

Big data, la donnée c'est toi !

Quelques repères sur la connexion
de nos subjectivités

Thomas Lemaigre

« Comment fonctionne internet, à l'attention de ceux qui n'y comprennent rien, mais qui se disent quand même qu'il se passe là quelque chose d'important ? » Tel aurait pu être le titre de cet article. Un quart de l'humanité vibronne sur ce réseau des réseaux. Internet est un continent économique, social et culturel à part entière, où les rapports de force ont diamétralement changé en un quart de siècle. En témoigne la maîtrise des données sur les utilisateurs : pour les capter, les stocker et les traiter, fabriquer internet revient désormais à doter nos subjectivités de multiples appareils connectés.

« La difficulté, c'est de transmettre à chaque homme la sensation que même au fond d'une cellule exigüe, il conserve une capacité de contrôle sur le monde qui est le sien », Gonçalo M. Tavares¹.

« We know where you are. We know where you've been. We can more or less know what you're thinking about », Eric Schmidt, ancien PDG de Google, 2010².

L'interconnexion d'une infinité d'ordinateurs, c'est la définition minimale que l'on peut donner d'internet. Cette interconnexion repose sur toutes sortes d'appareils : des câbles, des routeurs, des ordinateurs dédiés que l'on appellera des serveurs, éventuellement réunis en grand nombre dans des fermes comme le centre de données ouvert par Google

à Saint-Ghislain (Mons) depuis 2007. Et cette interconnexion devient interaction grâce à des langages informatiques, dont le plus visible est le HTTP, inventé en 1992. C'est bien comme une combinaison de machines et de langages qu'internet est inventé au tournant des décennies 1980-1990. Sans entrer dans les détails sur ces débuts, il est très instructif de se demander dans quels terreaux culturels ils se sont joués. Le premier est sans doute l'armée américaine, en quête d'avantages stratégiques avec un système de télécommunications

1 | *Apprendre à prier à l'ère de la technique*, éd. Viviane Hamy, 2010 (trad. D. Nédellec).

2 | « Nous savons où vous êtes. Nous savons où vous êtes allés. Nous pouvons à peu près savoir ce que vous pensez », interview vidéo du 4 octobre 2010 sur <http://read.bi/VLnvPc>.

si décentralisé qu'il resterait fonctionnel en cas de frappes nucléaires soviétiques. Le deuxième est la communauté scientifique qui y instille ses valeurs : « coopération, réputation entre pairs, autonomie, gratuité, consensus et liberté de parole. [...] Dépourvue de centre, (leur) architecture rend d'emblée difficile le contrôle et vise à favoriser la connectivité et l'expansion du réseau » (Richard, 2015). Troisième milieu qui féconde l'imaginaire d'internet : un ensemble épars d'entrepreneurs et de techniciens influencés par la contreculture post-hippie de la côte Ouest, à la fois radicalement individualiste et libertarienne, attirée par la technologie en tant que moyen d'émancipation en marge de la société, pour peu qu'on parvienne à la détourner et se l'approprier. Dans les années 1980 déjà, le numérique, c'était d'un côté l'informatique propriétaire avec le leader IBM, première firme à avoir tiré des profits importants de ces technologies en mettant sur le marché les premiers PC, et de l'autre la naissance du logiciel libre et des licences qui en empêchent la privatisation — celles-là mêmes qui ouvriront la voie aux *creative commons*, alternative aux formes usuelles de propriété intellectuelle, jugées trop restrictives.

Cette généalogie composite va structurer l'histoire d'internet et les récits à son propos, tendus entre ouverture et contrôle, liberté et domination, gratuité et libre entreprise ; elle est au cœur d'une dynamique de rapports de force entre ses acteurs et marque pour longtemps encore notre rapport ambigu au numérique. C'est ce qu'illustrera notre chronologie au galop en tentant de lier trois niveaux de réalités : l'évolution des techniques, l'évolution des usages et l'évolution des discours et des représentations.

Entre nouveau monde et nouvelle économie (1992-2000)

La première moitié des années 1990, internet reste dominé par les chercheurs, les techniciens et les artistes. On y navigue à l'aventure, en traçant son chemin de lien en lien et en suivant les recommandations d'annuaires et autres portails compilés bénévolement et plus ou moins spécialisés. C'est aussi par courrier électronique qu'on s'échange les adresses de sites web. « L'imaginaire qui se met en place dans ces années-là est fondamentalement ambigu. Le cyberspace peut être aussi bien libertaire que libéral » et se pense autour de termes comme la dématérialisation, la réinvention continue de soi, l'adaptation au changement, l'obsolescence des médias traditionnels et de l'État, « entre révolution et entreprise, nouveau monde et nouvelle économie » (Richard, 2015).

Fondatrice, cette époque a encore une place de premier plan dans la représentation que se fait tout un chacun d'internet. Mais elle sera brève. Entre 1997 et 2001, la première bulle spéculative gonfle autour des sociétés qui ont développé de nouvelles offres et activités en ligne, depuis la libéralisation des noms de domaine « .com » en 1995. Cette période voit internet se restructurer. Quelques sites commencent à concentrer des volumes de trafic très importants : certains portails et les premiers moteurs de recherche (Altavista). Le commerce en ligne s'impose avec Amazon, AOL et Ebay. Ces sociétés connaissent des capitalisations boursières records, tout comme quelques autres actives dans les logiciels (Netscape), la micro-informatique (Intel) et les communications en général (Alcatel, Lucent, Cisco), voire la production de contenus (Vivendi). C'est aussi dans cette ambiance que sont dérégulés les marchés des télécoms des pays de l'OCDE et privatisés ou autonomisés les monopoles historiques, à l'instar de

Belgacom. L'accès aux capitaux leur permet, avec leurs concurrents, de développer et de démocratiser les réseaux à haut débit (ADSL et télédistribution) et la téléphonie mobile.

Beaucoup de monde se fait beaucoup d'argent grâce à internet, parfois avec de très petites mises de départ, ce qui enflamme les fantasmes d'entrepreneurs de tout poil et nourrit les ambitions d'acteurs des marchés financiers comme les fonds d'investissement en capital-risque. C'est la première bulle internet, qui finit par exploser avec le krach de mars 2000 : il n'est plus possible aux entrepreneurs du web de répondre aux attentes de gains de leurs apporteurs de capitalisation boursière ou leurs créanciers.

Nombre de nouveaux services en ligne reposent bientôt sur une petite invention : le *cookie*. Il s'agit d'un simple fichier texte qu'un site web dépose sur votre ordinateur et qui va y rester plus ou moins longtemps. Quand vous repasserez sur ce site, son *cookie* pourra, avant de se remettre à jour, lui rendre des informations, par exemple le contenu de votre panier d'achats pour un site de vente en ligne. Si l'idée de base est d'authentifier l'accès à votre compte et de fluidifier votre navigation, par exemple en préremplissant des formulaires en ligne, le *cookie* est un outil très puissant et très peu transparent : non seulement il renseigne les entreprises du web sur vous (données personnelles, numéro de carte de crédit, codes d'accès) et sur vos pratiques de navigation (lieux et temps de connexion, matériel, autres pages visitées), mais il peut aussi être déposé dans votre machine par un site web que vous n'avez jamais visité (mais qui incruste de la publicité dans un site que vous visitez) — c'est cette technique qui vous fait envoyer de la publicité ciblée. Si le *cookie* peut être un mouchard très efficace avec les raffinements technologiques qui se succéderont, ces *cookies*

particuliers dits de tierce partie transforment vos sites préférés en rien de moins que des détourneurs de temps de connexion. Ils monétisent votre attention, soit en louant à un tiers l'accès à votre session, soit en revendant à des intermédiaires spécialisés les données collectées à votre sujet, le tout à votre insu. Ils transforment le face-à-face avec votre machine en deux choses : une marchandise et un acte qui n'a plus rien d'intime ou de discret. De façon générale, la gratuité d'internet n'aura désormais plus rien de désintéressé : le service va être pensé comme un moyen de vendre de l'audience à des annonceurs.

Le désenchantement du web (2000-2007)

La décennie qui s'ouvre va voir au moins trois évolutions cruciales d'un internet qui n'est pas encore le fameux « web 2.0 » : la suprématie de Google, l'apparition des *big data* et le retour de l'État.

Google s'impose comme moteur de recherche à partir de 2000. Pour le sociologue Dominique Cardon, il s'agit d'un changement de paradigme : internet fonctionnait sur des logiques de popularité, avec des techniques de mesure d'audience similaires à celles de la presse traditionnelle. Avec Google, il bascule dans une logique d'autorité. Le critère de base, à l'instar des citations dans la littérature scientifique, est le nombre de liens qui pointent vers un site web. Telle est la clé de l'un des plus fameux algorithmes : le *pagerank*, qui calcule l'ordre d'affichage des résultats d'une requête. D'autres critères du même type sont pris en compte, notamment en enregistrant et en traitant les milliards de requêtes reçues. Google va vite diversifier ses services pour devenir l'acteur dominant d'internet (cartographie, courriel, pub, blogs, agenda, traduction et mesure d'audience, c'est-à-dire accès aux données de trafic d'innombrables sites tiers). Google, suivi par d'autres, s'est

en réalité tôt mis à traiter des énormes quantités de données sur ses utilisateurs pour personnaliser les pages que ses sites affichent et les liens publicitaires qu'elles leur propulsent sous les yeux.

Cette évolution favorise et est attisée par une accélération spectaculaire, autour de 2002, des capacités de stockage de l'information, de la puissance des processeurs et de la diminution de leurs prix. Tout peut désormais être stocké et traité. À ces ressources, Google ajoute deux atouts de taille en matière de qualité des données : la gratuité de leur captation et leur diversité. Les *big data* seront désormais centraux dans le fonctionnement d'internet, basculement discret qui n'est porté à la connaissance du public qu'à postériori.

Troisième tournant important : l'arrivée de l'État sur la scène internet dans la foulée des attentats du 11 septembre 2001. Le *Patriot Act* d'octobre 2001 ouvre aux administrations américaines des facilités juridiques et des budgets considérables pour surveiller les communications, au point qu'il faudra plus de dix ans pour qu'on se rende compte des proportions que cette activité a prises. Ici aussi, les *big data* sont un enjeu central et plus que discret alors que les usages d'internet continuent à se démocratiser et à se banaliser.

Keep smiling (2007-aujourd'hui)

Une nouvelle forme d'interaction en réseau, très décentralisée, s'était installée et banalisée au début de la décennie 2000 avec Napster et les applications d'échange pair-à-pair de fichiers multimédias. Il s'agissait ici d'échapper à la fois au marché et à l'État, au droit d'auteur et aux agences supposées le faire respecter. Cette nouvelle utilisation des réseaux va très vite déstabiliser les industries culturelles et être interdite (Hadopi, par exemple). Sans avoir d'effet massif, ces interdictions ouvrent la voie à une économie pirate en ligne, d'un côté, et, de l'autre, incitent de nou-

veaux acteurs économiques à récupérer ces logiques de partage et de communautés dans des entreprises marchandes. Entre utopies libertariennes, contrôle et — encore — liberté d'entreprise.

On l'a compris, c'est la naissance des réseaux sociaux, avec le premier d'entre eux, Facebook, qui, à partir de 2007, se répand comme une trainée de poudre au-delà des États-Unis. Ici, toujours selon Cardon, apparaît une troisième génération d'algorithmes : ceux qui produisent de la réputation. Ce à quoi l'on porte attention acquiert de la valeur et cette valeur permet d'attirer l'attention. Tout cela se calcule, toujours de façon invisible, pour susciter un maximum de passages à l'acte (clics, *likes* et autres) permettant de capter un maximum de données. L'exemple par excellence de ce type de dispositif est l'algorithme de recommandation de sites comme Amazon : si vous achetez tel ouvrage, il vous en propose trois ou quatre autres, ceux qui ont également été vendus à l'acheteur du premier ouvrage. Si vous avez consulté telle et telle autre page juste avant, le choix qui vous est proposé sera encore plus pointu... et infallible.

Les réseaux sociaux sont le fer de lance d'une évolution plus large, celle du web 2.0 : les techniques d'édition de pages html et de communication entre sites internet permettent désormais de mettre en scène des possibilités d'interaction autour des contenus en ligne. On passe de flux d'information essentiellement à sens unique à un internet pensé et utilisé comme un véritable espace de socialisation. Et de nouveau, chaque interaction produit de la donnée.

Ce basculement est soutenu par la vague d'objets nomades qui déferle sur les marchés, avec, par exemple, l'iPhone 3G d'Apple lancé mi-2008. Ces appareils sont accessibles partout et tout le temps et permettent des interactions en ligne beaucoup plus conviviales. Le

smartphone n'est en effet rien d'autre qu'un minuscule ordinateur, connecté à internet et intégré à un téléphone portable complété par des systèmes de prise de vue (photo et vidéo), de géolocalisation (GPS) et autres. Très vite naissent des tas de variantes (Ipod, tablettes, etc.), mais surtout une énorme diversité d'applications, ces petits logiciels permettant d'activer une offre de service en ligne. Il en résulte une augmentation spectaculaire des temps d'utilisation et du trafic, et donc des volumes de données échangées — captées — et de leur variété. Il se vend aujourd'hui plus d'un milliard de smartphones par an et toutes les « apps » qu'on peut y installer ne sont disponibles que sur... trois plateformes (Itunes d'Apple, Googleplay et le Store de Microsoft). Chacun de nous ayant plusieurs points d'accès à internet, il stocke de plus en plus ses données en ligne plutôt que dans la mémoire des appareils — vous avez reconnu ce qu'on appelle le *cloud*, le nuage, en clair les serveurs de ces mêmes multinationales... qui ne doivent même plus se donner la peine de venir chercher vos données chez vous.

Tous ces flux de données sont de mieux en mieux monétisés par des intermédiaires spécialisés et très discrets. Par exemple, la page d'accueil du site du journal *Le Soir* vous connecte à au moins dix-huit autres sites ou serveurs tiers³, soit autant d'entreprises qui vont transformer en marchandise votre généreux et innocent passage. La marchandisation de la culture, dénoncée dans les années 1980 par les opposants à la publicité à la télévision, apparaît rétrospectivement comme de bien innocentes semonces...

3| Chacun peut facilement remplacer son navigateur propriétaire (Chrome, Safari ou Explorer) par le programme libre Firefox, qui permet d'installer des modules complémentaires comme Ghostery ou Ublock Origin, qui permettent de détecter ou d'empêcher ces interactions non sollicitées.

Big Brother is hijacking you

La crise des subprimes de 2008 ne mettra pas de frein à l'évolution d'internet. Au début de la décennie 2010, la quantité et la variété des contenus qui y circulent sont inégalées, les « tuyaux » qui les véhiculent ont des capacités jamais atteintes (norme 4G de débit des données de téléphonie mobile, fibre optique) et l'écosystème économique lié est extrêmement concentré autour de quelques multinationales de droit américain — on parle désormais des Gafam (Google, Amazon, Facebook, Apple, Microsoft) comme d'un quasi-cartel.

Avec l'économie de plateforme que génèrent les applications de tout poil, les impacts d'internet sur la vie quotidienne sont de plus en plus visibles (Szoc, 2015). Airbnb et l'hôtellerie, Uber et le transport de personnes, etc., tous ont des impacts économiques qui se mettent à marquer directement les territoires. Internet n'est plus un monde à part, soi-disant virtuel. Les jeux vidéo, les logiciels pédagogiques, les liseuses, les *chatbots* (voir la nouvelle « Ma femme 3.0 » dans la rubrique « Italique »), structurent la vie quotidienne de concitoyens chaque jour plus nombreux. Les données qui circulent ne se limitent plus à votre identité ou vos temps de connexion : il s'agit aussi de votre santé, votre carrière, votre pouvoir d'achat et vos consommations, vos déplacements, vos centres d'intérêt, vos opinions, votre réseau de relations. Et cette numérisation du quotidien n'a même pas encore pris sa véritable dimension puisque la fin de la décennie verra vraisemblablement l'essor des objets connectés à internet : jouets, vêtements, domotique, véhicules, drones, etc. Déjà votre téléviseur transmet des données sur ce que vous regardez, mais en plus, pour peu qu'il soit équipé d'une commande vocale, il transmet aussi des enregistrements de votre voix — afin d'améliorer sans cesse les logiciels de reconnaissance vocale, bien sûr... (Manise, 2016). « À la limite, pourquoi

développer des techniques de surveillance directe [...] ? On n'aurait plus qu'à coller son oreille sur la coquille électronique et à écouter le son de l'océan des données » (Chamayou, 2015).

On n'est plus très loin des « télécrans » de 1984 d'Orwell, mais avec plusieurs dimensions en plus : ce n'est pas Big Brother qui vous a imposé l'omniprésence de l'appareil, c'est vous qui l'avez acheté ; personne ne vous oblige à lire en ligne tel quotidien, mais personne ne l'oblige non plus à vous dire à combien de « partenaires » il propose de capter votre temps de navigation. Le tout dans la convivialité, la même que dans les shows télévisés qui servent à vendre votre « temps de cerveau » à telle ou telle marque. Bref, le cauchemar d'Orwell sans doute, mais aussi celui de Guy Debord, pourfendeur de la publicité et de la société du spectacle⁴. Sans oublier une ultime couche, plus préoccupante encore, celle des biais plus ou moins intentionnels introduits en toute opacité dans notre perception du monde ainsi appareillée, comme on commence à peine à l'entrevoir avec les travaux de Citton, Stiegler ou Rouvroy (PointCulture, 2016 — voir les autres contributions de ce dossier).

Tout va donc bien dans un monde de participation et de partage, les souris cliquent, les serveurs chauffent, les capitaux se concentrent et sont généreusement rémunérés⁵. On se rassure en se disant que les données collectées sont si foisonnantes que nul n'est capable de les traiter, ou alors si mal (Chamayou, 2015). Un quart de la population mon-

diale utilise internet. Le consommateur conscientisé, lui, peut toujours modifier les réglages par défaut de ses applications pour se protéger de la publicité ou naviguer discrètement. Et les utopistes, les libertariens ou les défenseurs des communs conquièrent eux aussi de nouveaux territoires, avec l'encyclopédie collaborative Wikipedia, le site Wikileaks qui sert de portevoix et de couverture aux lanceurs d'alertes, ou encore le Bitcoin et les monnaies sans autorité d'émission⁶.

Tout va bien... jusqu'au 6 juin 2013, jour où commence un nouveau retournement. Le lanceur d'alerte Edward Snowden se met à distiller dans la presse des documents classifiés à propos des dispositifs de surveillance des communications électroniques des États-Unis et du Royaume-Uni. On savait depuis le *Patriot Act* que de tels dispositifs existaient, et avaient déjà donné des résultats tant sur le territoire américain que sur des théâtres d'opérations extérieures (Pakistan). On savait qu'« On ne surveille pas les terroristes présumés, on surveille tout le monde dans l'hypothèse où se cacherait un terroriste dans la foule » (Campion, 2013). Mais ce qui frappe, c'est plutôt l'incroyable servilité des multinationales, Gafam et autres, qui développent les applications en ligne, celles qui fournissent les accès à internet et celles qui possèdent ses infrastructures (Pesnot, 2014). L'exemple le plus frappant est celui des portes dérobées, des morceaux de programme installés secrètement dans des logiciels (par exemple de gestion de routeurs ou de chiffrement) pour transmettre les informations de ses utilisateurs, en temps réel et sans intermédiaire, à des agences de renseignement. Un autre enseignement des révélations de Snowden concerne les nouvelles méthodes de traitement des données collectées. Que l'on soit une multinationale

4| Dès les années 1960, Debord reformule la théorie marxiste à la lumière des médias de masse et de l'industrialisation de la culture, et cherche les formes de résistance possible à ces nouvelles formes de disparition du sujet.

5| On pourrait ici noyer le lecteur sous un torrent de chiffres illustrant la dimension financière indispensable aux évolutions décrites. Il suffit de relever l'évolution de la valeur de marché des actions en circulation de Google : en dix ans (2004-2014), elle a augmenté de 1294 %. La personne qui aurait investi 1 000 \$ dans ces actions au moment de leur introduction en Bourse le 19 août 2004 se trouvait dix ans plus tard avec 13 945 \$!

6| Sans oublier les *makers*. Voir l'article éponyme de Lionel Maes, *La Revue nouvelle*, n° 5, 2015, www.revue-nouvelle.be/Makers.

du web ou un service de renseignement, il y a moyen de prendre d'immenses quantités de données et de les traiter pour en faire émerger des récurrences ou des « signatures ». C'est ce que l'on appelle l'apprentissage profond ou l'apprentissage semi-supervisé par la machine⁷ — tout l'inverse de ce que vous avez appris en matière de calcul de probabilités dans vos premières années d'université, où il fallait toujours partir d'hypothèses, bref commencer par savoir ce que l'on cherchait. Ici, ce sont ces algorithmes, que Cardon appelle prédictionnels, qui vont produire des analyses de la réalité à partir de données brutes, et affiner leurs propres règles en fonction de ce qu'ils découvrent.

C'est une évolution récente, apparue cette décennie et pas encore arrivée à maturité. C'est surtout un nouveau moment de bascule dans la nature même d'internet puisque se démultiplient par là les usages et la valeur que l'on peut tirer des *big data* captés. Bref, alors qu'internet s'est organisé comme une vaste machine de surveillance publique et privée, que le pouvoir s'y est concentré à outrance, en Europe continentale pas moins qu'ailleurs, l'évolution des technologies et des usages dote les acteurs dominants d'outils de maîtrise de l'information qui désormais dépassent littéralement l'entendement.

Face à tout cela, le *web business* continue comme depuis un quart de siècle, ce qui n'est pas sans poser de profondes questions sur la (dé)mobilité des consciences chez nos contemporains. Heureusement, contrepropositions et contrepouvoirs se multiplient, certes traversés par les tensions congénitales d'internet que nous avons effleurées. Les logiques solidaires et non marchandes commencent elles aussi à se nourrir des *big data* et trouvent dans internet de quoi se réinventer et faire mouvement — sans avoir encore trouvé d'effets de seuil.

7| Voir un exemple de manuel dans les documents publiés par Snowden, le « Data Mining Research Problem Book » des services anglais (2011), <http://bit.ly/2ff87bp>.

Bibliographie

Campion B. (2013), « Espionnage numérique : le trou noir de la surveillance privée », *La Revue nouvelle*, n° 9, septembre, <http://bit.ly/2fmcu10>

Chamayou Gr. (2015), « Dans la tête de la NSA », *Revue du Crieur*, n° 1, juin, p. 20-39

Citton Y. (2014), *Pour une écologie de l'attention*, coll. « La couleur des idées », Seuil

Manise J.-L. (2016), « Vie privée : votre meilleur ennemi est dans la poche droite », *Secouez-vous les idées*, n° 107, Cesep, septembre

Pesnot P. (2014), « L'affaire Edward Snowden », émission Rendez-vous avec X, 24 mai, France Inter, <http://bit.ly/2eLjJcK>

PointCulture Bruxelles (2016), « Pour un numérique humain et critique », cycle de conférences, captations vidéo sur <http://bit.ly/2fm7mL5>

Richard Cl. (2015), « Penser internet. Une histoire intellectuelle et désenchantée du réseau », *Revue du Crieur*, n° 2, octobre, p. 144-159

Szoc E. (2015), « Du partage à l'enchère : les infortunes de la "sharing economy" », association Joseph Jacquemotte, <http://bit.ly/2fDwfBo>

La quasi-totalité des dates et données chiffrées proviennent de Wikipedia, consulté en octobre 2016.