

n'est pas de mise car chaque institution choisit son propre modèle d'archivage.

Les paradoxes d'un dépôt légal partiellement sélectif

En matière de web, l'exhaustivité – but originel du dépôt légal – n'est plus un objectif accessible : il n'est pas possible de capturer chaque site à chaque mise à jour. La BnF vise donc la représentativité : il s'agit de constituer une image, un « instantané » de l'internet français, qui prenne en compte tous les types de publications (du site officiel à la plate-forme de diffusion de vidéos ou aux parties publiques des réseaux sociaux), et tous les contenus, du plus sérieux au plus dérisoire.

À cette fin, la BnF conjugue deux modèles de collecte : le premier est la collecte « large », réalisée une fois par an, qui concerne tous les sites qui ont été automatiquement identifiés comme français – soit plus de

quatre millions à ce jour. Les collectes « ciblées » en revanche concernent des sites à capturer plus fréquemment (jusqu'à une fois par jour) ou plus profondément (jusqu'à plusieurs centaines de fichiers par domaine) ; il peut également s'agir de ressources à collecter en raison de leur lien à un événement donné (élections, festivals, rencontres sportives...). Ces sites-là, près de trente mille à ce jour, sont identifiés individuellement, soit par des bibliothécaires de la BnF (une centaine d'agents de la BnF, répartis dans les différents départements thématiques, participent à la sélection), soit par des partenaires (bibliothèques, centres d'archives, laboratoires de recherche). Prenons l'exemple de la littérature contemporaine : le département Littérature et art a identifié des sites de référence en matière d'analyse ou de critique, il a également collaboré avec l'Association pour l'autobiographie³ pour recenser plusieurs centaines de

blogs, équivalents en ligne des journaux personnels⁴.

Le « modèle intégré » adopté à la BnF vise donc à conjuguer les avantages de la logique du dépôt légal (constituer un « miroir » de la production et de la consommation culturelles françaises) et ceux de la sélection documentaire (conserver les pans les plus dynamiques et novateurs de l'internet). Comparée à ses homologues internationaux, la BnF se situe donc à mi-chemin, certaines institutions reposant exclusivement sur des collectes automatiques, tandis que d'autres ne conservent que les segments qu'ils jugent les plus essentiels de leur web national. L'expérience de l'archivage du web montre ainsi qu'au-delà des moyens techniques et des ressources mis à disposition, ce sont le cadre juridique, les missions et les traditions scientifiques qui priment dans la constitution des collections nationales.

Régime juridique des archives numériques des utilisateurs du web stockées par des entreprises privées

SÉBASTIEN FANTI (AVOCAT)

Introduction

Lorsque je me remémore la période durant laquelle il m'a été offert le privilège de lire et de décortiquer l'ouvrage prophétique de Georges Orwell *1984*, différentes questions surgissent immédiatement. Qu'est-ce qui a pu générer une telle réflexion, une prospection d'une rare intensité, finesse et justesse ? Quels facteurs ont été déterminants pour l'anticipation, substantifique moelle de cet ouvrage ? En bref, comment a-t-il fait, alors que la technologie n'était encore qu'à ses soubresauts, à l'aune de la fulgurance actuelle ? Si Orwell était encore parmi nous aujourd'hui, il aurait à sa disposition une friche de prédiction extraordi-

naire. Son atrabile demeurerait certainement identique à ce qu'il nous a légué : l'abhorration de toute forme de totalitarisme déshumanisant. Albert Einstein, un autre visionnaire, disait en 1921, qu'« *il est hélas devenu évident aujourd'hui que notre technologie a dépassé notre humanité* ». Lui aussi serait certainement interloqué devant la profusion de domaines de la vie courante où la cohabitation engendre des difficultés croissantes.

Sans avoir la prétention de, ne serait-ce qu'effleurer, la virtuosité de

ces auteurs et acteurs majeurs du xx^e siècle, il convient de nous interroger sur les défis qu'engendre la planétisation du numérique, laquelle à l'instar d'une machine de Gramme a donné naissance à un *iWorld*, dont on peine à imaginer tant les limites, que les écueils. Dans ce contexte, les archives numériques sont, à l'évidence, une source intarissable pour les géants du Web et elles peuvent à l'ère de l'Open Data constituer un enjeu économique majeur. Il paraît utile de rappeler que ce sont les gouvernements qui ont, les premiers, recueilli des informations à large échelle. Leur somme excède celle des données stockées par des acteurs majeurs du secteur du web et leur diffusion n'est pas sans danger. Le recoupement des bases de

³ <http://autobiographie.sitapa.org/>.

⁴ Sur ce sujet, voir Gildas Illien, « Les mémoires de la Toile, l'archivage d'Internet à la BnF », dans *La Faute à Rousseau*, n° 45, 2007.

données publiques permet d'établir aisément un lien avec un individu. Le risque de compromission de la sphère privée est encore accru si une combinaison intervient avec les données du Big Data (paiement, téléphone, réseaux sociaux...). Finalement, une fois que les données publiques telles que les archives sont offertes en libre accès, il n'existe plus de pouvoir effectif de contrôle.

Le contexte actuel (Big Data World ou le déluge annoncé des données)

Le Big Data (ou *données massives*) désigne la collecte, l'exploration et l'analyse de grandes masses de données. Des chiffres, des textes ou des images, mais aussi des gènes, des étoiles, des particules ou des traces de trafic routier... Les Big Data se réfèrent à ce qui peut être accompli à grande échelle et ne peut pas l'être à une échelle plus petite, en matière d'extraction de nouvelles connaissances ou de création de nouvelles formes de valeur, avec impact sur la transformation des marchés, des organisations, de la relation entre les citoyens et les gouvernements, et bien plus encore¹. À titre exemplatif, Google traite quotidiennement plus de 24 pétaoctets de données (24 millions de milliards d'octets ou environ 200 millions de milliards de bits). Ces données sont le pétrole du XXI^e siècle.

Le Préposé à la protection des données et à la transparence a déclaré vouloir s'atteler à cette problématique². Il craint en effet « *le fait que, grâce aux algorithmes, l'on puisse désormais combiner des bases de données entre elles pour*

dresser le portrait d'un individu et prédire son comportement et ses besoins futurs ». Cette lutte contre le volume exponentiel des données générées par les réseaux sociaux, la téléphonie ou tout simplement Internet est inégale, respectivement ambiguë. Chacun reconnaît que le Big Data peut s'avérer un facteur de progrès, par exemple en permettant une analyse en temps réel de l'évolution d'une maladie infectieuse. Le paradigme consiste donc à faire coexister les avantages indéniables mis en exergue et une sphère privée intacte. Et à éviter la dictature des données, car celles-ci, ne l'oublions jamais, ne reflètent pas toujours la réalité.

Les limites légales en droit suisse

Il existe différentes normes qui peuvent trouver application dans un contexte tel que celui qui vient d'être exposé. Toutefois, ces règles sont limitées par le principe de territorialité. Cela signifie concrètement que la société contre laquelle vous entendez agir doit avoir un siège en Suisse pour que la démarche soit couronnée de succès (ce qui n'est pas le cas de Facebook par exemple). L'effectivité de la protection s'en trouve évidemment substantiellement réduite.

Il s'agit tout d'abord d'une règle (article 13 de la Constitution fédérale) qui figure dans notre charte fondamentale et qui garantit le droit au respect de la vie privée en ces termes : « *toute personne a le droit d'être protégée contre l'emploi abusif des données qui la concernent* ». Tout est dit. Mais rien n'est concrétisé. Cette règle n'a pas de portée absolue et elle doit se concilier avec la liberté d'expression et la liberté des médias qui ont aussi rang constitutionnel. Une pesée des intérêts en cause devra être diligentée entre la nécessité d'information et les droits de la personne intéressée ; il faudra également examiner si les objectifs poursuivis par l'auteur, de même que les moyens utilisés sont dignes de protection.

S'agissant des relations entre les particuliers, les articles 28 et suivants du Code civil trouvent application en cette matière. Ils permettent de proscrire, juguler et sanctionner en espèces sonnantes et trébuchantes les violations de la sphère privée qui inclut le secret de la correspondance électronique. À titre exemplatif, ils pourraient être invoqués par une personne qui souhaite obtenir la suppression d'un article de journal litigieux publié sur un blog.

Une disposition de la loi fédérale sur la protection des données (art. 15 LPD) vient compléter l'arsenal législatif, si la publication de faits qui relèvent de la sphère privée peut être considérée comme un traitement de données personnelles. Le demandeur peut alors requérir la rectification ou la destruction des données. La diffusion d'informations véridiques peut s'avérer problématique, quand les personnes intéressées ne veulent pas les voir publiées (faillite, appartenance à une secte, condamnation, etc.).

Conclusions prospectives

Actuellement, la cuirasse des règles juridiques exposées ne permet pas de garantir une protection efficace du citoyen contre les phagociteurs de données que sont Facebook, Google ou encore de manière générale Internet. Cette défaillance a permis à ces world company d'occuper le terrain, ce qui signifie de capter toutes vos données. Mais il y a plus. Ces opérateurs globaux tentent d'imposer des conditions générales comportant des règles très libérales qui impliquent de facto une renonciation à la protection offerte par nos normes (avec par exemple la nécessité d'ouvrir une action aux USA pour Facebook). Cette privatisation du droit doit faire rapidement l'objet d'adaptations législatives. Une loi-cadre technologique pourrait permettre de pallier les manques les plus criants. À défaut, il faudra apprendre à vivre avec moins de sécurité et de prévisibilité juridiques.

¹ Viktor Mayer-Schönberger, Keneth Cukier, *Big Data, la révolution des données est en marche*, Robert Laffont, Paris, 2014, p. 15.

² *Le Temps* du 6 mai 2013, « On ne mesure pas l'impact du Big Data sur la sphère privée », http://www.letemps.ch/Page/Uuid/9b77f946-b581-11e2-abad-bb9758babf60/On_ne_mesure_pas_l'impact_du_Big_Data_sur_la_sphère_privée.