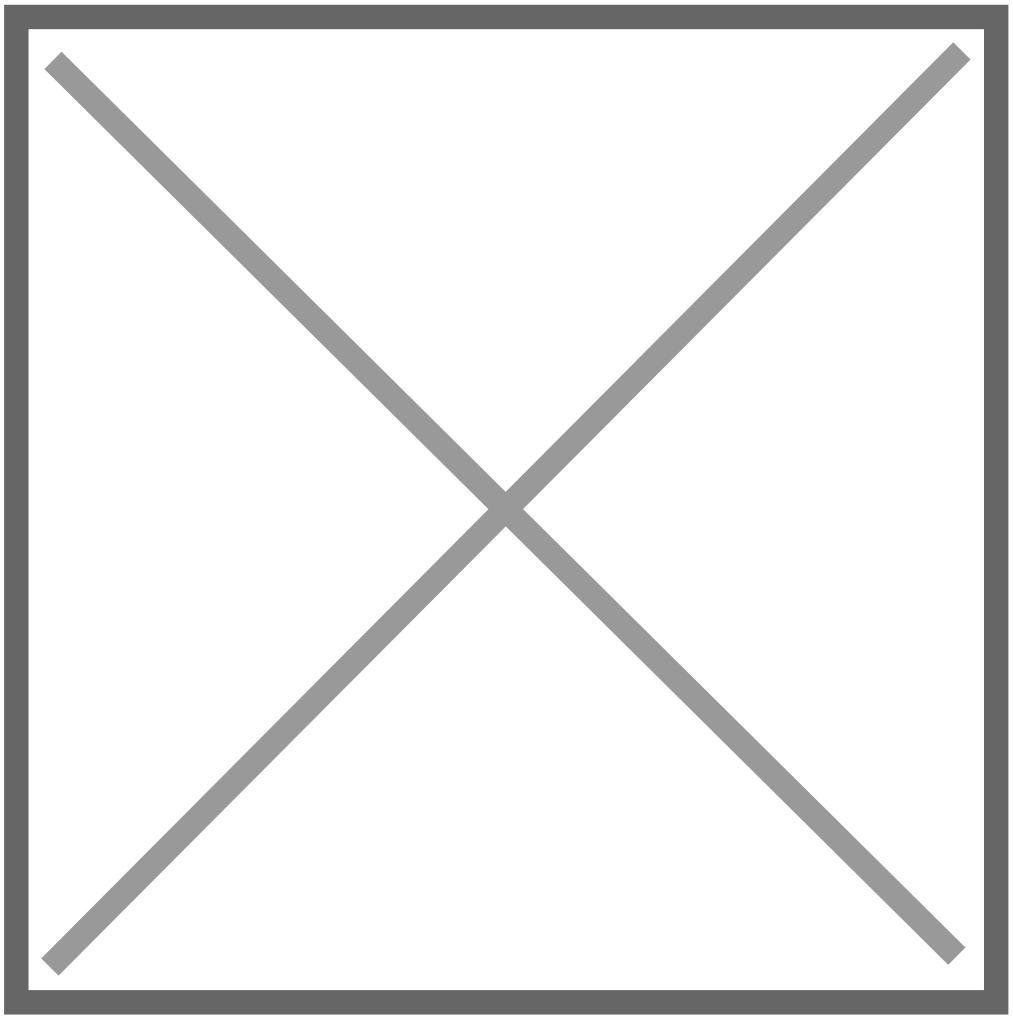
- la liberté d’étudier le fonctionnement du programme et de le modifier pour qu’il effectue vos tâches informatiques comme vous le souhaitez (liberté 1) ; l’accès au code source est une condition nécessaire ;

- la liberté de redistribuer des copies, donc d’aider votre voisin (liberté 2) ;

- la liberté de distribuer aux autres des copies de vos versions modifiées (liberté 3) ; en faisant cela, vous donnez à toute la communauté une possibilité de profiter de vos changements ; l’accès au code source est une condition nécessaire.

Les principes des logiciels libres, à leurs débuts, avaient surtout un objectif éthique. Il ne s’agissait donc pas tant de produire des logiciels puissants et agréables à utiliser, mais plutôt d’assurer la capacité de pouvoir partager ses connaissances (le code source des logiciels) avec ses pairs. Au fil des années, la popularité des logiciels libres s’est accrue, si bien qu’au tournant du millénaire, ce modèle a attiré l’attention des entreprises qui y voyaient avant tout une façon plus économique de développer des logiciels pérennes et efficaces. C’est ainsi que le terme « *open source* » (à code source ouvert) est apparu, afin de proposer une terminologie plus séduisante pour les entreprises que le terme « *free software* », trop souvent associé à la notion de gratuité en anglais. Cette tension existe encore aujourd’hui, à tel point que Richard Stallman, le fondateur des logiciels libres, considère qu’il s’agit de deux camps politiques au sein d’une même communauté.

Contrairement à ce qu’affirmaient les premiers observateurs des communauté de logiciels libres, celles-ci sont loin de fonctionner sans règles. Les modalités de production des logiciels libres peuvent être décrites par le terme d’ « innovation par l’usage », ou d’ « innovation ascendante ». Selon Dominique Cardon, l’innovation ascendante se fait autour de trois cercles : le cercle des « innovateurs », qui a démarré le projet, la nébuleuse des « contributeurs » qui apportent des contributions minimales au projet et un cercle intermédiaire, le cercle des « réformateurs », qui inclut des acteurs qui réforment ces contributions.



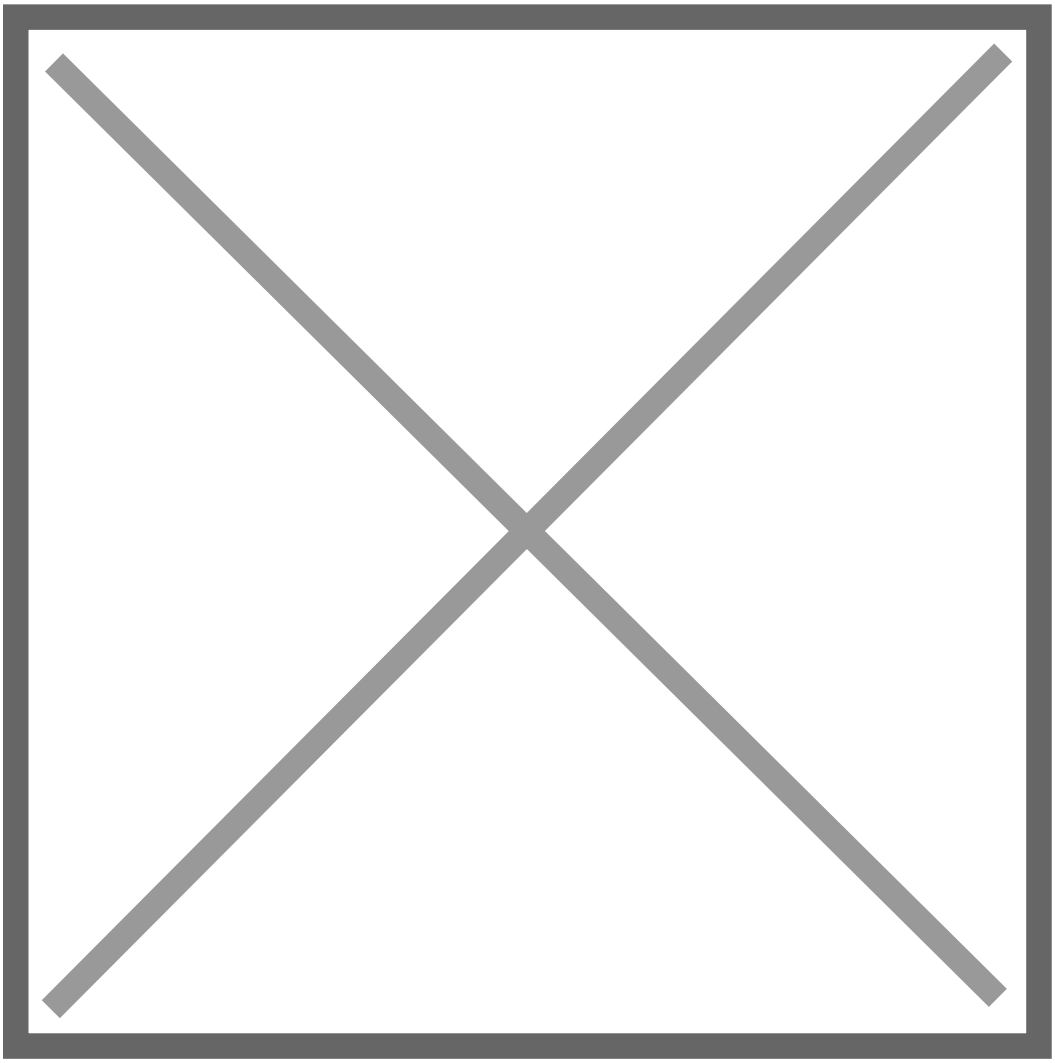
Source : Cardon, Dominique. 2005. « Innovation par l'usage ». In *Enjeux de mots : regards multiculturels sur les sociétés de l'information*, sous la dir. de Alain Ambrosi, Valérie Peugeot et Daniel Pimienta. Caen (France) : C & F Éditions. En ligne <

http://vecam.org/article588.html>

Bien qu'il existe une multitude de projets de logiciels libres très peu actifs et développés sur une base individuelle, la plupart des projets matures et collectifs peuvent être décrits par ce modèle d'innovation ascendante. Cependant, il existe des différences importantes en termes de gouvernance et d'organisation du développement. Je m'attarderai ici plus particulièrement sur deux cas de logiciels libres, aujourd'hui emblématiques : le logiciel *Linux*, et le système d'exploitation *Debian*.

Le noyau Linux [2]

Linux est sans doute l'un des logiciels libres les plus connus au point qu'il fait parfois figure d'emblème des logiciels libres. On fait cependant souvent l'erreur de considérer Linux comme un système informatique complet et autonome (comparable à Windows par exemple), alors que Linux se limite en fait au noyau d'un système d'exploitation. C'est d'ailleurs pour cette raison que beaucoup de militant(e)s du logiciel libre insistent souvent pour parler de GNU/Linux plutôt que simplement de Linux. Si Linux est encore aujourd'hui utilisé de manière marginale sur les bureaux de travail, il est en revanche très largement majoritaire pour certains usages, comme celui des téléphones *Android*, qui constituent la plus grande part du marché des téléphones intelligents.



Logo de Linux

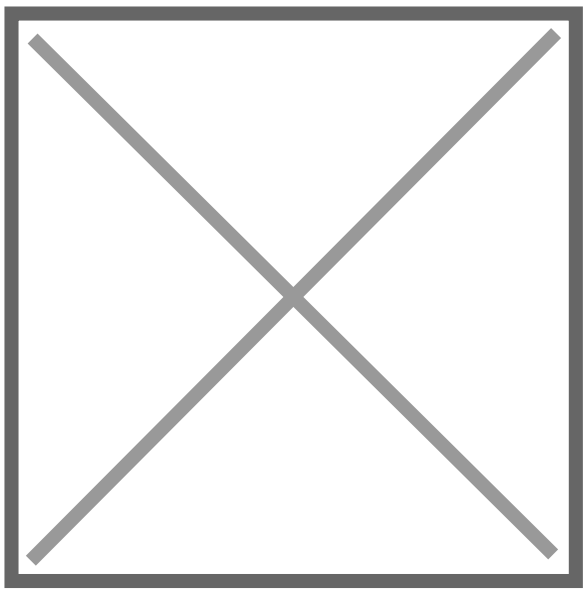
Linux a été créé en 1991 par Linus Torvalds, alors étudiant en informatique. L'objectif de Torvalds à ce moment n'était pas tant de contribuer aux biens communs, mais simplement de s'amuser (« *just for fun* »). Cette initiative a cependant créé un intérêt assez soudain chez beaucoup d'informaticiens, un succès qui a d'ailleurs plus tard amené Eric Raymond, l'un des créateurs du terme *open source*, à décrire le modèle de Linux comme celui d'un bazar où le coordonnateur (Linus Torvalds) puise différentes pièces pour les assembler dans un tout cohérent [3]. Cette description est parfois considérée comme la première analyse sociologique du modèle du logiciel libre.

Le développement de Linux est généralement vu comme étant dirigé d'une main de fer par son fondateur, Linus Torvalds, et fortement influencé par les entreprises. Torvalds est d'ailleurs souvent considéré comme le « dictateur bienveillant » de la communauté. Il est accompagné d'une petite équipe autour de lui, ainsi que de plusieurs centaines de contributeurs qui proposent des modifications mineures au projet. Cette dynamique correspond assez bien au modèle d'innovation ascendante décrit précédemment. Plus précisément, le modèle de gouvernance et de développement de Linux s'articule autour du cycle de production des nouvelles versions. Ainsi, le leader du projet (et dictateur bienveillant) est responsable de la production, environ tous les trois mois, d'une nouvelle version appelée « *vanilla* » qui intègre des nouvelles fonctionnalités. D'autres versions dites « *stables* » et maintenues sur le long terme relèvent quant à elles de la responsabilité de son équipe rapprochée, appelée « équipe de stabilité ». Une version de « *développement* » est aussi disponible, et les nouvelles modifications y sont intégrées chaque jour.

Mentionnons également que les entreprises commerciales participent fortement au développement de Linux en assignant des salarié(e)s à la programmation de certaines fonctionnalités ou en contribuant financièrement. La Fondation Linux regroupe des membres corporatifs, tels que IBM, Intel, Samsung et Google. Elle a pour mission de soutenir le développement de Linux, notamment en employant son fondateur, en fournissant les infrastructures nécessaires à son développement et en soutenant les individus et les entreprises dans leur appropriation de Linux. La participation des entreprises leur permet d'orienter davantage le développement du logiciel selon leurs propres intérêts et objectifs. Ceci n'est d'ailleurs pas sans semer l'inquiétude parmi certains défenseurs du logiciel libre qui y voient de nouvelles formes de privatisation du bien commun [4].

Le projet Debian

Debian est ce que l'on appelle une « distribution », c'est-à-dire un système d'exploitation et un ensemble de composantes logicielles basées sur le noyau Linux. Quoique peu populaire, la distribution Debian est significative car elle est la source d'autres distributions, en particulier Ubuntu, beaucoup plus populaire et souvent utilisée par les débutant(e)s. Il est intéressant de présenter ici Debian en le comparant à Linux, car les modes d'organisation des deux projets sont assez différents et en fait, beaucoup plus communautaires dans le cas de Debian. Debian se démarque également par des documents qui forment le socle de gouvernance du projet : le contrat social, la constitution et la charte.



Logo du projet Debian

Le contrat social est en quelque sorte le document fondateur de la communauté. Il définit les principes moraux et inaliénables de la communauté, expose sa propre définition de ce qui constitue un logiciel libre et stipule notamment que toutes les composantes du système doivent rester libres. Le contrat affirme également un devoir de transparence par rapport à la mémoire de la communauté, par exemple en conservant une trace des débats et des bogues qui ont eu lieu dans le passé.

La constitution [5] définit les modalités de décision et de résolution de conflits, ainsi que le pouvoir des membres. Elle énonce par exemple que la personne responsable du projet est élue par une majorité après une période d'élection de six semaines. Elle stipule également qu'une décision de la personne responsable peut être suspendue si un certain nombre de développeurs le demande [6]. La constitution elle-même peut être amendée par une majorité de 3 pour 1. La charte de Debian stipule quant à elle les règles techniques pour qu'un « paquet » (un logiciel ou une composante logicielle) soit intégré dans la distribution Debian. La charte précise par exemple que tous les logiciels inclus dans Debian doivent être libres (au sens de la *Debian Free Software Guidelines*) et stipule certaines règles pour choisir le nom et établir la version d'un paquet. La charte est un document beaucoup plus flexible que la constitution, et évolue continuellement. Voici comment le guide de l'administrateur décrit le processus éditorial de la charte :

« Tout le monde peut proposer une modification de la charte Debian : il suffit de soumettre un rapport de bogue de « gravité » [...] sur le paquet *Debian-policy*. Le processus qui débute alors est documenté dans /usr/share/doc/debian-policy/Process.html : s'il est reconnu que le problème soulevé doit être résolu par le biais d'une nouvelle règle dans la charte Debian, la discussion se poursuit sur debian-policy@lists.debian.org jusqu'à l'obtention d'un consensus et d'une proposition. Quelqu'un rédige alors la modification souhaitée et la soumet pour approbation (sous la forme d'un correctif à relire). Dès que 2 autres développeurs approuvent le fait que la formulation proposée reflète bien le consensus ayant émergé de la discussion précédente (en anglais, le verbe consacré est "*to second*"), la proposition peut être intégrée au document officiel par un des mainteneurs du paquet *debian-policy*. Si le processus échoue à l'une des étapes, les mainteneurs fermeront le bogue en classant la proposition comme rejetée [7] ».

Il est par ailleurs intéressant de noter la présence, au sein du projet, du groupe *Debian-Women*, qui vise à faciliter une meilleure intégration des femmes. Ceci montre autant une certaine politisation du projet qu'un désir d'aborder une pluralité de dimensions de la participation.

Le logiciel libre comme bien commun ?

On voit donc que les logiciels libres, dans leur expression concrète, répondent bien aux critères du bien commun, à savoir qu'ils ne sont ni des biens privés, ni des biens publics. Il est toutefois intéressant de remarquer les différences de gouvernance entre les projets, plus autoritaire (voire « dictatoriale ») dans le cas de Linux, et plus démocratique dans le cas de Debian.
Il faut également mentionner les liens étroits, dans le monde du logiciel libre, entre les dimensions de bien commun, de bien public et de bien privé. Comme nous l'avons montré, si Linux reste un bien commun, il est cependant fortement investi par des grandes entreprises commerciales et réutilisé dans la production de biens privés, ou semi-privés comme par exemple l'écosystème Google. De la même manière, Debian, bien que sa nature de bien commun est beaucoup moins ambiguë, est à son tour réutilisé par la distribution Ubuntu, qui est développée et soutenue par une entreprise commerciale, *Canonical*. Cependant, ces liens ne sont pas à sens unique, car les entreprises privées contribuent également au bien commun, même si cette contribution est elle-même souvent intéressée et parfois source d'inquiétude parmi les défenseurs du logiciel libre. Par ailleurs, l'intérêt croissant des administrations publiques pour les logiciels libres rend également compte des relations entre le bien commun et le bien public. On le voit, les relations entre bien commun, bien public et bien privé dans le cadre des logiciels libres sont complexes, mais il ne fait pas de doute que les logiciels libres conservent encore aujourd'hui leur nature de bien commun.

Notes

[1] http://www.gnu.org/philosophy/free-sw.fr.html

[2] Voir plus loin :

http://en.wikipedia.org/wiki/Linux_kernel#Development_model

https://github.com/torvalds/linux

[3] Raymond, Eric Steven. 2002. « The Cathedral and the Bazaar ».
http://www.catb.org/~esr/writings/cathedral-bazaar/cathedral-bazaar/.

[4] Voir par exemple cet article très populaire à propos de la main-mise de Google sur le système *Android* – un dérivé de *Debian* – et la manière dont ce système est développé afin d'être, à toute fin pratique, non-fonctionnel sans l'intégration de plusieurs composantes propriétaires :
http://arstechnica.com/gadgets/2013/10/googles-iron-grip-on-android-controlling-open-source-by-any-means-necessary

[5] http://www.debian.org/devel/constitution

[6] Dans un style très « geek », le projet spécifie que le nombre de développeurs nécessaires pour suspendre une telle décision est de 2K, où K est le minimum de Q et de 5, et Q est la moitié de la racine carrée du nombre courant de développeurs.

[7] Voir cette page pour une excellente présentation du fonctionnement de la communauté Debian : http://debian-handbook.info/browse/fr-FR/stable/sect.debian-internals.html

Les données ouvertes : un bien commun

Peut-on articuler données et biens communs? En quelques lignes, pour débattre de cette publicisation

Source : https://www.ritimo.org/Les-donnees-ouvertes-un-bien-commun

L'open data, ou l'ouverture des données publiques, consiste en la mise à disposition proactive de données issues principalement des administrations publiques, afin de favoriser la création de valeur par les citoyens et les entreprises, une plus grande transparence de l'action publique ou encore la participation citoyenne. Juridiquement, l'ouverture exige que les données soient accompagnées de licences dites ouvertes permettant à quiconque d'utiliser les données à la seule condition d'indiquer leur source et, éventuellement, de partager les améliorations avec la même licence. Puisque l'ouverture des données a vocation à devenir une pratique courante dans les administrations, assiste-t-on à l'émergence d'un bien commun numérique ? Quelles sont les communautés qui exigent leur ouverture et défendent leur gestion en bien commun ?

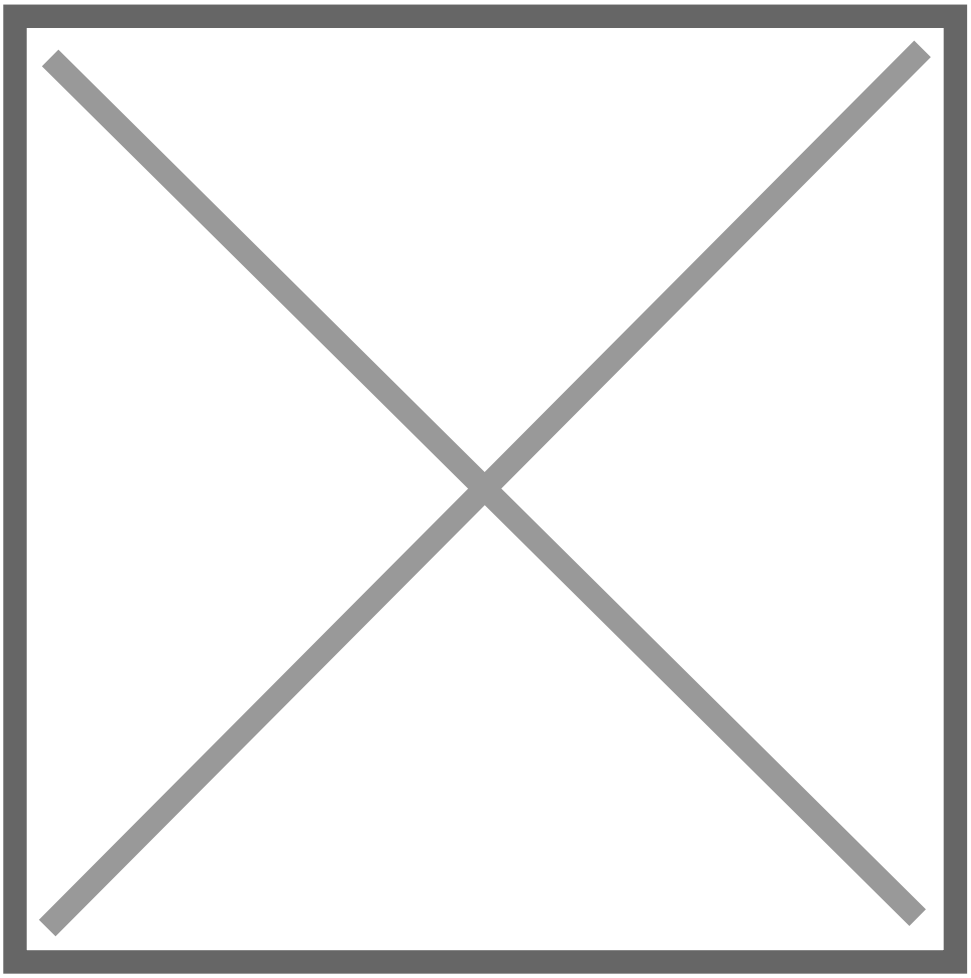
Exiger l'ouverture des données publiques

Rappelons d'abord que l'*open data* s'inscrit dans des pratiques anciennes de diffusion et de partage de l'information. L'accès à l'information détenue par l'État fait partie des revendications centrales de la Révolution Française. La Déclaration des droits de l'homme et du citoyen de 1789 stipule dans son article 15 que « la société a le droit de demander compte à tout agent public de son administration. » Après la Seconde Guerre mondiale, cette exigence de redevabilité (*accountability*) des agents publics a connu un nouvel essor avec l'émergence du concept d'*open government* qui exige que les citoyens aient accès aux secrets de l'État. Faisant face aux critiques de l'opacité de l'armée lors de la guerre du Vietnam, le président Johnson adopte en 1966 le *Freedom of Information Act* (FOIA) qui donne à tout citoyen le droit d'exiger les informations dont dispose une administration, sous réserve de certaines conditions liées à la confidentialité et à la sécurité nationale. Les dispositions du FOIA ont été reprises dans la plupart des démocraties et constituent un droit fondamental quant à la transparence de l'action publique.

Les premières pratiques de partage volontaire de données ont eu lieu dans le domaine des sciences. Avec le développement des réseaux informatiques et des capacités de calcul, le partage de données scientifiques a connu un essor considérable. Un exemple particulièrement illustratif est le projet de décodage du génome, qui a nécessité le partage de données entre laboratoires car le séquençage exigeait des infrastructures de recherche spécialisées. Bruno Strasser, dans son étude du partage des données génétiques [1], souligne la vigueur du débat sur la nécessité du partage

des données lors de la création de la première base de données du génome, GenBank. En effet, selon lui, le partage des données scientifiques s'inscrit dans une « économie morale » où les chercheurs ne diffusent leurs données que si un échange de capitaux symboliques compense leur partage. Il cite le cas de GenBank qui est parvenue à s'imposer comme la principale base de données génétiques parce qu'elle ne considérait pas les données comme sa propriété, contrairement au projet concurrent, l'*Atlas of Protein Sequence*. Depuis 1990, les scientifiques doivent publier sur Genbank les articles utilisant des données sur le génome. Ces pratiques scientifiques qui inaugurent le partage de données à grande échelle n'étaient pas encore qualifiées d'*open data*. C'est en 1995 que le terme apparaît dans un rapport de la *National Academy of Science* des États-Unis, intitulé « *On the Full and Open Exchange of Scientific Data* » et dans lequel l'académie réclame le partage des données recueillies par les satellites.

Alors que les pratiques d'ouverture de données sont déjà établies dans les sciences, l'open data en tant que revendication politique fait son apparition au Royaume-Uni. En 2006, le *Guardian* publie une tribune exigeant l'ouverture des données détenues par l'État britannique. Intitulée « *Give us back our crown jewels* », la tribune revendique en particulier l'ouverture des données de l' *Ordnance Survey*, l'institut géographique du Royaume-Uni, dont l'utilisation est soumise au paiement d'une redevance. L'argument, qui deviendra par la suite un fondement central de l'*open data*, consiste à revendiquer l'ouverture de ces données en vertu du fait qu'elles sont produites à partir de l'argent des contribuables. En 2007, le mouvement *open data* entre dans le débat politique aux États-Unis, à la suite d'une réunion qui se tient à Sebastopol, en Californie, et qui regroupe des activistes numériques bien connus comme Lawrence Lessig, Tim O'Reilly et Aaron Swartz, en vue de formuler des principes susceptibles d'être repris par les candidats à l'élection présidentielle américaine. Ces principes exigent la libération des données publiques dans leur intégralité dès leur production et telles qu'elles sont collectées, assortie à une licence ouverte permettant leur réutilisation par quiconque. Cette demande de données « brutes » vise l'émergence d'une nouvelle ère de la transparence en réduisant les asymétries d'information entre l'administration et les citoyens. Elle vise également l'émergence d'une vague d'innovation via la réutilisation des données par les développeurs et les entreprises ainsi qu'un meilleur usage des ressources publiques. L'injonction à l'ouverture des données publiques a engendré une prolifération de portails qui diffusent ces données partout dans le monde avec, au niveau national, le lancement de data.gov aux États-Unis en 2008, data.gov.uk en Royaume-Uni en 2009 et data.gouv.fr en 2011. En 2013, les principes de l'*open data* sont repris par le G8 qui, dans sa charte sur l'*open data*, se fixe les mêmes objectifs de transparence et de création économique, et établit par ailleurs que l'ouverture des données deviendra la pratique par défaut des administrations des huit pays signataires.



Acteurs et bénéficiaires de l’ouverture des données publiques, image extraite de la vidéo « l’open data à la loupe » de l’association Libertic

Source Kiosque Paris, Licence : Art Libre

Une ressource qui n’est pas gérée comme un bien commun

L’injonction à l’ouverture des données publiques s’appuie sur l’argument selon lequel les données produites par les administrations et financées par l’argent public constituent un bien commun qui doit être partagé avec la société. Mais les données sont-elles gérées comme un bien commun ? Existe-t-il des mécanismes de gouvernance partagée qui permettraient à la société de gérer les données ouvertes comme une ressource commune ?

En exigeant des données « brutes », les militants de l’ouverture des données ont attiré l’attention sur des fichiers qui n’étaient jusqu’alors pas partagés et servaient au travail quotidien des agents de l’administration. Leur publication est donc inédite pour les administrations et cela amène naturellement des résistances et des questionnements. La plupart des données publiées sur les portails *open data* sont des fichiers administratifs et non des données statistiques produites

spécifiquement pour créer un savoir général sur la société. Ces données sont en effet produites pour répondre à la mission d'un service administratif et non en vue de devenir une ressource susceptible d'être utilisée par des citoyens. La sélection des données qui vont être publiées sur ces portails *open data* fait aussi l'objet de négociations entre les personnes en charge d'ouvrir les données et leurs producteurs. Ce travail de sélection, que nous avons observé lors d'une enquête dans des administrations françaises ayant libéré des données publiques [2], va à l'encontre des principes fondamentaux de l'*open data*, définis en 2007 à Sébastopol, qui exigent l'ouverture de toutes les données publiques dès leur production. Dans la pratique, les données sont au contraire sélectionnées par l'administration selon des critères variés comme la facilité techniques de leur extraction, l'intérêt pour des ré-utilisateurs potentiels ou encore la publication d'une donnée similaire sur d'autres portails *open data*. De plus, la question du degré de sensibilité des données, c'est-à-dire du risque que leur ouverture représenterait pour le public mais également pour l'administration qui les produit, est aussi souvent cruciale. Enfin, les données sont travaillées en vue de leur publication ; elles ne sont pas brutes au sens d'inaltérées. Les producteurs de données les éditent pour enlever des éléments problématiques, nettoyer leur mise en forme ou encore effacer les commentaires, acronymes et colonnes qui n'ont pas d'intérêt hors de leur contexte de production.

Toutes ces étapes qui sont cruciales dans la construction des données publiques échappent au regard de leurs utilisateurs. La production, la sélection et la modification des données publiques sont donc invisibles aux citoyens alors même que c'est dans les coulisses de ces administrations que se fabrique la transparence de l'État. Les citoyens n'ont pas non plus la possibilité de participer à la gouvernance de ces données en contrôlant leur production et leur diffusion. Pourtant, une telle régulation existe pour les données des organismes statistiques dont la mission est de produire un savoir général sur la société. En France, les statistiques publiques font l'objet depuis 1972 d'une gouvernance partagée par un organisme de concertation, le Conseil National de la Statistique (CNIS), composé de représentants des institutions, des syndicats, des organisations patronales, de la recherche et de la société civile. Le CNIS peut ainsi délibérer sur la nécessité de la production, de la diffusion ou encore sur les catégories en vigueur concernant les données produites par les organismes statistiques. Une telle gouvernance partagée n'existe pas encore pour les données des administrations dont la gestion échappe au contrôle des citoyens.

Les commoneurs [3] : producteurs et gestionnaires de données comme biens communs

En l'absence d'une gouvernance effective et redevable des données publiques, des commoneurs se sont saisis des données publiques pour préserver leur caractère de biens communs. L'*Open Knowledge Foundation*, un réseau mondial militant pour l'ouverture des contenus et des données, a publié en 2009 la licence *Open Database License* (ODbL) qui applique le principe du *copyleft* aux données ouvertes, ce qui était auparavant impossible car la licence *Creative Commons* ne s'appliquait pas aux bases de données. La licence ODbL exige le partage avec la même licence des données réutilisées, préservant ainsi leur caractère de biens communs. Des administrations comme celle de la ville de Paris ont utilisé cette licence pour éviter une éventuelle *enclosure* de leurs données, craignant que leur exploitation par des acteurs privés ne bénéficie pas à la

collectivité en retour. Des commoneurs ont aussi contribué à faire émerger les données publiques comme un bien commun en republiant leurs améliorations. En France, l'association Regards Citoyens extrait automatiquement les données issues des sites de l'Assemblée Nationale et du Sénat pour les exploiter sur son observatoire citoyen de l'activité parlementaire, nosdeputes.fr et nossenateurs.fr. Ces données sont republiées en *open data* dans des formats exploitables par les machines et ont permis le développement de nombreux autres outils de surveillance de l'activité parlementaire. De manière similaire, le site *OpenCorporates* regroupe les données des registres des entreprises de 75 juridictions dans le monde et permet à quiconque de les exploiter, même pour des usages commerciaux, tant que la clause de partage à l'identique garantissant le caractère de biens communs des données est respectée.

Par ailleurs, des bases de données gérées comme un bien commun sont aussi produites par les citoyens. Le cas le plus connu est celui d'OpenStreetMap (OSM), une base de données géographiques mondiale qui a été créée en 2004 par un chercheur britannique à la suite du refus de l'*Ordnance Survey* de partager ses données gratuitement. OSM repose sur les contributions des citoyens qui éditent le « Wikipedia de la carte ». La base de données couvre désormais la plupart des pays du monde et concurrence le géant Google Maps. En 2013, OSM a célébré son millionième contributeur. Dans les sciences, TeleBotanica est un réseau de botanistes, professionnels et amateurs, qui créent collaborativement une base de données partagée sur la faune et la flore partout dans le monde. On peut aussi citer OpenFoodFacts qui permet à chacun de scanner les données nutritionnelles présentes sur les emballages des aliments, ou encore OpenMeteoData pour le partage libre de données météorologiques. Tous ces projets reposant sur la participation des citoyens à la création de bases de données sous licence libre complètent les données publiques, et ils nous permettent d'envisager une gouvernance partagée de ces ressources numériques essentielles.

Notes

[1] Bruno J Strasser, "The Experimenter's Museum GenBank, Natural History, and the Moral Economies of Biomedicine," Isis, 2011, vol. 102, no. 1, p. 60–96.

[2] Jérôme Denis et Samuel Goëta. « La fabrique des données brutes. Le travail en coulisses de l'open data », in *Penser l'écosystème des données. Les Enjeux Scientifiques et Politiques des données numériques*, dir. Mabi Plantin Monnoyer- Smith, éditions FMSH, à paraître. Accessible en ligne : http://halshs.archives-ouvertes.fr/halshs-00990771

[3] Les commoneurs sont des individus qui produisent ou entretiennent des biens communs dont ils participent collectivement à la gestion. (voir David Bolier, *La Renaissance des communs. Pour une société de coopération et de partage*, Traduit de l'américain par Olivier Petitjean, C&F éditions, 2014)

Un documentaire et un débat pour questionner la monétisation des données

Ce documentaire visibilise la monétisation et utilisation des données. Suivies d'un débat / et ou mise en perspective avec un groupe , c est un outil de sensibilisation adéquat

Pour se décentrer de l’argument souvent entendu Moi mais j'ai rien a cacher !

Guide d'autodéfense intellectuelle face à l'information

Comment cultiver son hygiène préventive au jugement et réagir face aux flux d'information. Quel attitude et méthode adopter pour éviter de se faire avoir par une fausse informations. Quels outils numériques pour nous aider ?

Suivez le Guide ici.

Un documentaire pour questionner les données et leur €

Un documentaire a visionner et ensuite à mettre en débat avec un groupe.

Il s'agit avec ce documentaire de déconstruire certains arguments souvent entendu : comme je n'ai rien a cacher, je m en fiche de cette loi / conditions de vente / conditions d utilisation... Et ainsi prendre conscience des enjeux systémiques : individus et sociétaux.

Rien a cacher ?