

Qu'est-ce que l'identité numérique ?

Olivier Ertzscheid, *Qu'est-ce que l'identité numérique ? Enjeux, outils, méthodologies*, Encyclopédie numérique | 1, 2013.

À la recherche des premières traces de l'identité numérique

En mars 1974, le projet français de système automatisé pour les fichiers administratifs et le répertoire des individus (SAFARI) visait à permettre l'interconnexion des administrations à partir d'un matricule commun. Le rejet de ce projet fut à l'origine de la loi « Informatique, fichiers et libertés » du 6 janvier 1978. Un débat, récurrent depuis, se fait jour autour du partage administratif de données de plus en plus étendues sur les individus, dans le cadre de l'administration électronique.

En 2002, lors de la remise officielle du rapport *Administration électronique et protection des données personnelles*¹, Michel Sapin, ministre de la Fonction publique et de la Réforme de l'État se prononce contre un identifiant unique de l'utilisateur.

Nous devons partir du principe que l'identité numérique n'est pas et ne peut pas être unique, pas plus que l'identité au sens traditionnel des relations « papier » avec l'administration. De la même façon que nous disposons aujourd'hui, entre autres, d'un numéro de Sécurité sociale, d'un numéro fiscal (le SPI), d'une carte d'identité, d'un passeport, autant « d'identifiants » distincts les uns des autres, nous aurons demain plusieurs identifiants électroniques [...].

Identifiants, identification, identité(s)

Jusque vers le milieu des années 1990, l'identité numérique est d'abord une question « d'identifiants numériques », parmi lesquels la sécurisation des données est prédominante et concerne principalement les entreprises et les administrations. L'écho sociétal de ces questions est quasi inexistant.

De la fin des années 1990 au début des années 2000, commence à apparaître la problématique de la « vie privée » tandis que les grands acteurs du Web (France Télécom, Cisco, Sun, eBay) réfléchissent à « des normes mondiales pour la gestion des données personnelles et les procédures d'authentification »². À partir de 2003, on assiste au lancement des grands réseaux sociaux, dont MySpace, Friendster et LinkedIn³. Dans le même temps, Google est devenu l'outil incontournable que nous connaissons aujourd'hui et le premier miroir identitaire grand public en dépassant le cap des trois milliards de pages indexées fin 2002. Dès 2005, les premières traces sociétales de la problématique de l'identité numérique apparaissent, deux ans après le lancement des réseaux sociaux, au moment où les usages se sont massifiés et banalisés.

Définition de l'identité numérique ?

L'identité numérique est constituée de la somme des traces numériques se rapportant à un individu ou à une collectivité : des traces « profilaires » correspondant à ce que je dis de moi (qui suis-je ?) ; des traces « navigationnelles » qui renseignent sur les sites que je fréquente et sur lesquels je commente ou j'achète (comment je me comporte) ; enfin des traces inscriptibles et déclaratives – ce que je publie sur mon blog par exemple – qui reflètent directement mes idées et mes opinions (ce que je pense).

De manière plus circonscrite, l'identité numérique peut être définie comme la collection des traces (écrits, contenus audios ou vidéos, messages sur des forums, identifiants de connexion, etc.) que nous laissons derrière nous, consciemment ou inconsciemment, au fil de nos navigations sur le réseau et le reflet de cet ensemble de traces, tel qu'il apparaît « remixé » par les moteurs de recherche.

Mon identité numérique c'est : adresse IP, cookies, courrier électronique, nom, prénom, pseudos, coordonnées (personnelles, administratives, bancaires, professionnelles, sociales), photos, avatars, logos, tags, liens, vidéos, articles, commentaires de forums, données géolocalisées, etc.

Identité ou réputation ?

Complément parfois pesant de l'identité numérique, la réputation numérique ou « e-réputation » correspond à ce que l'on dit de moi. Elle

peut également constituer ma « marque » (on parle alors de *personal branding*). Elle est nécessairement subjective et fluctuante. Reposant sur l'image perçue mais également sur la confiance ou la crédibilité accordée, elle peut se déconstruire aussi rapidement qu'elle est longue à établir et à instaurer.

Où sont mes identifiants ?

À la multiplication des services en ligne correspond l'accroissement exponentiel de nos identifiants, c'est-à-dire des couples associant un nom d'utilisateur et un mot de passe, et permettant d'accéder aux services. Il n'est pas rare qu'un même individu possède plusieurs dizaines d'identifiants et autant de mots de passe différents. À tel point que l'on a vu émerger des logiciels dédiés à la seule gestion de ces identifications multiples.

Aujourd'hui, de plus en plus de services et d'acteurs du Web ont choisi de simplifier la tâche de leurs utilisateurs en adhérant à un système d'identifiant unique baptisé OpenID : une paire unique identifiant/mot de passe permet ainsi d'accéder à ses courriels, à l'administration de son blog ou bien encore à son compte Facebook ou Twitter.

Les grands navigateurs (Firefox, Google Chrome mais également Internet Explorer) travaillent de leur côté à une gestion directe et « native » des identifiants de connexion depuis leur interface. Soucieux, en tout cas en apparence, de mettre en avant la protection de la vie privée de leurs utilisateurs, ils proposent également des options de navigation anonyme, c'est-à-dire sans stocker les habituelles données de connexion⁴.

Les enjeux

Machineries et fausses transparences identitaires

L'interrogation de Google nous donne l'illusion d'une page blanche, composée de la seule zone de recherche. Pourtant, dans nos pratiques quotidiennes, nombre d'entre nous commencent leur journée de travail par le relevé de leur courrier électronique, souvent sur la plate-forme Gmail du même Google. À partir de ce moment, et lorsque nous rebasculons vers le moteur de recherche, nous sommes nominativement identifiés par Google qui dispose ainsi d'une transparence totale sur l'ensemble de nos requêtes et a accès à la totalité des messages que nous avons envoyés et que nous avons reçus.

L'exemple de Gmail

Quand Google ouvre, le 1^{er} avril 2004, son service de courrier électronique en ligne (Gmail) il augmente très significativement l'espace de stockage offert aux internautes et, surtout, il y incorpore la logique et la recette de sa régie publicitaire AdSense : l'ensemble des messages de notre correspondance privée seront considérés comme des pages web et à ce titre, scannés et indexés par les algorithmes et les *crawlers* de Google.

L'objectif : associer aux mots-clés caractéristiques de nos conversations des publicités les plus ciblées et contextuelles possibles. Chaque message est ainsi indexé mais Google est également en capacité d'analyser les thèmes ou sujets de conversation dont nous parlons le plus souvent, les personnes avec lesquelles nous communiquons le plus et les sujets que nous abordons, de manière à rendre sa régie publicitaire encore plus efficace. Si tous les autres webmails concurrents affichaient, en guise de modèle économique, des bannières d'annonceurs, c'est avec Gmail que notre correspondance privée est pour la toute première fois indexée de la même manière, avec les mêmes techniques et dans l'optique du déploiement d'un modèle économique qui était jusque-là réservé aux pages web publiques. Et s'il est possible, à l'aide de services tiers, de masquer l'affichage des dites publicités ciblées, il est en revanche impossible de bloquer l'indexation de nos courriels.

Le prix à payer ?

S'identifier sur un service avant d'en utiliser un autre – particulièrement un moteur de recherche – revient trop souvent à mettre sa conscience et sa vigilance en sommeil durant tout le temps de sa navigation. Voilà pourquoi, dans la définition de l'identité numérique, les traces que nous laissons inconsciemment ou de manière non délibérée sur le réseau sont

centrales : elles incarnent ces machineries de l'identification, au service d'une ingénierie de la transparence identitaire.

Est-ce le prix à payer pour une navigation facilitée et des résultats de recherche plus pertinents ? L'identification et la traçabilité de l'utilisateur sont défendues par la plupart des sociétés internet comme le seul moyen d'offrir une expérience enrichie de navigation, la seule possibilité permettant de personnaliser le service rendu. Si l'argument se justifie techniquement, il n'exonère pas les mêmes sociétés de fournir des garanties sur la durée de conservation de ces données personnelles ainsi que sur l'usage qu'elles en feront.

Un cercle vicieux ?

Le cercle vertueux de l'identité numérique se décline comme suit : avant d'accéder aux ressources d'un système (autorisation), je dois d'abord dire qui je suis (identification) et une vérification doit être effectuée, soit par une procédure technique – *a minima* un mot de passe – (authentification), soit par un tiers – protocole OpenID par exemple – (certification).

Mais ce parcours se mue trop souvent en un cercle vicieux dans lequel l'autorisation est celle faite au système d'accéder à mes ressources (documents, courriels, amis, etc.) voire d'écrire dans mes espaces (Facebook ou compte Twitter). L'identification est persistante et souvent transparente (voir *supra*), l'authentification rendue délicate par la multiplication des pseudonymes et la certification devient, *in fine*, une sorte de cheval de Troie permettant à des tiers (choisis ou non) d'accéder à certaines de mes données personnelles. C'est le cas de certaines applications Facebook, plébiscitées parce que ludiques et « conviviales ». Lorsque nous les installons dans nos profils, elles permettent à des annonceurs de prendre connaissance de nos préférences, de nos goûts, de toute une partie de notre vie privée pouvant parfois aller jusqu'à notre vie intime.

L'offre et la demande

Nous documentons littéralement, de manière persistante et de plus en plus transparente, nos identités numériques, c'est-à-dire la part numérique de nos productions documentaires et ce qu'elles disent de nous une fois captées et remixées dans l'interface des moteurs de recherche et des réseaux sociaux.

La demande

En tant qu'individus, nous nous percevons comme le plus petit commun multiple de l'ensemble de nos traces documentaires numériques (identitaires ou non). Notre demande consiste à pouvoir le plus aisément possible rassembler, « tenir ensemble » nos traces numériques éparpillées.

L'offre

Pour les moteurs et les réseaux sociaux, c'est une logique inverse : pour mieux nous connaître, pour mieux personnaliser les services offerts, pour nourrir ce qu'un analyste américain a baptisé leur « base de données des intentions », ils ont besoin de nous offrir un éventail de services le plus large possible, ils se positionnent donc comme les plus grands dénominateurs communs de nos identités numériques.

Notre empreinte numérique : l'homme qui valait 3 milliards de documents

Empruntons un instant une analogie avec une autre notion très à la mode, celle de notre empreinte écologique. Nous produisons et reproduisons chaque jour une quantité phénoménale de traces numériques. À l'image du volume de la masse documentaire, il y a de cela déjà quelques années, on observe aujourd'hui une autre explosion : celle de notre empreinte identitaire.

En amont de nos pratiques, nous mettons des photos sur Flickr ou PicasaWeb, des remarques quotidiennes sur Facebook ou sur Twitter, des vidéos sur YouTube, nous partageons des signets dans Delicious ou Diigo, des documents de travail dans Google Documents, des courriels dans des webmails, des billets sur nos blogs, des articles dans des journaux de type *Agoravox*, des publications scientifiques dans des archives ouvertes et des revues en ligne, des livres sur des sites d'éditeur, etc.

La complexité qui préside à toute tentative de calcul de notre empreinte identitaire numérique est énorme et tient à plusieurs points distincts. D'abord, en aval de nos usages, nous nous abonnons, nous absorbons, nous consultons, nous « souscrivons », grâce à la syndication de contenus

(fils RSS), à des ensembles de traces documentaires numériques produites par d'autres. Cette documentation informe en retour notre propre rapport à l'information. Dans le même temps, le Web nous propose une fragmentation toujours plus grande de ce qui peut constituer un document : c'est l'exemple de l'essor de la « statusphère » dans laquelle 140 signes sur Twitter ou un statut Facebook deviennent autant d'unités minimales de signification, autant d'éclats de verre du gigantesque miroir identitaire qu'est le Web. Nos usages sont également démultipliés du fait des stratégies de synchronisation proposées par la plupart des grands écosystèmes du Net : je commence à lire un document sur ma tablette tactile dans le métro et je continue de lire le même document, à l'endroit où je m'étais arrêté, depuis mon ordinateur personnel chez moi, *via* un service d'abonnement comme GooglePlay ou encore iTunes. Enfin, la possibilité constante et presque désormais consubstantielle à l'ensemble des contenus proposés – si fragmentaires soient-ils – de pouvoir les éditer (wiki), les commenter, les recomposer, les rediffuser (*retweets* et autres « partages » sur les réseaux sociaux), les re-documenter, ajoute sans cesse de nouveaux niveaux de profondeur à toute tentative de saisie ou de mise à plat de notre empreinte identitaire.

L'identité : bernard-l'hermite ou *Caulerpa taxifolia*

Ainsi cohabitent en permanence une identité numérique perçue, celle du bernard l'hermite à l'abri dans sa coquille, une identité que l'on aimerait rassemblée, protégée, étanche aux autres si nous le souhaitons, confinée, possiblement confidentielle ou suffisamment « floutée » et une identité numérique vécue, celle de l'envahissante algue *caulerpa taxifolia*, identité difficilement contrôlable, éparpillée dans les moteurs et les réseaux sociaux et fonctionnant sur un mode de propagation qui est celui de la viralité, d'une incontrôlable viralité.

L'un des enjeux de premier plan de ce que l'on appelle « la société de l'information » est de permettre à chacun d'inverser la tendance entre l'identité numérique vécue et celle perçue, de reprendre le contrôle, de mesurer l'étendue de l'ensemble de ses traces identitaires et d'en circonscrire, si on le souhaite, le périmètre.

Éléments contextuels

Pulsions scopiques

L'identité numérique est indissociable du désir de voir, que les psychologues qualifient de « pulsion scopique ». Conjuguées au désir de voir sans être vu, ou à l'inverse, d'être vu pour exister, ces pulsions sont, de la simple imprudence à la vraie impudeur, à l'origine des traces que nous laissons en ignorant ou en refusant de voir leur potentiel de nuisance.

Ces pulsions trouvent un abri idéal dans les nouveaux panoptiques des réseaux sociaux comme Facebook, lesquels vont instrumentaliser cette part pulsionnelle et permettre à chacun de voir et d'être vu, de voir sans être vu. Mais seul Facebook et ses partenaires commerciaux disposeront d'une vue d'ensemble sur les communautés, les voisinages, et le bouillon pulsionnel de la première communauté humaine numérique de la planète : le milliard d'utilisateurs du site.

Identité in vitro et post mortem

L'identité numérique dépasse aujourd'hui les frontières de notre temps biologique.

D'avant la naissance...

On connaissait déjà le cas de ces parents qui réservaient un nom de domaine correspondant à celui de leur enfant ou lui ouvraient une adresse de courrier électronique avant même que celui-ci soit né ou en gestation. On sait désormais, selon une étude de la société AVG d'octobre 2010^z, que 81 % des enfants de moins de deux ans (74 % en France) ont déjà une présence numérique, c'est-à-dire une ou plusieurs photos postées sur les réseaux sociaux par leurs parents.

... à après la mort

Sur Facebook, lorsqu'un membre du réseau social décède, un formulaire est disponible pour en avvertir ses « amis ». Son profil peut alors être gelé et transformé en un « mémorial numérique » sur lequel les mêmes « amis » peuvent venir déposer des messages. D'autres sociétés font ouvertement commerce de la gestion de notre identité numérique *post mortem*.

Lorsque tout devient indexable

Les frontières jadis distinctes entre le Web public, le Web privé et le caractère intime de certaines conversations tenues sur le Web sont aujourd'hui de plus en plus floues et peuvent être parfois complètement abolies dans certains environnements informationnels. Les moteurs de recherche ont la capacité d'indexer aussi bien les ressources du Web public, nos blogs les plus personnels ou le contenu de nos courriels. De leur côté, les réseaux sociaux, Facebook en tête, ont choisi d'ouvrir leur base de données de profils pour que les mêmes moteurs puissent venir les indexer. C'est à une nouvelle dérive des continents documentaires que nous assistons, mais une dérive inverse de celle que nous enseignent la géologie et sa tectonique des plaques : aujourd'hui ne subsiste qu'une même pangée, qu'un même territoire uniformément indexable par quelques-uns, au nom de tous les autres.

L'informatique en nuages

Vous viendrait-il à l'idée de laisser en permanence vos documents d'identité à un tiers ?

Le *cloud computing* – ou informatique en nuages – désigne le processus qui consiste à laisser nos données et nos documents sur des serveurs hôtes de différentes entreprises : YouTube pour nos vidéos familiales, Flickr ou Facebook pour nos photos de vacances, Google Documents pour nos documents de travail, Dropbox pour les fichiers de notre ordinateur, etc. Nos données et informations identitaires se trouvent aujourd'hui pareillement « dans les nuages », à disposition de ceux qui en sont aujourd'hui les hébergeurs exclusifs et qui en seront demain, peut-être, les principaux prédateurs.

Trop de connectivité met en danger l'identité

Dans le monde d'avant-hier, un monde majoritairement déconnecté, notre identité semblait protégée. Dans le monde d'hier, un monde majoritairement connecté, notre identité était parfois dangereusement exposée. Le monde d'aujourd'hui est un monde de l'hyperproximité, de la connexion permanente, de l'informatique nomade et ubiquitaire. Dans ce monde-là, notre identité est, si nous ne mettons pas en œuvre des garde-fous, en danger. Plus exactement, elle est en permanence susceptible de représenter un danger pour et « sur » nos sociabilités dans le monde physique comme dans le monde numérique. En attestent notamment les nombreux cas de licenciements ayant comme origine une utilisation trop naïve du premier des réseaux sociaux.

Les « pourquoi » et les risques

Pourquoi une identité numérique ? Pour les usagers

Du côté des usagers, la pyramide des besoins, telle que définie par le psychologue Abraham Maslow, permet de rendre compte du parcours d'une construction identitaire type. Elle peut être transposée au numérique. On commence par un besoin de « sécurité », qui est celui du choix d'un identifiant. On passe ensuite au besoin d'amour et d'appartenance qui peut se décliner sur tout site de nature communautaire (dont les sites de rencontre comme le célèbre Meetic.com). Vient alors le besoin d'estime des autres sur lequel vont se déployer les stratégies de construction et de gestion d'une « réputation » numérique. Vient enfin le besoin d'estime de soi, c'est-à-dire la dimension narcissique qui préside et achève toute présence ou stratégie identitaire sur le Web. Alors, seulement, peut s'incarner le besoin qui clôt la pyramide de Maslow, celui de l'accomplissement personnel que l'on peut ici définir comme l'adéquation entre la réputation perçue (d'un individu et d'une entreprise) et son identité voulue.

Pour les acteurs du marché

Du côté des acteurs du marché de l'identité numérique, moteurs et réseaux sociaux, il s'agit d'abord et avant tout de conforter un modèle économique, celui d'un service gratuit financé par la publicité. Il leur faut pour cela disposer d'un graphe, sans cesse réactualisé, de l'ensemble des contenus accédés sur le Web, mais également – ce fut l'atout des réseaux sociaux – du graphe de nos relations sociales et de nos intérêts, la conjugaison des deux permettant d'alimenter leur fonctionnement de régie publicitaire et de revendre aux annonceurs des données très précieuses sur des échantillons immenses mais toujours davantage segmentés et représentatifs. Au royaume d'une économie dont

l'attention, notre attention, est la ressource la plus rare et donc la plus chère, le profilage et la segmentation marketing règnent en maîtres.

La pyramide de Maslow



Les risques encourus pour les individus

Pour les individus, et au-delà du seul péché d'orgueil, baptisé *ego-googling* ou *vanity search* et consistant à taper son nom dans un moteur de recherche, ces risques sont assez divers :

- traçage et ciblage comportemental qui permettent – aujourd'hui avec l'aide active des outils de géolocalisation – de faire fonctionner les industries de la recommandation et un marketing publicitaire de plus en plus « personnalisé » et contextuel ;
- risques de contrôle et de fichage permis par l'interconnexion de fichiers étatiques ou commerciaux ;
- risques, enfin, liés à l'expression de soi, qui peut se faire de manière soit totalement transparente (se décrire ou s'enregistrer sous son vrai nom), soit partiellement dévoilée (avatars, pseudonymat), soit totalement dissimulée (anonymat, usurpation ou détournement d'identité).

L'une des questions que pose l'identité numérique sur le Web est de savoir quelles stratégies identitaires l'emporteront dans l'éventail qui va de l'anonymat complet au dévoilement le plus transparent. Sur ce sujet, Clay Shirky¹⁴ déclarait en 2010 : « Dans vingt ans, il y aura des archipels d'utilisateurs identifiés, lesquels retireront beaucoup de bénéfices, de notoriété de leur implication dans cet écosystème, mais ces archipels demeureront noyés dans un océan d'anonymat » (notre traduction).

Pour les entreprises, les risques réputationnels recouvrent d'autres enjeux, principalement financiers. Christophe Asselin et la société Digimind ont produit un tableau synoptique les résumant parfaitement.

1 Pierre Truche, Jean-Paul Faugère, Patrice Flichy, *Administration électronique et protection des données personnelles – Livre Blanc*, <http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/024000100/0000.pdf>, consulté le 30 octobre 2012.

2 Rémi Vallet, « Données personnelles : France Télécom épouse l'Alliance », *Transfert.net*, 21 décembre 2001, <http://www.transfert.net/a7953>, consulté le 30 octobre 2012.

3 MySpace : <http://www.myspace.com/> Friendster : <http://www.friends.com/> LinkedIn : <http://www.linkedin.com/>

4 Le standard « Do Not Track », développé par le W3C (<http://www.w3.org/Tr/tracking-dnt/>) à l'attention des navigateurs commence à être adopté, notamment par Microsoft qui a choisi, à la différence de Firefox, de l'activer par défaut. Mais cette adoption est loin de faire l'unanimité : « En bloquant par défaut les régies publicitaires qui souhaitent poser leurs cookies sur le navigateur de l'internaute, Microsoft miserait sur le fait que les utilisateurs acceptent de faire une exception au DNT pour accéder à ses propres services. Or une fois l'exception pour Microsoft acceptée, toutes les publicités ciblées gérées par Microsoft seraient autorisées. » (Guillaume Champeau, « Do Not Track : pourquoi Microsoft vous veut du bien », *Numerama*, 11 juin 2012, <http://www.numerama.com/magazine/22853-donot-track-pourquoi-microsoft-vous-veut-du-bien.html>, consulté le 30 octobre 2012.).

5 John Battelle's searchblog, <http://battellemedia.com>.

6 Sigmund Freud, *Trois essais sur la théorie sexuelle*, Paris, Gallimard, 1987, p. 65-68.

7 Johanna Godet, « 81 % des enfants de moins de deux ans ont déjà une empreinte numérique », *L'informaticien*, 11 octobre 2010, <http://www.linformaticien.com/actualites/id/9202/81-des-enfants-de-moins-de-deux-ans-ont-deja-une-empreinte-numerique.aspx>, consulté le 30 octobre 2012.

8 Voir par exemple : <http://www.laviedapres.com>.

9 YouTube : <http://www.youtube.com>, Flickr : <http://www.flickr.com>, Google Documents : <http://www.docs.google.com>, Dropbox : <http://www.dropbox.com/>

10 Sur cette question, il est possible de consulter la page de la Commission nationale de l'informatique et des libertés (CNIL) faisant référence à la jurisprudence de 2011 : <http://www.cnil.fr/la-cnil/actu-cnil/article/article/maitriser-les-informations-publiees-sur-les-reseaux-sociaux/>, consultée le 30 octobre 2012 et l'article de Christelle Dardant, « Incertitudes autour de la jurisprudence "Licenciements Facebook" », *Institut de recherche et d'études en droit de l'information et de la communication (IREDIC)*, 31 janvier 2012, <http://junon.univ-cezanne.fr/u3iredic/?p=8378>, consulté le 30 octobre 2012.

11 « Pyramides des besoins de Maslow », *Wikipédia*, http://fr.wikipedia.org/wiki/Pyramide_des_besoins_de_Maslow, consulté le 30 octobre 2012.

12 Les suggestions d'achat sur Amazon en sont un exemple ; elles reposent à la fois sur l'historique de vos achats et de vos consultations sur le site, ainsi que sur la statistique globale des achats sur la plateforme.

13 Voir les débats autour des textes de loi Loppsi, Hadopi, fichier Edwige, etc.

14 « Clay Shirky », *Wikipédia*, http://fr.wikipedia.org/wiki/Clay_Shirky, consulté le 30 octobre 2012.

15 « *It will be an archipelago of named users, who get a lot of value from participating in that part of the ecosystem, but still set in an ocean of anonymity.* » Voir Janna Quitney Anderson, Lee Rainie, « The Future of The Internet », *Pew Internet*, <http://www.pewinternet.org/~media/Files/reports/2010/Future%20of%20internet%202010%20-%20AAAS%20paper.pdf>, consulté le 30 octobre 2012.

Une typologie des risques d'e-réputation

