

Le cas du projet *Domesday* de la BBC (société britannique de radiotélévision) fournit un exemple de ces difficultés techniques. En 1986, soit quelque 900 ans après une entreprise similaire (voir l'encadré page 90) ordonnée par Guillaume le Conquérant (1027-1087), la BBC a lancé un recensement et un sondage sur l'ensemble du Royaume-Uni. Les données, recueillies par plus d'un million de volontaires dans les 23 000 cellules territoriales définies pour l'occasion, constituaient une sorte d'immense instantané de la société britannique; elles furent mises à la disposition du public au moyen d'un vidéodisque de type *LaserDisc*, un support des données dont on considérait à l'époque qu'il serait durable. Or, dès 1998, il fut supplanté par les DVD, de sorte qu'à partir du début des années 2000, on ne trouvait déjà pratiquement plus de lecteurs de *LaserDisc*, ni d'ordinateurs où pouvait tourner le logiciel nécessaire pour présenter le contenu du disque. Résultat: plusieurs équipes durent fournir un important effort de recherche afin d'émuler (d'imiter) le fonctionnement des lecteurs de *LaserDisc* dans un environnement informatique actuel.

Ce cas montre que pour préserver des logiciels, il importe de maintenir les connaissances nécessaires pour les faire exécuter par une machine, et donc de disposer de leur code source et de leur environnement d'exécution. Dans le cas du *LaserDisc*, quelques années avaient suffi pour que tout cela disparaisse! C'est bien à cause de telles situations qu'aujourd'hui, on reconnaît de plus en plus que l'usage de formats de données ouverts (qui peuvent être librement utilisés et modifiés, et dont les spécifications

sont publiques et sans restriction d'accès) et de logiciels libres (dont le code source est disponible pour tous) est indispensable à la pérennité des informations.

Mais une fois le problème de la lecture des *LaserDisc* du projet *Domesday* résolu, on s'aperçut que d'épineux problèmes légaux restaient à démêler. Selon le droit d'auteur en vigueur, la réalisation d'une copie exigeait que l'on recueille d'abord les autorisations des millions d'individus ayant contribué au projet, ainsi que celles des auteurs des logiciels utilisés... À une époque où reproduire l'information est devenu banal, l'imposant arsenal juridique des droits d'auteurs nous interdit nombre de répliques, et ce sur des durées de plus en plus longues... Si le droit protégeant la propriété intellectuelle sert aujourd'hui essentiellement à préserver les modèles économiques ayant précédé l'ère d'Internet, il a souvent pour effet de bloquer le développement des modèles nécessaires à l'économie numérique.



2. EN 2011, 1,8 ZETTAOCTET de données a été produit, et diverses estimations en annoncent huit pour 2015. Un zettaoctet représente 1 000 milliards de gigaoctets (c'est-à-dire 10^{21} octets). Ainsi, en quelques jours, l'humanité accumule aujourd'hui plus d'information qu'elle n'en a produit en deux millions d'années avant l'invention d'Internet.

Copier est-il hors la loi ?

Si nous voulons préserver les données, nous sommes donc confrontés à un dilemme: nous devons d'un côté les recopier sur de nouveaux supports (avant leur inévitable dégradation) et sous de nouveaux formats (avant leur inévitable obsolescence); d'un autre côté, nous devons respecter des lois qui nous interdisent de réaliser ces copies... Notre civilisation est donc confrontée à la question suivante: qui doit se charger du renouvellement continu des données et dans quelles conditions? Plusieurs



3. LES SUPPORTS DE STOCKAGE se suivent et ne se ressemblent pas, tant en termes de capacité que de format d'enregistrement. Ci-dessus, quelques-uns des formats datant des 30 dernières années: à gauche, une disquette 5 pouces 1/4, au milieu une disquette 3 pouces 1/2 de haute densité, puis un CD-ROM et, à droite, une clef USB actuelle.



4. WORDSTAR est un logiciel de traitement de texte introduit en 1978, qui fonctionnait sous DOS, le système d'exploitation pour ordinateurs personnels de Microsoft. Il occupa une position dominante au début des années 1980, mais fut supplanté par *Word* dans les années 1990, puis abandonné et oublié, au point de devenir inutilisable aujourd'hui.

réponses sont possibles. Chacune a ses avantages et ses inconvénients.

Une première possibilité serait de sous-traiter la préservation des données à quelques grands opérateurs chargés de les recopier régulièrement sur les nouveaux supports avant que les anciens ne se dégradent et, éventuellement, de les convertir dans les nouveaux formats. À cet égard, les offres d'hébergement gratuit ne manquent pas, qu'il s'agisse d'abriter nos photographies, vidéos, documents bureautiques ou encore notre courrier. Le plus souvent, elles sont le fait de grands fournisseurs de services sur Internet, tels *Google, Amazon, Facebook, Apple, Microsoft*, etc. Toutefois, les récentes révélations sur le programme de surveillance PRISM de l'Agence américaine de sécurité soulignent un phénomène général : nous payons ces services en fournissant involontairement des informations très détaillées sur nos activités, préférences, relations personnelles, achats, voyages, etc.

Ces données sont employées à notre insu, ce qui se traduit non seulement par une surveillance occulte, mais aussi de plus en plus souvent par des violations à grande échelle de la vie privée, par exemple par des vendeurs de toutes sortes de produits.

D'autres grands acteurs ont lancé un gigantesque effort de préservation des données disponibles sur Internet ou sur certains réseaux sociaux : fondée en 1996, l'*Internet Archive* fait régulièrement des copies des sites disponibles sur la Toile et conserve une vaste collection d'images, de vidéos, de livres et de logiciels. S'agissant des œuvres littéraires, musicales ou cinématographiques, de grandes sociétés se réservent le droit exclusif de les archiver, et interdisent au public de les copier ou de les échanger ; en outre, elles mettent parfois en place des mesures de protection contre la copie qui entravent la conservation de ces données à long terme.

Aussi, tous ces acteurs, aussi puissants soient-ils, finissent-ils par constituer des points de fragilité comparables à la grande bibliothèque d'Alexandrie des temps hellénistiques. Malgré les investissements faits pour créer des centrales de serveurs plus sûres et mieux organisées, la concentration de grandes masses de données sous le contrôle d'une seule organisation est toujours source de fragilité. Et au-delà du problème technique de la sauvegarde des données contre les pertes accidentelles, il



5. L'AGENCE AMÉRICAINE DE SÉCURITÉ (NSA ou *National Security Agency*) exploite la concentration des données personnelles chez les grands fournisseurs de services sur Internet pour surveiller la population. Dans le projet PRISM, révélé par le lanceur d'alerte Edward Snowden, elle se fonde sur la loi américaine régulant les activités d'espionnage pour se fournir en données à décrypter chez les gros opérateurs, tels *Google, Amazon, Facebook*, etc.

■ L'AUTEUR



Roberto DI COSMO, professeur d'informatique à l'Université Paris-Diderot, est membre du Laboratoire Preuves, programmes et systèmes du CNRS.

s'agit aussi de trier les informations qui méritent d'être préservées et celles qui peuvent être effacées. Or le fait qu'une seule organisation, ayant sa logique et ses intérêts particuliers, en décide, introduit manifestement un énorme risque de destruction. Aurions-nous aujourd'hui des éditions complètes initiales des *Fleurs du mal* de Charles Baudelaire (1821-1867) ou de *Madame Bovary* de Gustave Flaubert (1821-1880), si, il y a plus d'un siècle, alors que ces chefs-d'œuvre étaient censurés, leur conservation avait été confiée à une organisation étatique centralisée ?

L'échange entre pairs

Une solution alternative de conservation des données de l'humanité à long terme provient du très grand nombre des ordinateurs personnels et de la croissance exponentielle de la capacité de stockage des mémoires disponibles pour le grand public. Le déploiement d'un système d'échange pair à pair (de l'anglais *peer to peer*) rend envisageable la transformation du parc mondial d'ordinateurs personnels en une gigantesque base de données distribuées, puis de laisser chacun y répliquer et y partager ses données.

Sous réserve que les ordinateurs personnels continuent à exister (!), cette solution aurait l'avantage de faire l'impasse sur une autorité centrale fixant seule ce qui mérite d'être conservé. Cela permettrait à