
LES ENJEUX STRATEGIQUES DU STOCKAGE DE DONNEES

DECEMBRE 2018

PROMOTION MSIE29 - Mathilde P. - Gilles C. - Céline G. - Louis J.
ECOLE DE GUERRE ECONOMIQUE



« Les principes simples de la recherche scientifique.

Vérifier toute intuition par l'expérience et l'observation.

Théoriser les intuitions et les valider.

Et rejeter celles qui ne le sont pas.

Interpréter les résultats sans à priori.

Et ne rien tenir pour acquis. »

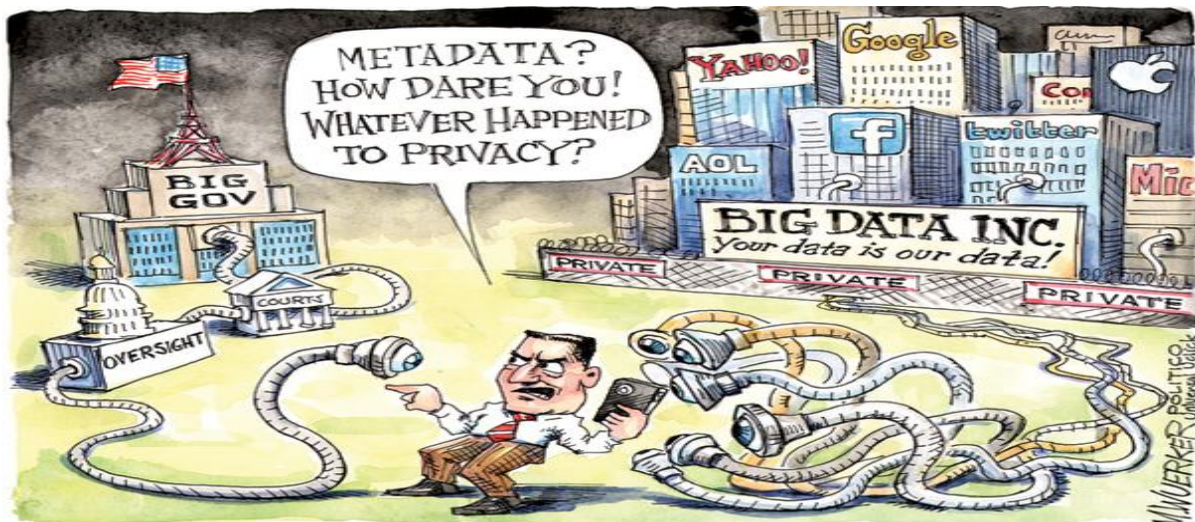
Neil de Grasse Tyson.

Table des matières

.....	0
EXECUTIVE SUMMARY	4
LA DONNEE A TRAVERS LES AGES.	5
LA DONNEE, TOUTE UNE HISTOIRE	8
1 DEFINITIONS	15
2 LES DONNEES SONT LE NOUVEL OR NOIR.	19
2.1 DE LA DONNEE AU BIG DATA : CHANGEMENT DE PARADIGME	19
2.2 BIG DATA.....	21
2.3 MODUS OPERANDI DU RECUEIL ET DU TRAITEMENT DE LA DONNEE PAR LES ENTREPRISES.	23
2.4 COMMENT LES ENTREPRISES UTILISENT-ELLES VOS DONNEES ?	24
2.5 UTILISATION DES DONNEES POUR SECURISER LES DONNEES	25
3 LA SECURISATION DE LA DONNEE.....	28
3.1 REPOSE JURIDIQUE	28
3.1.1 LE CLOUD ACT.....	28
3.1.2 EUROPE ET RGPD.....	29
3.1.3 CLOUD ACT ET RGDP : QUELLES CONSEQUENCES ?	35
3.1.4 UN RGPD AMERICAIN.....	36
3.1.5 DONNEES PERSONNELLES VERSUS DONNEES IDENTIFIABLES, L'ENJEU DE L'ANONYMISATION	40
3.1.6 ETUDE DE CAS	41
3.2 TECHNOLOGIQUE	45
3.2.1 L'EVOLUTION DU STOCKAGE.....	46
3.2.2 QUI EST CONCERNE ?	49
3.2.3 LES ACTEURS.....	49
3.2.4 POUR QUELLES RAISONS PASSER AU CLOUD ?.....	50
3.3 LE STOCKAGE, UN ENJEU STRATEGIQUE SOUS-ESTIME POUR LA SECURISATION DE DONNEES.....	51
3.3.1 LA QUESTION DE LA RESPONSABILITE DES DONNEES RESTE FLOUE	52
3.3.2 RAPPELER LA REGLE 3-2-1	52
3.3.3 LES ACTEURS DU CHANGEMENT DANS LA COURSE A LA DIGITALISATION	52
3.3.4 LA DEPENDANCE AUX FOURNISSEURS DE SERVICES CLOUD.....	55
4 LES EVOLUTIONS TECHNOLOGIQUES DE HYPE CYCLE DE GARTNER	57
5 GEOPOLITIQUE DU NUMERIQUE	61
5.1 ETAT-NATION CONTRE ETAT-PLATEFORME	61

6	CHINE, LA DATA AU CŒUR DE L’ECONOMIE.....	64
6.1	NUMERO UN DES TELECOMMUNICATIONS.....	64
6.1.1	LE MARCHÉ FLEURISSANT DU TOUT MOBILE	65
6.1.2	CAP SUR LA 5G.....	66
6.2	POLITIQUE VOLONTARISTE POUR PROMOUVOIR LES NORMES CHINOISES.....	67
6.3	BIG DATA et IOT.....	68
6.3.1	VOLONTE POLITIQUE.....	68
6.3.2	MARCHÉ ET PERSPECTIVES	69
6.3.3	UN CLOUD CHINOIS PRIVE, LA TERRITORIALISATION DES DONNEES	70
6.4	UN ETAT DERRIERE SES ENTREPRISES : BATX A L’ASSAUT DU MONDE.	71
6.4.1	LA TERRITORIALISATION DES DONNEES.....	73
6.4.2	CREER UNE SOCIETE DE LA DATA.....	73
7	LA SOUVERAINETE A LA FRANCAISE.....	75
7.1	LA VOLONTE DE CREER UN CLOUD SOUVERAIN.	75
7.2	LE PLAN CLOUD COMPUTING	75
7.3	2018 : LE VIRAGE DU BIGDATA ET DE L’IA.....	76
7.3.1	CLOUD SOUVERAIN	76
7.3.2	PLAN D’INVESTISSEMENT EN INTELLIGENCE ARTIFICIELLE (IA)	77
8	CONCLUSION – DE L’INFINIMENT GRAND A L’INFINIMENT PETIT.....	79
9	BIBLIOGRAPHIE.....	83

EXECUTIVE SUMMARY



LA DONNEE A TRAVERS LES AGES.

Les données représentent notre passé, notre présent et notre futur.

Notre passé :

- Soit il y a 3,5 millions d'années, nos ancêtres, ont laissé des traces. Durant la plus grande partie de l'histoire de l'humanité soit environ 40 000 générations, nous vivons en nomades, au sein de petites tribus de chasseurs cueilleurs. Nous fabriquons des outils, apprivoisons le feu et inventons le langage.
- Il y a 30 000 ans, nous dessinons sur les parois des grottes. C'est à ce moment que nous inventons l'astronomie. Notre survie dépend de la lecture des astres. Il faut pouvoir prédire l'arrivée de l'hiver et la migration des troupeaux.
- Puis il y a 10 000 ans une révolution bouscule notre mode de vie. Nos ancêtres apprennent à modeler leur environnement, ils domestiquent les animaux, cultivent la terre et se sédentarisent. Rien n'est plus pareil !

Pour la première fois de notre histoire, nous avons trop de possessions pour tout transporter avec nous. Il nous faut un moyen pour faire le compte.

- Il y a environ 6000 ans nous inventons l'écriture.

Très vite, nous notons bien d'avantage que des quantités de grains.

L'écriture nous permet de consigner nos pensées et de les diffuser dans l'espace et le temps. Grâce à des inscriptions sur des tablettes d'argile, nous pouvons prétendre à l'immortalité. Le monde en est bouleversé.

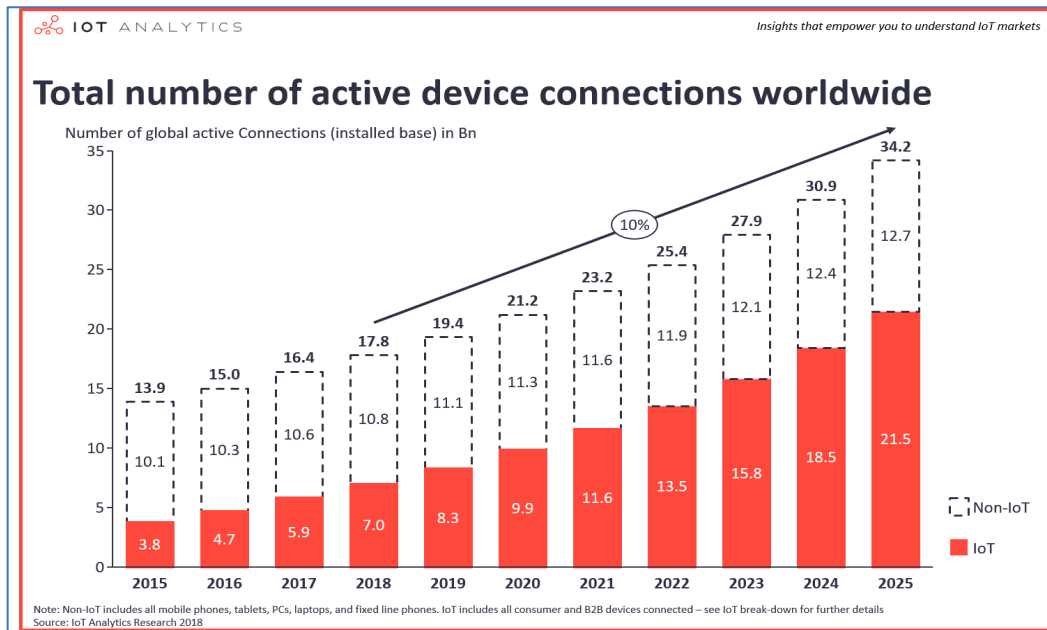
Notre présent : Nous laissons délibérément ou non une quantité massive d'informations.

Nul besoin d'être écrivain et d'avoir accès à la publication dite traditionnelle pour laisser aujourd'hui une trace dans l'histoire.

Internet a changé le monde de la communication.

Sites, blogs, réseaux sociaux hébergent nos données : celles que nous laissons volontairement en écrivant une tribune, un post, un avis, mais aussi celles que nous ne voulons pas forcément laisser lors d'une visite de site.

Les fonctions de navigations privée et d’effacement l’historique ne permettent qu’un « nettoyage » surfacique de votre passage. En effet, les logs de notre adresse IP, notre empreinte numérique (navigateur utilisé et son référeur, la configuration de système, le temps de navigation) permettent de tracer nos passages sur les sites via les hébergeurs. Et ceci ne se cantonne pas uniquement à nos téléphones et ordinateurs, mais à tous les objets connectés – pas moins de [7 milliards annoncés en 2018](#) qui ont envahi notre quotidien.



Notre futur :

Du prédictif au prescriptif, nos données sont collectées puis traitées aux fins de mieux cerner nos besoins, anticiper les offres commerciales voire prédire quand nous serions malades, un mouvement de foule ou lors d’une séparation dans une vie de couple. Il s’agit du prédictif. L’étape suivante consiste non seulement à prédire l’avenir mais aussi de soumettre des propositions de décisions. C’est l’heure du prescriptif.

Place au Big Data et à l’Intelligence Artificielle. Si ces termes sont passés du statut de la science fiction, puis de l’anticipation à ceux du marketing, il faut y voir outre la disruption technologique un **changement de rapport de force entre les États nations et les géants du numériques** dont les données en sont la pierre angulaire.

Qu’ils soient politiques, économiques et sociétales, la sécurité de nos données passe par les protections réglementaires et les choix technologiques de nos institutions, de nos entreprises souffrant d’absence de stratégie.

AVANT PROPOS

LA DONNEE, TOUTE UNE HISTOIRE

« Comment nous autres, humains, qui ne vivons rarement plus d'un siècle, pouvons-nous concevoir le temps qui a passé de notre univers ?

Selon la théorie de la création de l'univers par le Big Bang, L'univers serait vieux de treize milliards huit cents millions d'années et serait en expansion.

1. Pour mesurer l'échelle des temps cosmiques, rapportons cette durée aux douze mois d'une seule année :

- Notre calendrier cosmique débute le 1^{er} janvier avec la naissance de notre univers, il y est inscrit tout ce qui s'est passé depuis jusqu'à maintenant c'est à dire jusqu'au 31 décembre, à minuit. A cette échelle, chaque mois symbolise environ à 1 milliards d'années et chaque jour correspond à presque 40 millions d'années... Dans cette immensité temporelle que représente notre calendrier, si le 1^{er} janvier correspond au big bang, l'apparition de l'humanité ne commence qu'à la dernière heure du dernier jour de cette année cosmique.

- Il est 23h59 et 46 secondes, la totalité de l'histoire des Hommes ne couvre que 14 secondes... En changeant d'échelle, il est 21h45, en ce soir du 31 décembre soit il y a 3,5 millions d'années, nos ancêtres, ont laissé des traces. Durant la plus grande partie de l'histoire de l'humanité soit environ 40 000 générations, nous vivons en nomades, au sein de petites tribus de chasseurs cueilleurs. Nous fabriquons des outils, apprivoisons le feu et inventons le langage. Tout cela au cours de la dernière heure de l'année cosmique.

Changeons encore d'échelle. La dernière minute du dernier soir de l'année.

- Il y a 30 000 ans, nous dessinons sur les parois des grottes. C'est à ce moment que nous inventons l'astronomie. Notre survie dépend de la lecture des astres. Il faut pouvoir prédire l'arrivée de l'hiver et la migration des troupeaux.
- Puis il y a 10 000 ans une révolution bouscule notre mode de vie. Nos ancêtres apprennent à modeler leur environnement, ils domestiquent les animaux, cultivent la terre et se sédentarisent.

Pour la première fois de notre histoire, nous avons trop de possessions pour tout transporter avec nous. Il nous faut un moyen pour faire le compte.

- Il y a environ 6000 ans soit 14 secondes avant minuit, nous inventons l'écriture.

Très vite, nous notons bien d'avantage que des quantités de grains.

L'écriture nous permet de consigner nos pensées et de les diffuser dans l'espace et le temps. Grâce à des inscriptions sur des tablettes d'argile, nous pouvons prétendre à l'immortalité. Le monde en est bouleversé.

Moïse naît il y a 7 secondes, Bouddha 6 secondes, Jésus 5 secondes, Mahomet 3 secondes. Il n'y a que 2 secondes que les civilisations européennes et amérindiennes n'entrent en contact pour le meilleur et pour le pire. Ce n'est qu'à la dernière seconde de notre calendrier cosmique que nous commençons à nous servir de la science pour comprendre les lois de la nature. L'avènement de la méthode scientifique est si radical que 4 siècles seulement séparent la première utilisation d'un télescope par Galilée de la première expédition habitée sur le sol lunaire. **Grâce à la science, nous savons à quel endroit de l'espace et du temps, nous nous situons dans l'univers. »**

Neil de Grasse Tyson

2. Présentation des phases principales de l'évolution du concept de donnée depuis 4000 ans avant Jésus Christ jusqu'à nos jours.

Si la sémantique est importante, la notion de temporalité permet de « donner » un sens différent pour la définition des mots.

Le participe passé féminin substantivé du verbe donner a d'abord eu le sens d' « aumône, distribution » (1200). Il s'est limité à quelques sens spécialisés en mathématiques (1755) et psychologie. On l'utilise aussi en informatique et statistique, traduit de l'anglais data, pluriel du supin, première forme, du verbe latin dare, « donner » (voir Rey et al., 1993). Dans cette brève présentation de son histoire, nous distinguerons quatre périodes distinctes

Du sumer au 16^e siècle

Les données produites pendant cette période concernent essentiellement deux opérations : le dénombrement et l'observation de phénomènes astronomiques.

Le dénombrement a toujours été une opération importante de l'activité humaine. Le recensement des populations en est son expression statistique la plus visible. Les premiers témoignages de mise en œuvre de cette méthode de collecte de données sont gravés sur des tablettes d'argile sumériennes et babyloniennes.

- Dès le début de son utilisation, le recensement s'est avéré être un outil de gestion apprécié des puissants. Si les Mésopotamiens y ont recouru très tôt, on en trouve aussi trace dans l'Égypte ancienne, dès la fin du troisième millénaire avant notre ère. Ces peuples avaient bien saisi l'intérêt de recenser les populations pour savoir combien d'hommes pouvaient participer à la construction des temples, palais, pyramides... ou encore d'utiliser cette technique à des fins fiscales.
- Plus une population est nombreuse, plus le recensement s'avère utile. C'est ce qu'avaient compris aussi les empereurs chinois. Quelle que soit l'époque concernée, ceux-ci avaient doté la Chine d'une structure administrative consacrée à ce thème, dirigée par des directeurs des multitudes aux pouvoirs affirmés. Pendant plus de 2000 ans, le recensement a constitué un outil au service de l'administration chinoise.
- L'Inde est un autre pays d'Asie qui a procédé dès le 4^e siècle avant notre ère au dénombrement de sa population. Elle a même été plus loin que cet objectif fondamental en prônant une politique planificatrice d'expansion territoriale et économique basée sur une connaissance approfondie de sa population.
- Un traité a défini la façon d'y parvenir, l'Arthasastra, rédigé par Kautilya, ministre de l'empire indien des Maurya. Il s'agit d'une méthode dont la minutie est remarquable, tant dans la manière de définir les caractères de la population prise en compte que de la quantité importante de données relevées (voir Hecht, 1987). On imagine sans difficulté que la mise en œuvre de ce type de relevé reposait sur un soutien administratif très dense, encadré par un contrôle policier explicite qui n'incitait pas aux non-réponses. Cette manière d'agir a mis plus de temps pour être appliquée en Occident. La civilisation grecque accorda moins d'importance à la recherche du nombre d'habitants qu'à celui de la composition idéale de la Cité, chère à Platon,

même si Aristote, dans sa politique, s'attarda à réaliser des travaux de statistique descriptive et comparative.

- Les Romains reprirent les objectifs originaux : réaliser des recensements périodiques grâce à une structure administrative bien organisée afin de contrôler et d'administrer toutes les composantes de leurs territoires. Initiés sous Servius Tullius au 5^e siècle avant notre ère et réalisés jusqu'en l'an 73 sous Auguste (année du dernier recensement de l'empire romain), les dénombrements ont fait de la fonction de censeur, comme en Chine, un privilège recherché.
- La période du déclin de l'Empire romain et le Haut Moyen Age n'ont pas constitué un terrain fertile pour l'organisation de recensements.
- Ce n'est qu'entre le 14^e et le 16^e siècle, que l'on ressent à nouveau le besoin d'informations, que ce soit au niveau des rôles fiscaux ou à celui des relevés d'ordre religieux

Dans un autre domaine, l'astronomie, les Babyloniens ont observé les mouvements du soleil et des planètes à intervalles réguliers, obtenant ainsi plusieurs observations d'un même phénomène.

Nous ne connaissons malheureusement pas la manière dont ils ont remplacé ces observations multiples par des « valeurs de compromis ». On possède plus d'informations sur les travaux ultérieurs des astronomes grecs (voir Droysbeke et Saporta, 2010).

- Ainsi, Ptolémée, astronome du 2^e siècle, utilisa les relevés antérieurs d'Aristarque de Samos et surtout d'Hipparque et proposa, en présence de plusieurs observations d'un même phénomène, de conserver une seule valeur accompagnée de mesures de variation basées, semble-t-il, sur l'étendue des observations, c'est-à-dire l'écart entre la plus grande et la plus petite d'entre elles. Jusqu'au 16^e siècle, on préféra retenir une « bonne valeur » en omettant souvent de justifier l'adjectif utilisé que recourir à une synthèse systématique. Le premier qui utilisa une moyenne comme « outil de synthèse » est probablement l'astronome Tycho Brahé (1546-1601) dont les nombreuses données sur le mouvement des planètes permirent à Johannes Kepler (1571-1630) d'énoncer les lois qui portent son nom.
- En cette fin du 16^e siècle, on construit des données primaires et individuelles, semblables à celles élaborées dans l'Antiquité depuis l'apparition de l'écriture. Pour ce qui concerne les données résultant d'observations répétées d'un même phénomène, les faibles progrès techniques réalisés dans la recherche d'une plus grande précision des instruments de mesure ont fait croire longtemps qu'une « bonne mesure » était meilleure qu'une agrégation dont on ne soupçonnait pas l'intérêt.

Les 17^e et 18^e siècles

Depuis le 15^e siècle, de nombreuses villes ont recensé leurs habitants. Les Etats tendant à se centraliser et à se doter d'une administration solide, le besoin de dénombrer se fait à nouveau sentir, même si la pratique est souvent défailante.

- Jusqu'à la fin du 17^e siècle, les registres sont en effet loin d'être parfaits ! Ce 17^e siècle voit trois courants distincts se développer en Europe : la Staatkunde allemande, les enquêtes de l'administration française et l'arithmétique politique anglaise. La

Staatkunde allemande trouve ses racines dans les travaux d'Aristote. Pour ses défenseurs

La statistique est la science de l'Etat.

Purement descriptive, elle ne fait pratiquement jamais appel à des données chiffrées. Son influence est cependant significative, surtout en Europe centrale, et perdurera jusqu'au 19^e siècle.

En France, on plaide toujours pour les dénombrements comme outils de gouvernement. Deux hommes se sont particulièrement illustrés dans le recours à des enquêtes en raison des contraintes économiques : Colbert (1619-1683) qui développe une stratégie de dénombrement des villes et des régions, et Vauban (1633-1707), auteur d'une Méthode générale et facile pour faire le dénombrement des peuples en 1686.

Mais c'est en Angleterre qu'un mouvement novateur se répand avec l'arithmétique politique due principalement à Graunt (1620-1674) et Petty (1623-1687). Comme le dira Charles Davenant (1656-1714), émule de Petty, « l'arithmétique politique est l'art de raisonner par des chiffres sur des objets relatifs au gouvernement ».

On y trouve les fondements de la méthode du multiplicateur qui a marqué les techniques de dénombrement des 17^e et 18^e siècles, provoquant une mise à l'ombre certaine de la Staatkunde allemande en Europe occidentale.

La méthode du multiplicateur repose sur l'idée suivante :

il existe des quantités qui sont en rapports simples et relativement constants avec la population d'un pays. Si ces quantités sont plus simples à dénombrer (nombre de maisons, feux (foyers), ou encore nombre de naissances, de décès... dans l'année), il suffit de multiplier leur nombre par un multiplicateur adéquat pour obtenir une estimation du nombre d'individus dans la population.

Pour les responsables politiques de l'époque, le recensement d'une population présente des désavantages certains (réactions de méfiance des enquêtés, coûts de mise en œuvre trop élevés...); mais d'un autre côté, le choix d'une entité plus simple à dénombrer et la détermination d'un multiplicateur unique posent aussi des problèmes de fiabilité.

Une des caractéristiques du 18^e siècle

Le siècle des Lumières est le triomphe de l'esprit de calcul. Il faut dire que les progrès réalisés par les mathématiques sont alors considérables et la loi des grands nombres de Bernoulli vient ajouter sa pierre à l'édifice. L'époque est cependant marquée par de nombreuses imprécisions sur les estimations fournies par les uns et les autres.

Il n'est donc pas étonnant de constater que les recensements sont revenus en force au 19^e siècle avant de connaître une stagnation puis un déclin au 20^e siècle, dû notamment à l'introduction de registres administratifs performants et au développement des techniques de sondage. Mais cela, c'est une autre histoire sur laquelle nous reviendrons ci-dessous.

Parmi les développements qui contribuent significativement à l'évolution de l'histoire des données, il faut souligner l'amélioration des instruments de mesure. Celle-ci est essentielle car elle permet aux hommes de s'aventurer sur les mers en s'assurant une meilleure qualité des moyens de se guider. Par ailleurs, si les mesures astronomiques constituent toujours une manière incontournable malgré leurs imprécisions de savoir où l'on se trouve, un autre instrument de connaissance permet de mieux maîtriser le sol sur lequel on vit : la géodésie.

Cette discipline et l'astronomie constituent deux domaines privilégiés d'un traitement de données qui se cherche.

Un exemple remarquable est celui de la mesure d'un arc de méridien, au centre d'une question primordiale à l'époque : quelle est la figure de la terre ? Le recours à la technique de triangulation est à l'origine d'aventures multiples de cette mesure dans diverses régions du globe qui permettront de résoudre la question (voir Droysbeke et al. , 2016).

Le besoin de mesurer est partagé par de nombreux savants qui bénéficient d'instruments de mesure de plus en plus précis. On devient exigeant à propos de la qualité des observations effectuées et l'erreur de mesure devient un souci essentiel.

Deux approches coexistent pendant de nombreuses années.

- La première milite pour la recherche d'une bonne mesure, entachée d'une erreur limitée, inférieure à une erreur maximale, acceptable ou en tout cas à craindre. Dans cette optique – défendue par Leonhard Euler (1707-1783) prendre en compte d'autres mesures en plus de la bonne ne peut que faire croître l'erreur globale, notamment en utilisant les observations les plus mauvaises.
- Un deuxième modèle nous intéresse davantage ici. Il repose sur l'hypothèse que l'utilisation de toutes les observations permet des compensations dont on peut espérer qu'elles réduisent l'erreur résultante. C'est en recherchant des modèles appropriés de distribution des erreurs que de nombreux scientifiques contribueront à la consolidation d'une théorie qui sera qualifiée en 1765 de théorie des erreurs par Johann-Heinrich Lambert (1728-1777).

La multiplicité des observations et le besoin de s'interroger sur le comportement des erreurs de mesure n'amènent pas seulement de nombreux scientifiques à vouloir modéliser cette erreur pour mieux la dompter ; cette question comporte aussi des aspects politiques et commerciaux qui constituent autant d'enjeux importants pour l'époque. Les données se multiplient et se contredisent.

Il faut en comprendre la raison, les gérer pour en tirer profit. Le calcul des probabilités vient en aide à ceux qui affrontent ce problème. Il en résulte une conséquence à trois facettes dont les effets seront durables : la loi des erreurs, qui sera qualifiée de « normale » à la fin du 19^e siècle, devient une loi de référence, la moyenne s'avère être le mode de synthèse privilégié et un critère d'ajustement devient incontournable : le critère des moindres carrés (voir Droysbeke et Tassi, 2015). Deux hommes jouent un rôle central dans cette histoire : Pierre Simon de Laplace (1749-1827) et Carl Friedrich Gauss (1777-1855). Les données individuelles deviennent plus fiables ; elles font place aux données agrégées et aux données transformées (voir Droysbeke et Vermandele, 2016).

Quelques points forts du 19^e siècle

Le 19^e siècle occupe une place très importante en statistique. S'il fallait retenir cinq caractéristiques essentielles de ce siècle dans l'histoire des données, notre choix serait le suivant :

- L'application des trois outils utilisés en astronomie (loi normale, critère des moindres carrés et moyenne) à l'étude des populations et de leurs caractéristiques humaines, permettant à Adolphe Quetelet (1796-1872) de créer une théorie des moyennes aux

accents multiples (voir Académie Royale de Belgique, 1997, Desrosières, 1993 ou encore Droesbeke et Vermandele, 2016).

- Le développement de la statistique comme outil de gestion des Etats, au niveau économique et social (voir Desrosières, 1993). Les tables statistiques et les représentations graphiques deviennent un outil important d'analyse et de communication
- Le remplacement du rôle central de la moyenne par celui de la dispersion dans les préoccupations des savants de tous bords.
- Le déplacement du centre de gravité de la statistique vers Londres et l'émergence des concepts de corrélation et de régression (voir Droesbeke et Tassi, 2015 et Droesbeke et Vermandele, 2016).
- L'émergence d'une nouvelle méthode de recueil des données : les sondages (voir Droesbeke et Tassi, 2015).

Les données individuelles se répandent ; la manière de les produire se diversifie. Elles deviennent nombreuses : il faut les montrer et les résumer.

Le 20^e siècle et le début du 21^e siècle.

Il est difficile de détailler les développements de la statistique au 20^e siècle, tant les innovations sont nombreuses et diversifiées.

L'inférence statistique est au centre de ces dernières, avec ses deux problèmes centraux, l'estimation de paramètres d'une population et les tests d'hypothèses réalisés à partir d'un échantillon.

Nous ne pouvons expliciter ici toutes les ouvertures nouvelles du 20^e siècle et du début du siècle actuel, qui traitent des données : elles s'appellent plans d'expérience, méthodes de sondage, analyse statistique bayésienne, analyse exploratoire des données, analyse robuste... Parallèlement des stratégies d'analyse ont vu le jour ainsi que des procédures de diffusion des résultats d'analyse appropriées.

Les données sont à présent multi variées, elles sont qualitatives ou quantitatives, il en est de manquantes et d'extrêmes, elles deviennent de plus en plus nombreuses. En ce début de 21^e siècle, les données massives ou Big Data nous lancent des défis de toute nature :

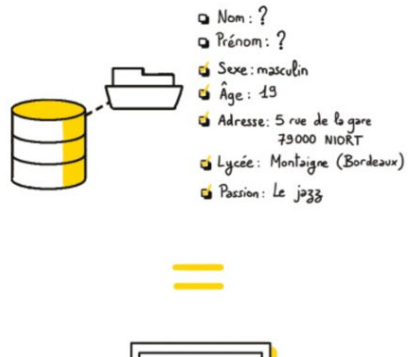
Méthodologiques, techniques, juridiques... Le statisticien de demain se doit de s'ouvrir encore davantage de la science des données.

1 DEFINITIONS

- **Big Data.** Source CNIL. « On parle depuis quelques années du phénomène de big data que l'on traduit souvent par « données massives ». Avec le développement des nouvelles technologies, d'internet et des réseaux sociaux ces vingt dernières années, la production de données numériques a été de plus en plus nombreuse : textes, photos, vidéos, etc. Le gigantesque volume de données numériques produites combiné aux capacités sans cesse accrues de stockage et à des outils d'analyse en temps réel de plus en plus sophistiqués offre aujourd'hui des possibilités inégalées d'exploitation des informations. Les ensembles de données traités correspondant à la définition du big data répondent à trois caractéristiques principales : volume, vitesse et variété. »
- **Clauses contractuelles types.** Source CNIL « Il s'agit de modèles de clauses contractuelles adoptés par la Commission européenne permettant d'encadrer les transferts de données personnelles effectués par des responsables de traitement vers des destinataires situés hors de l'Union européenne. Elles ont pour but de faciliter la tâche des responsables de traitement dans la mise en œuvre de contrats de transfert. On distingue les transferts de responsable de traitement à responsable de traitement et les transferts de responsable de traitement à sous-traitant. Il existe donc deux types de clauses afin d'encadrer chacun des transferts »
- **Donnée.** Le terme donnée est un terme générique utilisé dans plusieurs domaines (Source Wikipedia)
 - En informatique et en statistique, un jeu de données est une collection de valeurs structurées.
 - En statistique, une donnée est étudiée dans le cadre de l'analyse des données.
 - En informatique, une donnée fait l'objet de traitements de données.
- **Données anonymes:** données rendues anonymes d'une manière telle que la personne concernée n'est plus identifiable (considérant 26 dir. 95/46)
- **Donnée personnelle**
 - Art. 4 du règlement général sur la protection des données (RGPD) :
"Toute information se rapportant à une personne physique identifiée ou identifiable (ci-après dénommée « personne concernée ») ; est réputée être une « personne physique identifiable » une personne physique qui peut être identifiée, directement ou indirectement, notamment par référence à un identifiant, tel qu'un nom, un numéro d'identification, des données de localisation, un identifiant en ligne, ou à un ou plusieurs éléments spécifiques propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale ;".
« La personne concernée par un traitement de données à caractère personnel est celle à laquelle se rapportent les données qui font l'objet du traitement.
Les données à caractère personnel doivent être collectées pour une finalité (but, objectif) déterminée, explicite et légitime. Article 5 du Règlement européen sur la protection des données personnelles ».

- Point de vue de la CNIL

Qu'est-ce qu'une donnée personnelle ?



La notion de « données personnelles » est à comprendre de façon très large

Une « donnée personnelle » est « toute information se rapportant à une personne physique identifiée ou identifiable ».

Une personne peut être identifiée :

- **directement** (exemple : nom, prénom)
- **ou indirectement** (exemple : par un identifiant (n° client), un numéro (de téléphone), une donnée biométrique, plusieurs éléments spécifiques propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale, mais aussi la voix ou l'image).

« Les données sont considérées "à caractère personnel" dès lors qu'elles concernent des personnes physiques identifiées directement ou indirectement. Une personne est identifiée lorsque par exemple son nom apparaît dans un fichier.

Une personne est identifiable lorsqu'un fichier comporte des informations permettant indirectement son identification (ex. : adresse IP, nom, n° d'immatriculation, n° de téléphone, photographie, éléments biométriques tels que l'empreinte digitale, ADN, numéro d'Identification Nationale Étudiant (INE), ensemble d'informations permettant de discriminer une personne au sein d'une population (certains fichiers statistiques) tels que, par exemple, le lieu de résidence et profession et sexe et âge,...).

Des données que vous pourriez considérer comme anonymes peuvent constituer des données à caractère personnel si elles permettent d'identifier indirectement ou par recoupement d'informations une personne précise. Il peut en effet s'agir d'informations qui ne sont pas associées au nom d'une personne mais qui permettent aisément de l'identifier et de connaître ses habitudes ou ses goûts. En ce sens, constituent également des données à caractère personnel toutes les informations dont le recoupement permet d'identifier une personne précise. (ex. : une empreinte digitale, l'ADN, une date de naissance associée à une commune de résidence ...).

- **Donnée publique** : « La notion de donnée publique caractérise une donnée devant être transparente et connue du plus grand nombre. Sa diffusion est en général du ressort des activités de puissance publique, notamment dans le cadre de la liberté d'accès aux documents administratifs. »
- **Donnée biométrique**. Source CNIL : « Caractéristique physique ou biologique permettant d'identifier une personne (ADN, contour de la main, empreintes digitales...). »
- **Données structurées** Les données structurées sont des informations organisées et classées en vue de faciliter leur lecture et leur traitement.

- **Données non structurées:** source CNIL. Les données non structurées sont une désignation générique qui décrit toute donnée extérieure à un type de structure. Les données non structurées textuelles sont générées par les courriels, les présentations PowerPoint, les documents Word, ou encore les logiciels de collaboration ou de messagerie instantanée.
- **Donnée sensible.** : Source CNIL. « Information concernant l'origine raciale ou ethnique, les opinions politiques, philosophiques ou religieuses, l'appartenance syndicale, la santé ou la vie sexuelle. En principe, les données sensibles ne peuvent être recueillies et exploitées qu'avec le consentement explicite des personnes. »
- **Droit à l'information.** Source CNIL : « Toute personne a un droit de regard sur ses propres données ; par conséquent, quiconque met en œuvre un fichier ou un traitement de données personnelles est obligé d'informer les personnes fichées de son identité, de l'objectif de la collecte d'informations et de son caractère obligatoire ou facultatif, des destinataires des informations, des droits reconnus à la personne, des éventuels transferts de données vers un pays hors de l'Union européenne. »
- **Fichier:** Tout ensemble structuré de données à caractère personnel accessibles selon des critères déterminés, que cet ensemble soit centralisé, décentralisé ou réparti de manière fonctionnelle ou géographique ' - Art. 4, 6, RGPD
- **Stockage de données.** Source Wikipedia « Le stockage d'information est aujourd'hui assuré par un support d'information électronique qui, vu de l'utilisateur, peut être physique (disque dur, clé USB, etc.) ou virtuel (Internet dénommé le nuage ; cloud en anglais), mais qui en l'état de la technologie est enregistré sur un support physique (disque dur, CD/DVD, bande magnétique, etc.). Ainsi, le terme de « dématérialisation » employé pour désigner le passage d'un support d'information papier à un support électronique est peu approprié, puisque ce dernier est lui aussi matériel.

Le choix de la méthode de stockage se fait selon plusieurs critères :

- la fréquence d'utilisation ;
- les besoins capacitaires de l'information (taille) ;
- la criticité de l'information (coût, sécurité) ;
- La capacité de stockage et la vitesse d'accès à l'information.

L'évolution des techniques de stockage est rapide, et tend vers plus de capacité, plus de vitesse, plus de fiabilité, tout en étant moins cher à capacité équivalente. Les types de média sont variés et changent souvent.

- **Traitement de données à caractère personnel.** Source CNIL : « Toute opération, ou ensemble d'opérations, portant sur de telles données, quel que soit le procédé utilisé (collecte, enregistrement, organisation, conservation, adaptation, modification, extraction, consultation, utilisation, communication par transmission diffusion ou toute autre forme de mise à disposition, rapprochement ou interconnexion, verrouillage, effacement ou destruction, ...

2 LES DONNEES SONT LE NOUVEL OR NOIR.

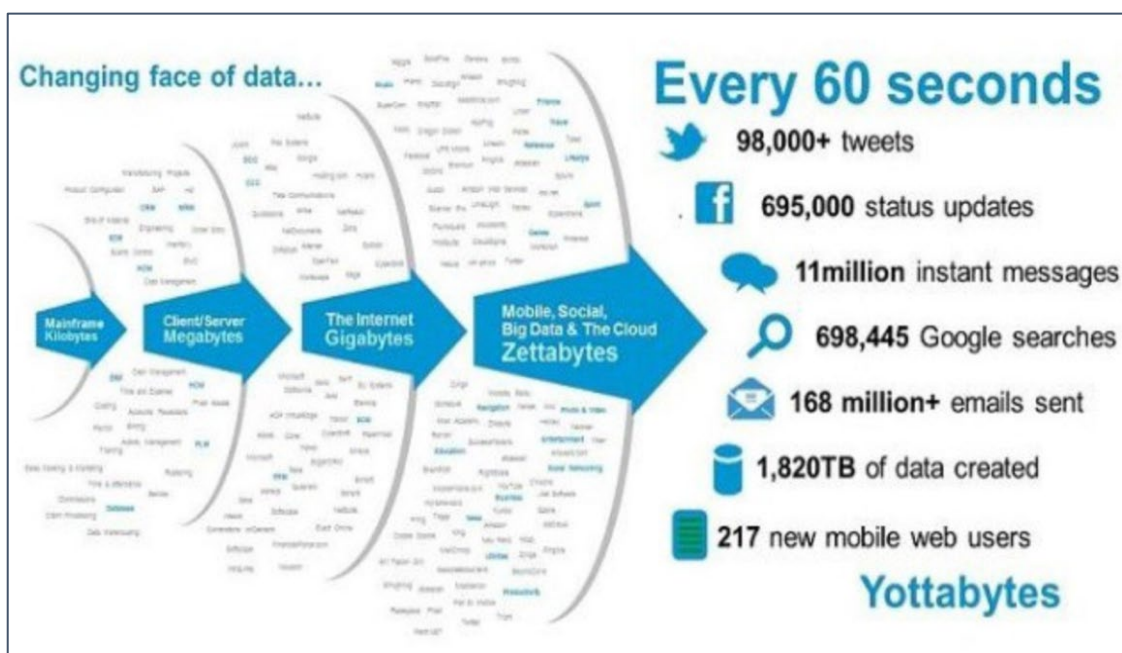
Qu'elles soient personnelles ou identifiables, publiques ou privées, fournies avec le consentement de l'utilisateur ou non, les données n'ont plus uniquement une fonction de pouvoir pour les états mais surtout un modèle économique sans précédent pour les entreprises de technologies de l'information.

Nos datas nourrissent un circuit économique détenu par les GAFAM, BATX et toutes les autres structures auxquelles nous souscrivons un contrat à travers l'acceptation des conditions générales d'utilisation.

Ces dernières sont assez illisibles et non négociables entraînant ainsi un état de fait ou nous nous dépossédons d'une partie non négligeable de notre personnalité.

Nos datas sont alors agrégées, exploitées et souvent revendus en échange d'une gratuité illusoire, où nous, utilisateur en sommes à la fois matière première, produit et consommateur.

2.1 DE LA DONNEE AU BIG DATA : CHANGEMENT DE PARADIGME



Le magazine *Emerging Technology Trends* publie le dernier rapport du Gartner sur les technologies phares aux enjeux stratégiques. Le Hype Cycle retrace ainsi une trentaine de technologies en vogue, selon leur niveau de maturité et les classe en cinq grandes catégories et explique :

Gartner prédit que “les technologies d’IA seront pratiquement partout **au cours des 10 prochaines années.** “Ces technologies, qui permettent aux ‘early adopters’ de s’adapter à de nouvelles situations et de résoudre de nouveaux problèmes, vont être mises à la disposition des masses – démocratisées. Des tendances comme le Cloud Computing, la communauté des ‘makers’ et l’open source vont propulser l’IA **entre les mains de tout le monde** ».



Figure : Les Grandes Technologies émergentes et stratégiques de 2018

Source : <https://www.silkhom.com/les-35-technologies-a-forts-enjeux-strategiques-selon-gartner/>

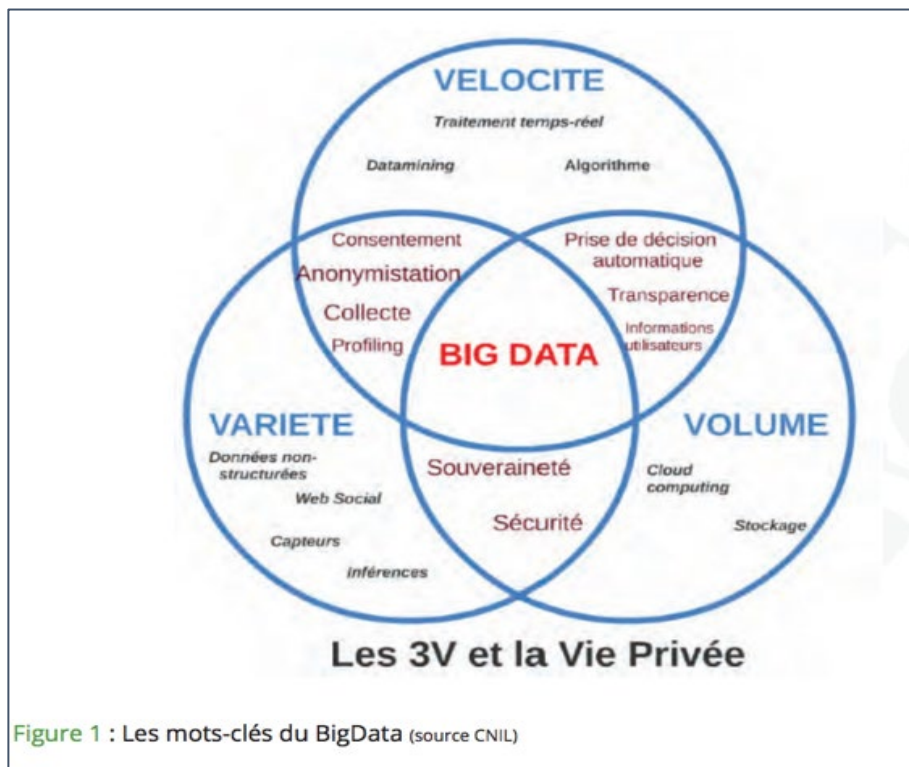
2.2 BIG DATA



L'expression « Big Data » date de 1997 selon l'Association for Computing Machinery. En 2001, l'analyste du cabinet Meta Group (devenu Gartner) Doug Laney décrivait les Big Data d'après le principe des « trois V » :

- Le Volume de données de plus en massif
- La Variété de ces données qui peuvent être brutes, non structurées ou semi-structurées
- La Vélocité qui désigne le fait que ces données sont produites, récoltées et analysées en temps réel.

Certaines entreprises ajoutent un quatrième « V » à cette définition pour la Véracité qui évoque la nécessité de vérifier la crédibilité de la source et la qualité du contenu afin de pouvoir exploiter ces données.



Cependant aucune définition « officielle » n'est aujourd'hui plébiscitée. En voici quelques-unes :

- Définition Wikipedia : « Le big data, littéralement « grosses données », ou mégadonnées (recommandé³), parfois appelées données massives⁴, désigne des ensembles de données devenus si volumineux qu'ils dépassent l'intuition et les capacités humaines d'analyse et même celles des outils informatiques classiques de gestion de base de données ou de l'information⁵. Divers experts, grandes institutions (comme le MIT aux États-Unis, le Collège de France en Europe), administrations et spécialistes sur le terrain des technologies ou des usages considèrent le phénomène big data comme l'un des grands défis informatiques de la décennie 2010-2020 et en ont fait une de leurs nouvelles priorités de recherche et développement, qui pourrait notamment conduire à l'Intelligence artificielle en étant exploré par des réseaux de neurones artificiels autoapprenants. »
- Définition LeBigdata
« Littéralement, ces termes signifient mégadonnées, grosses données ou encore données massives. Ils désignent un ensemble très volumineux de données qu'aucun outil classique de gestion de base de données ou de gestion de l'information ne peut vraiment travailler. En effet, nous procréons environ 2,5 trillions d'octets de données tous les jours. Ce sont les informations provenant de partout : messages que nous nous envoyons, vidéos que nous publions, informations climatiques, signaux GPS et bien d'autres encore. Ces données sont baptisées Big Data ou volumes massifs de données. Les géants du Web, au premier rang desquels Yahoo, Google et Facebook, ont été les tous premiers à déployer ce type de technologie.

Cependant, **aucune définition précise ou universelle ne peut être donnée au Big Data.** Etant un objet complexe polymorphe, sa définition varie selon les communautés qui s'y intéressent en tant qu'utilisateur ou fournisseur de services. Une approche transdisciplinaire permet d'appréhender le comportement des différents acteurs : les concepteurs les catégories d'utilisateurs (gestionnaires, responsables d'entreprises, chercheurs), les acteurs de la santé et les usagers. »

- **NIST** (National Institute of Standards and Technology.) NIST argues that big data is data which “*exceed(s) the capacity or capability of current or conventional methods and systems.*” In other words, the notion of “big” is relative to the current standard of computation.”
- **Futura Tech:** “Les big data ou mégadonnées désignent l'ensemble des données numériques produites par l'utilisation des nouvelles technologies à des fins personnelles ou professionnelles. Cela recoupe les données d'entreprise (courriels, documents, bases de données, historiques de processeurs métiers...) aussi bien que des données issues de capteurs, des contenus publiés sur le web (images, vidéos, sons, textes), des transactions de commerce électronique, des échanges sur les réseaux sociaux, des données transmises par les objets connectés (étiquettes électroniques, compteurs intelligents, smartphones...), des données géolocalisées, etc.

Et ajoute *L'essor des big data a suivi l'évolution des systèmes de stockage et de traitement des données avec notamment l'avènement du cloud computing et des supercalculateurs. Désormais, on parle de pétaoctets et de zettaoctets pour désigner les volumes que représentent les big data. Selon les prévisions du cabinet IDC, le volume de données produites dans le monde attendra les 40 zettaoctet en 2020.*

2.3 MODUS OPERANDI DU RECUEIL ET DU TRAITEMENT DE LA DONNÉE PAR LES ENTREPRISES.

Les données sont devenues une priorité majeure pour les entreprises de toutes tailles. Au fur et à mesure que les technologies de capture et d'analyse des données prolifèrent, les capacités des entreprises à contextualiser les données et à en tirer de nouvelles perspectives se multiplient. L'Internet des objets et l'intelligence artificielle sont deux outils essentiels pour les entreprises dans la capture et l'analyse des données, qu'il s'agisse de mieux comprendre les opérations quotidiennes, de prendre des décisions commerciales ou de mieux connaître leurs clients.

Les données clients sont un domaine d'intérêt à part entière. Du comportement des consommateurs à l'analyse prédictive, les entreprises capturent, stockent et analysent régulièrement de grandes quantités de données sur leur base de consommateurs chaque jour. Certaines entreprises ont même construit tout un modèle d'affaires autour des données des consommateurs, qu'elles créent des publicités ciblées où qu'elles les vendent à un tiers. Les données clients sont un business à part entière.

Voici quelques exemples de la façon dont les entreprises saisissent les données de leurs clients, ce qu'elles en font.

Comment les entreprises collectent-elles nos données ?

Les entreprises saisissent des données de nombreuses façons à partir de nombreuses sources. Certains processus sont de nature très technique, tandis que d'autres sont plus déductifs (bien que ces méthodes utilisent souvent des logiciels sophistiqués).

En fin de compte, cependant, les entreprises utilisent une pléthore de sources pour saisir et traiter les données des clients sur les paramètres, des données démographiques aux données comportementales.

Les données sur les clients peuvent être recueillies de trois façons - en demandant directement aux clients, en suivant indirectement les clients et en ajoutant d'autres sources de données sur les clients aux fichiers déjà existant d'une entreprise.

Une stratégie commerciale solide a besoin des trois.

Les entreprises sont capables d'extraire des données de presque tous les coins et recoins. Les endroits les plus évidents sont l'activité des consommateurs sur leurs sites Web et leurs pages de médias sociaux, mais il y a aussi des méthodes plus intéressantes au travail.

Par exemple, la publicité géo localisée, qui utilise l'adresse IP d'un appareil connecté à Internet (et les autres appareils avec lesquels il interagit) pour créer un profil de données personnalisé. Ces informations sont ensuite utilisées pour cibler les appareils des utilisateurs avec des publicités hyper-personnalisées et pertinentes.

Les entreprises fouilleront également dans leurs propres dossiers de service à la clientèle pour voir comment les clients ont interagi avec leurs services de vente et de soutien dans le passé. Ici, ils intègrent une rétroaction directe sur ce qui a fonctionné et ce qui n'a pas fonctionné, ce qu'un client a aimé et n'a pas aimé, sur une grande échelle.

En plus de recueillir des données, les entreprises peuvent également les acheter ou les vendre à des tiers. Une fois saisies, ces informations changent régulièrement de mains dans un marché de données à part entière.

Transformer les données en connaissances.

La capture de grandes quantités de données pose le problème du tri et de l'analyse de toutes ces données. Aucun humain ne peut raisonnablement s'asseoir et lire ligne après ligne les données des clients toute la journée, et même s'il le pouvait, il ne ferait probablement pas grand-chose. Heureusement, les ordinateurs sont beaucoup mieux adaptés à ce type de travail que les humains, et ils peuvent fonctionner 24 heures sur 24, 7 jours sur 7 et 365 jours par an sans prendre de pause.

Au fur et à mesure que les algorithmes d'apprentissage machine et d'autres formes d'IA prolifèrent et s'améliorent, l'analyse des données devient un domaine encore plus puissant pour décomposer la mer de données en éléments gérables d'informations exploitables. Certains programmes d'IA signaleront les anomalies ou offriront des recommandations aux décideurs d'une organisation en fonction des données contextualisées.

Sans de tels programmes, toute la saisie de données dans le monde serait totalement inutile.

2.4 COMMENT LES ENTREPRISES UTILISENT-ELLES VOS DONNEES ?

Il y a plusieurs façons dont les entreprises utilisent les données sur les consommateurs qu'elles recueillent et les idées qu'elles tirent de ces données :

Améliorer l'expérience client.

Pour de nombreuses entreprises, les données sur les consommateurs offrent un moyen de mieux comprendre les exigences de leurs clients et d'y répondre. En analysant le comportement des clients, ainsi que de nombreux commentaires et réactions, les entreprises peuvent modifier leur présence numérique, leurs biens ou services pour mieux s'adapter au marché actuel.

Non seulement les entreprises utilisent les données des consommateurs pour améliorer l'expérience des consommateurs dans son ensemble, mais elles utilisent également les données pour prendre des décisions sur un plan individuel.

La source la plus importante d'intelligence marketing provient de la compréhension des données des clients et de leur utilisation pour améliorer la fonctionnalité des sites Web. L'expérience client a été améliorée en créant des promotions personnalisées et des offres spéciales basées sur les données des clients. Puisque chaque client aura ses propres préférences, la personnalisation est la clé.

Affiner la stratégie marketing.

Les données contextualisées peuvent aider les entreprises à comprendre comment les consommateurs s'engagent et réagissent à leurs campagnes de marketing, et à s'adapter en conséquence. Ce cas d'utilisation hautement prédictif donne aux entreprises une idée de ce que les consommateurs voudront en fonction de ce qu'ils ont déjà fait. Comme d'autres aspects de l'analyse des données des consommateurs, le marketing est de plus en plus sur la personnalisation en conséquence.

Il est maintenant essentiel de cartographier les trajets des utilisateurs et de personnaliser leur voyage, non seulement par le biais des sites Web, mais aussi sur des plateformes comme YouTube, LinkedIn, Facebook ou tout autre site Web.

La segmentation des données vous permet de ne commercialiser que les personnes que vous savez les plus susceptibles d'intéresser. Cela a ouvert de nouvelles opportunités dans des industries auparavant très difficiles à commercialiser.

Transformer les données en flux de trésorerie.

Les entreprises qui capturent des données peuvent également en tirer profit. Les courtiers en données, ou les entreprises qui achètent et vendent de l'information sur leurs clients, sont devenus une nouvelle industrie à côté des grandes données. Pour les entreprises qui capturent de grandes quantités de données, cela représente une opportunité pour une nouvelle source de revenus.

Pour les annonceurs, le fait d'avoir cette information disponible à l'achat est extrêmement précieuse, de sorte que la demande pour de plus en plus de données ne cesse d'augmenter. Cela signifie que plus les courtiers en données peuvent puiser dans des sources de données disparates pour produire des profils de données plus complets, plus ils peuvent gagner de l'argent en vendant cette information entre eux et aux annonceurs.

2.5 UTILISATION DES DONNEES POUR SECURISER LES DONNEES

Certaines entreprises utilisent même les données des consommateurs pour sécuriser des informations plus sensibles. Par exemple, les institutions bancaires utilisent parfois des données de reconnaissance vocale pour autoriser un utilisateur à accéder à ses renseignements financiers ou pour le protéger contre les tentatives frauduleuses de vol de ses renseignements.

Ces systèmes fonctionnent en mariant les données provenant de l'interaction d'un client avec un centre d'appels et des algorithmes d'apprentissage machine qui peuvent identifier et signaler les tentatives potentiellement frauduleuses d'accès au compte d'un client.

Cela élimine une partie du travail de devinette et de l'erreur humaine ayant pour but d'éviter l'usurpation d'identité ou toutes autres formes d'escroquerie.

À mesure que les technologies de saisie et d'analyse des données deviendront plus sophistiquées, les entreprises trouveront des moyens nouveaux et plus efficaces de recueillir et de contextualiser les données sur tout, y compris les consommateurs. Pour les entreprises, il est essentiel de le faire pour rester compétitives à long terme ; ne pas le faire, c'est comme courir une course les jambes liées. La perspicacité est roi, et la perspicacité dans l'environnement commercial moderne est glanée à partir de données contextualisées.

3 LA SECURISATION DE LA DONNEE.

Aux fins de protéger les données et de réguler leurs accès, il est nécessaire d'apporter des réponses d'ordre **juridique** et **technologique**.

Le rapport de force entre les Etats et les GAFAM a vu une mise à jour de l'arsenal réglementaire.

L'année 2018 a été marquée par les réformes importantes que sont le CLOUD ACT en mars par les Etats-Unis puis par le RGPD en Europe en mai.

Nous allons développer ici les différents textes et mettre en lumière les travaux de recherche et d'investigations démontrant la nécessité d'anonymisation des données.

3.1 REPONSE JURIDIQUE

3.1.1 LE CLOUD ACT



Le « *Clarifying Lawful Overseas Use of Data Act* » ou Cloud Act a été adopté par le Congrès américain le 23 mars 2018. Ce texte a été intégré au Consolidated Appropriations Act de 2018. Le nom de *Clarifying Lawful Overseas Use of Data Act (Cloud Act)* peut être traduit par la « *loi clarifiant l'usage légal des données hébergées à l'étranger* ». Ce texte marque un changement significatif dans l'encadrement légal de l'accès aux données induit par le Cloud Act.

Principalement, cette loi permet au gouvernement américain d'accéder à des données à caractère personnel stockées à l'étranger (paragraphe 2713 du Cloud Act).

Elle marque un changement majeur dans l'encadrement légal de l'accès aux données : Elle permet en premier lieu au gouvernement américain d'accéder à des données à caractère personnel stockées à l'étranger (paragraphe 2713 du Cloud Act).

L'affrontement entre Microsoft et le gouvernement américain

Elle marque un tournant dans l'affaire opposant la société Microsoft aux autorités américaines depuis 2014 (Affaire 14-2985). Pour rappel, l'affaire s'est portée sur la question de l'application du « *Stored Communications Act* ». Ce texte régit la conservation des télécommunications, des communications et des transactions électroniques, et octroie au gouvernement fédéral le pouvoir de soumettre un fournisseur de services de communication à lui partager le contenu d'une communication téléphonique ou électronique.

A l'origine de cette affaire, dans le cadre d'une affaire liée au trafic de drogue, le gouvernement américain a demandé à la société Microsoft de partager les données personnelles d'un compte messagerie utilisateur.

La première réaction de l'entreprise a été le refus, répondant que les datacentres situés en Irlande, les données n'étaient pas stockées sur le territoire national. Le gouvernement a donc compris que Microsoft ne suivait pas la réglementation du *Stored Communication Act*. **L'affaire est en discussion** devant la Cour Suprême des Etats-Unis depuis février 2018.

En réponse directe à cette affaire, le gouvernement a lancé le Cloud Act. En effet, cette loi conduit expressément à l'application de l'extraterritorialité du droit américain. L'objectif de cette loi est de permettre l'application de sa loi en dehors des frontières américaines. En clair, les opérateurs de télécommunication et de communications électroniques pourront être amenés à divulguer des informations que celles-ci « *soient situées dans le territoire des Etats-Unis ou dans le territoire d'un autre pays* ».

Quelle réponse possible pour les opérateurs ?

Les opérateurs de télécommunications auront la possibilité de contester les demandes du gouvernement à conditions de réunir ces deux conditions :

- Si la personne recherchée « n'est pas un citoyen américain, un résident permanent des États-Unis en situation régulière, une association composée majoritairement de citoyens américains ou de résidents permanents ou une corporation américaine » ;
- Si la divulgation de ces informations conditionne un « risque matériel de violation de la législation d'un gouvernement étranger ayant conclu un « *accord exécutif* » avec le gouvernement américain ».

3.1.2 EUROPE ET RGPD



Historiquement, en France, les données personnelles sont protégées par le cadre juridique de la loi n°78-17 du 6 janvier 1978 dite « loi Informatique et Libertés », qui évolue avec l'entrée en vigueur en mai 2018 du règlement européen n°2016/679 du 27 avril 2016 sur la protection des données personnelles (RGPD).

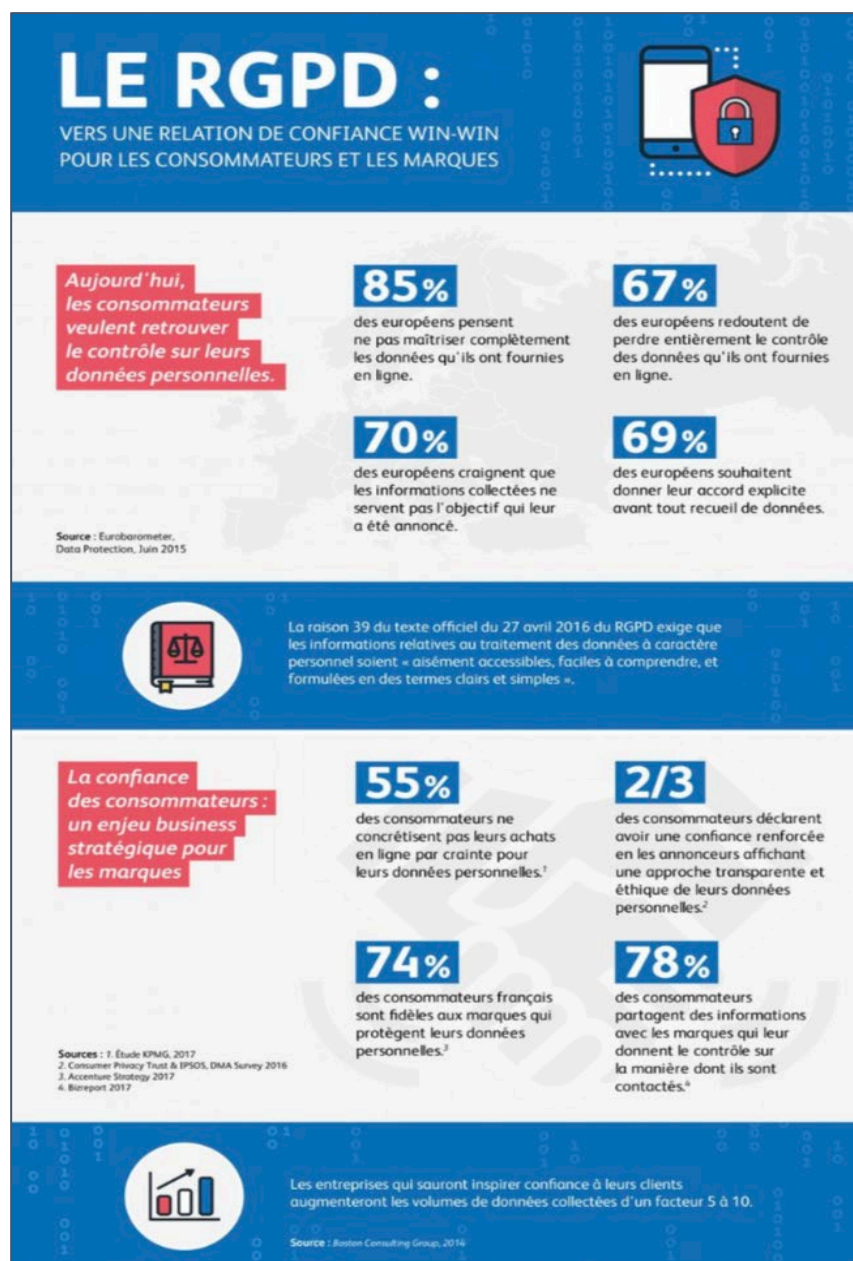
Le RGPD est un texte européen, commun à tous les pays membres de l'Union européenne, qui concerne tous les organismes, tant publics que privés, et tous les secteurs d'activité. Il renforce les droits des personnes et accroît les obligations des responsables de traitement et des sous-traitants. Il s'applique aux traitements de données personnelles, réalisés sur support informatique (logiciels, applications, bases de données, sites web...), mais également sur support papier (Source CNIL).

3.1.2.1 QU'EST-CE QUE LE RGPD ?

L'acronyme RGPD signifie « Règlement Général sur la Protection des Données » (en anglais « General Data Protection Regulation » ou GDPR). Le RGPD encadre le traitement des données personnelles sur le territoire de l'Union européenne.

Le contexte juridique s'adapte pour suivre les évolutions des technologies et de nos sociétés (usages accrus du numérique, développement du commerce en ligne...).

Ce nouveau règlement européen s'inscrit dans la continuité de la Loi française Informatique et Libertés de 1978 et renforce le contrôle par les citoyens de l'utilisation qui peut être faite des données les concernant. Il harmonise les règles en Europe en offrant un cadre juridique unique aux professionnels. Il permet de développer leurs activités numériques au sein de l'UE en se fondant sur la confiance des utilisateurs.



3.1.2.2 QUI EST CONCERNE PAR LE RGPD ?

Tout organisme quels que soient sa taille, son pays d'implantation et son activité, peut être concerné.

En effet, le RGPD s'applique à toute organisation, publique et privée, qui traite des données personnelles pour son compte ou non, dès lors qu'elle est établie sur le territoire de l'Union européenne, ou que son activité cible directement des résidents européens.

Par exemple, une société établie en France, qui exporte l'ensemble de ses produits au Maroc pour ses clients moyen-orientaux doit respecter le RGPD.

De même, une société établie en Chine, proposant un site de e-commerce en français livrant des produits en France doit respecter le RGPD. Le RGPD concerne aussi les sous-traitants qui traitent des données personnelles pour le compte d'autres organismes.

Ainsi, si vous traitez ou collectez des données pour le compte d'une autre entité (entreprise, collectivité, association), vous avez des obligations spécifiques pour garantir la protection des données qui vous sont confiées.

3.1.2.3 DES RECOMMANDATIONS POUR PROTEGER SES DONNEES.

Le rapport RGPD émet un certain nombre de recommandations pour protéger les données des entreprises.



Ceci peut se résumer en 5 étapes (source CNIL). Rappel des bonnes pratiques :

3.1.2.4 CREER UN REGISTRE DES TRAITEMENTS DE DONNEES DE L'ENTREPRISE

Ce document permet de recenser tous vos fichiers et d'avoir une vision d'ensemble. L'objectif est d'identifier les activités principales de l'entreprise qui nécessitent la collecte et le traitement de données. Exemples : le recrutement, la gestion de la paye, formation, les outils de force de ventes etc.

Dans le registre, il faut créer une fiche pour chaque activité recensée, en précisant :

- L'objectif poursuivi (la finalité - exemple : la fidélisation client) ;
- Les catégories de données utilisées (exemple pour la paie : nom, prénom, date de naissance, salaire, etc.) ;
- Qui a accès aux données (le destinataire - exemple : service chargé du recrutement, service informatique, direction, prestataires, partenaires, hébergeurs) ;
- La durée de conservation de ces données (durée durant laquelle les données sont utiles d'un point de vue opérationnel, et durée de conservation en archive).

Le registre est placé sous la responsabilité du dirigeant de l'entreprise.

3.1.2.5 FAIRE LE TRI DANS SES DONNEES

La constitution du registre vous permet de vous interroger sur les données dont votre entreprise a réellement besoin.

Pour chaque fiche de registre créée, vérifiez que :

- Les données que vous traitez sont nécessaires à vos activités (par exemple, il n'est pas utile de savoir si vos salariés ont des enfants, si vous n'offrez aucun service ou rémunération attachée à cette caractéristique) ;
- Ne traiter aucune donnée dite « sensible » ou, si c'est le cas, que vous avez bien le droit de les traiter (voir la fiche « Traitements de données à risque : êtes-vous concerné ? ») ;
- Seules les personnes habilitées ont accès aux données dont elles ont besoin ;
- Ne pas conserver les données au-delà de ce qui est nécessaire.

3.1.2.6 RESPECTEZ LES DROITS DES PERSONNES

Le RGPD renforce l'obligation d'information et de transparence à l'égard des personnes dont vous traitez les données (clients, collaborateurs, etc.). A chaque fois que vous collectez des données personnelles, le support utilisé (formulaire, questionnaire, etc.) doit comporter des mentions d'information.

Vérifiez que l'information comporte les éléments suivants :

- Pourquoi vous collectez les données (« la finalité » ; par exemple pour gérer l'achat en ligne du consommateur) ;
- Ce qui vous autorise à traiter ces données (le « fondement juridique » : il peut s'agir du consentement de la personne concernée, de l'exécution d'un contrat, du respect d'une obligation légale qui s'impose à vous, de votre « intérêt légitime ») ;
- Qui a accès aux données (indiquez des catégories : les services internes compétents, un prestataire, etc.) ;
- Combien de temps vous les conservez (exemple : « 5 ans après la fin de la relation contractuelle ») ;
- Les modalités selon lesquelles les personnes concernées peuvent exercer leurs droits (via leur espace personnel sur votre site internet, par un message sur une adresse email dédiée, par un courrier postal à un service identifié) ;
- Si vous transférez des données hors de l'UE (précisez le pays et l'encadrement juridique qui maintient le niveau de protection des données).

Pour éviter des mentions trop longues au niveau d'un formulaire en ligne, vous pouvez par exemple, donner un premier niveau d'information en fin de formulaire et renvoyer à une **politique de confidentialité / page vie privée** sur votre site internet. **À l'issue de cette étape, vous avez répondu à votre obligation de transparence.**

Les personnes dont vous traitez les données (clients, collaborateurs, prestataires, etc.) ont des droits sur leurs données, qui sont d'ailleurs renforcés par le RGPD : droit d'accès, de rectification, d'opposition, d'effacement, à la portabilité et à la limitation du traitement.

Vous devez leur donner les moyens d'exercer effectivement leurs droits. Si vous disposez d'un site web, prévoyez un formulaire de contact spécifique, un numéro de téléphone ou une adresse de messagerie dédiée. Si vous proposez un compte en ligne, donnez à vos clients la possibilité d'exercer leurs droits à partir de leur compte.

Mettez en place un processus interne permettant de garantir l'identification et le traitement des demandes dans des délais courts (1 mois au maximum).

À l'issue de cette étape, vous serez en capacité de répondre aux demandes des personnes concernées.

3.1.2.7 SECURISER LES DONNEES DE L'ENTREPRISE

Si le risque zéro n'existe pas en informatique, vous devez prendre les mesures nécessaires pour garantir au mieux la sécurité des données. Vous êtes en effet tenu à une obligation légale d'assurer la sécurité des données personnelles que vous détenez.

Vous garantissez ainsi l'intégrité de votre patrimoine de données en minimisant les risques de pertes de données ou de piratage.

Les mesures à prendre, informatiques ou physiques, dépendent de la sensibilité des données que vous traitez et des risques qui pèsent sur les personnes en cas de d'incident.

Des réflexes doivent être mis en place : mises à jour de vos antivirus et logiciels, changement régulier des mots de passe et utilisation de mots de passe complexes, ou chiffrement de vos données dans certaines situations. En cas de perte ou vol d'un outil informatique, il sera plus difficile pour un tiers d'y accéder.

BONNE PRATIQUE

Pour évaluer le niveau de sécurité des données personnelles dans votre entreprise, voici quelques questions à se poser :

- Les comptes utilisateurs internes et externes sont-ils protégés par des mots de passe d'une complexité suffisante ?
- Les accès aux locaux sont-ils sécurisés ?
- Des profils distincts sont-ils créés selon les besoins des utilisateurs pour accéder aux données ?
- Avez-vous mis en place une procédure de sauvegarde et de récupération des données en cas d'incident ?

BONNE PRATIQUE

L'approche assurantielle au-delà du RGPD :

Cette démarche d'anticipation sur le niveau global de sécurité peut être complétée par une approche assurantielle. Renseignez-vous auprès de ces professionnels sur le contenu possible des polices d'assurance (responsabilité civile, dommages couverts...) ET surtout sur les services à l'assuré (notamment l'assistance en cas de sinistre, de gestion de crise...). Signalez à la CNIL les violations de données personnelles

Votre entreprise a subi une violation de données (des données personnelles ont été, de manière accidentelle ou illicite, détruites, perdues, altérées, divulguées ou vous avez constaté un accès non autorisé à des données) ?

Vous devez **la signaler à la CNIL** dans les 72 heures <https://www.cnil.fr/fr/violations-de-donnees-personnelles-1er-bilan-apres-lentree-en-application-du-rgpd> si cette violation est susceptible de représenter un risque pour les droits et libertés des personnes concernées. Cette notification s'effectue en ligne sur le site internet de la CNIL.

Si ces risques sont élevés pour ces personnes, vous devrez les en informer.

À l'issue de cette étape, vous serez en capacité d'assurer une protection des données personnelles en continu et de faire face aux incidents. »

Bilan provisoire



Depuis la mise en vigueur du RGPD, la CNIL a sorti un premier [rapport d'activité officiel en Octobre 2018](#) : On y apprend qu'entre le 25 mai et le 1er octobre, la Cnil s'est vu notifier 742 violations qui dans une immense majorité (695 cas) avaient trait à la confidentialité des données. Second motif de notification de violation, l'atteinte à la disponibilité des données personnelles (71 cas) devance de peu l'atteinte à l'intégrité des données (50 cas). Il est à noter que notification ne correspond pas forcément à violation avérée. En effet, ce processus enclenche une procédure de vérification opérée par la Cnil, qui, ensuite décide de lancer les actions et sanctions.

Ces quelques 742 violations présumées représentent tout de même 33,7 millions de personnes touchées en six mois.

3.1.3 CLOUD ACT ET RGDP : QUELLES CONSEQUENCES ?

Deux puissances, deux réformes pour la protection des données à caractère personnel dans un monde ultra-connecté et bel et bien mondialisé.

Le Règlement général sur la protection des données, entré en application le 25 mai dernier. Il vise à définir un niveau élevé de protection des données à caractère personnel des personnes au sein de l'Union européenne, notamment en garantissant leur sécurité et leur confidentialité.

Le RGPD encadre également de manière stricte les transferts de telles données en dehors de l'Espace économique européen. En effet, ces transferts ne peuvent sauf exception avoir lieu s'ils ont pour conséquence de diminuer le niveau de protection dont dispose la personne concernée initialement.

Il est intéressant de rappeler que la Commission européenne a proposé **le 17 avril** de nouvelles règles visant à permettre aux autorités policières et judiciaires d'accéder plus facilement et rapidement aux preuves électroniques, notamment celles stockées en ligne.

La confidentialité des données a même atteint les États-Unis sous la forme de la California Consumer Privacy Act (CCPA). L'ACFPC est, à certains égards, semblable au GDPR, mais elle diffère en ce sens qu'elle exige que les consommateurs se retirent de la collecte de données et qu'elle désigne l'État comme l'entité chargée d'élaborer les lignes directrices, plutôt que les décideurs internes de l'entreprise.

La réglementation sur la protection des données modifie la façon dont les entreprises saisissent, stockent, partagent et analysent les données des consommateurs. Les entreprises qui ne sont pas encore touchées par la réglementation sur la protection des données peuvent s'attendre à ce que ces types de lois prolifèrent à mesure que de plus en plus de consommateurs réclament le droit à la vie privée. Il est toutefois peu probable que la collecte de données par les entreprises privées disparaisse ; elle changera simplement de forme à mesure que les entreprises s'adapteront aux nouvelles lois et réglementations.

3.1.4 UN RGPD AMERICAIN

3.1.4.1 LE CONTEXTE.

Suite au scandale de Facebook et de Cambridge Analytica, les législateurs européens ont pu justifier et éprouver le Règlement Général de la Protection de la Donnée, outil législatif luttant contre le traitement et la circulation des données à caractère personnel ; Ces informations sur lesquelles les entreprises s'appuient pour proposer des services et des produits.

Les États-Unis n'ignorent pas ce qui se passe en Europe en matière de vie privée. Alors que le RGPD est désormais en place, Washington consulte sur l'opportunité d'avoir une loi du même ordre.

Comme le suggère [le laboratoire d'innovation numérique de la CNIL](#), en pointant diverses initiatives à l'international, la mise en place du RGPD peut conduire d'autres régions du monde à se doter d'un texte équivalent.

« Ainsi, aux Etats-Unis, certains élus de New-York et de Californie semblent vouloir s'inspirer du texte européen. Quant à la Chine, elle planifie un texte à horizon de 5 ans. » (CNIL)

3.1.4.2 LES ACTEURS.

L'identification des protagonistes tant physique que moral nous désigne :

- [Douglas MacMillan](#) est journaliste au bureau technique du Wall Street Journal à San Francisco. Il couvre Google et ses nombreuses ramifications. Auparavant, il a parlé de l'essor des start-ups comme Uber et Airbnb et du déclin de Yahoo, pionnier de l'internet vieillissant. Avant de rejoindre WSJ en 2013, MacMillan a couvert la technologie pour Bloomberg News et Bloomberg Businessweek. Il a obtenu la bourse Knight-Bagehot de l'Université Columbia en économie et en journalisme d'affaires en 2017.

- Le [Wall Street Journal](#) est un [quotidien](#) national [américain](#) détenu par News Corp, qui traite de l'actualité [économique](#) et [financière](#), créé à [New York](#) par [Dow Jones and Company](#). Il est le quotidien économique et financier le plus vendu au monde et sa diffusion aux États-Unis a désormais dépassé celle du quotidien généraliste [USA Today](#) depuis octobre 2009.

- [News corp](#), entreprise américaine de [média](#) issue de la scission de [News Corporation](#) le 28 juin 2013. Elle regroupe les anciennes activités de presse, alors que l'autre entité résultante de la fusion, 21st Century Fox, est spécialisée dans la télévision et le cinéma. News Corp possède plusieurs actifs importants : [Dow Jones & Company](#), [HarperCollins](#), [News International](#), [New York Post](#).

- [21st Century Fox](#), entreprise américaine de [média](#) issue de la scission de [News Corporation](#), faite le 28 juin 2013. Elle regroupe les anciennes activités de cinéma et de télévision de [News Corporation](#), c'est-à-dire [Fox Entertainment Group](#) qui comprend [20th Century Fox](#), ainsi que [STAR TV](#), 39,14 % de [British Sky Broadcasting](#) et de [Sky Italia](#), et 54,5 % de [Sky Deutschland](#).

Pour rappel, la Fox Entertainment Group est conservatrice, donc « Pro Trump ».

Catherine Gouësset, a rédigé cet [article](#) dans l'Express « Avec Trump, Fox News a enfin trouvé son président idéal » et précise que « La chaîne ultra-conservatrice Fox News a beaucoup contribué à la victoire de Donald Trump à la présidentielle américaine. »

- [Abigail Slater](#), avocate générale auprès de l'Internet Association, et assistante spéciale du Président TRUMP pour les technologies, les télécommunications et la politique de cybersécurité au sein du Conseil économique national de la Maison-Blanche.

Dans ses fonctions actuelles, Mme Slater représente les membres de l'Intelligence Artificielle - y compris Google et Facebook - sur des questions de politique publique devant les tribunaux et les organismes de réglementation.

Elle s'est jointe à l'I.A après avoir travaillé à la Federal Trade Commission, où elle a été avocate chargée de l'application de la loi et conseillère juridique du commissaire.

- La [Business Roundtable](#) est un [lobby](#) conservateur des dirigeants des grandes entreprises américaines créé en 1972 par John Harper pour faire pression sur les [politiques publiques](#) du [Gouvernement fédéral des États-Unis](#).

- [Dean Garfield](#), président et chef de la direction d'ITI depuis 2009. En 2015, Dean Garfield a été nommé meilleur lobbyiste par The Hill pour son leadership au sein d'ITI et pour son travail avec les entreprises et les intervenants afin de faire de la technologie une industrie plus diversifiée et inclusive.

- [Information Technology Industry Council \(ITI\)](#) ou Conseil de l'industrie des technologies de l'information est le principal organisme de défense des intérêts et d'élaboration de politiques des grandes entreprises d'innovation du monde. Nous avons gagné la confiance des marques technologiques les plus reconnues dans le monde pour résoudre leurs problèmes politiques les plus complexes.

3.1.4.3 CE DERNIER SEMESTRE.

De façon anti chronologique :

Le 26 septembre 2018, audition programmée de Google par le auprès du Senate Commerce Committee. Les entreprises Apple, Twitter et AT&T ont également été convoquées.

Le 3 juillet, l'article du WSJ de Douglas McMillan qui met à mal Google.

Le 20 juin 2018, le site américain d'informations spécialisé Axios publie un [article](#) « La Maison-Blanche cherche à coordonner le plan de protection de la vie privée en ligne ».

Jeudi 14 juin 2018, [Abigail Slater](#) a rencontré les membres du comité technologique de la [Business Roundtable](#) pour discuter des options possibles, selon deux personnes ayant connaissance de la réunion.

Par ailleurs, Abigail Slater a parlé de la mise en œuvre du RGPD et de la protection de la vie privée avec M. Dean Garfield, PDG du Conseil de l'industrie des technologies de l'information (ITI), l'association professionnelle représentant les entreprises de technologie, dont Apple, Google, Facebook, IBM, Microsoft et d'autres.

Pourquoi c'est important : Les conversations préliminaires montrent que la Maison-Blanche veut avoir voix au chapitre dans le débat national et mondial sur la façon de protéger la vie privée des consommateurs en ligne. Jusqu'à présent, la stricte réglementation européenne sur la protection des données a fixé les termes du débat, ce qui a alarmé les entreprises et certains organismes de réglementation aux États-Unis.

Ce qui se passe : Abigail Slater, assistante spéciale du président Trump for tech, telecom et cyber-policy au Conseil économique national de la Maison-Blanche, a rencontré des groupes industriels pour discuter des moyens possibles de mettre en place des garde-fous pour l'utilisation des données personnelles, selon de multiples sources familières avec le sujet.

Ce qu'il en ressort : Des sources ont souligné plusieurs résultats possibles des conversations que la Maison-Blanche a eues.

L'une des options est un décret exécutif ordonnant à un ou plusieurs organismes d'élaborer un cadre de protection de la vie privée. Cela pourrait inciter l'Institut national des normes et de la technologie, une branche du ministère du Commerce, à travailler avec l'industrie et d'autres experts pour élaborer des lignes directrices, selon deux sources. (Le NIST a refusé de commenter.)

Le processus serait probablement similaire à celui que l'administration Obama a lancé sur la cybersécurité en 2013. Le Congrès a par la suite appuyé la poursuite des travaux sur les normes, et le président Trump a exigé l'utilisation du cadre pour aider à gérer le risque cybernétique des organismes gouvernementaux.

Un décret exécutif pourrait également donner le coup d'envoi d'un partenariat public-privé pour établir des pratiques exemplaires volontaires en matière de protection de la vie privée, qui pourraient devenir des normes de facto, selon les sources.

Cependant, il reste des gaps tels que : comment un décret exécutif affecterait la Federal Trade Commission, qui est un organisme indépendant chargé de surveiller les questions de protection de la vie privée dans le secteur privé.

La Maison-Blanche n'a pas répondu aux demandes de commentaires.

Les pressions s'intensifient sur les organismes de réglementation et les législateurs pour qu'ils prennent des mesures en matière de protection de la vie privée aux États-Unis.

Le RGPD de l'Europe a fixé un seuil élevé pour les entreprises qui traitent les données des utilisateurs, car l'UE peut imposer de lourdes amendes aux entreprises qui ne s'y conforment pas. Slater a indiqué que l'administration américaine est intéressée à développer un contrepoids au RGPD pour qu'il ne s'agisse pas de la norme mondiale de facto, selon une source lors de la réunion du Business Roundtable.

En Californie, les électeurs évaluent un régime strict de protection de la vie privée en ligne - la California Consumer Privacy Act - dans le cadre d'une initiative de vote de novembre 2018. On craint que les lois des États ne créent une mosaïque confuse de règles de protection de la vie privée.

Cet automne, les organismes de réglementation européens et américains devront revoir l'accord sur le bouclier de protection de la vie privée, qui régit la façon dont les entreprises américaines traitent les données des citoyens européens. Des sources de l'industrie affirment qu'il existe un désir de faire passer la protection de la vie privée avant cet examen.

Le scandale de Cambridge Analytica sur Facebook a sensibilisé les consommateurs et les législateurs aux questions de protection de la vie privée.

"Nous discutons de ce que l'administration pourrait et devrait faire en matière de protection de la vie privée ", a déclaré Mme Slater lors d'une conférence organisée le mois dernier par la National Venture Capital Association, mais il a ajouté qu'il n'y a aucun désir de créer un clone américain des règles européennes.

Mme Slater précise qu'elle pense que "donner aux consommateurs plus de contrôle sur leurs données" et "plus d'accès à leurs données" sont de bons principes contenus dans le RGPD. Mais elle a aussi soulevé le fait que certaines parties, comme le " droit d'être oublié ", peuvent ne pas fonctionner en vertu de la loi américaine.

Elle a souligné que les règles sont lourdes pour les petites entreprises et créent de l'incertitude pour les entreprises qui tentent de se conformer - une préoccupation soulevée par le secrétaire au Commerce, Wilbur Ross, dans un récent éditorial du Financial Times.

Les données sur les consommateurs sont la monnaie la plus précieuse dans l'économie numérique en pleine croissance. Bien que les entreprises de technologie s'inquiètent de la perspective de nouvelles règles en matière de protection de la vie privée, bon nombre d'entre elles veulent aussi avoir une certitude quant aux données des utilisateurs afin de se protéger contre les amendes et les questions juridiques.

3.1.5 DONNEES PERSONNELLES VERSUS DONNEES IDENTIFIABLES, L'ENJEU DE L'ANONYMISATION



La [directive 95/46/CE](#) mentionne l'anonymisation pour exclure les données anonymisées du champ d'application de la législation sur la protection des données:

«considérant que les principes de la protection doivent s'appliquer à toute information concernant une personne identifiée ou identifiable; que, pour déterminer si une personne est identifiable, il convient de considérer l'ensemble des moyens susceptibles d'être raisonnablement mis en œuvre, soit par le responsable du traitement, soit par une autre personne, pour identifier ladite personne; que les principes de la protection ne s'appliquent pas aux données rendues anonymes d'une manière telle que la personne concernée n'est plus identifiable; que les codes de conduite au sens de l'article 27 peuvent être un instrument utile pour fournir des indications sur les moyens par lesquels les données peuvent être rendues anonymes et conservées sous une forme ne permettant plus l'identification de la personne concernée».

L'objectif de l'anonymisation de détruire toute possibilité de pouvoir identifier à quel individu appartiennent les données personnelles trouvées. La méthode consiste à modifier le contenu ou la structure de la donnée visée, afin de rendre celle-ci illisible ou non identifiable après traitement. Cependant, ce procédé reste compliqué à mettre en œuvre dans un contexte de big data : plus le volume de données est important, plus le risque de recoupement est élevé. Par l'analyse des comportements – habitudes de navigation sur internet, achats en ligne, activités et préférences sur les réseaux sociaux - il devient alors possible d'identifier la personne.

La CNIL rappelle que *« pour qu'une solution d'anonymisation soit efficace, elle doit empêcher toutes les parties d'isoler un individu dans un ensemble de données, de relier entre eux deux enregistrements dans un ensemble de données (ou dans deux ensembles de données séparés) et de déduire des informations de cet ensemble de données »*. Implémenter un processus

d'anonymisation complet de toutes les données d'une entreprise devient alors complexe à mettre en œuvre. Beaucoup ont alors recours à la seconde option proposée par le RGPD : la **pseudonymisation**.

- La réglementation RGPD introduit un nouveau concept de protection des données à échelle européenne - article 4 du RGPD :

« Pseudonymisation », le traitement de données à caractère personnel de telle façon que celles-ci ne puissent plus être attribuées à une personne concernée précise sans avoir recours à des informations supplémentaires, pour autant que ces informations supplémentaires soient conservées séparément et soumises à des mesures techniques et organisationnelles afin de garantir que les données à caractère personnel ne sont pas attribuées à une personne physique identifiée ou identifiable ;

La pseudonymisation consiste à séparer les données de leurs propriétaires respectifs pour que tout lien avec une identité ne soit possible sans une information supplémentaire. En clair, il s'agit d'une technique d'amélioration de la vie privée : les données d'identification directes sont conservées séparément et en toute sécurité à partir des données traitées, afin de garantir la non-attribution.

Dans ce contexte, les données ne sont pas totalement anonymes ni totalement identifiables pour autant. Le point faible majeur de cette option est la clef d'identification. En effet, pour assurer la sécurité des données, une clef d'identification est générée pour établir un lien entre les personnes et ses informations. Mal protégée, cette clef permet à un attaquant de recomposer le puzzle et rattacher les personnes à leurs informations. Si le recours au chiffrement de données sensibles est nécessaire, l'entreprise doit mettre en place des mécanismes de sécurisation adéquats pour limiter le risque lié au vol de données sensibles.

3.1.6 ETUDE DE CAS

L'ouvrage L'Empire des données d'Adrien Basdevant et Jean-Pierre Mignard explique que chaque humain doté d'objets munis de capteurs, chaque robot émettent de manière passive ou active, de façon intentionnelle ou involontaire, une quantité incroyable d'informations sur leurs caractéristiques objectives : identité civile, taille, poids, casier judiciaire, solde du compte bancaire et subjectives : orientation sexuelle, préférences, opinions politiques, nombre d'amis.

Ces données sont individuelles dans la mesure où elles sont propres à chaque homme.

Elles peuvent être anonymes ou personnelles, gratuites ou payantes, privées ou publiques en fonction des modalités d'accès. Elles cessent d'être considérées comme anonymes et deviennent personnelles lorsqu'elles sont liées à un élément – adresse postale, adresse mail, numéro de téléphone, numéro de permis de conduire, numéro bancaire permettant l'identification directe ou indirecte de la personne concernée.

Ainsi, l'article 4-1 du nouveau Règlement Européen sur les données personnelles définit les données à caractère personnel comme « Toute information se rapportant à une personne physique identifiée ou identifiable (ci-après dénommée « personne concernée ») ; est réputée être une « personne physique identifiable » une personne physique qui peut être identifiée, directement ou indirectement, notamment par référence à un identifiant, tel qu'un nom, un numéro d'identification, des données de localisation, un identifiant en ligne ou à un ou

plusieurs éléments spécifiques propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale. »

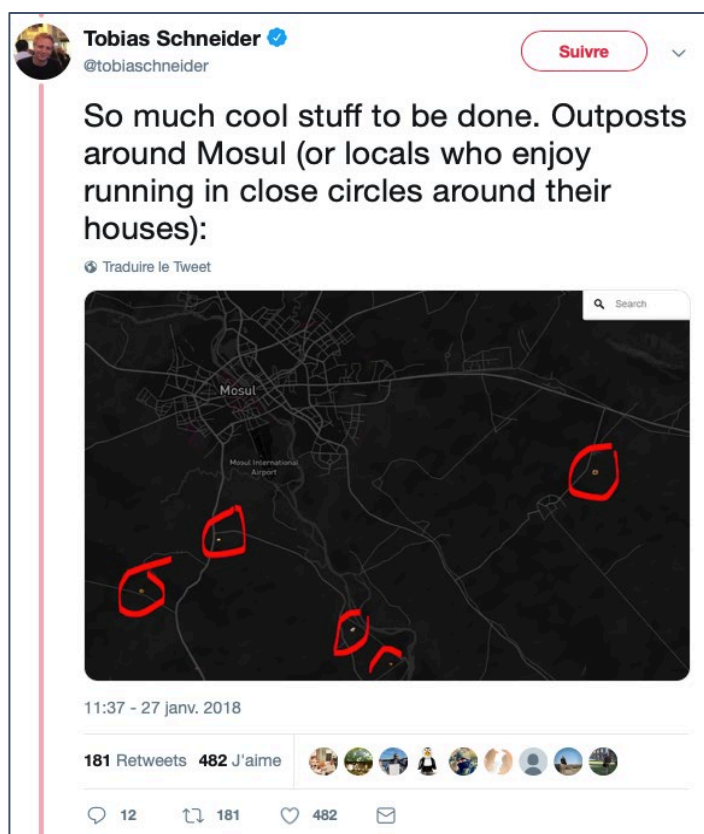
Pour rappel et pour exemples, les récents scandales de Facebook, de Cambridge Analytica concernant la campagne d'élection américaine de Donald Trump en sont le parfait exemple, mais aussi le cas de l'application de running Strava relatée comme suit :

“Fin janvier, un chercheur australien avait ainsi mis au jour la position de certaines bases militaires secrètes de l'armée américaine en Afghanistan, en Irak et en Syrie en repérant des tracés lumineux témoignant d'une activité inhabituelle dans des zones où aucune activité sportive n'était recensée par l'application. L'observation des tracés montrait également une forte activité autour de la base française de Madama, au Niger. La carte thermique laisse aussi apparaître les pérégrinations des agents du renseignement extérieur autour du siège de la DGSE, au 141 boulevard Mortier, dans le XXe arrondissement de Paris.



La carte thermique de l'application Strava montre des allées et venues d'utilisateurs au siège de la DGSE

Crédits : Strava | Date : 21/02/2018



Un agent localisé en planque en Irak

Plus inquiétant, *Le Canard Enchaîné* explique que l'application enregistre un certain nombre de données relatives aux parcours qui permettent de remonter assez facilement à l'état civil des agents. En passant au crible les itinéraires à proximité du siège de la DGSE, il est possible de repérer "des joggeurs qui, le midi, quittaient et regagnaient en petites foulées des sites appartenant à la DGSE", écrit l'hebdomadaire satirique.

Inscrits sous pseudos, ils ne se contentent pas d'archiver leurs footings quotidiens. Leurs performances lors de compétitions officielles comme le marathon de Paris sont également affichées. Il suffit alors de les recouper avec les résultats accessibles sur Internet pour connaître le nom, l'âge et parfois la photo de certains espions français.

Le Canard Enchaîné cite le cas d'un agent qui "a pris comme couverture le nom d'un personnage de dessin animé et qui, depuis deux ans, met malgré lui sa vie en vitrine". "L'appli affiche ses parcours autour du domicile de sa mère, lors des réveillons 2015, 2016 et 2017. Elle détaille aussi ses petites foulées encerclant le fort de Noisy-le-Sec (Seine-Saint-Denis), siège du Service action de la DGSE". L'application a aussi permis de le localiser alors qu'il était en planque en Irak.

D'autres agents ont été repérés sur l'une des plus importantes stations d'écoutes des services du renseignement extérieur. Les recherches ont également permis de localiser une station d'écoute française en Afrique dont l'emplacement est classé top secret. La DGSE n'a pas commenté ces révélations. Fin-janvier, le ministère des Armées avait mis en garde ses troupes contre les dangers représentés par les applications de running et rappelé en interne la nécessité de respecter les règles élémentaires de sécurité.

Pour appuyer nos propos, une étude du prestigieux Massachusetts Institute of Technology (MIT) publiée par Larry Hardesty le 29 janvier 2015 démontre que les chercheurs du MIT signalent que seulement quatre éléments d'information tels que les dates et lieux de quatre achats sont suffisants pour identifier 90 % des personnes d'un ensemble de données enregistrant trois mois de transactions par carte de crédit par 1,1 million d'utilisateurs. Les chercheurs ont également tenu compte des données sur les prix des achats, seulement trois points de données étaient suffisants pour identifier un pourcentage encore plus élevé de personnes dans l'ensemble de données. Cela signifie qu'une personne avec des copies de seulement trois de vos reçus récents - ou un reçu, une photo Instagram où vous prenez un café avec des amis et un tweet sur le téléphone que vous venez d'acheter - aurait 94 % de chances d'extraire vos relevés de carte de crédit de ceux d'un million d'autres personnes. C'est vrai, disent les chercheurs, même dans les cas où personne dans l'ensemble de données n'est identifié par son nom, son adresse, son numéro de carte de crédit ou tout autre renseignement que nous considérons habituellement comme des renseignements personnels.

L'étude vient environ deux ans après une analyse antérieure des enregistrements des téléphones mobiles qui avait donné des résultats très similaires. (<http://news.mit.edu/2013/how-hard-it-de-anonymize-cellphone-data>)

"Si nous le montrons à l'aide de quelques ensembles de données, il est plus probable que ce soit vrai en général", affirme Yves-Alexandre de Montjoye, étudiant diplômé du MIT en Média Lab et en sciences qui est le premier auteur des deux articles. "Honnêtement, je pourrais imaginer les raisons pour lesquelles les métadonnées des cartes de crédit seraient différentes ou équivalentes aux données de mobilité."

De Montjoye est accompagné de son conseiller, Alex "Sandy" Pentland, titulaire de la chaire Toshiba en Média Lab et en sciences ; Vivek Singh, ancien postdoctorant du groupe de Pentland qui est maintenant professeur adjoint à l'Université Rutgers ; et Laura Radaelli, postdoctorante de l'Université Tel Aviv.

L'ensemble de données que les chercheurs ont analysé comprenait le nom et l'emplacement des magasins où les achats ont eu lieu, les jours où ils ont eu lieu et les montants des achats. Les achats effectués avec la même carte de crédit ont tous été étiquetés avec le même numéro d'identification aléatoire.

Pour chaque numéro d'identification - chaque client de l'ensemble de données - les chercheurs ont choisi les achats au hasard, puis ont déterminé combien d'autres clients avaient les mêmes points de données dans leur historique des achats. Dans des analyses distinctes, les chercheurs ont fait varier le nombre de points de données par client de deux à cinq. Sans information sur les prix, deux points de données étaient encore suffisants pour identifier plus de 40 % des personnes de l'ensemble de données.

L'autre extrême, cinq points d'information sur les prix étaient suffisants pour identifier presque tout le monde. Les chercheurs ont caractérisé les prix de façon très grossière, traitant tous les prix qui se situaient dans quelques fourchettes fixes comme étant équivalents sur le plan fonctionnel. Ainsi, par exemple, un achat de 20 \$ dans un magasin à un moment donné de l'histoire d'une personne compterait comme un achat de 40 \$ par quelqu'un d'autre dans le même magasin le même jour, puisque les deux achats se situaient entre 16 \$ et 49 \$, ce qui constituait une tentative pour représenter l'incertitude d'une estimation du montant des

achats à partir de renseignements secondaires, comme une photo Instagram des aliments dans l'assiette. Les limites de chaque fourchette étaient fondées sur un pourcentage fixe de sa valeur médiane : La fourchette de 16 \$ à 49 \$, par exemple, est la valeur médiane des achats (32,50 \$) plus ou moins 50 %, arrondie au dollar près.

La préservation de l'anonymat dans les grands ensembles de données est une préoccupation urgente parce que les entités publiques et privées considèrent les données numériques agrégées comme une source d'idées nouvelles. Les détaillants qui étudient l'historique anonyme des cartes de crédit pourraient certainement apprendre quelque chose sur les goûts de leurs clients, mais les économistes pourraient aussi apprendre quelque chose sur la relation, par exemple, entre l'inflation ou les dépenses de consommation et d'autres facteurs économiques.

Les chercheurs du MIT ont donc également examiné les effets d'une vulgarisation des données - en les rendant intentionnellement moins précises, dans l'espoir de préserver la vie privée tout en permettant une analyse utile. Cela rend l'identification des individus plus difficile, mais pas à un rythme très encourageant.

Même si l'ensemble de données caractérisait chaque achat comme ayant eu lieu au cours d'une semaine dans l'un des 150 magasins situés dans les mêmes secteurs généraux, quatre achats (avec 50 % d'incertitude quant au prix) seraient suffisants pour identifier plus de 70 % des utilisateurs.

Néanmoins, de Montjoye et Pentland demeurent inflexibles quant à la nécessité de poursuivre l'utilisation socialement bénéfique des grandes données. "Sandy et moi croyons vraiment que ces données ont un grand potentiel et qu'elles devraient être utilisées ", dit de Montjoye. "Cependant, nous devons être conscients et tenir compte des risques de ré-identification."

De Montjoye, Pentland et d'autres membres du groupe Pentland ont commencé à développer un système (<http://news.mit.edu/2014/own-your-own-data-0709>) qui permettrait aux gens de stocker les données générées par leurs appareils mobiles sur des serveurs sécurisés de leur propre choix.

Les chercheurs à la recherche de tendances utiles dans les données agrégées enverraient des requêtes par l'entremise du système, qui ne retournerait que les données pertinentes - comme, par exemple, la quantité moyenne dépensée en essence pendant différentes périodes de temps.

3.2 TECHNOLOGIQUE

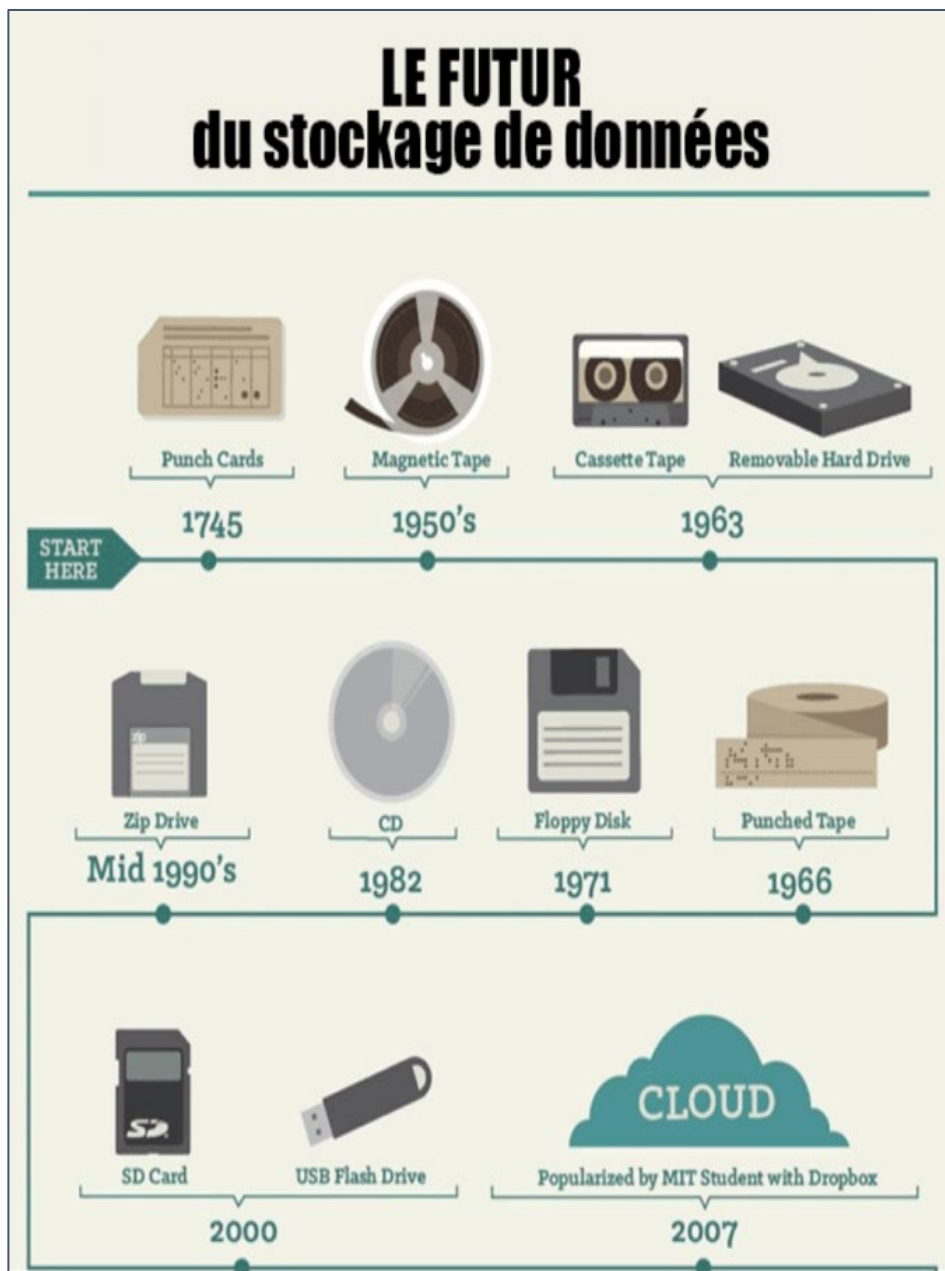
A l'heure où les textes de loi se concentrent sur les données à caractère personnel laissant un flou, un vide juridique sur les données identifiantes comme les chercheurs du MIT ont pu le démontrer ces dernières années (cf supra).

La digitalisation ou transformation digitale des entreprises, des institutions et des Etats apporte son lot d'innovation, voire de disruption contraignant les utilisateurs à adapter leur fonctionnement, leur économie. Cela implique une refonte de l'architecture informatique. Créant ainsi de nouveaux métiers mais aussi un déplacement de l'hébergement du stockage des données.

3.2.1 L'ÉVOLUTION DU STOCKAGE

En 1956, Le premier disque dur IBM RAMAC-305 dont la capacité est 5 Mo coûtait 50,000 dollars. En 2017, une clé USB 8Go se vend 5,9 euros en moyenne. Notre civilisation a produit plus d'informations durant ces trente dernières années que pendant les cinq mille ans qui les ont précédées. L'humanité produit actuellement plus de données tous les ans que depuis le début de sa création.

Le stockage a ainsi suivi l'évolution de la puissance de calcul des ordinateurs pour absorber une quantité grandissante d'informations. Mémoires flash ou encore stockage perpendiculaire permettent de stocker désormais des volumes considérables sur une petite surface.

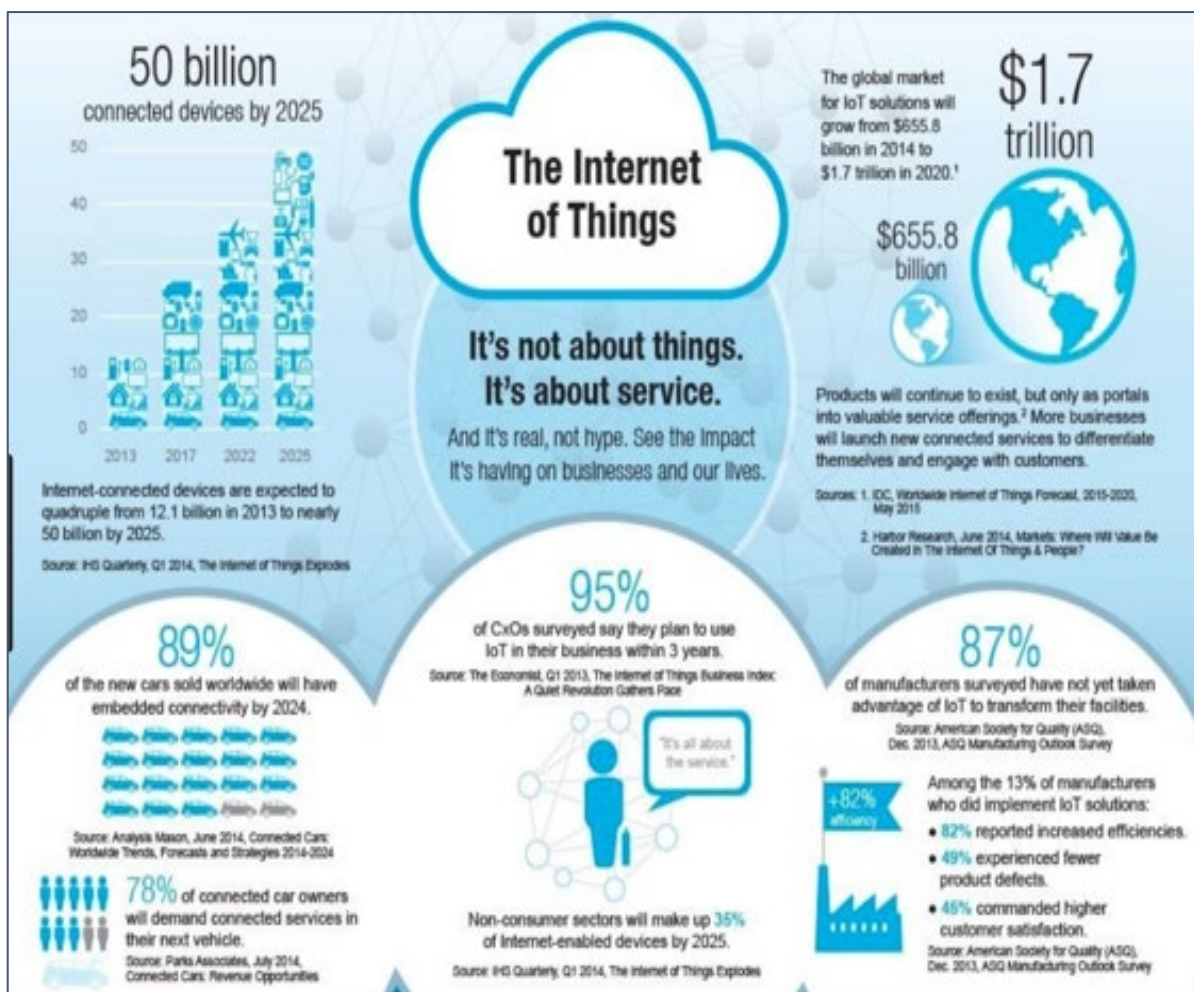


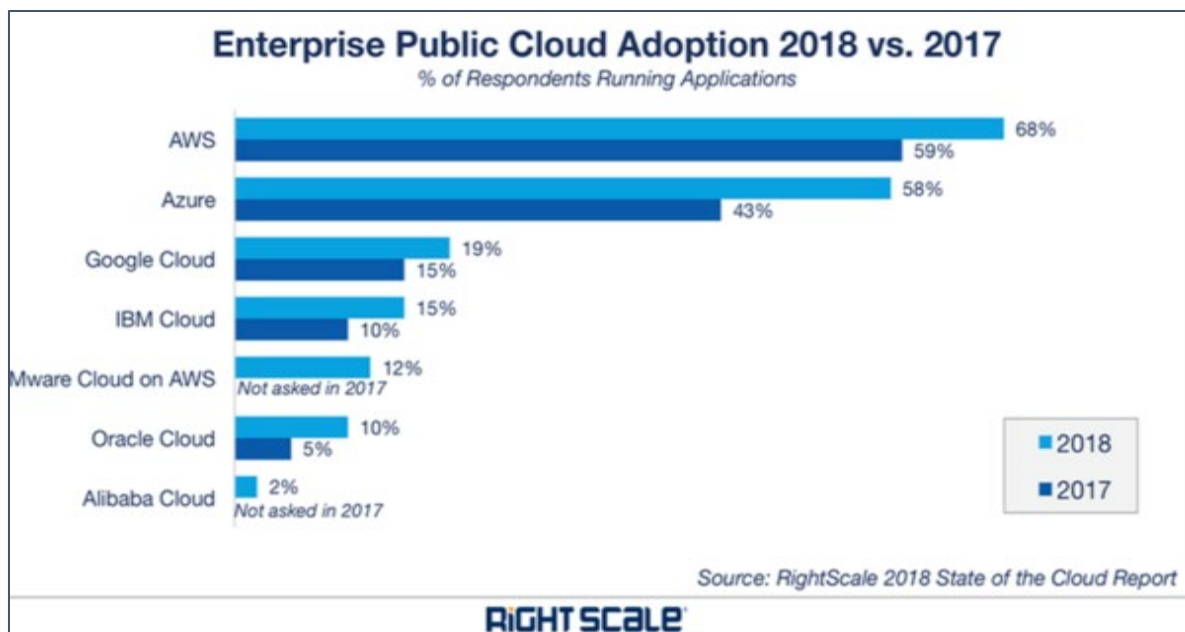
On distingue plusieurs générations de stockage :

- 1ère génération : les supports physiques : la carte perforée et le ruban perforé.
- 2ème génération : les supports magnétiques : la bande magnétique, sa petite sœur la cassette, le disque dur et la disquette.
- 3ème génération : les supports optiques : le disque compact (CD, CD-R ou CD-RW), le DVD (DVD-Rom, DVD-RW) ou le Blu-ray.
- 4ème génération : les supports amovibles la clé USB, la carte SD ou microSD.
- De nos jours : Cloud (Google Drive, iCloud, etc), RAID, NAS, SAN, ...

Le Cloud Storage, inspiré par “Cloud Computing” a été inventé dans les années 60 et devient de plus en plus populaire dans la vie professionnelle. Il permet le déploiement d'environnements de développements ou d'applications ainsi que la possibilité de stocker des données accessibles ou que l'on soit. Pour l'utilisateur, les avantages du stockage de données sur le cloud sont multiples : pas d'investissement matériel et logiciel initial, aucune compétence technique particulière requise, un déploiement facile, une gestion simplifiée, des SLA exigeants et un service continu, un paiement à la demande.

L'IOT et l'explosion du marché du Cloud Computing





Avec l'explosion des données et l'Internet of Everything, le Cloud est devenu le modèle économique privilégié pour les entreprises de toutes tailles au niveau mondial.

Dans son rapport Cloud Insights 2018, réalisé auprès de 550 entreprises de tailles diverses, IDG note une large adoption de l'informatique dans le nuage : 73% des responsables interrogés affirment héberger au moins une application dans le cloud ou une portion de leur infrastructure et 17% d'entreprises prévoient de passer au cloud d'ici début 2019.

Une adoption massive accompagnée de budgets en progression puisque la part allouée au cloud dépasse désormais 30% du budget IT total. Des investissements principalement placés dans le SaaS (Software as a Service) qui représente 48% du budget cloud contre 30% pour le IaaS (Infrastructure as a Service) et 21% pour le PaaS (Platform as a Service) qui, porté par le Serverless Computing, bénéficie cependant de la plus forte progression.

3.2.2 QUI EST CONCERNE ?

Cette explosion technologique concerne tous les pays, petits ou grands avec une forte croissance attendue en Asie Pacifique.

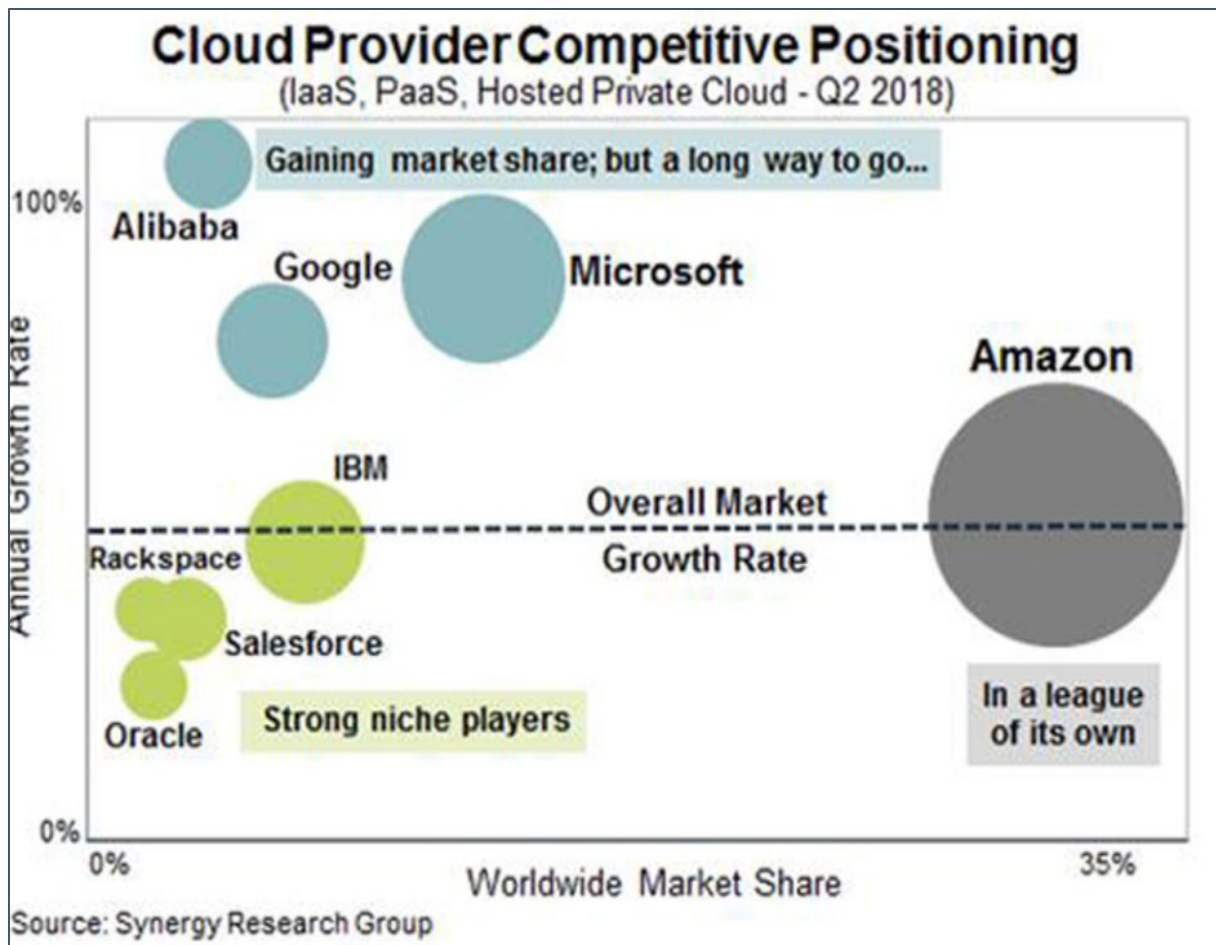


La Chine bien sûr : Elle représente la plus forte croissance et terre d'investissements en nouvelles technologies du globe, devant les États-Unis
L'Inde : Avec son programme gouvernement de transformation digitale, l'Inde s'ouvre elle aussi au cloud public. Les chiffres du [rapport de Janvier 2018](#) de Livemint affichent une croissance supérieure à 50% pour l'adoption des entreprises indiennes au cloud public entre 2018 et 2020.

Cette étude monte l'évolution du marché du Cloud « as a service » d'ici 2022. Le marché est en pleine croissance et les objets connectés en sont les premiers catalyseurs.

3.2.3 LES ACTEURS

Parmi les grands acteurs du marché, on retrouve principalement les GAFAM américains à savoir Amazon Web Services, Google, Microsoft, IBM, mais aussi les BATX chinois qui font leur apparition dans le peloton de tête, comme Alibaba



3.2.4 POUR QUELLES RAISONS PASSER AU CLOUD ?

Ce rapport montre que les entreprises n'adoptent désormais plus le cloud en priorité pour des questions de TCO (Total cost of ownership) ou de migration du Capex vers l'Opex de leurs dépenses informatiques. Ces critères sont désormais secondaires. Les raisons principales de l'adoption du cloud sont :

- La mise à disposition des services IT (les fameux SLA) pour 71% des réponses reçues
- Une plus grande flexibilité permettant une meilleure réactivité face aux évolutions du marché (à 63%)
- Le besoin d'assurer une continuité de services (à 58%)
- La pression des dirigeants (38%) : le rapport fait part de la pression subie par les DSI de la part de leurs cadres dirigeants pour migrer 100% des services IT dans le Cloud et ainsi offrir plus de résilience, d'agilité et de performance à leurs clients internes (branches métiers) et externes (consommateurs).

Enfin il est à noter que 2018 marque l'arrivée d'un modèle hybride, dit multicloud. Des besoins d'indépendance, de flexibilité, d'optimisation de coûts et de sécurité, les entreprises envisagent un modèle hétérogène, où leurs données sont distribuées entre plusieurs fournisseurs de cloud publics mais également privé pour leurs données sensibles.

Cette donnée accroît significativement la complexité des environnements de travail et la difficulté de sécurisation de l'ensemble de ces données.

3.3 LE STOCKAGE, UN ENJEU STRATEGIQUE SOUS-ESTIME POUR LA SECURISATION DE DONNEES

83% des entreprises stockent des données sensibles dans un cloud public

(source LeMondelInformatique, 24 Avril 218 sur la base d'une étude de Bitglass et McAfee)

- La première limite affichée du modèle est **la sécurisation des données**.

Si la donnée est le nouvel or noir, la conserver comme l'asset principal de l'entreprise, la protéger contre toute éventuelle intrusion ou vol, devient clé.

Du fait de leur constitution physique, presque tous les supports de stockage ont une durée de vie limitée, entraînant un risque de perte d'information. La duplication des données es donc utilisée pour assurer l'intégrité et la disponibilité de ces données.

Le tout est de ne pas tomber « dans le piège » comme dirait Richard Stallman (fondateur de la « Free Software Foundation » et créateur du projet GNU).

En effet, comment garantir la sécurité des données stockées dans le Cloud ? Quels sont les risques d'intrusion, de perte ou de dégradation de données ? Le risque est-il largement supérieur au stockage de données en local dans un datacentre dur à maintenir, vieillissant ? Tout spécialiste en cybersécurité vous répondra de la même façon : Qu'elle que soit votre infrastructure et vos choix technologiques, la faille principale restera toujours l'humain « Problem Exists Between Keyboard and Chair (PEBKAC) ». Aucune solution technologique ne peut s'affranchir de décision humaine pour le moment.



Ces données sont transmises de façon chiffrée et sécurisée, elles ne sont pas à la disposition de tout le monde, il y a un risque infime pour que celles-ci soient piratées. Il faut noter également que la plupart de nos boites mails aujourd'hui fonctionnement déjà sur ce principe.

3.3.1 LA QUESTION DE LA RESPONSABILITE DES DONNEES RESTE FLOUE

Le cloud est tellement paré de toutes les vertus que rappeler les réticences justifiées des uns ou des autres n'est jamais inutile. Selon McAfee, 97% des entreprises utilisent des services cloud, publics ou privés et 59% utilisant un modèle hybride. Pourtant les risques et difficultés sont avérés : **une entreprise sur quatre s'est déjà fait voler des données, une sur cinq a déjà vu un cloud public être attaqué.** Malgré tout, 83% stockent dans un cloud public des données sensibles alors que 69% seulement lui font confiance en matière de sécurité...

La question de la responsabilité des données reste floue. Selon Vanson Bourne, **69% des entreprises pensent que la protection, la confidentialité et la conformité des données relèvent de la responsabilité des fournisseurs de service cloud.** Et 51% estiment que le fournisseur de service cloud est responsable de la sauvegarde des volumes chargés dans le cloud !

3.3.2 RAPPELER LA REGLE 3-2-1

Pour exemple, Microsoft protège Office 365 au niveau de l'infrastructure. Cela comprend la sécurité physique des datacenters et l'authentification et l'identification dans les services cloud, ainsi que les contrôles des utilisateurs et des administrateurs intégrés à l'interface utilisateur d'Office 365. Pour sa part, **le service informatique est responsable de la sécurité au niveau des données.** Il existe une longue liste de risques de sécurité internes et externes, comme la suppression accidentelle, les administrateurs malveillants qui abusent de leurs accès et les ransomware pour n'en citer que quelques-uns.

Bien évidemment, **la question de la protection des données ne se limite pas au cas d'Office 365. Elle concerne tous les services dans le cloud public.**

« rappeler la règle 3-2-1 : disposer de trois copies de vos données au moins, stocker ces copies sur deux supports différents et conserver une copie de la sauvegarde hors site. Cloud ou pas, la règle 3-2-1 reste d'actualité ! »

3.3.3 LES ACTEURS DU CHANGEMENT DANS LA COURSE A LA DIGITALISATION

Véritable acteur de la transformation numérique dans l'entreprise, le Directeur Financier consulte le Responsable des Systèmes de Sécurité et d'Information pour trouver la solution la plus sécuritaire dans son projet de réduction de coût au niveau de l'infrastructure.

Pascal Corcos, associé chez PWC, et responsable des activités de Conseils pour les Directions Financières déclare :

« La transformation de la Fonction Finance est un impératif et la clé du succès est la dimension humaine :

Le rythme effréné des ruptures technologiques, les changements très réguliers de réglementations (IFRS, fiscalité, ...) et les axes stratégiques définis par le Management imposent aux Directions Financières une agilité forte pour s'adapter à ces évolutions. Les équipes sont de fait de plus en plus sollicitées mais de nombreux domaines d'opportunités

s'ouvrent également à elles. Illustrons brièvement ces propos à l'aide de sujets détaillés dans cette étude : Cloud, Robotisation et Prédicatif.

Cloud ou non Cloud ? Manifestement la question ne se pose plus pour plus de 83% des Directeurs Financiers. Le Cloud sera un incontournable du SI Finance dans tous les domaines : dématérialisation de documents et de dossiers de clôture, gestion de trésorerie, fiscalité, mais aussi ERP (progiciel de gestion intégré), EPM (management de la performance), BI (informatique décisionnelle)...

Si les avantages sont nombreux (projet court, robustesse de la solution, intégrité et sécurité, transparence de coûts, partage de bonnes pratiques), les outils Cloud ouvrent la voie à de nouvelles disciplines, de nouvelles compétences et à une approche projet agile, collaborative mettant l'accent sur le métier et le renforcement de la communication avec les autres fonctions de l'entreprise au-delà de la Finance.

« **Robotisation** » du métier de comptable, la fin d'une fonction support ? Question choquante, voire effrayante pour certains, mais il faut constater que vous êtes près de 50% à vous interroger sur le lancement d'un projet d'automatisation comptable d'ici 2020 ! « C'est inévitable » nous dites-vous car cette rupture technologique majeure répond à trois enjeux : réduction des coûts de traitements administratifs, vérification de la conformité et soutien des opérations. L'automatisation de tâches répétitives via des algorithmes va notamment permettre de focaliser les énergies humaines sur l'analyse d'indicateurs de performance. Cependant, la performance ne doit pas desservir la conformité. La fonction reste donc fondamentale mais elle doit évoluer et faire monter en compétence les équipes.

« **Prédicatif** » ou comment faire parler les données ! Alors que la fameuse notion de « Big Data » pouvait encore faire peur l'année dernière, il s'agit maintenant pour vous d'oser utiliser des machines (exploitants des données internes et/ou externes, financières et/ou extra-financières) pour calculer des prévisions de stocks, de ventes, de risques clients, de trésorerie, de risques de non conformité. Mais les perspectives sont encore plus larges qu'il n'y paraît car l'enjeu est de renforcer la robustesse et l'intérêt pour les opérations des informations communiquées par les financiers aux opérationnels et au management.

En conclusion, dans les deux études précédentes nous constatons que le digital était perçu essentiellement comme un sujet « informatique » que vous aviez du mal à appréhender. Cette année, bien que certaines questions se posent encore à vous (ROI, concurrence avec d'autres projets, etc.), vous semblez être convaincus de sa nécessité. Il y a une prise de conscience des opportunités induites en y associant les dimensions organisationnelle et humaine. »

Dématérialisation, Cloud et Robotisation : 3 usages numériques au service de la transformation

Dématérialisation : thème prioritaire, notamment grâce aux outils de pilotage de clôture



66%

des Directeurs Financiers estiment que la **dématérialisation** permet à la fonction Finance d'être plus efficace

Priorités pour les outils de pilotage



Automatiser le processus de revue des écritures manuelles



Lisser le pic d'activité pendant la clôture



Standardiser le calendrier de clôture

Ambition pour le Cloud



Réduire les coûts informatiques



Fiabiliser les données



Déployer les outils



des Directeurs Financiers pensent que le **Cloud** sera un incontournable du SI Finance. La question de la sécurité des données reste une préoccupation forte mais n'est plus un frein

Un intérêt très marqué pour la Robotisation au service de la Fonction Finance



48%

des Directeurs Financiers, d'entreprise de toutes tailles, considèrent que la **robotisation** permettra à la Fonction Finance de gagner en efficacité et d'améliorer le contrôle interne

Quelques exemples de robotisation

- Administrer le référentiel (création, mise à jour)
- Administrer les commandes et/ou litiges clients
- Contrôler les interfaces
- Calculer et enregistrer les écritures comptables
- Identifier et analyser des alertes

83 % des entreprises migrent sur les Clouds public comme Amazon et Google, mais comment garantir la confidentialité et la souveraineté des données hébergées ?

3.3.4 LA DEPENDANCE AUX FOURNISSEURS DE SERVICES CLOUD

Un aspect limitant du modèle de cloud public est également la situation de dépendance dans laquelle l'entreprise se place vis-à-vis des fournisseurs cloud. Si un environnement hétérogène type multicloud peut répondre partiellement à cette problématique, il n'en reste pas moins que si le provider décide un jour de multiplier ses prix, l'entreprise n'aura d'autre choix que suivre cette directive. Les données étant l'asset principal d'une entreprise numérique, personne ne voudrait la laisser partir.

4 LES EVOLUTIONS TECHNOLOGIQUES DE HYPE CYCLE DE GARTNER

Le Hype Cycle est une description graphique du phénomène qui advient chaque fois qu'une innovation est révélée. « **Hype** » en anglais désigne le battage médiatique : car le Cycle du Hype se concentre avant tout sur la communication autour des nouvelles technologies. Il propose une évaluation des technologies innovantes à la fois hardware et software en termes d'impacts business, de rythme d'adoption, de niveau de maturité afin d'accompagner les décideurs dans leurs investissements.

En Juillet dernier, Gartner a publié son Hype cycle sur les tendances à venir dans l'évolution des technologies de stockage de données.

En synthèse, le marché du stockage de données traditionnel atteint son niveau de maturité : stabilisation des acteurs du marché et tendance à la baisse des prix de ces technologies.

L'explosion de données et la nécessité de réduction de coûts tout en améliorant les performances amènent les entreprises à se tourner vers des solutions plus innovantes. Ce positionnement des entreprises est d'ailleurs renforcé par les politiques publiques mises en place pour la protection des données. Les entreprises ont de plus en plus recours à la déduplication et compression de données en ligne, au stockage objet et archivage évolué d'informations en entreprises pour optimiser leurs capacités de stockage.

Gartner conseille de passer dans un modèle plus agile, comme le Software defined storage ou encore un modèle "public cloud « as a service, comme le montre cet extrait du rapport : *“ Organizations should then shift their focus to innovation areas such as software-defined storage, NVMe and NVMe-oF, **public cloud storage**, container-native storage and hybrid cloud storage as these technologies mature.”*

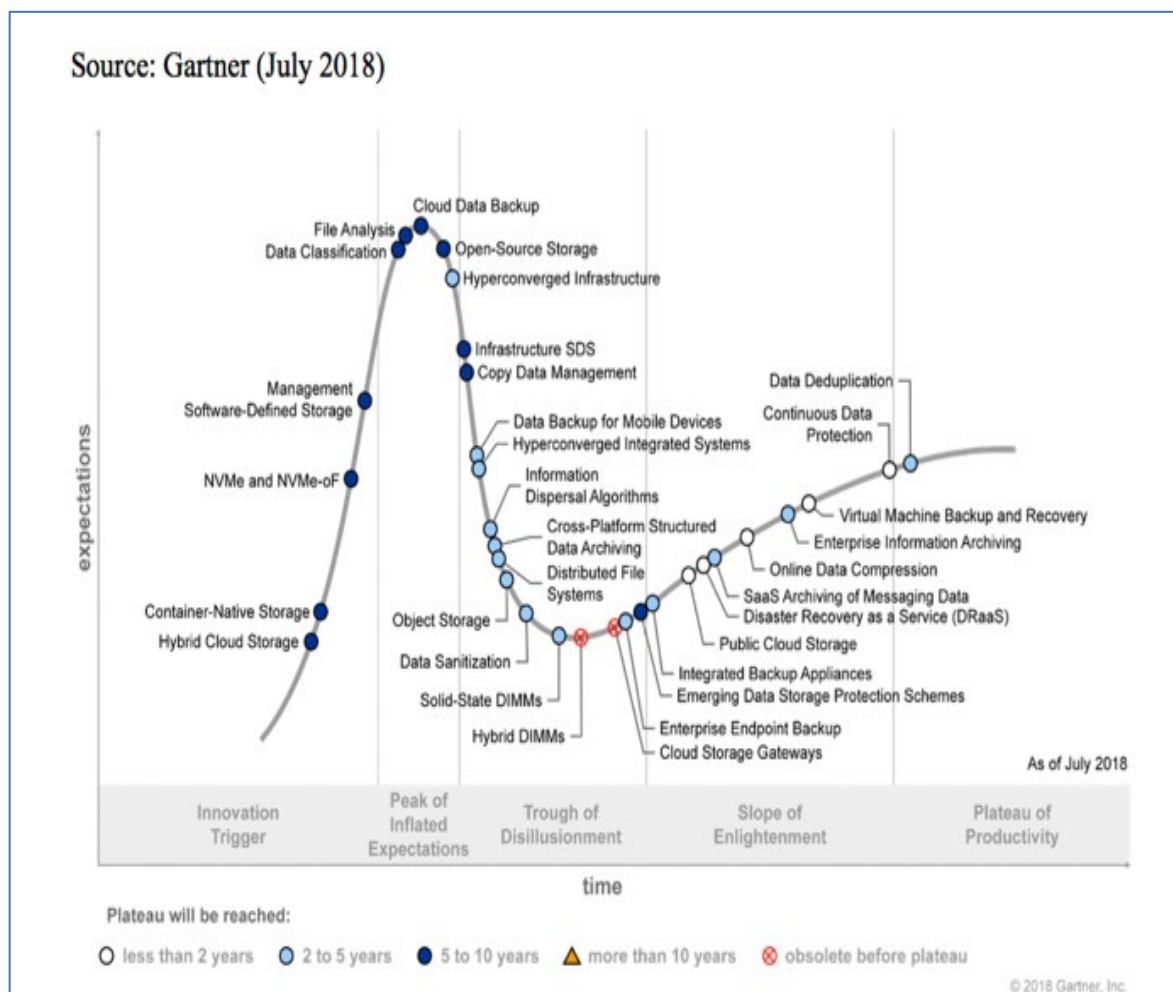


Figure : Hype Cycle for Storage technologie

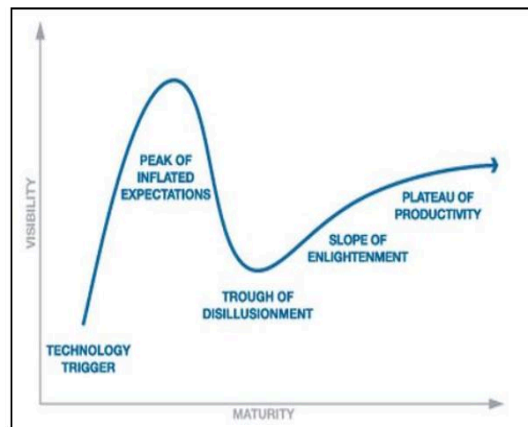
Le Cloud Computing apparait comme le modèle innovant, mature vers lequel se tournent nos entreprises à l'échelle mondiale.

Pour autant, ce modèle connaît des limites, notamment sur les aspects sécuritaires et de dépendance vis-à-vis de quelques entreprises mondiales qui dominent le marché.

Que vont-elles faire de nos données ?

Si certaines proposent de les commercialiser pour en faire profiter les particuliers à la fois utilisateurs et générateurs de ces mêmes données, il n'en reste pas moins la question fondamentale de la gouvernance : les états et les utilisateurs peinent à retrouver une gouvernance des données face aux géants commerciaux.

Savoir lire le Hype Cycle de Gartner



Un **Hype Cycle** comprend 5 phases :

1. **"Technology Trigger"** — La première phase d'un "Hype Cycle" correspond à l'arrivée sur le marché d'un nouveau produit ou d'une nouvelle technologie.
2. **"Peak of Inflated Expectations"** — Dans la phase suivante, un emballement généralisé aboutit souvent à des attentes exagérées et non réalistes (Buzz). Un certain nombre de sociétés mettront en oeuvre cette technologie avec succès, mais beaucoup d'autres termineront en échec.
3. **"Trough of Disillusionment"** — Cette technologie entrera dans le « creux de désillusion » parce qu'elle ne parvient pas à répondre aux attentes et devient vite démodée. Par voie de conséquence, la presse se détourne généralement de ce sujet et de cette technologie.
4. **"Slope of Enlightenment"** — Bien que la presse ait peut-être cessé de couvrir cette technologie, certaines entreprises continuent à travers la « pente de l'illumination » et l'expérimentent pour comprendre ses avantages et ses pratiques d'application.
5. **"Plateau of Productivity"** — Une technologie atteint le "plateau de productivité" lorsque les bénéfices qu'elle procure deviennent largement démontrés et acceptés. La technologie devient de plus en plus stable et évolue dans une deuxième puis troisième génération. La hauteur finale du plateau est variable suivant que la technologie est largement applicable ou au contraire ne bénéficie qu'à un marché de niche.

5 GEOPOLITIQUE DU NUMERIQUE

L'analyse d'Adrien BASDEVANT et de Jean-Pierre MIGNARD démontre que la géopolitique du numérique fait face à trois difficultés pour comprendre le rapport de l'Etat-nation à l'Etat-plateforme.

- La course que mène chaque Etat-nation pour affirmer son influence dans le cyberspace.
- La nouvelle opposition entre ces Etats-nations et les nouveaux acteurs du numérique, les plateformes. Ces derniers entretiennent des relations subtiles et ambivalentes avec les pouvoirs traditionnels.
- Le soft power, enjeu décisif, permettant d'imposer sa norme, son pouvoir.

5.1 ETAT-NATION CONTRE ETAT-PLATEFORME

Le concept politique d'Etat-nation moderne caractérise un processus d'unification de populations dans un territoire sur lequel des institutions exercent souverainement leur pouvoir.

La compétence de l'Etat-nation est aujourd'hui remise en cause par la mondialisation, l'Internet, les flux financiers voire la crise climatique. Là où l'Etat-nation prétendait pouvoir exercer sa souveraineté sans partage, il est contraint de mutualiser ses efforts, de partager ses compétences, voire peut-être d'abdiquer sa souveraineté.

Soit les Etats sont contraints de transiger avec d'autres Etats sur leur conception, soit avec de entreprises privées : tels sont les géants du numérique, que certains considèrent comme de nouveaux Etats-plateformes.

Les démocraties libérales permettent l'essor d'acteurs économiques privés qui rivalisent avec elles en puissance. Quid d'un Etat communiste à parti unique ?

Les précédents existent avec le pétrole, le gaz, et l'armement. Ce qui est inédit, c'est l'influence qu'exercent ces géants du numérique sur les populations, et l'impact sur les finances des Etats de leur évitement fiscal. Tout récemment l'ouverture d'un Apple Store sur les Champs - Elysées à Paris.



Manifestation du mouvement Attac – Paris le 18 novembre 2018

Ces plateformes constituent-elles des Etats par la force dont elles disposent ? Supérieures économiquement à la majeure partie des Etats composant la communauté internationale, elles détiennent des capacités d'adhésion plus importantes que la majorité d'entre eux.

La culture empruntée au marketing entraîne un sentiment d'adhésion de nature identitaire de millions de personnes à leur réseau affinitaire.

Toujours est-il que ces plateformes ne remplissent à l'heure actuelle aucune mission de service public, et ne devraient donc pas être confondues avec des Etats-nations ni même être qualifiée d' « Etats »-plateformes.

Depuis l'avènement d'internet, les Etats et les institutions font en tout cas face à un choc de souveraineté.

Aux lois nationales, les géants numériques opposent leurs propres constitutions, ces fameuses CGU (les conditions générales d'utilisation, que nous avons déjà évoquées).

Facebook rallie aujourd'hui plus d'un milliard d'utilisateurs, bientôt deux, alors que les Etats-Unis sont constitués d'à peine plus de 325 millions de citoyens.

Ce nouveau choc de souveraineté s'inscrit dans un contexte d'internationalisation des échanges économiques, où l'intervention institutionnelle et publique ne fait plus l'unanimité, particulièrement en matière de technologie.

A la suite des révélations d'Edward Snowden de l'existence de plusieurs programmes américains de surveillance de masse, les utilisateurs des plateformes se sont réfugiés dans les solutions technologiques susceptibles de garantir leur sécurité.

Les affaires FBI vs Apple, FBI vs Microsoft illustrent parfaitement la lutte entre Etats-nations et les plateformes.

Selon Peter Swire, professeur du programme Law and Ethics du Georgia Institute of Technology, nous avons plutôt affaire de « sécurité nationale contre sécurité informatique », dans tous les cas, la sécurité des citoyens.

6 CHINE, LA DATA AU CŒUR DE L'ÉCONOMIE.



En Juillet 2015, la Chine a émis un grand plan technologique pour devenir l'acteur numéro un de l'économie digitale, donc de la data appelé « **Internet Plus** ».

Ce plan a pour objectifs :

- Intégrer Internet Mobile, Cloud Computing, Big Data et Internet of things dans l'industrie
- Encourager le développement du e-commerce, les réseaux industriels, et la banque sur Internet
- Aider les sociétés Internet à augmenter leur présence à l'international
- Promouvoir le développement et l'innovation par des financements spécifiques, en plus des 40 Md yuans de financement déjà en place pour les industries émergentes.

6.1 NUMERO UN DES TELECOMMUNICATIONS

Les télécoms sont la pierre angulaire de notre révolution digitale. Selon une étude conjointe d'Accenture et du Forum économique mondial, l'économie digitale pourrait générer près de 100 000 milliards de dollars de valeur pour les entreprises et la société d'ici à 2025.

Avec un flux annuel qui ne cesse de croître et devrait atteindre plusieurs dizaines de zettaoctets en 2020, les opérateurs et entreprises telecoms doivent amorcer au plus vite leur transformation interne : mettre à niveau leurs infrastructures, leurs applications, leurs services et leurs business models.

La politique du gouvernement chinois dans le secteur des télécoms est basée sur deux objectifs distincts et complémentaires :

- Utiliser les télécoms comme relais de croissance de l'économie chinoise
- Faire émerger la Chine comme un champion mondial du secteur.

En 2018, la Chine est aujourd'hui première sur le marché mondial des télécommunications.

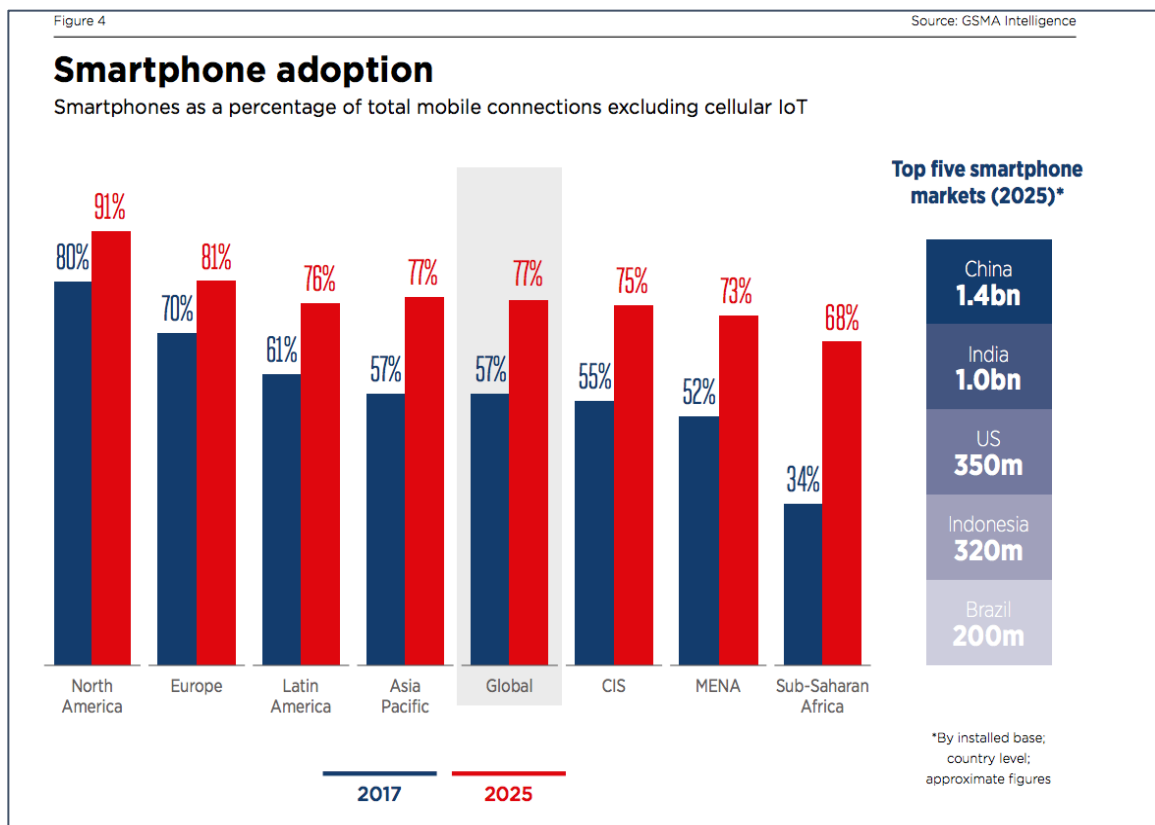
Au-delà de l'évolution des usages, elle a opéré un rattrapage industriel et technologique notable, et tient désormais une place incontournable dans la définition des standards internationaux et le développement de nouvelles technologies.

6.1.1 LE MARCHÉ FLEURISSANT DU TOUT MOBILE

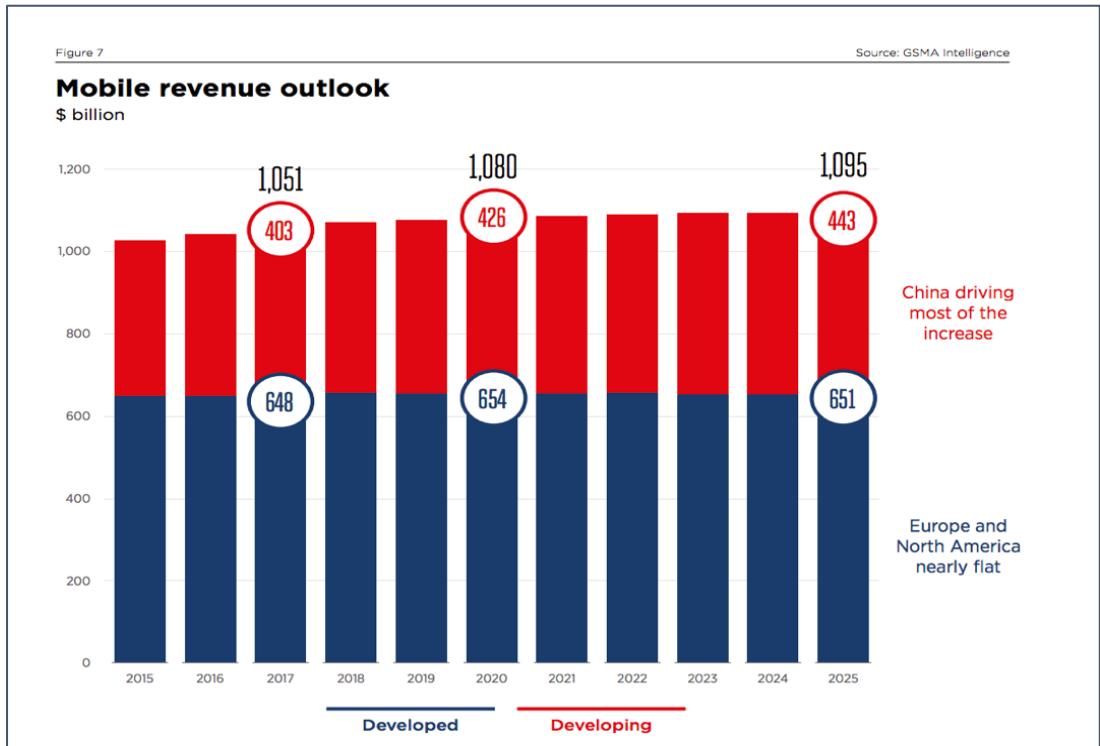
Sur le marché du mobile, en pleine croissance, la Chine tire son épingle du jeu.

Avec un peu plus 1,4 milliard d'utilisateurs mobile, dont plus de 60% sont abonnés à la 4G, 460 millions de smartphones s'y sont vendus en 2016, utilisés par 95% de la population connectée comme moyen d'accès privilégié à Internet.

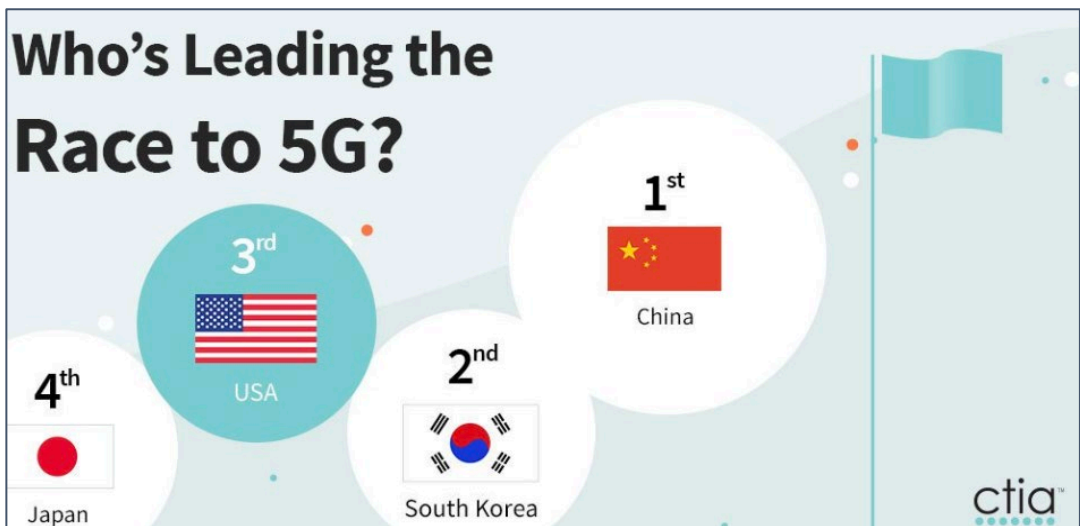
Sur le marché du mobile, en pleine croissance, la Chine tire son épingle du jeu. D'après une étude de GSMA Intelligence, du mois d'octobre 2018, l'Empire du milieu représente le marché le plus en croissance.



D'après une étude de GSMA Intelligence, du mois d'octobre 2018, la Chine représente à la fois le marché le plus prometteur mais également le plus actif au niveau international, tirant l'économie numérique.



6.1.2 CAP SUR LA 5G



La Chine mise sur les réseaux mobiles, du terminal à l'infrastructure centrale.

La croissance illustre d'entreprises telles que Huawei ou ZTE témoigne de cette ambition politique affichée. Pour prospérer, ces deux entreprises ont su s'appuyer sur le marché intérieur chinois et s'exporter à l'international.

Le marché local est dominé par des entreprises chinoises, malgré son ouverture aux entreprises étrangères. Par exemple, l'analyse des appels d'offres pour le déploiement d'infrastructures 4G (Base stations) lancés par les opérateurs chinois montre que les grands équipementiers nationaux - Huawei et ZTE- décrochent en moyenne plus de 70% des contrats

attribuables. Les équipementiers étrangers, eux, obtiennent au mieux 30% des appels d'offre, souvent répartis à parts presque égales entre les géants Nokia et Ericsson.

Sur le marché international Huawei et ZTE concentrent la vaste majorité des exportations chinoises dans le secteur. Huawei est le premier équipementier télécom mondial, avec un chiffre d'affaires en 2016 sur ce type d'équipements de 291 Milliards RMB (+24%), supérieur à celui de ses concurrents Ericsson (173 Mds RMB) et Nokia (180 Mds RMB). En outre, la majorité de son chiffre d'affaires est désormais réalisé hors de Chine dont un tiers en zone Europe Moyen Orient (EMEA).

Cependant, pour l'heure, les entreprises chinoises font face à certaines réticences et connaissent quelques difficultés notamment sur les marchés étatsuniens et Five Eyes (Canada, Australie, Royaume Uni, Nouvelle Zélande).

6.2 POLITIQUE VOLONTARISTE POUR PROMOUVOIR LES NORMES CHINOISES

La Chine veut jouer un rôle clé dans les prochaines générations de réseaux de télécommunications (5G d'abord mais aussi la 6G qu'elle prépare déjà), à la fois dans le développement de ses propres technologies et dans la définition des standards internationaux.

La Chine marque une forte présence et activité pour mettre en avant ses technologies et ses brevets dans l'objectif d'orienter le développement des standards mondiaux. Ces efforts sont répartis principalement sur quatre entités : l'UIT et le 3GPP au niveau mondial et l'ETSI et le 5G PPP au niveau européen, avec pour fer de lance Huawei.



- A l'UIT, les acteurs chinois ont largement investi les groupes du secteur de la normalisation (UIT-T) et engagé le processus de normalisation des architectures non-radio des IMT 2020. D'ailleurs, le groupe de normalisation spécialisé est présidé par un représentant de Huawei.
 - Au sein du 3GPP, l'organisation chinoise de normalisation (CCSA) a récemment révisé à la hausse sa contribution financière, prenant plus de poids vis-à-vis des autres partenaires, notamment européens, comme l'ETSI (qui reste encore majoritaire avec 56% des sièges). Par ailleurs, les délégations chinoises sont très actives notamment dans la production de travaux techniques pour diffusion dans les organismes de normalisation.
 - A l'ETSI20, comme toutes les parties prenantes, Huawei est représentée par un membre du conseil d'administration de l'ETSI et bénéficie d'un nombre de voix quasi-identique à ceux de Nokia ou d'Ericsson. L'entreprise pilote d'ailleurs des groupes d'études sur des sujets stratégiques comme l'utilisation des bandes de fréquence millimétrique ou l'introduction d'intelligence artificielle dans la gestion des réseaux.
- Enfin, il est important de noter que Huawei participe activement au 5GPPP, partenariat public-privé de la Commission Européenne dédié à la 5G, via sa filiale allemande. L'entreprise est donc membre du principal programme de recherche européen mené dans ce cadre, Horizon 2020, ce qui lui permet d'obtenir des subventions et un regard avisé sur les travaux de recherche au niveau européen.

6.3 BIG DATA ET IOT

6.3.1 VOLONTE POLITIQUE

Le China Computer Federation a créé une Task force Big Data dédiée au programme « Internet Plus ». Pilotée par l'Institute of Computing Technology (Académie des Sciences), elle compte aujourd'hui plus de 1000 membres qui animent des conférences et ateliers régulièrement.

Publié en octobre 2015, l'*Outline of Operations to Stimulate the Development of Big Data* liste différents sujets :

1. Créer des plateformes pour open data et données publiques institutionnelles
2. Promouvoir l'usage du big data au sein du gouvernement
3. Lancer un projet de « service public big data » (dans les domaines de la Santé ou des transports notamment)
4. Développer une application Big data pour le suivi de l'activité agricole et la ruralité
5. Lancer un projet Big data pour l'Industrie (retail, services et e-commerce)
6. Devenir leader mondial de la Recherche et Développement des technologies liées au Big Data
7. Compléter les systèmes de sécurisation du Big data.
8. Renforcer nos talents par la formation et le recrutement des cerveaux dans le monde

Par ailleurs, toutes les grandes universités ont créé des Instituts Data Science pour la Recherche mais surtout l'enseignement et le partage avec les meilleurs talents de la planète, par exemple:

- CAS : Key Lab on Network Data Science & Technologies
- Université de Tsinghua : Institute for Data Science, BigNet Center, FIT
- Shanghai :
 - o Université de Fudan – Big Data School & Institute for Big Data
 - o Université Jiao Tong – Big Data Engineering Technology Research Center

6.3.2 MARCHE ET PERSPECTIVES

Le marché Big Data en Chine a atteint 110 Md yuans en 2015 et devrait atteindre 879 Md yuan en 2020 .

Ce marché ne cesse de croître, en suivant les jalons du plan « Internet Plus » qui doit aboutir en 2025.

The image shows a screenshot of a news article from Chinanews.com. The main headline is "La Chine est en train d'investir massivement dans le monde". The article is dated Wednesday, July 6, 2016. The article title is "The perfect partner is just an ocean away". The author is Hu Yuan Yuan. The article text discusses Chinese businesses joining international match-making and the challenges of overseas mergers and acquisitions.

On the right side of the article, there is a vertical timeline illustrating major Chinese corporate acquisitions:

- 1992**: Shuangxing Group buys a stake in Intel's Red Hat, one of the earliest overseas mergers and acquisitions dealt by a Chinese company.
- 2005**: China National Offshore Oil Corporation bids \$18.5 billion for control of Unocal, the largest US oil and gas company. The bid fails.
- 2009**: Aluminium Group of China offers \$19.5 billion for a partnership with Rio Tinto Group. The offer is not accepted.
- 2014**: Anting Group buys the Hartford Financial Services Group, New York for \$1.95 billion.
- 2016**: China National Chemical Group acquires \$43 billion for Swiss agricultural giant Syngenta in a deal that would be the largest foreign investment by a Chinese company.

Le développement du Big Data est directement corrélé avec un autre axe stratégique du plan gouvernement : le Cloud Computing chinois.

6.3.3 UN CLOUD CHINOIS PRIVE, LA TERRITORIALISATION DES DONNEES

LE PROJET CHINOIS

Dès le départ, le gouvernement a clairement affiché la couleur : la Chine veut construire son cloud pour « contrer toute potentielle dépendance américaine » et devenir « l'Alternative ».

PLACE DE CHOIX SUR L'ECHIQUIER MONDIAL

Trois ans après cette annonce, quel résultat ?

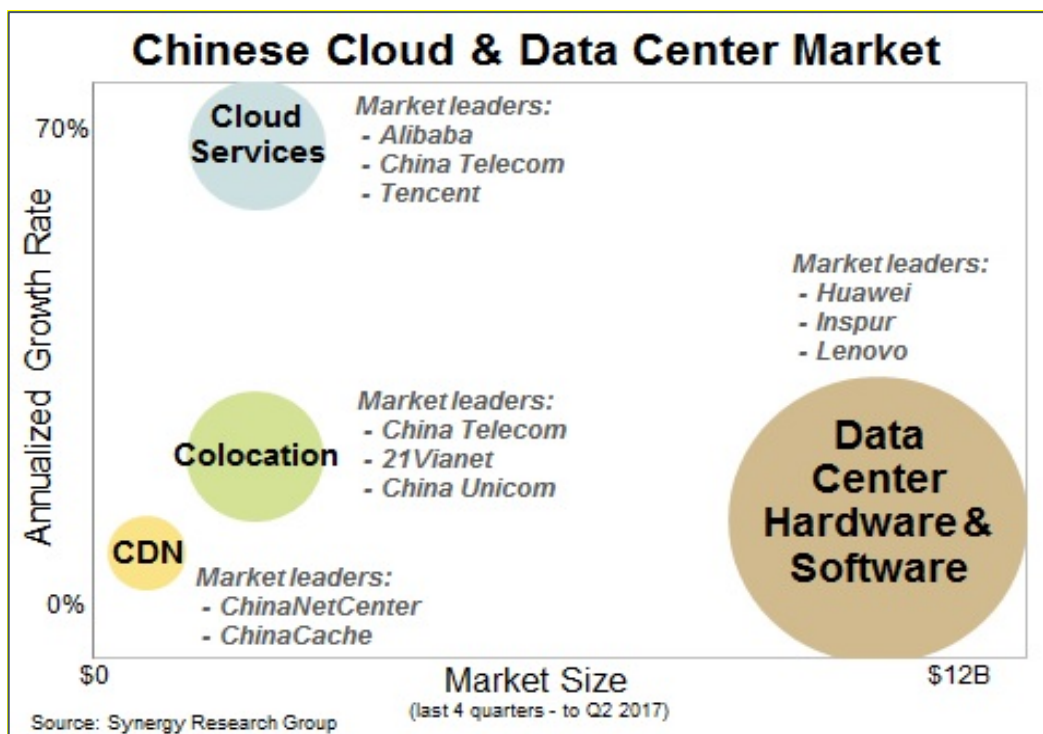
Dans un premier temps, les GAFAM (Google, Amazon, Facebook, Apple, Microsoft) à la place largement dominante dans le monde, ont tenté d'entrer en Chine.

Les trois premiers se sont heurtés à la vigueur respective de Baidu, Alibaba, Tencent.

Dans le domaine des mobiles, Huawei a su tirer son épingle du jeu, même si Apple a trouvé une petite place initialement.

Concernant les offres Cloud, les entreprises chinoises ont su contrer les offensives étrangères sur leur marché. Selon les derniers chiffres du cabinet Synergy Research Group, le marché chinois du data center et du cloud reste dominé par les entreprises chinoises. En effet, les entreprises chinoises dominent chacun des trois segments de marché clés :

- Matériel et logiciel pour datacentres : marché ouvert à la concurrence et assez hétérogène, la Chine représente environ 50% du marché.
- Services cloud : marché local largement par la Chine avec 80% du chiffre d'affaires
- Colocation et CDN (Content Delivery Network)



Au total, les revenus annuels de ces marchés dépassent maintenant 15 milliards de dollars par année, en augmentation de plus de 16% par an. Et ce n'est que le début.

Pour l'heure, comme dans le reste du monde, le marché du matériel et des logiciels des datacenters (serveurs, stockage, réseau, sécurité, OS, logiciels de virtualisation) est encore beaucoup plus important que le marché des services cloud (IaaS, PaaS, cloud privé hébergé), mais le marché du cloud en plein essor a un taux de croissance beaucoup plus élevé.

Les fournisseurs de services de colocation, qui hébergent des datacenters pour les entreprises et les fournisseurs de services de cloud, continuent de croître fortement et le marché du CDN continue lui aussi d'évoluer.

Au niveau global, sur ces marchés cloud, la Chine est le deuxième ou le troisième pays du monde en termes d'activité commerciale, mais première en terme de rythme de croissance.

« La différence entre la Chine et tous les autres pays est frappante »
précise John Dinsdale, analyste chez Synergy Research Group.

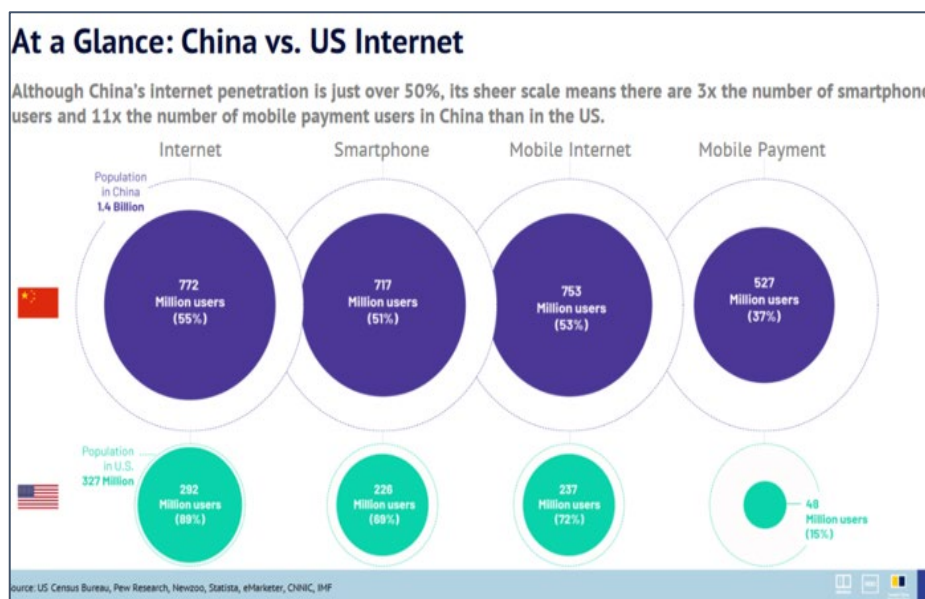
« Les marchés des services de cloud et d'infrastructures sont véritablement mondiaux et, dans toutes les régions, ils sont dominés par des sociétés américaines, mais la Chine est la seule grande exception.

À l'avenir, il est difficile de voir les entreprises américaines faire beaucoup de progrès en Chine. A l'inverse, il ne fait aucun doute que certaines entreprises chinoises auront un impact croissant dans les pays autres que la Chine. »

6.4 UN ETAT DERRIERE SES ENTREPRISES : BATX A L'ASSAUT DU MONDE.

Le gouvernement chinois, dans une politique volontariste, s'affiche derrière ses fleurons industriels.

Aussi les BATX sont-ils partis à la conquête du monde avec une célérité remarquable, et tiennent désormais importante dans l'échiquier économique mondial.





Les entreprises chinoises ont su s'adapter et innover pour proposer des alternatives aux offres actuelles – largement dominées par les entreprises américaines.

6.4.1 LA TERRITORIALISATION DES DONNEES

Si le marché local est en pleine croissance, la Chine souhaite rester maître de ses data. Si les entreprises américaines investissent en Chine sur les domaines technologiques liés à la data – comme Dell en 2015 sur l’IA- la Chine sait mettre les frontières virtuelles.

En Juin 2016, le gouvernement a sorti une loi de protection des données et cybersécurité. Celle-ci impose tout acteur étranger à stocker les données du marché chinois en Chine.

Apple s’est donc contraint à cette loi. Le 28 Février 2018, Apple a transféré l’ensemble des données utilisateurs de l’iCloud chinois à un partenaire local : photos, documents, messages des utilisateurs chinois d’Apple sont donc désormais stockées dans un datacenter spécialement créé pour cela dans la province centrale de Guizhou et géré par Guizhou-Cloud Big Data Industry Co, Ltd.

Apple a assuré :

Avoir « de bonnes protections pour garantir la sécurité et la confidentialité des données en place, et aucune porte dérobée ne sera créée dans nos systèmes »

Il est d’ailleurs à noter que le site de gestion des données ne sera pas géré par Apple. Longtemps, les géants du Web américain ont refusé de céder aux exigences chinoises. Mais, étant donné la taille du marché adressable, l’attractivité de la Chine les pousse à s’adapter. En Juin, Apple avait déjà annoncé retirer les applications de VPN permettant de contourner la censure chinoise de l’Apple Store chinois.

6.4.2 CREER UNE SOCIETE DE LA DATA.

L’objectif ultime du plan « Internet Plus » est de créer une société numérique de crédits sociaux basée sur le big data.

Ce plan – qui doit être appliqué ds 2020 dans le but de renforcer le contrôle sur Internet et la surveillance des données privées des citoyens - envisage de mettre en place un système de crédits sociaux permettant d’attribuer une note à chaque citoyen et entreprise du pays.

« Lors du Big Data Expo organisée à Guiyang, au sud-ouest de la Chine, un représentant de China Mobile a expliqué que la firme aide la police de Guiyang à suivre les mouvements de migrants en temps réel. Pour ce faire, China Mobile s’appuie sur les informations d’identité fournies par les clients lors de l’achat d’une carte SIM. L’opérateur téléphonique est en mesure d’obtenir des informations sur les citoyens au sein d’un périmètre délimité, mais aussi des détails sur leur genre, leur âge et leur provenance géographique », rapporte le magazine BigData.fr.

Les usages sont nombreux : de la protection du citoyen à l’amélioration des services de santé, ou encore l’éducation.

Toutefois, ce plan a suscité de nombreuses réactions et dissidences mais tient au partenariat entre les entreprises implantées en Chine et le gouvernement.

D'ailleurs, de nombreuses entreprises chinoises se joignent à cette expérience gouvernementale : Mobike, entreprise de location de vélo, offre des points de récompense aux utilisateurs qui inspectent volontairement les vélos garés et informent l'entreprise sur les mauvais comportements d'autres cyclistes.

Mais dans quel but ?

Si pour les entreprises internationales implantées en Chine, ce système de crédits sociaux représente un frein à l'entrée et des contraintes significatives, il permet aussi les entreprises chinoises de se démarquer.

Kweichom Moutai Group produit des vins de luxe en Chine. Elle existe via une application mobile, l'entreprise peut donc surveiller les achats sur internet et se positionner par rapport à la concurrence. Ensuite, elle peut proposer des promotions « personnalisées » en fonction des informations recueillies. Et ceci, à des fins commerciales et privées.

Après être devenue l'usine du monde, la Chine est en passe d'en devenir le laboratoire !

Et ce n'est que le début.

India to see over 1 million cloud computing job roles by 2022: Report

ET Bureau | Updated: Nov 23, 2018, 05.38 PM IST

L'Inde annonce un grand plan de transformation digitale du pays : de la numérisation des procédures administratives au stockage de données citoyens par la mise à disposition d'un [cloud](#) national.

Que reste-t-il ? Etant donnée les démographies et les rythmes des croissances économiques africaines, nul doute que le berceau de l'humanité sera le prochain théâtre d'opération de la conquête numérique.

7 LA SOUVERAINETE A LA FRANCAISE

7.1 LA VOLONTE DE CREER UN CLOUD SOUVERAIN.

La souveraineté numérique est une préoccupation sérieuse de la Présidence actuelle.

Dès 2010, l'État français s'est investi sur la création d'un Cloud souverain et la rénovation des datacenters institutionnels.

En 2012, avec le soutien financier de la Caisse des Dépôts, l'État a ainsi permis le lancement des sociétés CloudWatt et Numergy, en charge de développer un Cloud souverain et des offres numériques autour de Datacenters de nouvelle génération.

En 2013, la Direction Interministérielle des Systèmes d'Information et de Communication (DISIC) a réaffirmé son engagement dans l'adoption d'une politique de Cloud Computing pour « favoriser la convergence des systèmes d'informations des ministères et la consolidation des salles informatiques de l'État ». Le projet TCI (Transformation des Centres Informatiques) a pour objectif de consolider l'ensemble des infrastructures des administrations centrales sur dix datacenters français répartis sur l'hexagone.

En 2014, le plan Cloud Computing a inscrit le développement du Cloud dans les priorités stratégiques gouvernementales pour la Transformation digitale de la France.

7.2 LE PLAN CLOUD COMPUTING

L'État s'est particulièrement engagé dans le plan Cloud Computing, élaboré avec les chefs de projet Thierry Breton (Atos) et Octave Klaba (OVH), et annoncé le 4 juin 2014 dans le cadre des 34 plans de la Nouvelle France Industrielle et dont la feuille de route a été validée par le Gouvernement.

Le plan Cloud Computing comporte 10 propositions d'actions en faveur du développement des usages autour du Cloud Computing avec l'objectif de faire émerger des services Cloud à la française comme alternatives aux services de grands groupes américains et de développer l'économie du Cloud Computing en France.

Le développement de la filière implique une action coordonnée sur deux fronts :

- 1) Encourager la demande du secteur privé et participer à l'effort de développement à travers la commande publique
- 2) Soutenir et entretenir la richesse de l'offre proposée par les acteurs du marché.

En 2015, selon *les Echos* Numergy ne sait pas comment payer les salaires de ses 72 collaborateurs. Pourtant, en 2014, le chiffre d'affaires de l'entreprise atteignait les deux millions d'euros. CloudWatt fait face à des difficultés pour finalement renoncer. Projet trop ambitieux, mauvais moment – en fin de mandat présidentiel, offres inadaptées au marché, pas de capitalisation sur nos acteurs français du cloud – tels OVH, Gandi ou Ikoula., un échec retentissant

7.3 2018 : LE VIRAGE DU BIGDATA ET DE L'IA



7.3.1 CLOUD SOUVERAIN

Mars 2018, le Président Macron a clairement annoncé vouloir faire « du Cloud et du Big Data sa priorité pour la Transformation digital de notre pays ». La France relance dans le cloud souverain avec une stratégie axée sur trois niveaux d'hébergement, selon le degré de sensibilité des données.

Trois solutions sont proposées :

- **Un cloud externe** dans lequel figureraient les données, les applications jugées peu sensibles. Elle reposerait sur un ou plusieurs offres de tiers accessibles sur Internet.
- **Un cloud intermédiaire** dit « cloud dédié » pour accueillir les renseignements et les outils de sensibilité moyenne. Sa conception reposera sur une offre externe mais « personnalisée pour les besoins de l'État et hébergé sur des infrastructures dédiées ».
- **Un cloud interne** pour accueillir toutes les données considérées comme sensibles. Il serait accessible à tous les ministères via un portail dédié et serait hébergé par l'administration. « Ce cloud sera conforme aux exigences régaliennes de sécurité »..

Pour les projets de cloud intermédiaire et interne, l'ANSSI – Agence Nationale de Sécurisation des Systèmes d'Information- tiendra son rôle de référent cybersécurité pour l'infrastructure et les procédures.

Présentée par le secrétaire d'État chargé du numérique, la stratégie gouvernementale sur le cloud contient une particularité : des ajustements règlementaires seront effectués pour que l'administration puisse « recourir à des offres cloud sécurisées avec un hébergement hors du territoire national », ceci « pour favoriser un cadre de compétition équitable entre les offres cloud et l'informatique traditionnelle ».

Le gouvernement a fixé un calendrier à trois ans pour le développement de l'usage du cloud par les administrations, les établissements publics et les collectivités territoriales.

7.3.2 PLAN D'INVESTISSEMENT EN INTELLIGENCE ARTIFICIELLE (IA)

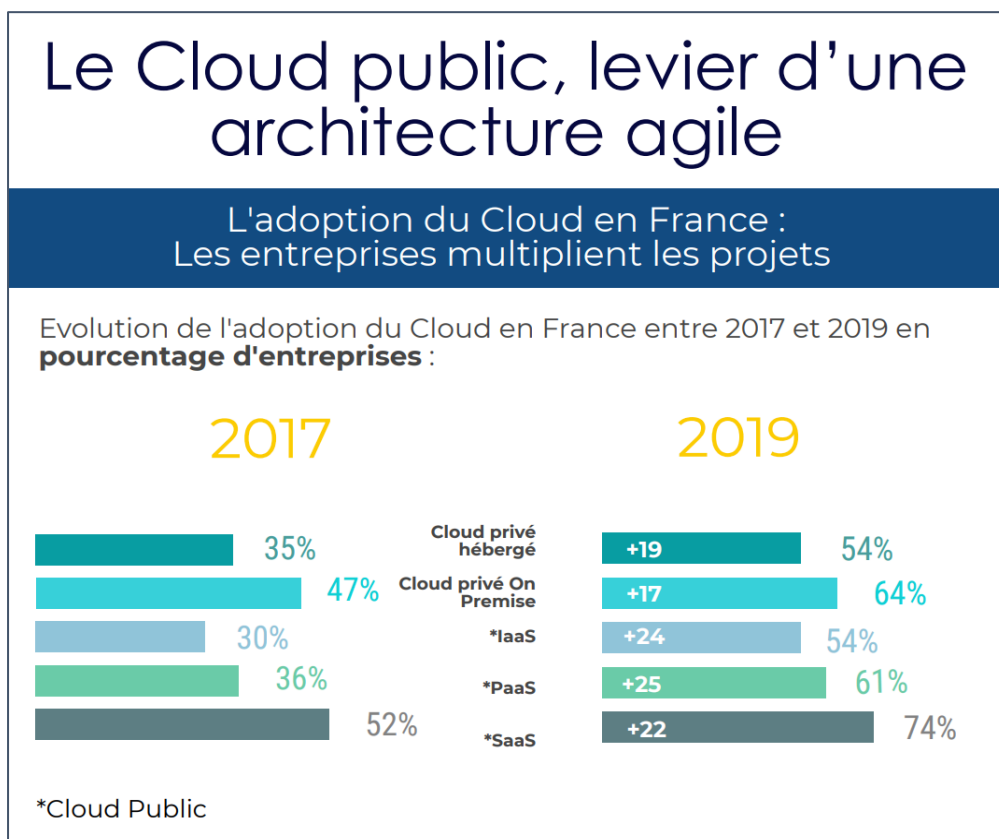
Mars 2018, le Président Macron veut investir 1,5 milliard de dollars dans les domaines de l'intelligence artificielle et du Big Data au cours de son quinquennat. Par ailleurs, le gouvernement souhaite également miser sur l'ouverture des données (open data), notamment dans le secteur de la santé et de la mobilité. Cet investissement passera par divers appels à projets : renforcer les filières de formation telles que les écoles de mathématiques, d'ingénierie, d'intelligence artificielle, mais aussi de Data Science, promouvoir les instituts et programmes de recherche fondamentale et appliquée et faire de la France le numéro un de l'Intelligence Artificielle.

Il est encore trop tôt pour tirer les conclusions de ce vaste programme, il en ressort une volonté étatique française de repenser la gouvernance de la donnée et revenir dans la course à l'innovation numérique et de l'IA.

Reste à traiter le « comment » et le plan d'actions. Saurons-nous apprendre de nos erreurs passées ?

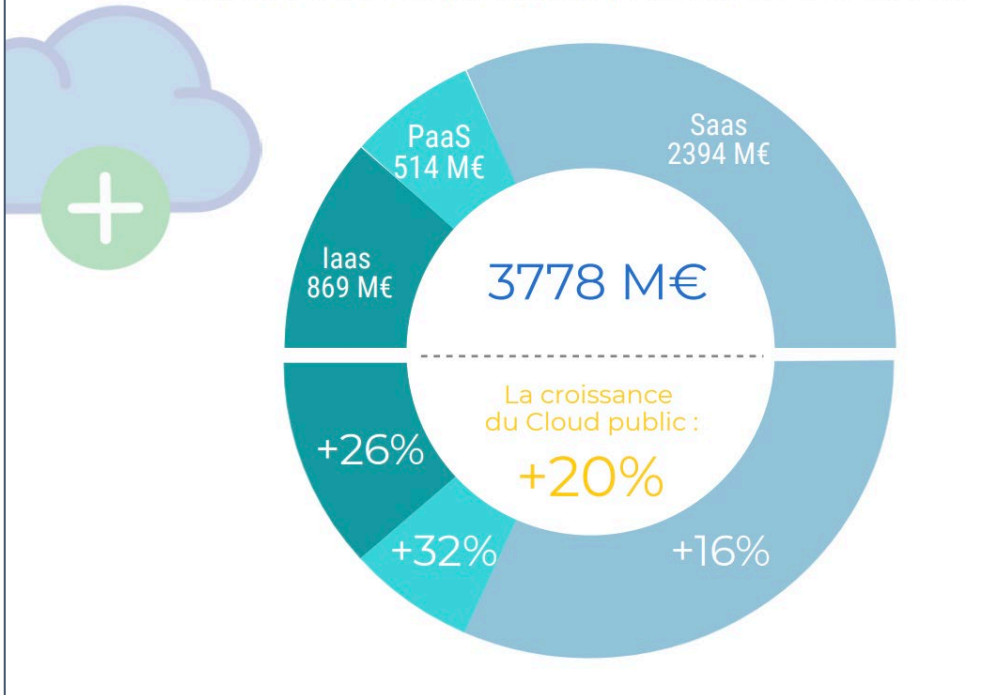
2- Le marché du cloud en France

Infographie avec les derniers chiffres des rapports IDC :



Le marché du Cloud en 2018

Infrastructure et plateforme prennent le relais du SaaS



8 CONCLUSION – DE L'INFINIMENT GRAND A L'INFINIMENT PETIT

La rédaction de ce mémoire a mis en lumière l'existence des rapports de force entre les Etats et les GAFAM et la transformation digitale des entreprises. Il a également mis en exergue que le RGPD occultait la notion des données identifiables à ce jour, et que la souveraineté des données dans les Clouds, publics et privés était remise en question tant du point de vue réglementaire (avec le Cloud Act) que sécuritaire avec l'externalisation du stockage.

Avec l'essor de l'internet mobile, des objets connectés, du *quantify environment*, la production de données actuelle est exponentielle. Nous avons atteint le point où nous collectons plus de données que nous ne pouvons en stocker. Cela nécessite plus de capacité de stockage avec le déploiement d'énormes datacentres équipés de milliers de disques durs, consommant de plus en plus de ressources en matières premières, en terres rares et en énergie.

Les data center sont responsables de 1,5 % de la dépense énergétique mondiale et 2% des émissions du carbone mondial.

La notion d'usure est également à prendre en compte, puisque pour rappel, la durée d'un disque dur est inférieure à dix ans.

Il est nécessaire d'anticiper les futurs besoins en inventant de nouvelles solutions durables de stockage d'information numérique.

Il ne s'agit donc pas de voir grand dans le futur, mais plutôt, de penser petit.

Nous avons abordé en introduction la cosmogonie de notre univers, pour présenter l'Histoire de la donnée à travers les âges, bien que nos scientifiques n'aient toujours pas trouvé l'équation reliant l'infiniment grand à l'infiniment petit. Et si la réponse était déjà en nous... Des millions d'années d'évolution nous ont donné un système de stockage biologique ultra perfectionné, dont nous sommes tous porteurs. L'ADN.

A la recherche de nouvelles technologies de stockage

Des chercheurs de l'Université de Californie du Sud ont calculé en 2011 que la planète avait produit un zetta-octet de données (mille milliards de milliards de caractères) en 2010. D'après leur prédiction, nous en créerons entre 300 et 700 000 Zo en 2040.

Le volume de data créé chaque seconde est estimé à 29 000 Go : le défi n'est plus seulement dans leur transmission, mais dans leur stockage.

L'objectif des recherches actuelles ? Une technologie permettant de stocker durablement un maximum d'informations sur des supports de plus en plus réduits requérant le minimum d'énergie.

De nombreuses pistes sont à l'étude actuellement, comme l'utilisation du diamant artificiel comme espace de stockage, mais le support le plus surprenant reste la molécule d'ADN.

L'ADN synthétique : un support prometteur.

Immense réservoir d'informations génétiques, l'ADN est depuis plusieurs années la source d'inspiration pour les chercheurs en technologies de stockage. En 2012 une équipe d'Harvard

réussit à transcrire le contenu d'un livre entier de 658 Ko dans un fragment d'ADN synthétique.

L'année dernière, Microsoft et l'Université de Washington avaient réussi à stocker 200 Mo. Le volume des informations retranscrites a donc été multiplié par 300 en quatre ans. Ces volumes paraissent modestes, mais avec un millimètre cube d'ADN, on estime la capacité de stockage à 1 milliard de Giga-octets!

Des chercheurs de Columbia et du New York Genome Center ont publié ce mois-ci les détails d'une nouvelle expérience dans la revue [Science](#). Ils ont pu inscrire dans une molécule d'ADN synthétique un système d'exploitation, un virus informatique et un film de 50 secondes : L'arrivée d'un train en gare de la Ciotat des frères Lumière. Et ce, sans altération des données une fois retraduites en langage binaire.

Des 0 et des 1 dans l'ADN

Quel est le point commun entre un bit et un nucléotide ? L'un est la base de l'information numérique. L'autre est la base de l'ADN : c'est à dire, de l'information génétique. Pour convertir les informations, le procédé est en théorie plutôt simple : l'information génétique est codée par l'ordre dans lequel s'enchaînent quatre nucléotides fonctionnant deux à deux (Adénine et Guanine, Cytosine et Thymine). L'information numérique, quant à elle, se lit dans l'ordre des bits, qui peuvent être un 0 ou un 1. Dans cette expérience, A et C ont donc servi à écrire les 0, et G et T, les 1.

Les informations numériques ont ensuite été introduites dans des cellules d'ADN synthétisées artificiellement, au moyen d'une imprimante à jet d'encre. Le tout est ensuite placé sur une puce de verre, qui constitue le support lisible par un ordinateur.

Au final, le procédé n'est donc pas très différent d'une copie de données sur disque dur... à la différence près qu'on ne peut écrire qu'une seule fois sur un brin d'ADN. Et donc, comment effacer nos données ? Il suffira de détruire leur support.

Les acteurs du changement

Deux sociétés d'origine Française : TWIST BIOSCIENCE et DNA SCRIPT ont été retenues comme seules finalistes dans le Projet très confidentiel du DNA DATA STORAGE de la DARPA. Les États-Unis sont leader sur le sujet en séduisant nos chercheurs par la reconnaissance de leurs travaux en utilisant les leviers que sont l'idéologie, l'égo et l'argent.

L'idéologie : En fédérant une recherche scientifique sur le bien commun, un monde meilleur et un cadre juridique autorisant la recherche sur le secteur.

L'égo : En reconnaissant la qualité de leurs travaux, les récompensant comme étant l'Innovateur de l'année par le prestigieux MIT, ou encore maître de conférences dans les meilleures universités américaines.

L'argent : Le capitalisme est par essence le fondement économique des États-Unis, la reconnaissance par la réussite financière est prônée sans complexe. Les montants alloués aux projets et les levés de fonds sont tellement importants pour nos chercheurs qu'ils se voient contraint de quitter la France pour poursuivre leurs recherches outre-Atlantique.

Bientôt un disque dur moléculaire ?

Alors, l'ADN est-il le futur du stockage d'informations ? On pourrait le croire, à voir les formidables capacités de stockage qu'il propose : un gramme d'ADN peut contenir jusqu'à **445 milliards de giga-octets**. C'est l'équivalent de 100 milliards de DVD ! A ce rythme, nos bonnes clés USB et disques durs pourraient être rapidement relégués au rang d'antiquités : les chercheurs affirment que d'ici à une dizaine d'années, l'ADN les aura remplacés.

L'ADN présente en outre de nombreux atouts :

- Facile à copier, il reste lisible après des milliers d'années, même conservé dans les pires conditions imaginables. Et contrairement à toute l'évolution technologique traditionnelle, aucune chance que le support de stockage soit obsolète au bout de quelques années : l'ADN est aussi vieux que la Terre.
- Présentes dans toutes les cellules vivantes, les molécules d'ADN renferment l'ensemble des informations nécessaires au développement et au fonctionnement d'un organisme. Mais que l'on se rassure, l'invention de ces chercheurs ne s'appliquera pas de sitôt à nos cellules humaines : il est en effet impossible, pour le moment, d'écrire dans une cellule vivante... sous peine de la modifier et de perturber son évolution.

Toutefois il reste de grands obstacles à franchir.

En effet, le processus de synthèse est extrêmement coûteux pour le moment.

La lenteur du cycle, encodage – synthèse – séquençage - décodage est également un facteur à améliorer. Les données étant confidentielles, nous ne pourrions pas développer ce point.

Le stockage par l'ADN est suivi avec grand intérêt par les Etats-nations (DARPA, NIST) mais aussi par MICROSOFT qui a commandé en 2016 10 millions de brins ADN pour y encoder des données numériques auprès de la société d'Emily LEPROUST, chercheuse française qui a créé son entreprise aux États-Unis : TWIST BIOSCIENCE.

9 BIBLIOGRAPHIE

- Académie Royale de Belgique (1997), Actualité et universalité de la pensée scientifique d'Adolphe Quetelet, Actes du colloque des 24 et 25 octobre 1996, textes rassemblés sous la direction scientifique de J.-J. Droesbeke, Mémoire de la Classe des Sciences, 3^e série, tome 13.
- DESROSIERES A. (1993), La politique des grands nombres. Histoire de la raison statistique, Paris, la découverte.
- DROESBEKE J.-J., MAUMY-BERTRAND M., SAPORTA G, et THOMAS-AGNAN Ch. Eds (2016)- Models choices and aggregations, Paris, technip (à paraître).
- DROESBEKE J.-J et SAPORTA G. (2010) les modèles et leur histoire, dans droesbeke J.-J. et saporta G. (éds), Analyse statistique des données longitudinales, Paris, technip, 1-14.
- DROESBEKE J.-J. et TASSI Ph. (2015), Histoire de la statistique, 2^e édition corrigée, collection que-sais-je ?, IAD, Paris, Presses
- DROESBEKE J.-J. et VERMANdelec. (2016), Les nombres au quotidien (à paraître). HECHT J. (1987), L'idée de dénombrement jusqu'à la révolution, dans Affichar, J. (éd.), Pour une histoire de la statistique, 1, Paris, economica, 21-81.
- REY A., TOMI M., HORDE Tet TANET Ch (1993), Dictionnaire historique de la langue française, 2^e édition, Paris, dictionnaire le Robert.

<https://www.cnil.fr/fr/violations-de-donnees-personnelles-1er-bilan-apres-lentree-en-application-du-rgpdes>

https://fr.wikipedia.org/wiki/News_Corp

https://fr.wikipedia.org/wiki/21st_Century_Fox

https://www.lexpress.fr/actualite/monde/amerique-nord/avec-trump-fox-news-a-enfin-trouve-son-president-ideal_1872966.html

<https://www.linkedin.com/in/abiAbigail-slater-8720188/>

https://fr.wikipedia.org/wiki/Business_Roundtable

<https://www.itic.org/about/staff/dean-c-garfield>

<https://www.itic.org/about/our-history>

<https://www.axios.com/scoop-the-white-house-looks-to-coordinate-online-privacy-plan-a51691cf-78d9-466e-8deb-27a66b1843c7.html>

<https://www.linkedin.com/in/abiAbigail-slater-8720188/>

https://fr.wikipedia.org/wiki/Business_Roundtable

<http://science.sciencemag.org/content/355/6328/950>

<https://bigdatawg.nist.gov/uploadfiles/NIST.SP.1500-3r1.pdf>

https://bigdatawg.nist.gov/Day2_01-Big_Data_analytics_for_NIST_by_GeorgeStrawn.pdf

<https://www.whitehouse.gov/sites/whitehouse.gov/files/images/NSCI%20Strategic%20Plan.pdf>

<https://www.the-scientist.com/cover-story/making-dna-data-storage-a-reality-30218>

<https://www.livescience.com/62812-dna-computers-government-iarpa.html>

<https://www.technologyreview.com/s/612225/the-dna-data-storage-machine-thats-the-size-of-a-school-bus/>

<http://vixra.org/pdf/1707.0371v1.pdf>

<https://lab.dessimoz.org/media/DNA-storage/Le%20Monde%20Informatique.pdf>

<https://lab.dessimoz.org/media/DNA-storage/2017-LeFigaro.pdf>

https://strbase.nist.gov/pub_pres/Kline-DARPA-Feb2012.pdf

<https://www.espci.fr/fr/innovation/creation-d-entreprises/dna-script>

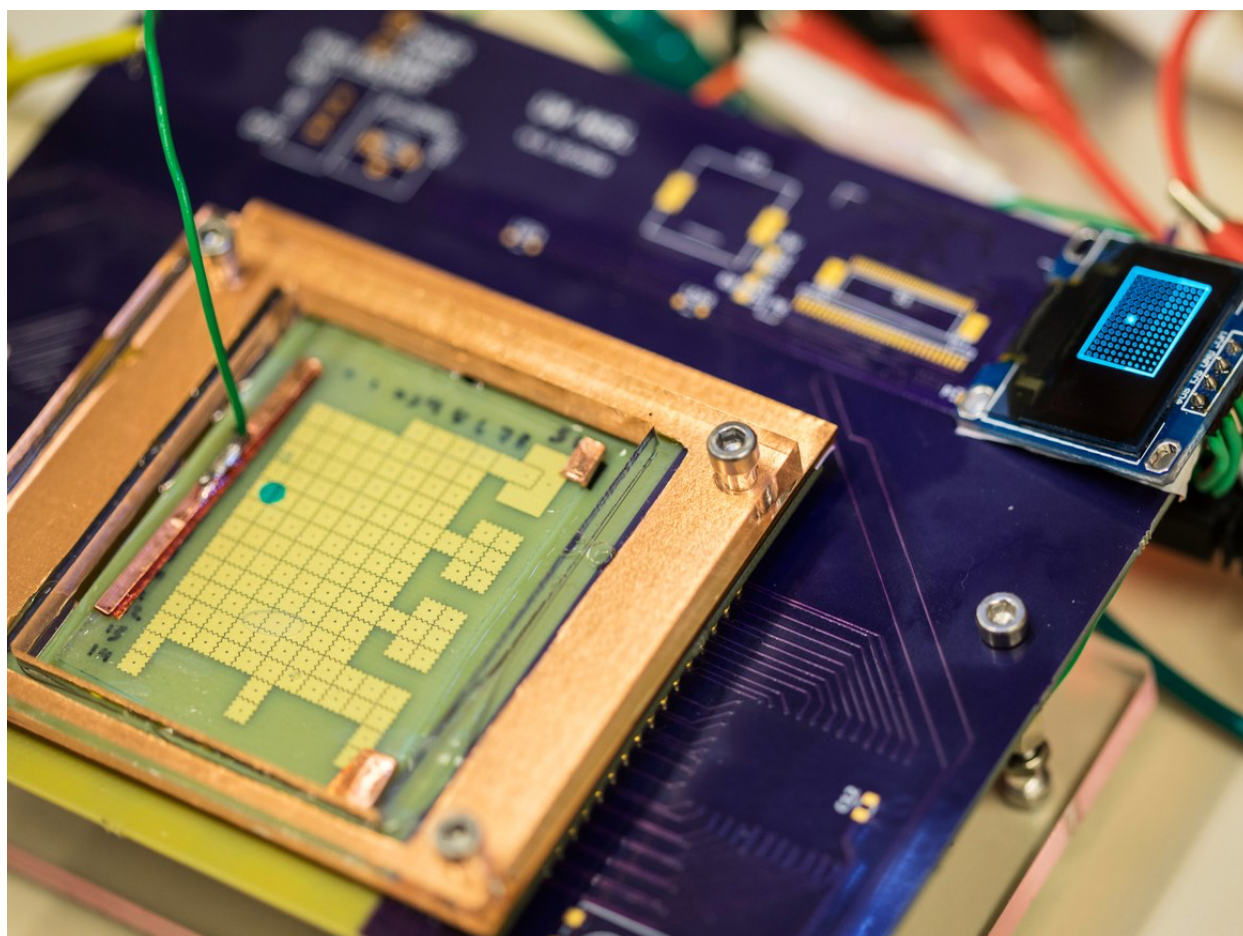
https://en.wikipedia.org/wiki/Commission_nationale_de_l%27informatique_et_des_libert%C3%A9s

<https://www.ssi.gouv.fr/en/actualite/european-secure-cloud-a-new-label-for-cloud-service-providers/>

<https://fr.wikipedia.org/wiki/Donnée>

ANNEXES

Darpa Wants to Build an Image Search Engine out of DNA



A digital microfluidics prototype automates DNA data-storage steps at UW's Molecular Information Systems Lab. A single droplet (here dyed blue) could carry multiple terabytes of information.

Dennis Wise/University of Washington

Most people use Google's [search-by-image feature](#) to either look for copyright infringement, or for shopping. See some shoes you like on a frenemy's Instagram? Search will pull up all the matching images on the web, including from sites that will sell you the same pair. In order to do that, Google's [computer vision algorithms](#) had to be trained to extract identifying features like colors, textures, and shapes from a vast catalogue of images. Luis Ceze, a computer scientist at the University of Washington, wants to encode that same process directly in [DNA](#), making the molecules themselves carry out that computer vision work. And he

wants to do it using your photos.

On Wednesday, Ceze's team at UW launched a social media campaign to collect 10,000 images from around the world and preserve their pixels in the As, Ts, Cs and Gs that make up the building blocks of life. They've done this sort of thing before; in 2016 [they encoded an entire OK Go music video](#)—setting the record for most amount of data stored in DNA. But this time they decided to crowdsource the data, building [a website where people can submit photos](#) and encouraging people to share their images on social media with the hashtag #MemoriesInDNA. "DNA can last thousands of years," says Ceze. "So this is essentially a time capsule. What do you want to preserve forever?"

UW's #MemoriesInDNA campaign might be a bit of a gimmick (there are plenty of available, high-quality image databases on which to train a molecular search engine). But the science behind it is a very real attempt to upend the last six decades of computing. DNA-based storage has so far been good [only for that](#): encoding pixels and locking them up in freeze-dried strands invisible to the human eye. So far, no one's figured out how to retrieve and *process* DNA-stored data—a necessary first step for creating any kind of serious molecular computing platform.

Who would want that, exactly? Well, Darpa for one.

In the last few months, the Department of Defense agency tasked with funding science's most far-out hopes has begun investing millions in discovering radical, non-binary ways to work with data. "Molecules offer a very different approach to 'computing' than the 0s and 1s of our existing digital systems," says Anne Fischer, program manager for [Darpa's Molecular Informatics program](#), which has so far awarded \$15.3 million to projects at Harvard, Brown, the University of Illinois, and the University of Washington. "The global community is creating data at a tremendous rate, and developing new approaches to access and process this information is critical to address looming shortfalls in storage capacity and

computational speed.”

An Illumina NextSeq flowcell, which Luis Ceze's group at UW uses to sequence DNA samples that contain digital data like photos and videos.

Dennis Wise/University of Washington

The digital age began with a simple act of delegation: man outsourcing memory to machine. First in vacuum tubes, then with transistors, tape discs, and flash drives. After more than 60 years, the essential logic-based architecture described by John von Neumann still undergirds modern computing infrastructures. And by any measure it has served humanity well. But its limits are becoming apparent as humans create ever more complex data.

“Moore’s Law has been all about miniaturization of devices,” says Karin Strauss, a senior scientist at Microsoft and collaborator on the UW project. “Electronics are great and will continue to exist, of course, but molecules are the final frontier when it comes to miniaturization.” Chemistry offers an untapped palette of molecular diversity—properties such as structure, size, charge, and polarity—that could be harnessed for information processing.

In the case of DNA, it’s the structure that does the heavy lifting. Strauss will be working with Ceze to first extract all the visual features from the crowdsourced images, and then map them into strings of As, Ts, Cs, and Gs. Each photo might get tens of thousands of unique DNA segments, each one encoding for a curve, or a vertical line, or a patch of blue. Then they can introduce a coded “query,” just the way you would type a few keywords into Google search. Except this query would be a string of DNA that corresponds to some of those visual features. And each query sequence would get a special coating of magnetic nanoparticles.

Drop a few of those in a microtest tube of DNA, where 10,000 images are stored in a few milliliters, and they’ll grab all the sequences that are a match. Then you just need a magnet to haul them out and a sequencer and some more algorithms to turn them back into visual images.

That's how they hope it will work, anyway. "The core of the Darpa project is figuring out which mechanisms are best equipped to do molecular processing," says Ceze. "We're focusing on visual data because it's by far the largest type of data in the world. And we think DNA's specific binding properties make it well-suited for that. But we'll see."

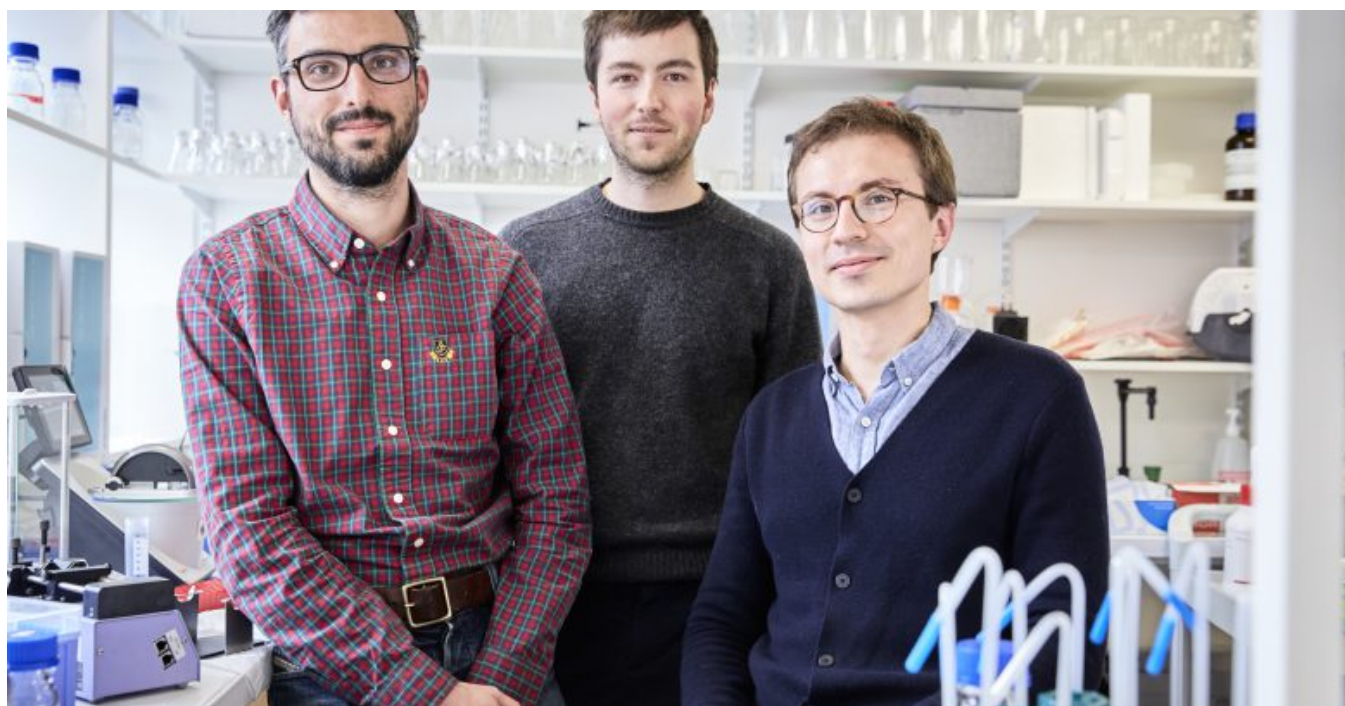
Other researchers are leveraging different physical properties of DNA to encode information. Olgica Milenkovic's group at the University of Illinois isn't manufacturing huge amounts of synthetic DNA, but rather making small cuts in naturally-occurring bacterial DNA. These changes can be counted, which makes them essentially addition and subtraction operators—one of the building blocks for programming languages like Java.

And DNA isn't the only molecule Darpa is interested in. Brenda Rubenstein is a theoretical chemist at Brown, where she's worked on quantum computing—encoding bits of information as either atoms, ions, photons, or electrons. But now she's extending that idea to organic compounds, specifically ones that have multiple places to attach R-groups—the variable parts of molecules, that lend them different physical and chemical properties. Running different reactions modifies those R-groups in predictable ways, which makes them good for computing basic linear algebra equations, says Rubenstein. "They have so many properties, there's an incredible capacity for storing and processing information," she says. "I think small molecules are almost an obvious choice for broadening the scope of computing."

Molecules, like DNA, might prove to have [some serious advantages](#) over the *in silico* state of the art; they've got way denser storage potential, last way longer, and may even be able to process way more in parallel. But they're not a silver bullet. DNA, like computer code, can [still be hacked](#). And it's hard to see how you'd get a soup of small molecule reactions crammed under the hood of your smartphone. But it's fun to at least

imagine that years from now the Department of Defense might be building underground bunkers, not for server farms, but for trays of microscopic glass beads; a nation's secrets held in freeze-dried DNA.

[INSIDERS] BioTech: la startup parisienne DNA Script ouvre une filiale aux États-Unis



DNA Script, spécialisé dans la synthèse enzymatique d'ADN, poursuit sur sa lancée. [Après une levée de 11 millions d'euros l'année dernière](#), la jeune pousse parisienne annonce l'ouverture d'une filiale aux États-Unis.



Crédit: Deepadesigns/Shutterstock

Elle en profite également pour renforcer son management avec une série de nominations: Jeffrey Jeddelloh, docteur en génétique moléculaire de l'Université Washington à Saint-Louis, devient vice-président en charge du développement et de la stratégie commerciale, Stephen Macevicz, docteur en biophysique et en droit de l'Université de Berkeley, est nommé vice-président en charge de la propriété intellectuelle et Christine Peponnet, docteur en génétique humaine de l'Université Paris VII, occupera le poste de vice-présidente responsable du développement technologique.

« Au cours de la dernière année, la société a porté son niveau de financement à 25 millions d'euros, s'est vu délivrer deux brevets, a déposé cinq nouvelles demandes de brevets et a atteint un effectif de 35 salariés. Ce n'est qu'un début : nous souhaitons vendre nos premiers produits commerciaux à nos partenaires d'ici 12 mois. Etant donné l'importance du marché américain, il est essentiel pour nous de s'y implanter et de renforcer notre équipe de direction avec des personnes ayant une expérience industrielle solide », explique l'entreprise.

Fondé en 2014 par Thomas Ybert, Sylvain Gariel et Xavier Godron, DNA Script développe une imprimante capable de synthétiser de l'ADN de haute qualité pour favoriser l'innovation dans les sciences et la technologie. Pour cela, cette technologie d'écriture utilise des enzymes génétiquement modifiées à la place de méthodes chimiques classiques. A travers cette technologie, DNA Script entend donner un second souffle à l'écriture des codes génétiques. Cette dernière considère que sa technologie va permettre d'accélérer le développement de nouveaux produits thérapeutiques, la production écologique de produits chimiques, ou encore l'amélioration du stockage de données sous forme d'ADN.

#MarTech: Splio met la main sur la startup Gowento

Pour sa première acquisition, [Splio](#), plateforme spécialisée dans la relation client et l'omnicanal, a mis la main sur la startup française Gowento, une solution d'engagement client sur mobile.



À travers ce rachat, Splio souhaite devenir l'acteur majeur du CRM mobile et moderniser la relation client en profondeur. « Avec l'acquisition de Gowento, Splio place les Mobiles Wallets au cœur des enjeux CRM et de néo-fidélisation, permettant notamment de digitaliser la carte de fidélité, d'envoyer des push notifications Wallet, et de déclencher un affichage géolocalisé et individualisé à proximité des points de vente », explique l'entreprise.

Fondé en 2015, Gowento revendique 50 marques clientes à l'instar de la Fnac, Dessange ou encore Nicolas et a déployé sa solution dans plus de 60 pays.

#Rakuten: de solides résultats sur 9 mois confortent les ambitions

AFP – Le groupe japonais de commerce et services en ligne Rakuten, qui a entrepris une vaste réorganisation de ses activités, a annoncé jeudi un bond de 49% de son bénéfice net sur les 9 premiers mois de l'exercice calendaire, grâce à de fortes ventes.



Rakuten a dégagé un gain net de 108 milliards de yens (831 millions d'euros), sur un chiffre d'affaires qui s'est élevé de 17% à 790,30 milliards de yens (6,09 mds EUR). Rakuten a un portefeuille de quelque 70 activités comprenant, outre le commerce en ligne, celles d'agence de voyage, d'opérateur mobile (il n'a pas encore son propre réseau mais cela devrait bientôt être le cas), de gérant de monnaie électronique, de maison de courtage, etc.

Rakuten, nouveau sponsor de l'équipe de football de Barcelone, encaisse surtout des recettes provenant de divers services Internet qu'il propose. Ces dernières ont augmenté de 15% et les bénéfices afférents de 11%. Bien que le groupe soit confronté à la forte concurrence d'Amazon au Japon, il a réussi à fidéliser des clients avec des événements promotionnels réguliers. Il tente continûment de renforcer l'attrait de sa galerie marchande virtuelle pour inciter les acheteurs à commander davantage avec leurs smartphones et tablettes. À l'étranger, il se félicite, comme à chaque fois, des performances de la société américaine Ebates, qui propose des rabais sur les achats en ligne. Rakuten est aussi propriétaire de la plateforme française de vente en ligne Priceminister et du canadien Kobo (liseuses et services d'ouvrages numérisés).

Dans le domaine des services financiers (fintech), son chiffre d'affaires a progressé de 24% et ses bénéfices de 14% grâce aux services bancaires, mais le groupe a quand même souffert de pertes dans le domaine de l'assurance en raison de plusieurs catastrophes naturelles survenues en septembre. Comme à son habitude, Rakuten n'a pas livré de prévisions chiffrées pour 2018. Le groupe est actuellement en train d'investir notamment pour accompagner des start-up. Il prépare aussi activement son propre réseau cellulaire afin de devenir le vrai 4e opérateur de télécommunications mobiles japonais, alors qu'il se contente jusqu'à présent de proposer des offres s'appuyant sur la location d'infrastructures aux acteurs en place.

US Spy Agencies Want to Store Data on DNA Computers

By [Rafi Letzter, Staff Writer](#) | June 13, 2018 01:15pm ET



Credit: vitstudio/Shutterstock

Government intelligence agencies have a plan to build computers that store information inside DNA and other organic molecules.

Intelligence Advanced Research Projects Activity (IARPA), a group within the Office of the Director of National Intelligence that develops technologies for U.S. intelligence services, [announced](#) plans to develop "tabletop"-sized machines that can store and retrieve data from large batches of [polymers](#) — a term that refers to a wide variety of long, stringlike molecules. Polymers can store data in the sequence of individual atoms or groups of atoms.

The project, which was reported [by Nextgov](#), is an attempt to solve a basic problem of the modern era: the vast and growing costs of [data storage](#). Datacenters around the world sucked up 416.2 terawatt hours of electricity in 2016. That's about 3 percent of the global supply, [according to a report in the Independent](#), and it accounts for 2 percent of global greenhouse gas emissions. Experts told the Independent that the world can't sustain the exponential rate of global data center growth.

A 2016 [paper](#) in the journal BioMed Research International found that DNA, in particular, could store computer information more densely, require less energy, and survive higher and lower temperatures than conventional hard drives. The authors of that paper reported on the successes of prototype DNA computers that used the genetic molecules for both long-term storage and random access memory (RAM). [[Humanoid Robots to Flying Cars: The 10 Coolest DARPA Projects](#)]

But no one has yet figured out how to implement DNA data storage on large scales.

IARPA officials said the new effort, called Molecular Information Storage, will be broken up into three chunks: a two-year program to figure out how to store data in DNA or other molecules at high speeds, a two-year program to figure out how to retrieve that data at high speeds, and a two-year effort to develop an operating system that can run on that DNA.

Many of the technologies IARPA wants to develop are untested at these scales, so it's unclear how far away that proposed "tabletop device" really is.

Originally published on [Live Science](#).