Dans un monde à la fois connecté et en tension, quels enjeux et quelles approches pour assurer valorisation et protection des données ?

Par Jérôme ANDRES
Systèmes d'information et communication sécurisés du groupe Thales

La révolution numérique intervenue au cours des trente dernières années est à l'origine de très nombreuses mutations au plan économique et dans notre vie quotidienne. Internet qui en est le socle s'est bâti sur l'échange de données toujours plus importantes, mais surtout sur leur valorisation dans tous les champs : scientifique, sociétal, politique, artistique, ludique... L'information et la connaissance qu'elles rendent possibles constituent un actif qu'il s'agit aujourd'hui de protéger au regard de leur valeur intrinsèque, mais aussi des conséquences que peut induire, directement ou indirectement, leur utilisation. À l'image du numérique et comme toute autre innovation, l'émergence des données n'est pas neutre, représentant tour à tour un risque ou une opportunité, un remède ou un « poison ». La cybersécurité est une pratique et un champ industriel qui peuvent permettre de mesurer et de canaliser ces différentes alternatives, mais qu'il faut aussi appréhender sur le plan politique, au niveau national comme international, pour tenter d'en réguler les équilibres.

Nouvel Eldorado, nouvel or noir ?

Les métaphores aurifères sont fréquentes concernant l'opportunité que revêt l'usage des données depuis une trentaine d'années et l'émergence d'Internet. À l'évidence, elles ont fait la fortune de quelques nouvelles entreprises nord-américaines (les GAFAM[1]), mais aussi chinoises (BATX[2]), qui ont su en quelques années seulement acquérir des positions dominantes inattendues bouleversant nos quotidiens. Aujourd'hui, ce sont nos *smartphones* qui en sont l'exemple le plus éclairant : ils nous sont aujourd'hui indispensables, que nous nous déplacions en recourant à un service de navigation, que nous suivions nos performances sportives ou physiologiques et, depuis deux ans, les courbes d'infection au Covid-19, les taux de vaccination ou d'hospitalisation et les clefs numériques de notre passe sanitaire, ou bien que nous consultions les derniers résultats sportifs, les annonces immobilières ou les biens d'occasion en vente, ou encore pour maintenir le lien avec nos proches au travers de réseaux sociaux ou avec des inconnus dont les centres d'intérêt nous sont similaires. Les applications qui rendent ces services possibles reposent bien sûr sur les technologies de l'électronique et des télécommunications, mais plus particulièrement sur un échange toujours plus varié de données, plus rapide d'informations et plus vertigineux de connaissances entre chacun d'entre nous ou les organisations au sein desquelles nous exerçons.

Hier, nous naviguions à partir d'ordinateurs fixes contraints par des débits bien moins importants. Demain, l'on nous fait la promesse que la 5G permettra de nous connecter à tous nos objets environnants, démultipliant encore nos échanges, telle une rhapsodie folle dans un rythme infernal, à la source de multiples inquiétudes (vie privée, réchauffement climatique, perte de souveraineté, espionnage, manipulation…). Tandis que les thuriféraires de l'innovation nous renvoient, hier comme aujourd'hui, à une seule crainte, celle du changement.

(1) GAFAM : Google, Amazon, Facebook, Apple et Microsoft, qui sont les cinq mastodontes américains de la nouvelle économie et qui sont plus puissants que bien des États. Il est à noter que les trois premiers cités n'existaient pas il y a vingt-cinq ans de cela.

(2) BATX : Baidu, Alibaba, Tencent et Xiaomi : les équivalents des GAFAM pour le Web chinois, lesquels sont des acteurs dominants en Chine et prépondérants en Asie. Leur création remonte à peine à vingt-quatre ans pour les plus anciens.

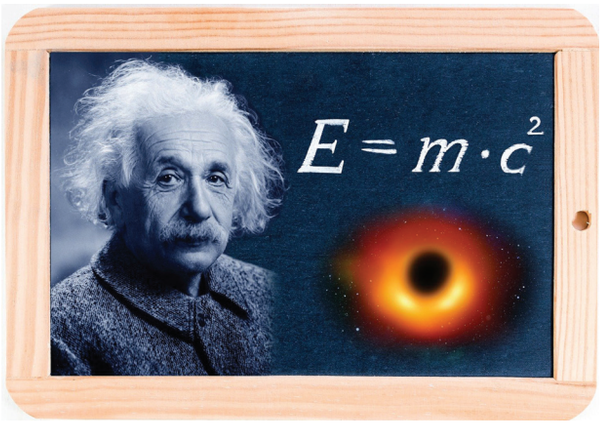


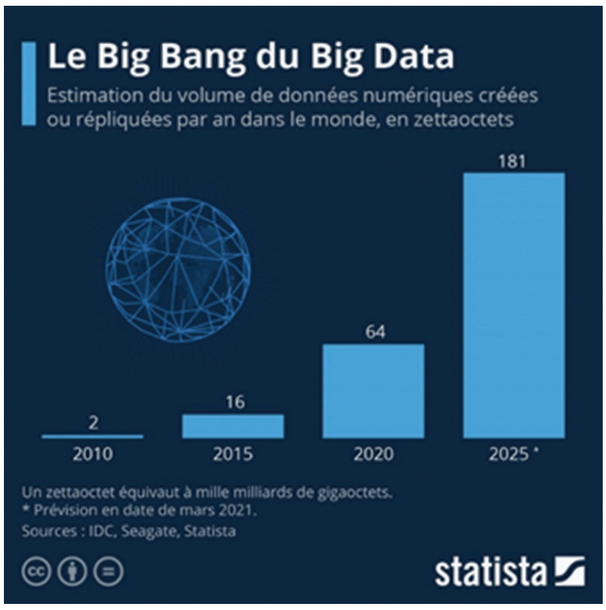
Photo © STUX

Photo 1 : La théorie de la relativité générale d'Albert Einstein, à la base de la découverte des trous noirs.
Source : Photo libre de droits téléchargée sur PIXABAY,

En adoptant une image plus financière, les données sont perçues comme un atout, ou encore un actif (le même mot « asset » recouvrant les deux sens en anglais) que l'on cherche à la fois à protéger et à valoriser. S'il s'agit de la manipuler tel un minerai précieux, il faut alors successivement la collecter, la transformer, la raffiner, la livrer et, enfin, la recycler en recourant à des outils et des processus adéquats et efficaces. À l'instar du pétrole et de la pétrochimie, l'émergence d'informations utiles passe par la combinaison de données, un alliage de celles-ci qui parfois est explosif socialement, politiquement ou économiquement parlant, rendant nécessaire la mise en place de cadres normatifs de différents ordres pour tenter de réguler les enjeux qui y sont associés.

Quelles sont les données à protéger ?

Toutes les données ne sont pas égales. Numérisées, ce sont toujours des suites de zéros et de symboles (bits) combinés en octets. Prises ensemble, elles peuvent représenter des volumes importants, comme une image ou une vidéo pour les plus massives que nous manipulons avec nos *smartphones*. On parle alors de résolutions pouvant aller de 330 x 480 pixels[3] pour les vidéos VHS, soit 0,16 Mpixels[4], jusqu'à 70 Mpixels pour l'IMAX qui a été utilisé par Christopher Nolan pour réaliser le film *Interstellar*, format actuel le plus fin en matière de caméra cinématographique. Cela est encore bien peu en comparaison des données accumulées pour aboutir à l'exploit scientifique rendu public le 12 mai 2022 : la reconstitution par l'EHT[5] du trou noir Sagittarius A* localisé au centre de la voie lactée. Cette prouesse a reposé sur la mobilisation de plusieurs



radiotélescopes répartis sur la surface du globe pour constituer un réseau d'antennes de la taille de la Terre, agissant comme une gigantesque lentille. Chaque campagne de recueil de données a permis d'amasser 7 pétabytes de données[6], sur cinq jours ; l'image divulguée en mai dernier a, elle, exigé la réalisation d'une multitude de calculs depuis 2019.

Ces chiffres vertigineux ne représentent qu'une partie des données numériques disponibles : IDC, Seagate et Statista estiment qu'en 2025, le volume des données créées ou répliquées sera égal à 181 zêta octets, soit mille milliards de Giga octets. Pour ne s'en tenir qu'à cette échéance…

Les données commerciales et industrielles

Les vidéos VHS des années 1970 à 1990 sont souvent les seuls souvenirs animés ou les seuls enregistrements des émissions télévisuelles de l'époque ; elles constituent un patrimoine familial, voire historique que l'on peut souhaiter préserver. S'agissant d'*Interstellar*, ce qui est ici protégé, c'est la propriété intellectuelle de C. Nolan et de ses ayants droit…, au point de ne pouvoir accéder sur Wikipédia qu'à une image d'un format 220 x 326, soit 0,07 Mpixels, donc une image bien moins précise que celle de notre bonne vieille cassette VHS. L'encyclopédie en ligne précise à ce propos que l'image de l'affiche précitée est présentée dans le cadre du *fair use* défini par la loi américaine du copyright : « tout usage ultérieur pouvant constituer une violation de ce droit », telle est l'infraction qui pourrait être imputée aux *Annales des Mines* en cas d'utilisation. Si la qualité en est suffisante et

(3) Un pixel, contraction de *picture element* en anglais, est l'unité de base mesurant la résolution d'une image numérique matricielle.

(4) Un méga pixel mesure une résolution d'image d'un million de pixel, soit, par exemple, une image matricielle de mille pixels de côté.

(5) Event Horizon Telescope est un consortium international de collaboration en matière d'observation des trous noirs grâce à l'utilisation combinée de plusieurs radiotélescopes. Dans les faits, il s'agit d'observer le flot de particules entourant le trou noir, qui ne délivre aucune information et dont aucun cliché ne peut être fait.

(6) Un pétabyte est une unité de mesure de volumes d'informations correspondant à un billion d'octets, soit la capacité de sauvegarde de 1 000 iPhone de dernière génération.

permet sans doute en outre de limiter les besoins de stockage afférents et par là même les coûts d'énergie et les émissions de gaz à effet de serre, la publication de cette image est source d'incertitudes juridiques : on ne sait quelle action pourrait engager les ayants droit, mais le risque pour l'éditeur des présentes *Annales des Mines* d'une exposition à des poursuites serait réel.

Le piratage massif de données dont a été victime en 2014 Sony Pictures Entertainment, la branche américaine de divertissement du géant japonais de l'électronique, a eu des retentissements et des conséquences bien plus cette fois-ci pour les ayants droit ou leurs intermédiaires. Un mois avant Noël, les employés de cette entreprise voient apparaître sur leurs écrans des images de squelettes et des messages de chantage exigeant la déprogrammation du film *L'interview qui tue !*, une comédie mettant en scène un projet d'assassinat du dictateur nord-coréen Kim Jong-Un. La Corée du Nord a toujours nié son implication dans ce piratage. Malgré l'appel au FBI et à Mandiant, une société de cybersécurité, cinq films inédits sont mis en ligne sur des sites pirates ainsi que le script du futur *James Bond*. Des courriers professionnels et confidentiels sont aussi divulgués.

Plus récemment, en 2020, les États et les acteurs mondiaux principaux de la pharmacologie engagés dans la recherche de vaccins contre le Covid-19 ont dû, semble-t-il, faire face, selon Microsoft, à des tentatives de piratage provenant de la Russie ou de la Corée du Nord, mais bien entendu sans reconnaissance explicite de celles-ci, comme c'est le cas pour la plupart des attaques étatiques supposées. À moins que ce n'aient été que de plus simples tentatives d'espionnage industriel, la valeur future attendue des informations de la recherche contre le Covd-19 représentant sans nul doute un énorme enjeu financier…

Les données personnelles

Les entreprises commerciales et industrielles ne sont malheureusement pas les seules à être victimes de vol de données. En marge des attaques qu'elle a subies en 2014, Sony Picture Entertainment a indiqué que des informations personnelles concernant ses employés ou leurs proches, comme leurs coordonnées bancaires, de cartes de crédit ou des mots de passe, avaient pu être dérobées.

Comme évoqué *supra*, le contexte du Covid-19 nous a aussi malheureusement rappelé que ces pratiques frauduleuses concernent aussi les particuliers, dont nombre d'entre eux ont vu leurs résultats de tests PCR divulgués. Les données de pas moins 1,4 million de personnes (identité, numéro de Sécurité sociale, coordonnées…) ont ainsi été dérobées en juin 2020. De manière intéressante, la « fuite » n'a pas émanée de l'application centrale SI-DEP[7] stockant les résultats des dépistages, mais d'une autre application[8] d'échange de fichiers, Dispose, qui joue un rôle de passerelle dans le cadre du traçage des contaminations. C'est cette dernière application qui, rendue vulnérable par une faille de sécurité inconnue à l'époque, a été à la source de la fuite.

Cela constitue un exemple fort intéressant de risques intrinsèques à la valorisation des données – ici le traçage des contaminations à partie des résultats positifs –, lorsque certaines mesures de précaution ne sont pas mises en œuvre – ici la sécurisation des échanges de données. Ce qu'illustre bien la reprise par le philosophe Bernard Stiegler, dans le contexte du numérique, du concept grec de *Pharmakon*[9], cher à Platon et Derrida, celui du double serpent du Caducée d'Hermès, qui est à la fois remède et poison. Fallait-il mettre en œuvre ce type de traçage pour tenter de casser le cycle des contaminations, au risque d'un éparpillement des données personnelles pouvant, par exemple, servir à des usurpations d'identité ou à du hameçonnage[10]? L'efficacité limitée du traçage en France pourrait incliner à regretter cette séquence malheureuse, mais pouvait-on en prévoir l'issue par avance ? À moins que l'on ne cherche un bouc-émissaire – la troisième face du *Pharmakon* antique… Le responsable du piratage et de la fuite des données s'est avéré être un opposant au passe sanitaire ; il s'est défendu en déclarant vouloir « démontrer la faiblesse et la faillibilité du système d'information de l'AP-HP »[11]. Mais il y a bien d'autres manières de rendre connues de telles vulnérabilités

D.R.

Photo 2 : Le philosophe Bernard Stiegler.

[7] Système d'informations de dépistage, qui est la « plateforme sécurisée où sont systématiquement enregistrés les résultats des laboratoires de tests Covid-19 ».

[8] Deux articles du journal *Le Monde*, datés des 15 (https://www.lemonde.fr/pixels/article/2021/09/15/covid-19-les-donnees-de-tests-de-1-4-million-de-personnes-derobees-aux-hopitaux-de-paris_6094806_4408996.html) et 21 septembre 2021 (https://www.lemonde.fr/pixels/article/2021/09/21/comment-les-donnees-de-1-4-million-de-franciliens-testes-pour-le-covid-19-se-sont-retrouvees-dans-la-nature_6095455_4408996.html).

[9] *Internet n'est pas neutre, Internet est un pharmakon. Ce qui nous arrive sur la Toile*, épisode du mardi 14 janvier 2014, par Xavier de La Porte, France Culture (https://www.radiofrance.fr/franceculture/podcasts/ce-qui-nous-arrive-sur-la-toile/internet-n-est-pas-neutre-internet-est-un-pharmakon-9357167).

[10] Technique de piratage, préalable à une usurpation d'identité numérique, consistant à se faire passer pour un tiers de confiance qui aurait un accès légitime aux données dérobées.

[11] Article du journal *Le Monde*, daté du 8 octobre 2021 (https://www.lemonde.fr/pixels/article/2021/10/08/vol-massif-de-donnees-de-sante-de-l-ap-hp-un-pirate-informatique-arrete_6097637_4408996.html).

informatiques : ainsi, les « Hackers éthiques » ont pris l'habitude d'en informer les éditeurs de logiciels incriminés ; autre exemple, celui de communautés comme les CERT (Computer emergency response team), qui traquent ces failles, facilitent l'élaboration des réponses à y apporter ou les conçoivent directement.

Le risque zéro n'existe pas, en tout cas, en ce qui concerne le numérique et le traitement des données, pas plus qu'en médecine ou en matière de transport. Tout est affaire d'anticipation, d'analyse et de compromis bien compris, mais il demeurera toujours une part d'incertitude.

Les acteurs de la cybersécurité s'appliquent à mettre en œuvre dans ce cadre une démarche dite d'analyse de risques, qui s'intéresse notamment :

• aux menaces pouvant entacher les processus supportés par les systèmes d'information ;
• aux risques proprement dits, à leur probabilité, à leur impact et à leur objet : la sécurité physique des personnes ou des biens, la perte ou le vol de données, leur modification, leur divulgation etc. ;
• enfin, aux moyens à mettre en œuvre pour réduire ces risques, diminuer la possibilité de leur survenue ou leurs conséquences, les délais et les coûts associés, à accepter ou à arbitrer.

Cela signifie qu'il existe toujours des risques dits résiduels, qu'il s'agit de documenter et, surtout, de surveiller ; autre domaine important de la cybersécurité.

Au niveau européen, la mise en place du RGPD (le Règlement général sur la protection des données), promulgué en 2016 et applicable depuis 2018, vise à définir le cadre d'exploitation des données à caractère personnel. C'est une référence juridique unique à l'échelle de l'Union européenne, qui, à elle seule, ne protège pas techniquement les usagers, mais donne un cadre à respecter par les entreprises et fait des émules à travers le monde, y compris en Californie, le berceau de la nouvelle économie d'Internet.

Photo © Gordon JOHNSON

Photo 3 : Le caducée, symbole de l'Ordre des médecins.
Source : Photo libre de droits téléchargée sur PIXABAY.

Données critiques et enjeux de souveraineté

L'impact d'une attaque sur les données peut-il toujours être circonscrit aux seules personnes ou entreprises qui en sont propriétaires ? Malheureusement, non. C'est dans ce cadre que l'on évoque la notion d' « opérateur d'importance vitale » (OIV), telle que définie dans le cadre du Code de la Défense en 2007. Si les entreprises sont invitées, ou plutôt incitées à protéger leurs données, si les particuliers sont sensibilisés à la nécessité d'être vigilants au regard de leurs traces numériques, certaines organisations, lorsqu'elles sont attaquées, peuvent être à la source, malgré elles, de conséquences bien plus larges.

La liste des opérateurs d'importance vitale n'est pas publique, mais ce sont pas moins de douze secteurs d'activité, relevant du public ou du privé, qui sont concernés depuis 2008. Depuis 2016, une série de neuf arrêtés ont été pris par le Premier ministre ; des arrêtés rédigés par l'ANSSI, l'Agence nationale de la sécurité des systèmes d'information[12], qui pour plusieurs sont au cœur des sujets abordés dans la présente revue : sous-secteurs de la recherche publique, de l'audiovisuel et de l'information, les communications électroniques et Internet, la santé, dont les données et leur protection sont considérées comme indispensables à la survie de la nation.

La recherche publique notamment, et toutes les connaissances qu'elle rend possibles, ont été jugées clefs par le législateur pour la pérennité de l'État et du corps constitué par les citoyens, mais aussi pour son rayonnement – et sans doute encore davantage pour l'innovation qu'elle permet en matière de santé, de matériaux, de composants, d'agriculture et d'environnement, et qui peut être à la source du développement de demain. Un laboratoire développant des vaccins est à cette aune sans doute plus « vital » qu'un éditeur de films ou de jeux.

L'audiovisuel traditionnel et, de plus en plus, les moyens numériques d'échange sont les médias de base de tout partage de données, mais aussi de formidables moyens d'influence. Le piratage de la chaîne TV5 Monde, le 8 avril 2015, a été un vrai coup de semonce : à la suite de la mise hors service en cascade d'équipements, la chaîne francophone doit interrompre toute diffusion ; en parallèle, ses comptes officiels Twitter et Facebook sont piratés et utilisés pour diffuser des messages en soutien à l'État islamique (EI). L'identité de proches de militaires français engagés dans la lutte contre l'EI est divulguée ; le président de la République François Hollande y est

[12] L'Agence nationale de la sécurité des systèmes d'information, service créé en France en 2009 à la suite de la création de la direction centrale de la Sécurité des systèmes d'information, « assure la mission d'autorité nationale en matière de sécurité des systèmes d'information. À ce titre, elle est chargée de proposer les règles à appliquer pour la protection des systèmes d'information de l'État et de vérifier l'application des mesures adoptées. Dans le domaine de la défense des systèmes d'information, elle assure un service de veille, de détection, d'alerte et de réaction aux attaques informatiques, notamment sur les réseaux de l'État » (https://www.ssi.gouv.fr/).

également mis en cause, quelques mois après la série d'attentats de janvier 2015. La piste de l'EI est progressivement abandonnée, en juin de la même année, les soupçons pointant alors vers le groupe de hackers AP28, proche du gouvernement russe ; l'attaque informatique a eu lieu quelques semaines seulement après l'annulation par la France d'un contrat de fourniture à la Russie de deux porte-hélicoptères Mistral, comme sanction à la suite de la crise ukrainienne d'alors, l'annexion de la Crimée par la Fédération de Russie.

En plein contexte de l'invasion récente de l'Ukraine, d'autres règlements européens ont été proposés, comme le DMA (Digital Markets Act), lequel, attendu pour 2023, vise à encadrer les pratiques des entreprises du numérique, ou encore le DSA (Digital Services Act) sur les termes duquel la Commission et le Conseil européens se sont mis d'accord, modernisant ainsi la direction dite « e-Commerce », afin de réguler les contenus numériques illégaux, la publicité et la désinformation.

Cette « crise » s'est en effet transformée depuis le début de cette année en une guerre – présentée comme une « opération spéciale » par les gouvernants russes – qui se déroule aux portes de l'Union européenne. Comme dans tous les conflits, la présentation des faits, les mots utilisés pour les nommer, le sens que l'on veut leur donner sont autant de moyens d'influence ou de contre-influence. Une des plus grandes surprises de ce conflit, et sans doute des succès enregistrés par l'Ukraine, concerne le domaine de « l'information », où l'on pouvait s'attendre à une bien plus grande domination de la part de la Fédération de Russie. Tous les champs de la communication ont été utilisés par les Ukrainiens et certains soutiens occidentaux :

• une protection Cyber plutôt très efficace des systèmes d'information ukrainiens, même si le piratage du satellite civil KA-SAT, dès le 24 février 2022, a eu des conséquences allant bien au-delà de l'Ukraine, pour les utilisateurs civils européens (un exemple typique, celui des opérateurs d'importance vitale) ;
• la divulgation par les Américains par avance des plans d'attaque russes, comme forme de « contre-propagande » très originale jouant sur la transparence de l'information ;
• l'utilisation de la constellation de satellites Starlink du milliardaire Elon Musk pour diriger les attaques dévastatrices de drones sur les colonnes de blindés russes ;
• la communication du président Zelensky devant les représentations nationales de très nombreux pays, la visite de Kiev devenant un lieu tendance pour les leaders internationaux ;
• la prise de contact avec les proches de soldats et de conscrits russes pour les informer du décès de leur fils ou de leur conjoint par le truchement de la reconnaissance faciale et l'analyse de données ouvertes afin de retrouver leur famille. Cette démarche accompagnée par la société Clearview AI a été jugée comme fort cavalière et bien peu respectueuse des droits humains et du droit de la guerre.

Ce sont autant d'exemples de valorisation de données sur tous les champs informationnels : télécommunications, médias, intelligence artificielle, transparence contre désinformation… Autant d'exemples de l'effet majeur de l'utilisation de la donnée et de l'information qu'elle permet de produire. Il faut donc tout à la fois pouvoir utiliser ces données, mais aussi les protéger, les tracer et en contenir l'exploitation.

Un autre parallèle pourrait être fait avec l'énergie nucléaire : à la différence des hydrocarbures, l'utilisation de cette énergie peut sembler être plus discrète, infinitésimale (au niveau de l'atome), moins directement polluante, mais surtout extraordinairement efficace, voire dangereuse si on libère sa « puissance » et qui peut nuire très longtemps, à l'image des données qui ne sont jamais réellement effacées. À moins qu'elles ne soient absorbées par un trou noir, mais peut-être pourront-elles s'en échapper par un trou de ver[13], le nouvel horizon de la recherche astrophysique.

[13] Concept théorique et encore hypothétique, suggéré par Albert Einstein en 1935, et dénommé ainsi en 1957 par Charles W. Misner et John A. Wheeler. S'il était avéré, il permettrait, en toute hypothèse, de rejoindre deux régions de l'espace-temps, le voyage temporel, et donc de libérer « ailleurs » l'information annihilée par le trou noir. L'hypothèse est utilisée par Christopher Nolan dans *Interstellar* comme dénouement de l'énigme des « fantômes » à l'origine de la chute de livres dans une bibliothèque. La future chercheuse, encore enfant, est sans le savoir en contact avec son propre père qui tente de lui faire passer des messages à travers ce trou de ver.