Comportements culturels et données personnelles au cœur du Big data

Entre la nécessaire protection et une exploitation au service des nouveaux équilibres économiques

EY
Building a better working world

FORUM 2013 D'AVIGNON

Sommaire

1. **Faire parler le Big data : un nouveau pouvoir dans le secteur médiatique et culturel** p. 7
   - Du Big data au Worthy data p. 8
   - La donnée personnelle culturelle au cœur de l'émergence d'un marché stratégique p. 14
2. **Pouvoirs en équilibre et responsabilités partagées : les nouveaux contours du Big data** p. 19
   - La réglementation, impulsion économique pour les acteurs de Big data en Europe p. 21
   - Quelle fiscalité 3.0 pour l'économie digitale ? p. 33
   - Responsabilités citoyennes p. 37
3. **Vers un nouveau paradigme économique propice à l'innovation et la création** p. 39
   - Opposer le principe de précaution au Big data : un risque réel pour l'innovation p. 40
   - Le temps de l'action p. 42
   - La nouvelle chaîne de valeur de la donnée personnelle culturelle p. 47

La donnée personnelle culturelle au cœur du Big data p. 50
Infographie

La présente publication peut être téléchargée sur les sites d'EY (www.ey.com/mediaentertainment) et du Forum d'Avignon (www.forum-avignon.org).

Avant-propos

6 ans de partenariat entre EY et le Forum d'Avignon

La vocation première du Forum d'Avignon est de se faire l'écho des enjeux qui comptent pour les industries de l'art et de la création. Le partenariat qui lie depuis maintenant six ans EY au Forum témoigne d'un engagement commun, aux côtés des grands acteurs de l'univers des médias et du divertissement.

Depuis 2008, EY analyse les grandes thématiques inscrites au programme du Forum, pour leur donner, en capitalisant sur son expertise et son expérience reconnue dans le secteur, une traduction concrète à travers l'identification d'enseignements majeurs.

Année après année, EY a observé et décrypté les mouvements à l'œuvre dans l'industrie des médias et des contenus, face à une révolution digitale qui a rebattu les cartes des forces en présence, à travers le prisme de la propriété intellectuelle *(« La propriété intellectuelle à l'ère du numérique »),* de la monétisation *(« Monétiser les médias numériques »)* ou encore de la vitesse de diffusion *(« Maîtriser le tempo, organiser la relation entre le temps et la valeur dans l'industrie des médias et du divertissement »).* L'écosystème qui s'est progressivement structuré semblait tendre jusqu'à présent vers un point d'équilibre entre les opérateurs de l'Internet, de télécomunications et les groupes médias. Néanmoins les enseignements de nos dernières études laissaient déjà entrevoir la déferlante Big data, qui pourrait introduire un nouveau facteur de déstabilisation, laissant un nombre restreint d'acteurs, capables de faire parler le Big data, détenir le pouvoir de contrôler et de prévoir.

La donnée personnelle culturelle, autour de laquelle se structure aujourd'hui un marché à la recherche de nouveaux équilibres, est à l'origine d'une ruée d'acteurs des secteurs numérique, médiatique et culturel, parce qu'elle vaut de l'or.

Méthodologie de l'étude :

Pour mieux comprendre, analyser et interpréter ces nouveaux enjeux de contrôle et la structuration d'un nouveau marché autour de la donnée personnelle culturelle, EY a rencontré et interrogé des dirigeants d'institutions et d'entreprises représentatives du secteur, dont la Réunion des musées nationaux - Grand Palais, l'INA, Solocal (PagesJaunes), Criteo, InterCloud, Kantar Media... Cette étude qui a mobilisé nos experts du secteur Médias et Divertissement, se fonde sur leurs points de vue, notre recherche sectorielle et nos propres analyses.

Remerciements

EY tient à remercier vivement pour leurs éclairages :
Roei Amit (Directeur adjoint chargé du numérique, Réunion des musées nationaux - Grand Palais), **Christophe Benavent** (Professeur, Responsable du Master Marketing opérationnel international, Université Paris Ouest), **Julien Billot** (Directeur général adjoint en charge du segment média, Solocal - ex PagesJaunes), **Jérôme Dilouya** (Fondateur et Président-directeur général, Intercloud), **Denis Gaucher** (Directeur exécutif Ad Intelligence Europe, Kantar Media), **Alban de Nervaux** (Directeur de la stratégie et du développement, Réunion des musées nationaux - Grand Palais), **Alexandra Pelissero** (Directrice de la communication, Criteo), **Stéphane Ramezi** (Responsable des éditions multimédia, INA).

Nous tenons aussi à remercier, pour leur apport lors des groupes de travail : **Benoît Tabaka** (Directeur des politiques publiques, Google France), **Pierre Geslot** (Responsable Projets Lectures numériques, France Télécom Group), **Laure Kaltenbach** (Directrice générale, Forum d'Avignon), **Olivier Le Guay** (Responsable éditorial, Forum d'Avignon).

Édito

Empreintes, fresques, statuettes, parchemins, hiéroglyphes et toute autre forme de production culturelle ou artistique sont autant de traces matérielles laissées à la postérité, de façon consciente ou non, qui nous ont permis de reconstituer des pans entiers de l'histoire de nos civilisations, pour forger notre savoir et notre culture.

C'est l'évolution de l'humanité tout entière qui est indissociable des traces, empreintes et autres indices que nous laissons derrière nous. Avec l'avènement de l'ère numérique et l'apparition des traces immatérielles dont nous marquons la toile et que nous pouvons désormais collecter, stocker et analyser à l'infini[1], nous est aujourd'hui donné un pouvoir inédit. Au pouvoir régalien, détrôné peu à peu par l'ouverture du savoir au plus grand nombre à travers l'imprimerie, la radio et la télévision - et duquel le règne d'Internet a semblé définitivement nous affranchir - succède aujourd'hui le pouvoir de contrôler et de prévoir, grâce à l'ouverture de données publiques et personnelles de tous à un nombre d'acteurs restreints, capables de faire parler le « Big data ».

Volume, variété et vélocité : c'est en ces trois mots que peut se résumer le Big data, pour exprimer le volume inédit de données produites et échangées par un nombre croissant de canaux (web, objets connectés au web et entre eux, plateformes), la variété de ces données (avec une part croissante de données non structurées et volatiles[2]) et enfin la vélocité, qui désigne la vitesse, toujours plus grande, de ces échanges.

Le Big data représente une formidable matière première pour qui saura en extraire la substantifique moelle, avec à la clé des opportunités de création de valeur qui pourront irriguer l'ensemble des secteurs d'activité de l'économie réelle… et en particulier celui de l'industrie médiatique et culturelle.

Car au cœur de cette masse de données vertigineuse brille une catégorie de données qui vaut de l'or : la donnée personnelle culturelle. Lorsque nous observons le bouleversement des rapports de force entre opérateurs Internet, opérateurs de réseaux et groupes médias, sous l'effet de l'explosion des traces et informations liées à la révolution digitale, force est de constater la ruée de tous ces acteurs vers cette donnée personnelle culturelle, nouveau sésame qui leur ouvrirait les portes de l'intimité de l'être.

À la fois miroir de nos goûts et de nos aspirations et reflet de l'image sociale que nous souhaitons renvoyer, la donnée personnelle culturelle représente en effet un fragment de notre identité. Une donnée d'autant plus précieuse qu'elle introduit un rapport inédit en réconciliant l'empreinte et le calcul[3] : si, à la manière d'une photographie, la donnée numérique conserve la trace de notre activité digitale, elle s'en distingue toutefois par sa disponibilité au calcul.

Contrairement à une simple photographie, qui comme le rappelait Roland Barthes[4], « ne se distingue jamais de son référent, de ce qu'elle représente. [Le référent] s'entête à être toujours là, il adhère », la donnée personnelle numérique est détachable et calculable.

1 Un yottaoctet est la capacité annoncée du nouveau *data center* de la NSA *(National Security Agency)* pour 2013, soit mille fois la totalité des données enregistrées en 2011 dans le monde - Sarah Belouezzane et Cécile Ducourtieux, « Vertigineux "Big data" », *Le Monde,* 26 décembre 2012

2 Données de géolocalisation, événementielles

3 Louise Merzeau « *Faire mémoire des traces numériques* », E-dossiers de l'audiovisuel, Sciences humaines et sociales et patrimoine numérique, INA, mis en ligne en juin 2012

4 Roland Barthes, *La chambre claire* - éd. Gallimard, 1980

En effet, la dimension révolutionnaire du *"Big data bang"* réside dans l'autonomisation des processus de production et d'échanges massifs, continus et toujours plus rapides de données ubiquitaires. Dans le domaine des contenus culturels, l'offre devient alors pléthorique. Mais à travers le téléchargement, le visionnage, l'échange de ces contenus culturels, les données personnelles du consommateur, tendent à se disperser de façon incontrôlée. Et alors qu'Internet peut apparaître comme un lieu de normalisation et de surveillance, se pose avec acuité la question du pouvoir de contrôle que permet la connaissance intime des comportements et des données personnelles culturels. Et avec elle, la question du respect de ces données et de la nécessité de développer des réponses concertées impliquant les acteurs de l'écosystème des contenus culturels numériques et des alliances autour de valeurs et impératifs communs : *"Big data is ethically neutral, the use of Big data is not*[5]*"*.

Ainsi, si le Big data apparaît comme une rupture majeure qui nous ferait définitivement quitter une ère, dont l'épuisement des ressources fait poindre les limites, pour entrer dans une économie du savoir et de la connaissance prometteuse, il est urgent d'apprendre à préserver la fragilité de cette ressource qu'est la donnée personnelle culturelle, dont la pérennité repose sur les équilibres subtils et les responsabilités partagées, qui jetteront les premiers jalons de ce nouveau marché en pleine structuration.

C'est à travers la protection de cet écosystème par un filtre de confiance que ses acteurs pourront conserver un équilibre entre liberté et contrôle. Un contrôle qui implique également de ne pas se soumettre à la tyrannie de la donnée, dans une tentative de profilage qui, poussée à l'extrême, enfermerait chaque individu dans des résultats qui ne laisseraient aucune chance à la sérendipité ; au risque de tomber sous la domination d'une culture unique… Des craintes qui, en leur temps, avaient déjà été imputées à l'imprimerie, ou au latin à la Renaissance[6]…

Au-delà des gains d'efficacité et de performance espérés, à travers une connaissance affinée du comportement de ses publics et usagers et donc d'une plus grande capacité d'anticipation de leurs attentes, les données personnelles culturelles représentent pour le secteur de l'industrie médiatique et culturelle une formidable opportunité de révolutionner leur modèle économique pour stimuler la création… à l'infini.

Bruno Perrin
Associé EY
Responsable du secteur
Technologies Médias Télécoms
en France

5 Kord Davis, "*Ethics of Big data – Balancing risk and innovation*" - ed. O'Reilly Media, Septembre 2012

6 René Trégouët, Sénateur, *« Des pyramides du pouvoir aux réseaux de savoirs - Tome 1 »,* Rapport d'information 331 - 1997/1998 - Commission des Finances, site internet du Sénat, consulté le 10 octobre 2013

1. Faire parler le Big data : un nouveau pouvoir dans le secteur médiatique et culturel

Du Big data au Worthy data

Des traces à l'ADN de l'être numérique

Une explosion des traces numériques

Depuis quelques années, les données numériques connaissent une croissance exponentielle et forment une masse gigantesque de *data*, autrement appelée Big data.

L'augmentation de cette masse de données s'explique par trois facteurs, connus sous la formule des 3V :

- Le boom du **volume** de données émises : l'individu, de plus en plus nomade, produit davantage de données, laissant derrière lui en temps réel des traces numériques toujours plus fraîches. En parallèle, les capacités de stockage augmentent également de façon exponentielle, ce qui permet de conserver un historique de toutes ces traces numériques. En 2013, les volumes de données créées ou manipulées auront dépassé les 4 zettaoctets[1], soit l'équivalent d'une pile de DVD mesurant la hauteur de... 4 millions de tours Montparnasse[2].
- La **variété** des outils connectés s'étend : du portable au réfrigérateur, de la tablette à la voiture intelligente en passant par la smart TV, le nombre d'objets connectés explose, comme en témoigne la demande croissante d'adresses IP. Capables de communiquer entre eux, ces objets qui forment « l'Internet des objets », peuvent tracer leurs utilisateurs et envoyer des informations précises sur leurs mouvements et habitudes, même quand ceux-ci ne les utilisent pas. Une variété qui qualifie également la masse non structurée des données produites par ces multiples sources, qui émettent des données présentant autant de codes, langages et formats différents.
- La **vélocité** de l'information, à savoir la vitesse à laquelle les données sont traitées simultanément, augmente elle aussi : ces données circulent toujours plus vite, émises par des sources toujours plus interconnectées et interdépendantes, dans des réseaux qui fonctionnent de moins en moins en silos.

En naviguant sur Internet, un utilisateur laisse des traces de natures variées. Ces données sont collectées selon différents modes opératoires et lui sont rattachées *a priori* ou *a posteriori*, selon qu'il se sera identifié volontairement, ou qu'il sera identifiable par recoupement (ex. : cookies, adresse IP, adresse MAC).

Aux 3V du Big data s'ajoute la **valeur** que représentent ces données, pour l'entreprise mais aussi pour l'utilisateur.
En effet, l'exploitation de ces « traces » numériques peut offrir aux utilisateurs un réel confort de navigation et leur fournir des services de qualité (mise à disposition d'une boîte mail, obtention d'applications gratuites, jeux gratuits, etc.), sans contreparties financières directes. Conscientes de la forte valeur de ces données, les entreprises tentent d'instaurer ou de gérer une relation donnant-donnant, visant à récompenser les utilisateurs qui transmettent leurs données personnelles, en leur offrant des produits ou services pour les inciter à révéler leurs préférences de consommation et des centres d'intérêt, *via*, par exemple, des commentaires sur les réseaux. Toutefois, proposer des services additionnels contre les données cédées suppose un équilibre délicat pour les entreprises : les consommateurs, pas toujours conscients de la portée de ce rapport donnant-donnant, peuvent se montrer réticents à l'idée d'être considérés comme des produits monétisables. D'où l'importance de porter à leur connaissance ce qu'ils sont en mesure de faire pour contrôler leurs données personnelles. ▸ En savoir plus p. 37
Cette masse de données numériques, à l'instar de la masse monétaire, aurait-elle atteint un seuil suffisamment critique pour faire fonctionner un système économique à part entière ? Existe-t-il une hiérarchie de valeur entre ces traces numériques ? Quelle est la valeur de cette donnée numérique ?

1 *« Big data, nouveaux défis »,* Revue de l'Association Telecom ParisTech Alumni, n° 169, juillet 2013

2 Une pile de DVD de la hauteur de la tour Montparnasse = 1 pétaoctet (équivalences : 1 zettaoctet = 10^{21} octets ou 10^{6} pétaoctets) - Sarah Belouezzane et Cécile Ducourtieux, « Vertigineux "Big data" », *Le Monde,* 26 décembre 2012

24h dans la vie du Big data

- **145 milliards** de mails envoyés
- **4,5 milliards** de recherches sur Google
- **104 000 heures** de vidéos mises en ligne sur YouTube
- **400 millions** de tweets postés
- **552 millions** d'utilisateurs se connectent à Facebook

Source : chiffres CNRS

De l'octet au yottaoctet, l'échelle des données

5 exaoctets de données générées depuis les premières mesures jusqu'en **2003**

- **En 2011,** il fallait **2 jours** pour les générer
- **En 2013,** il faut **10 minutes** pour les générer

Une page de texte Word	*Un morceau de musique*	*Un film de 2 heures*	*6 millions de livres numérisés*	*Une pile de DVD de la hauteur de la tour Montparnasse*			*La totalité des données enregistrées en 2011*	*Capacité du data center de la NSA*
30 Ko	5 Mo	1 Go	1 To	1 Po			1 Zo	1 Yo
Octet 1 o	**Kilo-octet (Ko)** 1 000 octets	**Mégaoctet (Mo)** 1 000 Ko	**Gigaoctet (Go)** 1 000 Mo	**Téraoctet (To)** 1 000 Go	**Pétaoctet (Po)** 1 000 To	**Exaoctet (Eo)** 1 000 Po	**Zettaoctet (Zo)** 1 000 Eo	**Yottaoctet (Yo)** 1 000 Zo

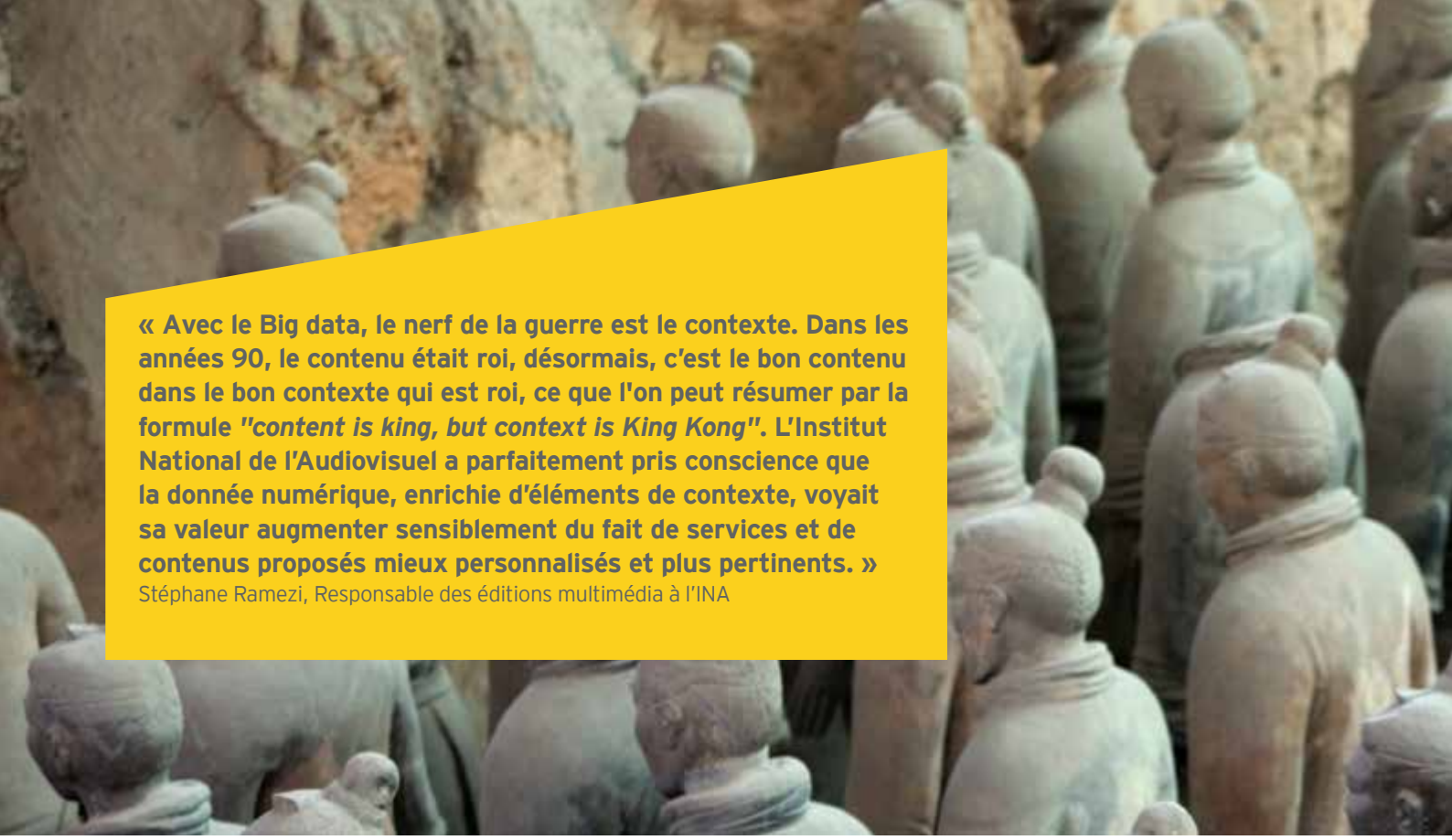
Source : CNRS

Quelles traces laissées par une simple recherche culturelle sur le web ?

Exemple : réservation d'une entrée pour une exposition au Grand Palais

Action utilisateur	Traces numériques *à caractère majoritairement personnel*	Services en contrepartie
Session utilisateur ▸ Moteur de recherche ▸ Messagerie ▸ Site tiers ▸ Réseau social…	▸ Données de contact (âge, sexe, coordonnées…) ▸ Données de connaissance (CSP, intérêts, relations, profil de consommation…)	▸ Rapidité de navigation ▸ Confort d'utilisation ▸ Richesse des informations et des services en ligne ▸ Gratuité des informations obtenues et accès aux services
Logiciel ▸ Navigateur ▸ Système d'exploitation	▸ Environnement logiciel ▸ Informations collectées par cookies ▸ Favoris, historique, paramètres, préférences de navigation	
Physique ▸ Équipement (smartphone/ tablette/ordinateur) ▸ Type de connexion ▸ Adresse IP/Adresse MAC ▸ Localisation (ADSL) : DSLAM ▸ Géolocalisation (mobile, wifi)	▸ Niveau d'équipement (type, résolution d'écran…) ▸ Vitesse et données techniques de connexion ▸ Statistiques/comportement de navigation ▸ Pays/zone géographique/localisation	

Source : EY ©



« Avec le Big data, le nerf de la guerre est le contexte. Dans les années 90, le contenu était roi, désormais, c'est le bon contenu dans le bon contexte qui est roi, ce que l'on peut résumer par la formule *"content is king, but context is King Kong"*. L'Institut National de l'Audiovisuel a parfaitement pris conscience que la donnée numérique, enrichie d'éléments de contexte, voyait sa valeur augmenter sensiblement du fait de services et de contenus proposés mieux personnalisés et plus pertinents. »
Stéphane Ramezi, Responsable des éditions multimédia à l'INA

Donner de la cohérence aux traces numériques

Individuelles, hétérogènes, multiples et éparses, les traces numériques collectées en temps réel n'ont, prises isolément, aucune valeur. Leur valeur vient du sens qu'on arrive à en tirer, en termes de corrélation ou de prédictibilité.

Il est possible de donner du sens à ces données en les rattachant à leur cause commune : le comportement d'un être humain. De cette façon, on peut non seulement espérer comprendre le comportement d'un individu à travers les traces qu'il laisse, mais aussi, *in fine,* recomposer l'ADN de son « être numérique ».

Il arrive que les images physique et numérique d'un individu se recoupent. Le simple achat d'une carte de transport par exemple : un individu qui passera, tous les soirs à 18h sauf le week-end, le portique du métro avec son titre de transport et laissera dans le même laps de temps des traces de recherches sur Internet à partir de son téléphone portable, sèmera suffisamment d'indices pour permettre de décrypter ses habitudes ; il s'agit sans nul doute de son trajet à la sortie du travail. Mais il arrive parfois que l'être physique et l'être numérique projettent des images différentes. On peut par exemple avoir une interprétation erronée d'une information transmise sur un réseau social, telle que « j'aime » ou « j'y étais » : prise isolément, une telle information sera non seulement trop parcellaire pour prétendre décrypter un comportement ou une personnalité, mais surtout, elle ne correspondra pas forcément à l'identité de l'être physique en termes de niveau social, d'éducation, etc.

Dans tous les cas, que l'image projetée de l'être numérique se recoupe ou non avec l'image réelle de l'être physique, l'image numérique aura un sens si elle s'inscrit dans une cohérence dans le temps. C'est à cette condition que les traces laissées par l'être numérique pourront donner un caractère prédictif aux modèles statistiques, et ainsi générer de la valeur.

Recomposer l'ADN de l'être numérique à travers des algorithmes toujours plus pointus

Au-delà de la collecte de données et de leur stockage, l'intelligence algorithmique est indispensable pour donner un sens à la masse de données que forment les « traces » laissées par chaque individu connecté. Cette intelligence algorithmique vise à regrouper et confronter des données numériques issues de sources diverses pour créer et caractériser l'ADN d'un être numérique, afin de décrypter et d'anticiper ses comportements dans des environnements différents.
Il faut donc contextualiser la donnée qui, seule, n'aura aucun sens, mais prendra toute sa valeur dans la mise en relation avec une multitude d'autres données. C'est ce pouvoir de contextualisation qui permettra à l'entreprise de proposer la bonne offre à la bonne cible, au bon moment, à travers le bon canal… en somme d'enrichir, personnaliser et valoriser son offre de contenus et services.

Prévoir le comportement de l'être numérique

Parallèlement, ou au-delà de la corrélation avec le contexte, l'intelligence algorithmique vise à établir des liens de cause à effet pour mieux prévoir les comportements de l'être numérique. C'est en ce sens qu'œuvrent les informaticiens et *data analysts* qui élaborent les algorithmes.

Ces modèles à visée prédictive sont déjà largement utilisés. Les recommandations de sites d'achats de biens et services culturels en ligne tels que Netflix ou Amazon reposent sur des modèles capables de prévoir ce qu'un individu serait en mesure d'apprécier au regard de ses achats antérieurs, mais aussi d'achats similaires effectués par d'autres consommateurs, afin de lui proposer des produits en conséquence.

Évolution des modes calculatoires

Le Big data implique le traitement de données volumineuses (nombreuses sources d'historiques, bases de corrélations, etc.) en un temps raisonnable, voire en temps réel.

Bien souvent, une combinaison de méthodes statistiques classiques (statistiques descriptives, segmentation, *scoring*, etc.) et de solutions de calcul permettent de résoudre ces difficultés. Par exemple, la parallélisation des calculs répète les mêmes calculs sur des groupes de données séparés, des séquences, avant de les réconcilier, afin qu'ils soient globalement effectués de manière plus rapide. Cette méthode de calcul est combinée avec des estimateurs statistiques pour converger vers une réponse la plus juste possible dans le délai imparti.

Il est à noter que les formes de statistiques descriptives auxquelles on aboutit aujourd'hui sont plus pures qu'à l'époque où l'on ne disposait que d'échantillons de données qu'il fallait extrapoler (du fait des coûts de récolte, de stockage et de traitement). Raison pour laquelle la quantité de données disponibles et leur traitement ne sont aujourd'hui plus une limite, permettant ainsi de travailler sur des données plus exhaustives.

La valeur d'une trace numérique : un arbitrage entre *cash-flow* futur et coût de collecte/analyse

Le décryptage de l'ADN d'un être numérique a un coût (collecte de la donnée, stockage, décryptage de l'information *via* des algorithmes performants).

Le défi majeur réside ainsi dans les coûts et investissements élevés que représentent les infrastructures de traitement de données volumineuses, en trouvant notamment un moyen de valoriser et monétiser les nouvelles analyses rendues possibles grâce aux données issues du Big data. Avec, à la clé, l'optimisation des *cash-flows* générés par la justesse des prédictions des comportements numériques, tout en gardant la maîtrise des coûts de développement d'algorithmes, d'achat informations, de collecte et de stockage.

Face à la déferlante des données et au phénomène de ruée vers ce nouvel or numérique, il faut garder à l'esprit que toutes ces données n'ont pas la même valeur. Une échelle de valeur va donc s'établir, la valeur d'une donnée variant selon ce qu'elle révèle ou non sur l'ADN de l'être numérique. Si, par exemple, une adresse ou un numéro de téléphone pouvaient avoir de la valeur à l'époque du télémarketing de masse, ils se révèlent aujourd'hui bien moins précieux que des données sur les centres d'intérêt ou les dernières recherches d'un individu. C'est pourquoi, dans cet amas enchevêtré de données de toutes natures, se détache aujourd'hui une pépite prometteuse de laquelle on pourra extraire des *worthy data* : il s'agit de la donnée personnelle culturelle.

La donnée personnelle culturelle : une *data* qui vaut de l'or

Comportement culturel de l'*homo conexus* à l'origine de la donnée personnelle culturelle

Depuis plusieurs années, le temps consacré à la consommation de contenus culturels ne cesse d'augmenter, une tendance amplifiée par le fait que le consommateur est désormais un *homo conexus*, connecté, mobile et multitâches.

Un phénomène à l'origine de l'explosion des traces personnelles culturelles laissées par l'*homo conexus*, parallèlement à l'augmentation des données émises par l'Internet des objets, qu'elles soient produites par son comportement culturel… :

- … numérique : moteurs de recherche, transactions de biens et services culturels (billeterie, livres, musique…), consultation et partage de vidéos/musique/séries en streaming, réseaux sociaux, portails médias, etc. ;
- … ou physique : une séance de cinéma ou la visite d'une exposition peuvent laisser des traces *via* les moteurs de recherche, les images et commentaires postés sur les réseaux sociaux, la géolocalisation. Mais aussi, une simple soirée devant la télévision : aujourd'hui, 52 % des commentaires sur Facebook portent sur les programmes diffusés à la télévision.

Une forte demande en données personnelles culturelles numériques

Bien plus que la consommation de biens et services matériels, qui satisfont les besoins fondamentaux et primaires, celle de biens et services culturels en dit beaucoup sur nos préférences, nos centres d'intérêt et nos aspirations. Elle touche ainsi à notre identité, non seulement dans ce qu'elle a de plus intime, mais aussi dans ce qu'elle a de plus social et communautaire : les biens et services culturels que nous consommons, ou nos comportements culturels, participent en effet largement de l'image que nous souhaitons renvoyer dans notre environnement social. C'est pourquoi nous partageons volontiers notre expérience et notre avis sur un livre, un film, un jeu, un concert ou une exposition avec nos amis ou communautés en ligne, tout comme avec les éditeurs de contenus en ligne.

La donnée personnelle culturelle renferme des informations contextuelles fortes et permet de qualifier de façon assez précise le pouvoir d'achat de l'être numérique mais aussi de prévoir son comportement. C'est pourquoi, en raison de sa valeur intrinsèquement élevée, elle figure aujourd'hui au centre de toutes les attentions des acteurs du Big data.

Il existe d'ores et déjà une demande sur le marché de la donnée personnelle culturelle, largement portée par des groupes tels que la Fnac et Amazon. Ils valorisent ces données personnelles culturelles pour packager, adapter et cibler leurs offres en conséquence. Mais, si la demande est bien là, qu'en est-il de l'offre ?

***Benchmark* de données : la donnée personnelle culturelle se distingue**

Source : EY ©

	Données publiques culturelles[1]	Données transactionnelles[2]	Données marketing[3]	Données personnelles culturelles[4]
	★☆☆	★★☆	★★☆	★★★
Degré de digitalisation	Faible	Moyen	Fort	Moyen
Multiplicité des sources	Fort	Moyen	Moyen	Fort
Variété de format	Fort	Faible	Moyen	Fort
Volume de données	Faible	Fort	Moyen	Fort
Intimité	Faible	Moyen	Moyen	Fort

Une production de données toujours plus exhaustive, pour une meilleure appréhension du comportement culturel de l'être numérique

Si, du côté de l'offre, la donnée personnelle culturelle est longtemps restée cantonnée à l'analyse des paiements, on observe aujourd'hui chez les entreprises et établissements du secteur médiatique et culturel un prolongement de l'expérience culturelle en-deçà et au-delà des transactions de paiement. Ces acteurs tendent à enrichir leurs offres de services culturels *via* un *continuum* de services.

Ceci permet de prolonger l'expérience culturelle des clients et de nouer une relation d'engagement avec eux. Une relation qui fournira de précieuses informations client, à condition d'être capable d'engranger et de lier efficacement une quantité et une variété inédites d'informations, générées dans des espaces et temporalités différents.

Les comportements culturels, davantage que les traditionnels CSP, apportent de la valeur aux profils numériques.

1 Selon une définition consacrée, est une « donnée publique culturelle » la donnée produite ou détenue par un établissement, organisme ou service culturel ayant une activité culturelle réelle et effective (Source : Guide Data Culture).
Ex. : inventaire du fonds artistique, horaire d'ouverture, catalogue d'exposition, œuvre du domaine public…

2 Données sur les actes d'achats de biens et services culturels.
Ex. : nombre d'entrées en salle, vente de disques, entrées aux musées, détenteurs de cartes d'abonnement, ventes Amazon…

3 Données de contact et qualification collectées au travers des opérations des acteurs de l'industrie culturelle.
Ex. : abonnés de l'espace personnel du Louvre, participants aux jeux-concours de promotion musicale, abonnés newsletters, données de navigation…

4 Ensemble des données permettant de connaître les préférences culturelles des consommateurs.
Ex. : discussions dans des forums, réseaux sociaux, avis et commentaires sur des œuvres/artistes…

La donnée personnelle culturelle au cœur de l'émergence d'un marché stratégique

Un marché en pleine structuration, sous l'effet de lourds investissements

La valeur élevée de la donnée personnelle culturelle numérique est source d'importants mouvements d'acteurs et fait l'objet de lourds investissements. Un marché hautement stratégique est en train de se dessiner à la faveur d'un double phénomène de rapprochement et de convergence, avec :

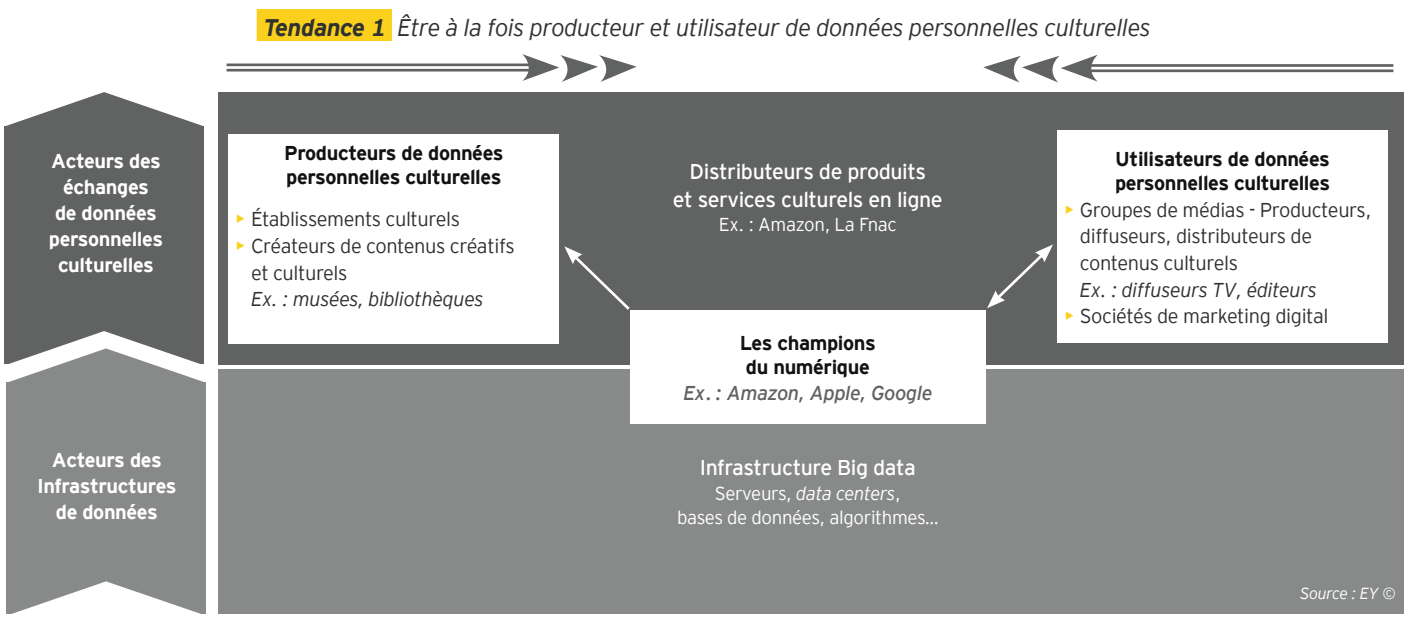
- D'une part **(tendance 1)**, les entreprises et établissements médiatiques et culturels qui innovent pour proposer un *continuum* de services, afin d'enrichir et de prolonger l'expérience culturelle en recueillant des informations précieuses sur les préférences de consommation de biens et services culturels.

Exemple : grâce au partenariat signé entre TF1 et Facebook en octobre 2013, le réseau social partagera ses données avec le groupe audiovisuel français qui bénéficiera des outils Facebook dédiés au suivi et à l'analyse des conversations autour des programmes TV. Ces outils permettront à TF1 de mesurer le nombre de publications autour d'un sujet précis et de les analyser, mais aussi d'afficher et de mesurer, pendant la diffusion de ses programmes, le taux de conversations en temps réel sur le réseau social, qui rassemble 26 millions d'utilisateurs en France et génère « 52 % des commentaires sur la télévision[1] ».

- D'autre part **(tendance 2)**, les leaders du numérique, les GAFA[2], qui possèdent déjà des infrastructures de données et sont utilisateurs de données personnelles culturelles, investissent pour se rapprocher de la production de données personnelles culturelles, mais aussi pour produire et diffuser eux-mêmes des contenus culturels.

Exemple : l'Institut culturel de Google intègre différents projets dont Google Art Project, qui permet de se promener dans les galeries de 151 musées ou lieux culturels en ligne à travers le monde (Versailles, Quai Branly, MoMa, The National Gallery *à Londres, Fondation Neslon Mandela, etc.). Ce service repose sur des bases de données utilisant différentes technologies de l'opérateur et collectant des données personnelles culturelles à l'échelle internationale, avec un effet multiplicateur fort* via *les réseaux sociaux. Et avec une volonté forte de développer des standards communs ou interopérables pour les musées à travers le monde. Cette invitation dans le monde la culture vient compléter les investissements du groupe dans les secteurs de la vidéo (rachat de YouTube pour 1,65 milliards de dollars en 2006), du livre (plusieurs centaines de millions d'euros[3]), des contenus audiovisuels (films et séries provenant des catalogues de grands studios américains ou de chaînes de télévision telle que BBC, etc.). Amazon s'invite également dans le marché de l'art en lançant, en août 2013, Amazon Art, une plateforme proposant à la vente plus de 40 000 œuvres en provenance de 150 galeries à travers le monde, des plus exceptionnelles aux plus accessibles, avec la volonté clairement affichée de toucher une audience plus large.*

Structuration d'un nouveau marché stratégique



1 « Social TV : Facebook partagera ses données avec TF1 et Canal+ », *Le Monde*, 7 octobre 2013

2 GAFA désigne les Big 4 du numérique : Google, Apple, Facebook, Amazon

3 Début 2012, Google avait numérisé plus de 20 millions d'ouvrages (Jennifer Howard, *"Google Begins to Scale Back Its Scanning of Books From University Libraries"*, 9 mars 2012, site visité le 10 octobre 2013) et le coût de numérisation de l'intégralité du catalogue de la Bibliothèque nationale de France (BNF), soit environ 15 millions d'ouvrages, est estimé à 750 millions (selon Yann Gaillard, rapporteur spécial de la Commission des finances pour la mission « Culture » et auteur du rapport intitulé : « La politique du livre face au défi du numérique. »)

Infrastructure Big data : de quoi parle-t-on ?

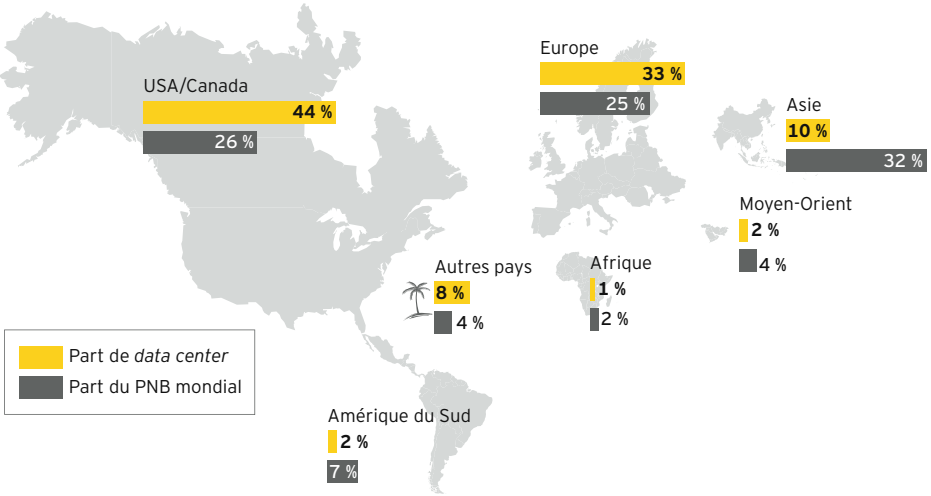
L'information sur la localisation géographique des *data centers* est très difficile à obtenir et jugée stratégique par les grands acteurs du secteur. Sur la base des informations publiques disponibles et pour les *data centers* vendant leurs services, EY a réalisé une carte anamorphose des *data centers*, qui montre l'avancée des États-Unis en matière de capacité de stockage de données, par rapport à l'Europe et aux autres régions du monde.

Les capacités de stockage dont se dotent les grands collecteurs de données sont bien supérieures aux besoins à très court terme comme en atteste la capacité du dernier *data center* de la *National Security Agency* (NSA), qui atteint 1 yottaoctet[4].

En outre, l'analyse des charges et revenus des GAFA montre que si la R&D, le stockage et le traitement des données constituent une part significative des coûts, l'exploitation directe des données Big data ne constitue qu'une faible partie des revenus, l'essentiel[5] de ces derniers provenant de la publicité ou de la distribution.

Carte en anamorphose des *data centers*

Source : EY ©



Si on compare, par grande zone géographique, les pourcentages de *data centers* dans le monde aux pourcentages de PBN mondial, on constate que certaines régions sont surreprésentées en *data centers* par rapport à leur poids économique (% de leur PNB/PNB mondial). Ainsi, l'Amérique du Nord a généré 26 % du PNB mondial en 2012 tandis qu'elle abritait 44 % des *data centers* proposant des services commerciaux.

4 Sarah Belouezzane et Cécile Ducourtieux, « Vertigineux "Big data" », *Le Monde*, 26 décembre 2012

5 Les revenus issus de l'exploitation directe des données Big data sont issus de la vente de services de stockage, de traitement et analyses de données Big data, Jeffrey Kelly, David Floyer, Dave Vellante, Stu Miniman, *"Big Data Vendor Revenue and Market Forecast 2012-2017"*, Wikibon, octobre 2013

« Les opérateurs tels que Google, Amazon ou Microsoft construisent leurs propres *data centers*. Google construit même les machines à l'intérieur de ses *data centers*. »
Jérôme Dilouya, Fondateur et Président-directeur général d'Intercloud

Caractéristiques du marché de la donnée personnelle culturelle

Pour mieux cerner le marché du Big data et en comprendre les règles du jeu, il convient d'analyser les caractéristiques des principales forces en présence.

Des coûts marginaux dégressifs

Les investissements initiaux sont principalement relatifs à « l'infrastructure Big data » : serveurs, *data centers*, algorithmes. Si la collecte massive de données personnelles, ainsi que le développement des algorithmes pour les exploiter nécessitent de lourds investissements de départ pour fournir le service au premier client, les coûts nécessaires à l'acquisition de clients supplémentaires sont dégressifs.

Un marché où seuls quelques acteurs pourront survivre à moyen terme

Toutes les entreprises n'ont pas les moyens d'investir dans cette infrastructure dont le niveau déterminera la qualité de l'exploitation des *data*. Ainsi, les chances de réussite d'un nouvel acteur sont loin d'être certaines, d'autant que celui-ci risque d'encourir des *sunk costs* importants.

Par ailleurs, l'acteur qui possède la plus grande masse d'informations (réseau social, application, *operating system*, etc.) et dispose des algorithmes les plus puissants attirera davantage de clients et collectera d'autant plus de données fiables additionnelles. Ceci aura pour conséquence de renforcer la qualité de ses analyses et de ses algorithmes et, *in fine*, sa position sur le marché.

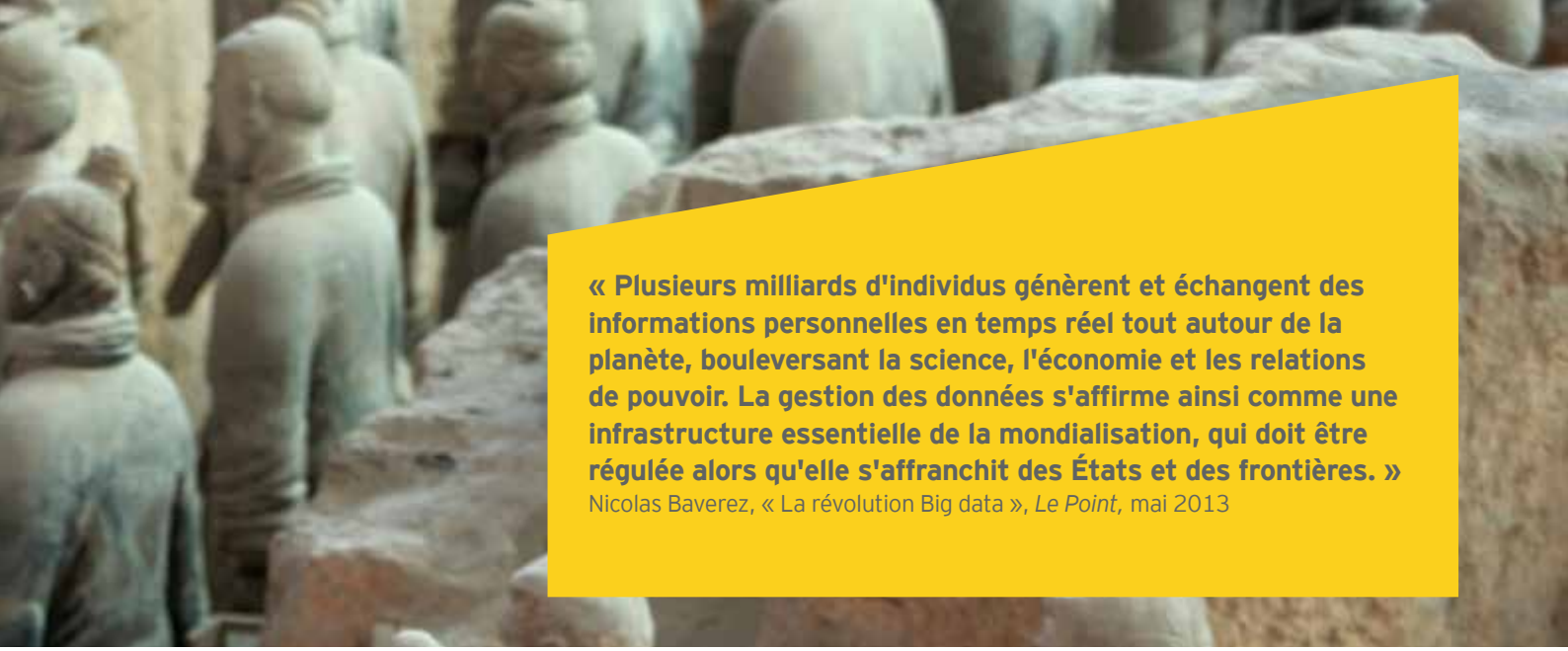
Cet effet « boule de neige » (*winner-take-all effect*) ajouté à la barrière du coût initial, aux fortes économies d'échelles et à un niveau important de *sunk costs* tend inévitablement à concentrer le marché à moyen terme.

Un marché touchant au « bien commun »

Le marché de la donnée personnelle culturelle est sensible car les informations concernant les comportements culturels d'une population donnée et leur exploitation sont souvent issues de statistiques ou d'enquêtes publiques et traitées/analysées à l'échelle nationale, par les instituts de statistiques publics. Rappelons par ailleurs que la propriété intellectuelle culturelle tombe dans le domaine public après une longue période d'exploitation par les auteurs et producteurs de ces données/contenus[1].

« Nous sommes face à l'enjeu suivant : qui va gagner la course au CRM du web ? [...] Celui qui aura atteint la masse critique de données aura gagné la course de vitesse du CRM digital. »
Roei Amit, Directeur adjoint chargé du numérique à la Réunion des musées nationaux - Grand Palais (Rmn-GP)

1 *« La propriété intellectuelle à l'ère du numérique – Défis et opportunités pour le secteur Médias et Divertissement »*, EY, Novembre 2011



« Plusieurs milliards d'individus génèrent et échangent des informations personnelles en temps réel tout autour de la planète, bouleversant la science, l'économie et les relations de pouvoir. La gestion des données s'affirme ainsi comme une infrastructure essentielle de la mondialisation, qui doit être régulée alors qu'elle s'affranchit des États et des frontières. »
Nicolas Baverez, « La révolution Big data », *Le Point,* mai 2013

Des caractéristiques de marché proches de celles d'une infrastructure essentielle ?

Investissements initiaux important puis coût marginal faible, *sunk costs* élevés, fortes économies d'échelle, *winner-take-all effects,* domaine touchant au bien commun, traditionnellement animé par des organismes publics, etc. : ces éléments, caractéristiques des monopoles naturels, sont longuement décrits dans la littérature économique.

Au regard des grandes tendances qui ont marqué certains secteurs avec monopoles naturels, comme les télécoms ou les chemins de fer, l'enjeu consiste à identifier l'infrastructure essentielle et ses marchés Amont et Aval afin de favoriser la concurrence et la diversité des acteurs sur ces marchés.

Dans cette optique, l'infrastructure et l'échange de données constituent-ils un marché pertinent ? Est-ce une infrastructure essentielle ? Existe-t-il des acteurs dominants sur ce marché qui doivent être régulés ?

De tels marchés ont besoin d'éléments de structuration pour arriver à leur point d'équilibre à moyen et long termes.

En outre, au cœur de ce marché en construction, ce sont les données de milliards d'individus connectés, générant des zettaoctets de données numériques collectées et utilisées par un nombre relativement restreint d'acteurs économiques, qui sont en jeu. Là aussi, dans un souci d'équilibre à long terme, la réglementation a un rôle à jouer pour instaurer une relation de transparence et de confiance entre les entreprises et les individus. C'est en leur donnant un droit de regard et de contrôle renforcés sur leur identité numérique que ces êtres numériques continueront à avoir une existence et à produire des données créatrices de valeur.

Les caractéristiques d'une infrastructure essentielle

- **Caractéristiques techniques**
  - Indivisibilité technique (indivisibilité des investissements)
  - Longue durée de vie
  - Produit faisant l'objet de peu d'échanges
  - Produit faisant souvent partie d'un tout, difficile à isoler techniquement
- **Caractéristiques économiques**
  - Fonction de coûts sous-additive (existence d'économies d'échelle)
  - Coûts irrécupérables importants (investissements non réversibles)
  - Externalités de réseau (prime au leader, effet boule de neige, importance d'une masse critique de consommateurs, anticipations auto-réalisatrices…)
  - Guerre des standards empêchant d'avoir des produits substituables
- **Caractéristiques socio-économiques**
  - Coordination et planification centralisées
  - Traditionnellement détenue par l'État
  - Traditionnellement considérée comme touchant au « bien commun »

2. Pouvoirs en équilibre et responsabilités partagées : les nouveaux contours du Big data

« Si l'on veut construire une innovation durable, qui ne soit pas rejetée par l'utilisateur, les entreprises doivent apporter des garanties en termes de protection des données personnelles. Ce n'est pas un coût, c'est un investissement. »
Isabelle Falque Pierrotin, *« Data, la nouvelle ruée vers l'or », Enjeux Les Échos*, mars 2013

La masse de traces numériques disponibles est désormais suffisante pour permettre de suivre et de modéliser des « êtres numériques » rationnels et cohérents. Dans cette course à l'extraction et à l'analyse de données pertinentes, se dessinent les contours d'un marché à part entière : il s'agit du marché des données personnelles culturelles numériques.

S'il ouvre des perspectives économiques prometteuses et étonnantes, ce nouveau marché doit cependant être accompagné dans sa structuration et son développement.

En effet, pour garantir son développement harmonieux à moyen et long termes et favoriser un nouvel écosystème économique innovant et créateur de valeur pérenne, il faut s'assurer que l'échange et l'utilisation massive des données personnelles culturelles se fassent dans le respect de la vie privée de chaque individu, premier producteur de données personnelles culturelles. Mais aussi favoriser les investissements des entreprises quant à la collecte et à l'organisation des données et enfin instaurer une saine concurrence économique.

La réglementation, impulsion économique pour les acteurs de Big data en Europe

La protection comme facteur de confiance

Un fort besoin de confiance quant à la gestion des données personnelles

Les réseaux sociaux se font souvent écho des manquements en matière de respect de la vie privée ou de sécurité informatique et propagent souvent le « buzz » autour de pratiques contraires à la loi sur la protection des données. La protection des données personnelles apparaît ainsi au centre des préoccupations des consommateurs, qui questionnent de plus en plus les entreprises et responsables du traitement sur les garanties et la sécurité qu'ils peuvent garantir aux données personnelles qu'ils leur confient. Les questions les plus fréquemment posées sont relatives à la localisation de ces données, aux mesures de sécurité mises en place pour les protéger contre des accès non autorisés, à l'usage qui en est fait par le responsable de traitement, à qui elles sont destinées, etc.

Dans un contexte où les questions de confidentialité et de vie privée suscitent un intérêt grandissant chez l'ensemble des parties prenantes, de plus en plus d'entreprises intègrent cette dimension dans leurs actions de communication, tentant ainsi de démontrer l'importance qu'elles attachent à la protection de la vie privée de leurs clients.

Au-delà de l'aspect juridique, la conformité à la réglementation en matière de protection des données personnelles devient un moyen efficace pour les entreprises de communiquer sur leur engagement éthique et sociétal. Le respect des bonnes pratiques en matière de protection des données à caractère personnel constitue indéniablement un avantage concurrentiel pour les entreprises, quel que soit leur secteur d'activité, mais également un moyen de se prémunir du risque de réputation.

Partant du postulat reconnu que la rétention d'un client requiert un investissement bien moindre que son acquisition, les entreprises déploient des stratégies pour maintenir et entretenir la fidélité de leurs clients. Celle-ci repose essentiellement sur la confiance, laquelle ne peut être construite qu'en s'appuyant sur les deux notions fondamentales que sont la connaissance et la reconnaissance. La connaissance enregistre « l'historique du client », c'est-à-dire ses interactions avec les différents services de l'entreprise, auxquelles s'ajoutent des informations relatives aux produits. La reconnaissance est la réponse personnalisée selon le client : il s'agit donc de s'appuyer sur la connaissance du client pour lui apporter une réponse ciblée, voire de devancer ses attentes.

L'évolution constante des technologies qui a rythmé les dix dernières années, à laquelle s'ajoute le décalage qui existe encore trop souvent entre les engagements, volontaires ou imposés, en matière de protection des données et les pratiques de traitement des données, ont pour effet de renforcer les attentes des consommateurs en matière de confiance et de transparence.

Comment définir aujourd'hui une donnée à caractère personnel ?

Définir la notion de « donnée à caractère personnel » n'est pas chose aisée en raison de son caractère mouvant, parfois subjectif ou relatif, mais avant tout contextuel.

La directive 95/46, qui constitue le socle communautaire en matière de protection des données personnelles, définit une « donnée à caractère personnel » comme « toute information concernant une personne physique identifiée ou identifiable (personne concernée) ; est réputée identifiable une personne qui peut être identifiée, directement ou indirectement, notamment par référence à un numéro d'identification ou à un ou plusieurs éléments spécifiques, propres à son identité physique, physiologique, psychique, économique, culturelle ou sociale ».

D'autres définitions existent : « nous ne devrions plus parler de données personnelles, mais de données relationnelles et transactionnelles. Nous devons abandonner la vision des données personnelles comme une chose définie.[1] »

En effet, les données qui, prises indépendamment, peuvent sembler insignifiantes, sont assemblées et analysées pour recréer des profils individuels ou définir des identités numériques à partir d'éléments de personnalité réels.

Souvent, il suffit d'une date et du lieu de naissance pour identifier un individu. Mais l'usage des réseaux sociaux, ainsi que des moteurs de recherche ou des services de messagerie peuvent donner une vision précise de cet individu à travers le prisme de ses goûts, de ses habitudes, de ses projets ou de ses croyances. Au risque de voir toutes ces données combinées afin d'identifier un individu, sans même connaître précisément son nom.
La quantité de données qui sont désormais collectées, traitées et stockées sur chaque personne permet d'aller beaucoup plus loin et contribue à un profilage très précis des individus avec la possibilité - en fonction des catégories de données traitées – de produire des modèles probabilistes pour en apprendre davantage sur leur croyance religieuse, leur opinion politique, leur mode de vie, leur orientation sexuelle et bien d'autres aspects de leur vie personnelle et intime.

La question se pose donc de redéfinir les catégories de données qui doivent être considérées comme sensibles et donc réglementées plus strictement par la loi. En effet, des données collectées à partir de ce que peut lire, écouter ou regarder un individu peuvent fournir des indications sur son orientation politique, ses croyances religieuses ou même son orientation sexuelle, et *de facto* relever du statut juridique protecteur des données sensibles. Force est de constater que les données culturelles, qu'elles répondent ou non à la définition des données sensibles, doivent faire l'objet d'une protection en cela qu'elles relèvent de l'intimité d'un individu.

Cette question est d'autant plus cruciale dans le contexte du développement de Big data et du *cloud computing* qui, par économie d'échelle, peuvent induire une augmentation de tous les risques soulevés par les questions suivantes :

- Où sont stockées les données personnelles ?
- Les données personnelles sont-elles sécurisées ?
- Est-ce qu'un individu possède encore le contrôle de ses données ?
- Comment un individu peut-il s'opposer au traitement de ses données ?
- Comment un individu peut-il récupérer ses données ?

La question de la confiance devient encore plus fondamentale lorsqu'il est question de bâtir une relation durable avec les utilisateurs et clients, afin de répondre à des doutes ou des craintes qui pourraient freiner le développement de ce qui semble être appelé à devenir un levier majeur de création de valeur dans l'industrie médiatique et culturelle.

Conjointement à ces initiatives privées, les acteurs publics doivent mener les actions qui s'imposent afin d'assurer un niveau adéquat de protection des données personnelles et soutenir la croissance du secteur, à l'image des politiques adoptées dans de nombreux pays pour offrir un cadre au commerce en ligne.

1 Dominique Boulier, *Vie Privée à l'Horizon 2020*, Cahier IP n° 1, CNIL, novembre 2012

Boîte à outils

Comment les données personnelles sont-elles collectées et traitées ?

ÉVALUER LE RESPECT DE LA VIE PRIVÉE

Privacy Score attribue une note aux sites web en fonction de leur niveau de conscience et du respect de la vie privée.

www.privacyscore.com

TRAQUER LES *TRACKERS*

The Guardian a mis en place une application en ligne afin de permettre aux internautes de comprendre comment ils sont suivis en ligne et par qui. Les cercles rouges sont les dix premières sociétés de *tracking* les plus prolifiques ; les cercles bleus se réfèrent aux 100 sites les plus populaires qui les utilisent.

www.theguardian.com

QUELLE EST LA VALEUR DES DONNÉES PERSONNELLES ?

Le site du *Financial Times* permet de calculer combien de commerçants feraient payer des données personnelles sur la base de nombreux critères comme l'âge, le travail, la famille, les maladies ou les activités potentielles.

www.ft.com

CONTRÔLER L'INFORMATION - CONFIGURER SON NAVIGATEUR WEB

Tous les navigateurs permettent le blocage des cookies de suivi.

Certains navigateurs comme Mozilla Firefox permettent de choisir une option spécifique, laquelle consiste à informer les sites web qu'on ne souhaite pas être suivi par des annonceurs et autres tiers. Respecter ce paramètre est facultatif, les sites web n'étant pas tenus de respecter la volonté des internautes.

Il est également possible de mettre en œuvre des *plug-in* supplémentaires pour surveiller l'intégralité des données recueillies à partir des appareils, la façon dont elles sont traitées et éventuellement, s'opposer à toute collecte de données.

États-Unis/Europe : deux approches différentes de la protection des données personnelles

En termes de politique publique, des approches différentes peuvent être mises en avant afin d'aborder la question de la vie privée et de la protection des données personnelles, à l'instar des États-Unis et de l'Europe qui ont une conception radicalement différente des concepts de « protection de la vie privée » et de « donnée à caractère personnel ».

La première différence entre ces deux approches réside dans la spécificité du système américain où les lois fédérales coexistent avec les lois de chacun des cinquante États, ce qui place les États-Unis au premier rang des pays ayant adopté le plus grand nombre de lois dans le domaine de la protection de la vie privée et des données à caractère personnel.

La deuxième différence consiste dans le fait qu'aux États-Unis, il n'existe pas un corps unique de règles protégeant la vie privée (à l'exception du *Privacy Act* de 1974 qui donne aux citoyens le droit de connaître les informations que le gouvernement fédéral détient à leur sujet et de corriger ou d'obtenir réparation si leurs données sont utilisées de manière non autorisée).

Aux États-Unis, la protection de la vie privée est déclinée par secteur d'activités, industrie, ou toute autre segmentation qui concernent par exemple le secteur de l'enfance (*"Children's Online Privacy Protection Act"* - COPPA) de 1998 qui protège les données personnelles des enfants de la collecte et du détournement de leurs données sur les sites internet commerciaux), le secteur financier (*"The Financial Services Modernisation Act"* ou *"Gramm-Leach-Bliley Act"* de 1999 qui règlemente les conditions de communication à des tiers d'informations personnelles détenues par les institutions financières) ou encore des lois qui protègent la santé comme le *"Health Insurance Portability and Accountability Act"* (HIPAA) de 1996 qui met en place des standards pour l'échange électronique d'informations médicales afin de protéger la vie privée des patients. De nombreux textes vont protéger les données mais de manière sectorielle et non de manière transversale. Dans le secteur des médias, on peut citer le *"Cable Communications Policy Act"*, le *"Telecommunications Act"* ou le "*Videotape Privacy Protection Act"*.

Au-delà de l'aspect législatif, la principale différence entre le droit à la vie privée aux États-Unis et au sein de l'Union européenne est d'ordre philosophique. Alors qu'aux États-Unis, la loi sur la vie privée est axée sur la protection du consommateur et vise à atteindre un équilibre entre la vie privée et l'efficacité de l'entreprise, l'Union européenne considère le respect de la vie privée comme un droit fondamental du citoyen, supérieur à tout autre intérêt commercial. La loi américaine sur la protection de la vie privée entend protéger le consommateur plutôt que l'individu et sanctionnera les manœuvres déceptives ou trompeuses plutôt que le non-respect de règles de protection des libertés publiques.

Enfin, une autre différence marquante est certainement l'accent particulier mis par la réglementation nord-américaine sur la protection de la sécurité des données, notamment à travers l'obligation de notifier les failles de sécurité. Plusieurs États nord-américains ont depuis longtemps édicté des lois qui mettent à la charge des organisations des obligations de notifier les failles de sécurité. Ainsi, la grande majorité des États américains dispose de lois qui vont imposer aux entreprises la mise en place de procédures particulières en cas de vol ou de perte de données personnelles. Cette réglementation contraint les entreprises à renforcer leurs mesures de sécurité internes, et cela afin d'éviter des procédures coûteuses et préjudiciables à leur image de marque en cas de publicité voulue ou subie.

Au niveau européen, l'obligation de notifier les failles de sécurité ne fait aujourd'hui pas encore partie du droit positif en matière de protection des données dans le cadre de la directive européenne 95/46, même si elle existe déjà pour les prestataires de communication électronique en application de la directive 2002/58 sur la vie privée. Toutefois, le projet de règlement pour la protection des données publié par la Commission européenne devant refondre le cadre européen de la protection des données prévoit une obligation similaire qui s'appliquera à tous les responsables de traitement et sous-traitants en Europe.

À cet égard, il est intéressant de souligner que les différences entre ces deux approches ont vocation à se réduire compte tenu des enjeux d'interopérabilité nécessités par l'augmentation croissante des flux transatlantiques de données.

En effet, au moment où la Commission européenne a publié le projet de règlement pour la protection des données le 25 janvier 2012, la Maison Blanche publiait un mois plus tard le projet de loi relatif au respect de la vie privée des consommateurs.

La comparaison entre ces deux projets de réforme constitue une brillante illustration des ambitions d'harmonisation entre les deux approches.

« L'Union européenne et les législations américaines commencent à utiliser le même langage en ce qui concerne la loi sur la protection des données, tant sur la définition juridique proposée que sur les grands principes mis en œuvre.[1] »

Donc si les deux projets de réforme sont issus de sources de droit différentes, ils convergent par l'émergence actuelle de dénominateurs communs et notamment la prise en compte de l'importance d'une responsabilisation des opérateurs (concept d'*accountability* qui va être introduit en droit européen) et la nécessité de fournir des garanties adéquates pour les personnes afin de maîtriser les menaces pour la vie privée issues des technologies.

RÉGLEMENTATION AUX ÉTATS-UNIS

Sources

- Absence de corps unique de règles protégeant la vie privée mais des réglementations sectorielles.
- Lois différentes dans les 50 États sur les questions de sécurité informatique et de protection de la vie privée.

Philosophie/Objectifs

- La protection du consommateur et la poursuite d'un équilibre entre protection de la vie privée et intérêt *business*.
- La sécurité informatique au cœur des dispositifs réglementaires de protection des données notamment au travers des obligations de notification des failles de sécurité imposées à toutes les entreprises.

RÉGLEMENTATION EN EUROPE

Sources

- Un corps unique de règles protégeant la vie privée et les données personnelles des citoyens.
- Directives transposées dans les 28 États membres.

Philosophie/Objectifs

- La défense du droit à la vie privée comme droit fondamental supérieur à tout autre intérêt commercial.
- Une prise en compte à géométrie variable des risques liés à la sécurité informatique, notamment s'agissant des obligations de notification des failles de sécurité aujourd'hui limitées aux seuls prestataires de communications électroniques.

1 Traduction libre de Gabriela Zanfir - *"European Integration Realities and Perspectives: EU and US Data Protection Reforms. A Comparative View."*

Ce qui va changer en Europe en matière de protection des données personnelles : des droits renforcés pour les individus, une responsabilisation accrue des organisations, des technologies respectueuses de la vie privée et plus de sanctions

Le projet de Règlement (proposition de la Commission européenne avant la prise en compte des amendements du Parlement européen) sur la protection des données personnelles renforcera les droits des citoyens en introduisant les règles suivantes :

- Création d'un « droit à l'oubli » pour aider les citoyens à gérer les risques en matière de protection des données en ligne. Lorsque la personne concernée ne voudra plus que ses données soient traitées et qu'il n'existe pas de motif légitime d'en maintenir le traitement, les données seront effacées. Ces règles ont pour visée de renforcer les droits des individus. Il ne s'agit pas de supprimer des événements passés ou de restreindre la liberté de la presse.
- Création d'un droit à la portabilité des données personnelles d'un fournisseur de service à un autre.
- Renforcement du principe du « consentement », lequel lorsqu'il est nécessaire doit être donné de manière explicite.
- Obligation mise à la charge des entreprises et des organisations d'avertir sans délai injustifié les particuliers des failles de sécurité relatives à leurs données personnelles qui seraient susceptibles de leur nuire. Ils devront également informer l'autorité compétente de protection des données.
- Amélioration des voies de recours administratives et judiciaires en cas de violation des droits de protection des données.
- Responsabilité accrue et application du principe d'*accountability* (« obligation de rendre compte aux parties prenantes ») des responsables de traitement - à travers des évaluations des risques en matière de protection des données, des délégués à la protection des données et à travers les principes de « *privacy by design* » et « *privacy by default* ».

Les démarches « *privacy by design* » et « *privacy by default* » vont également venir renforcer l'effectivité des règles européennes en matière de protection des données - ce qui signifie que les garanties de protection des données seront intégrées au sein des produits et des services, et que des paramètres respectant la protection des données personnelles deviendront la norme - par exemple sur les réseaux sociaux. Ces règles vont renforcer les droits des personnes d'une manière pratique.

Grâce à ces règles plus strictes de protection des données, la Commission européenne a pour objectif de contribuer à accroître la confiance dans les services en ligne, afin que les citoyens soient en mesure d'utiliser les nouvelles technologies avec plus de confiance en bénéficiant des avantages du marché intérieur.

Les autres avancées du projet de règlement sont les suivantes :

- Un seul corps de règles sur la protection des données, valable dans toute l'UE.
- Un interlocuteur unique : les entreprises n'auront en principe qu'à traiter avec une seule autorité de protection des données au niveau national, à savoir l'autorité compétente dans l'État membre où elles ont leur établissement principal.
- Les personnes auront le droit de se référer à leur autorité nationale de protection des données, même lorsque leurs données personnelles sont traitées en dehors de leur pays d'origine.
- Les règles de l'Union européenne s'appliquent aussi aux sociétés non établies dans l'Union européenne, si elles offrent des biens ou des services au sein de l'Union européenne ou si elles surveillent le comportement en ligne des citoyens.
- Des responsabilités accrues et l'avènement du principe d'*accountability* pour les responsables de traitement des données personnelles.
- Les contraintes administratives telles que les exigences de notification systématiques pour les entreprises responsables de traitement des données personnelles seront supprimées.

Source : Commission européenne - le projet est encore en discussion (processus de codécision entre le Parlement européen et le Conseil)

La protection et la valorisation des données personnelles en tant qu'actif des entreprises

Les données personnelles sont aujourd'hui au cœur de l'innovation et de la publicité en ligne et constituent « un type d'actif pour les entreprises[1] ». Appelées à devenir l'un des moteurs de l'économie numérique, elles font l'objet d'investissements massifs.

Paradoxalement, les outils juridiques à disposition des entreprises pour protéger et valoriser leurs données semblent aujourd'hui trop limités pour appréhender de manière adéquate cet actif stratégique.

Big data et protection des données personnelles culturelles

La question de la protection des données personnelles est au cœur du marché du Big data culturel, puisque les données personnelles culturelles reflètent la personnalité d'un individu. Le culturel, c'est la « donnée de l'intimité », pour autant la donnée culturelle n'est pas spécifiquement réglementée par le droit positif (directive européenne de 1995) et n'est pas appréhendée par le projet de Règlement sur la protection des données personnelles en cours de discussion à Bruxelles (le vote, initialement prévu au premier semestre 2014, pourrait être reporté en 2015 - Conseil européen, 25 octobre 2013). Par ailleurs, et c'est aussi un enjeu fondamental, il est important que le droit reconnaisse la valeur patrimoniale de ces données.
Le nouveau cadre réglementaire européen en matière de données personnelles demeurera certainement plus ambitieux par rapport aux règles en vigueur dans le reste du monde.
Il pourrait contribuer au développement des acteurs de Big data en Europe dans la mesure où il va renforcer le niveau de sécurité et confiance apporté que les acteurs européens seront tenus d'offrir aux individus dont ils traitent les données.

Les outils juridiques existants

La protection des bases de données

Les bases de données sont définies comme des recueils d'œuvres, de données disposées de manière systématique ou méthodique et individuellement accessibles par des moyens électroniques ou tout autre moyen.

La protection légale des bases de données est double :

- Le droit d'auteur protège la structure de la base de données si celle-ci peut être considérée comme originale au regard de l'organisation du choix des rubriques et de leur disposition. Il n'a pas vocation à protéger le contenu informationnel de la base.
- Le droit *sui generis* des producteurs de bases de données permet quant à lui de protéger l'investissement réalisé pour compiler le contenu de la base et pourrait donc palier, dans une certaine mesure, les limites du droit d'auteur. Pour autant, cette protection est limitée dans sa portée dans la mesure où elle n'offre que la possibilité d'interdire la reprise d'une part quantitativement ou qualitativement substantielle du contenu de la base.

Ces deux régimes de protection ont des objets différents (originalité du contenant c'est-à-dire de la structure, de la présentation de la base vs. investissement dans le contenu c'est-à-dire dans la constitution de la base). Pour autant, leur articulation ne permet pas de couvrir réellement l'intégralité de la valeur de l'information même structurée au sein d'une base de données :

- Le droit d'auteur protège une « coquille vide », la structure et l'organisation de la base et non pas l'information qu'elle contient (sauf si celle-ci est protégeable indépendamment de son inclusion dans la base, par exemple des titres d'œuvres protégées par le droit d'auteur).
- Le droit *sui generis* permet la protection indirecte de l'information dans la mesure où elle octroie au producteur de la base de données le droit d'en interdire l'exploitation par un tiers, mais sous réserve d'apporter la preuve du caractère (qualitativement ou quantitativement) substantiel de la réutilisation ou de l'extraction du contenu de la base de données et surtout de la consistance des investissements réalisés en amont pour la constituer.

Le droit des bases de données ne permet donc pas la protection de tout type d'information, des informations potentiellement stratégiques peuvent ne pas être éligibles à la protection prévue par le droit des bases de données.

1 *Competition and personal data protection*, Joaquin Almunia, *Vice President of the European Commission responsible for Competition Policy, Privacy Platform event: Competition and Privacy in Markets of Data, Brussels, 26 November 2012, SPEECH/12/860*

La protection des informations par le biais du savoir-faire

Le savoir-faire est une notion à géométrie variable, dont la définition peut beaucoup varier d'un pays à l'autre.

En France par exemple, il n'existe pas de définition juridique précise de cette notion. Elle se trouve donc limitée à une valeur qui peut faire l'objet d'accords de licence ou de cession auprès de tiers souhaitant bénéficier des connaissances qu'il recouvre. Pour autant, elle n'est pas directement protégeable par un titre de propriété industriel qu'il serait aisé de défendre.
Le savoir-faire est un bien économique pouvant être valorisé dans le patrimoine d'une entreprise. Mais il n'est pas un bien juridique et ne peut être considéré isolément comme l'objet d'un droit privatif.

Le savoir-faire entendu comme connaissance technique transmissible et non immédiatement accessible au public n'est visé par les textes que sous l'aspect pénal très particulier de la violation du secret de fabrique. Or cette notion ne désigne que les seuls secrets utilisables dans le domaine de l'industrie, sous forme de méthodes, de procédés ou matières utilisées *(Cass. Crim., 24 juin 1985, n° 83-92.873).* En revanche, elle n'appréhende pas les méthodes commerciales ou les secrets de commerce (fichiers clients par exemple).

La valeur économique résultant des investissements effectués afin de développer un savoir-faire n'est donc protégeable que sur le terrain du droit commun de l'action en concurrence déloyale ou en parasitisme, qui sont susceptibles de révéler deux types de difficultés :

- La nécessité de faire la preuve d'une faute, d'un préjudice et d'un lien de causalité entre la faute et le préjudice (à l'inverse la violation d'un droit de propriété intellectuelle qui ne requiert que la preuve de l'existence du droit et de l'acte de violation).
- L'octroi de dommages-intérêts ne couvrira que très rarement l'intégralité des préjudices subis par le titulaire du savoir-faire et cela d'autant qu'une fois le savoir-faire divulgué, il perdra une grande partie de sa valeur et ne bénéficiera plus d'aucune protection.

La faiblesse du régime de protection du savoir-faire, des innovations non éligibles à la protection par le droit d'auteur ou la propriété industrielle est un frein à la valorisation du patrimoine immatériel des entreprises. Pourtant les accords ADPIC qui ont pour objet de définir les règles minimales de protection de la propriété intellectuelle par les pays membres de l'OMC opèrent un rapprochement entre secret et savoir-faire et couvrent la notion plus large de « renseignement non divulgué ».

Cette notion permet de fonder une interdiction générale d'usurpation et d'usage d'informations, de renseignements confidentiels ou de techniques et procédés ayant une valeur commerciale, qui ne sont pas généralement connus et ne sont pas aisément accessibles.

Si la conception française du secret est très étroite et que des incriminations pénales ne sont retenues qu'en ce qui concerne les secrets de fabrique, parallèlement, d'autres pays ont retenu une notion plus proche de la lettre des accords ADPIC, permettant ainsi une protection plus large du patrimoine informationnel des entreprises, comme par exemple :

- En Italie : protection des informations relatives à l'entreprise et à son expérience technico-industrielle, y compris son expérience commerciale si ces informations sont secrètes et ont une valeur économique.
- Aux États-Unis : protection des informations financières économiques ou commerciales.

De nouveaux outils juridiques à développer pour protéger les actifs des entreprises

Une réflexion globale sur la protection juridique des investissements effectués pour valoriser le patrimoine informationnel des entreprises doit être conduite, afin d'élaborer des outils juridiques permettant aux acteurs de Big data, y incluant les entreprises du secteur culturel, de valoriser les investissements réalisés autour de Big data pour mieux connaître leurs clients et leur proposer de nouveaux services.

Une logique de droit d'auteur pour valoriser les actifs résultant des données personnelles

Outre la question de la protection purement juridique, plusieurs réflexions ont été menées à un niveau gouvernemental sur la distorsion existante entre la valeur réelle et la valeur comptable des données personnelles.

Du point de vue des États, cette distorsion crée aujourd'hui un obstacle du point de vue de la fiscalité des GAFA dont le modèle économique repose essentiellement sur l'exploitation des données de leurs utilisateurs et dont les profits ne sont pas toujours correctement appréhendés par les États où leurs services sont distribués.

Nicolas Collin et Pierre Colin, les auteurs du rapport sur la fiscalité du numérique, proposent de calquer la protection des données personnelles sur le régime du droit d'auteur.

Pour mémoire, le droit d'auteur présente deux composantes distinctes :

- Des droits moraux d'autoriser ou d'interdire la diffusion d'une œuvre.
- Des droits patrimoniaux garantissant aux auteurs une juste rémunération au titre de l'exploitation commerciale de leurs œuvres.

Sur cette base, le rapport propose de reconnaître la composante patrimoniale des données à caractère personnel en permettant aux personnes concernées d'appréhender la valeur réelle des données les concernant comme contrepartie des services « gratuits » disponibles sur Internet pour le grand public. La composante droit moral serait, quant à elle, calquée sur la réglementation en matière de protection à caractère personnel. Ainsi, les entreprises qui pourraient démontrer le respect de la loi Informatique et Libertés, pourraient bénéficier de l'application d'un taux de fiscalité réduit. ▸ En savoir plus p. 37

Droit de la concurrence et structuration des marchés sur l'échange de données personnelles

Un marché « biface » selon la Commission européenne

Jusqu'à présent, du point de vue du bon fonctionnement des marchés, le point de convergence entre les données personnelles et le droit de la concurrence a rarement été traité.
En effet, le terme d'« actif » pour qualifier les données personnelles a été utilisé pour la première fois par la Commission européenne dans l'affaire de la fusion entre Google et DoubleClick en 2008[1].

Pour la Commission européenne, les acteurs du marché de la publicité en ligne, tels que les moteurs de recherche, opèrent sur un marché biface[2]. D'un côté du marché, le fournisseur de services en ligne propose des services gratuits (moteurs de recherche, e-mails, le contenu, etc.) aux utilisateurs.

De l'autre côté du marché, le fournisseur de services en ligne propose des services payants pour les annonceurs. En d'autres termes, sur le premier côté du marché, les utilisateurs bénéficient de services gratuits en échange du recueil de leurs données et de l'autre côté du marché, ces données sont monétisées auprès des annonceurs[3]. Les profits des nouveaux acteurs du numérique ne proviennent pas directement des services qu'ils fournissent aux utilisateurs mais sont issus de leur activité publicitaire ou de la distribution. ▸ En savoir plus p. 15

Un marché à analyser selon le droit de la concurrence

À l'heure actuelle, une grande majorité des données personnelles sont collectées par une poignée de fournisseurs de services en ligne qui, en tant que leaders de l'innovation sur leur marché, ont la capacité de recueillir de grandes quantités de données.

La législation relative à la protection des données à caractère personnel et le droit de la concurrence ne poursuivent pas les mêmes fins : la réglementation relative à la protection des données vise à la protection de la vie privée alors que le droit de la concurrence vise à assurer une concurrence effective sur le marché.

En conséquence, la possibilité de contrebalancer la montée en puissance des principaux fournisseurs de services en ligne quant à la collecte et à l'usage de données personnelles doit être recherchée tant dans le droit de la concurrence que du droit de la protection des données.

Diverses pratiques pourraient être qualifiées, en vertu du droit de la concurrence, d'ententes anticoncurrentielles ou d'abus de position dominante. De telles pratiques peuvent se manifester au moment de l'acquisition de données personnelles ou lorsqu'une entreprise empêche d'autres d'acquérir de telles données. Pourrait également être qualifié d'abus le fait, pour une entreprise, de détenir à elle seule des données considérées comme indispensables pour les autres opérateurs (selon la doctrine des « infrastructures essentielles ») et de s'en réserver l'usage, en se prévalant éventuellement du droit de la propriété intellectuelle, notamment ici du droit spécifique sur les bases de données. Enfin, les données personnelles, comme « actifs », peuvent être prises en compte dans l'analyse d'impact sur la concurrence d'une opération de fusion ou de rapprochement de deux ou plusieurs entreprises, analyse dite de « contrôle des concentrations ».

1 Décision de la Commission européenne du 11 mars 2008 déclarant une opération de concentration compatible avec le marché commun et le fonctionnement de l'accord EEE, Affaire n° COMP/M.4731 - Google/DoubleClick

2 Commission européenne, Affaire n° COMP/M.5727 – Microsoft/Yahoo! Search Business, 18/02/2010, §100

3 *Personal data, will Competition Law collide with privacy? – Competition law and personal data: Preliminary thoughts on a complex issue,* D. Geradin and M. Kuschewsky, *Concurrences* n° 2-2013

Pratiques anticoncurrentielles dans l'acquisition de données : empêcher d'autres opérateurs d'acquérir ces données

Deux exemples de pratiques anticoncurrentielles peuvent être mis en avant : tout d'abord, des accords d'exclusivité conclus entre les fournisseurs de services en ligne et les éditeurs avec pour effet d'empêcher d'autres fournisseurs de services en ligne de recueillir des données et ensuite le fait d'empêcher les utilisateurs de transférer leur données d'un fournisseur de services en ligne à l'autre.

Les accords d'exclusivité

Les fournisseurs de services en ligne peuvent conclure des accords avec des éditeurs prévoyant que l'éditeur utilise exclusivement les services du fournisseur, comme un moteur de recherche, sur son site Internet[3]. Conformément à ces accords, l'éditeur ne peut pas recourir aux services prestés par d'autres fournisseurs de services en ligne. En conséquence, surtout lorsqu'une multitude d'accords d'exclusivité est conclue entre éditeurs de services et un fournisseur de services en ligne unique, les autres fournisseurs de services en ligne sont empêchés de recueillir des données auprès des éditeurs qui, liés par leurs accords exclusifs, ne peuvent pas conclure d'autres accords[3].

Ces accords d'exclusivité peuvent également être conclus entre des fournisseurs de services en ligne et des fournisseurs de logiciels ou de matériels informatiques lorsque les services prestés par le fournisseur de services en ligne sont paramétrés par défaut sur ce type de matériel[4].

Interdiction de la portabilité des données

L'interdiction de la portabilité des données fait référence à la pratique mise en œuvre par les prestataires de services en ligne qui consiste à interdire aux utilisateurs de transférer leurs données à caractère personnel à un autre fournisseur de service en ligne. La portabilité des données constitue l'une des nouveautés du projet de Règlement sur la protection des données générales (article 18)[5].

L'interdiction de la portabilité des données pourrait empêcher les annonceurs d'exporter les données d'une campagne de publicité d'une plateforme à une autre[3].

Dans l'analyse des fusions

Dans le cas de la fusion entre Google et Doubleclick, la Commission européenne a examiné l'effet de l'accroissement de la quantité de renseignements personnels obtenus par l'entité issue de l'opération.

Dans ce cas, « l'enquête a révélé que la combinaison des informations sur les comportements de recherche et le comportement de navigation web ne donnerait pas un avantage concurrentiel dans le secteur de la publicité tel qu'il ne pourrait être reproduit par d'autres concurrents qui ont accès à des données d'utilisation du web similaires ».

Si cet élément a bien été pris en compte dans l'analyse de l'opération de fusion impliquant notamment Google, il n'a pas abouti ici à la conclusion que la fusion pourrait avoir un effet anticoncurrentiel.

4 Ces accords d'exclusivité « sont susceptibles d'exclure les concurrents sur le marché, surtout quand ils sont conclus par des entreprises en position dominante », et *a fortiori* si un ensemble d'accords de cette nature a été conclu

5 *Proposal for a regulation of the European parliament and of the council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), 2012/0011, 25/01/2012*

Quels facteurs d'équilibre possibles ?

Dans le cas où des acteurs dominants seraient identifiés sur ce marché, comme ayant des caractéristiques proches d'une « infrastructure essentielle », certains outils existent pour réguler ces situations et ont été largement testés dans d'autres secteurs, comme celui des télécoms. Une Autorité peut, par exemple, réguler le marché *ex-ante* en obligeant la publication d'offres de référence, fixant certains tarifs (ou les fixant à moyen terme), en favorisant l'entrée d'acteurs sur des marchés de niche, en favorisant la portabilité des données ou en obligeant les acteurs dominants à réaliser certaines séparations fonctionnelles.

On peut imaginer également que le régulateur mette l'accent sur les utilisateurs, obligeant les acteurs à une concurrence par la qualité. Pour ce faire, l'Autorité de la concurrence a identifié des mesures en vue de renforcer l'usage des données personnelles comme levier de différenciation concurrentielle :

- Le droit à la portabilité des données peut limiter le risque de blocage par la promotion d'une adhésion cumulative ou alternative aux divers réseaux sociaux par exemple.
- La durée de stockage des données doit être proportionnée à l'objectif poursuivi par le responsable du traitement et raisonnable ; la limiter pourrait abaisser les barrières à l'entrée.
- La transparence sur la nature des données collectées, sur la finalité du traitement et les destinataires de l'information donnerait aux utilisateurs les moyens de comparer les offres sur la base du critère de la protection des données personnelles, leur donnant le pouvoir de contrôler l'utilisation de leurs données personnelles[1].

1 *« Données personnelles, le droit de la concurrence doit-il prendre en compte la protection de la vie privée ? – Le point de vue de l'Autorité française de la concurrence »,* Bruno Lasserre, Président de l'Autorité française de la concurrence, *Concurrences* n° 2-2013, p. 28

Quelle fiscalité 3.0 pour l'économie digitale ?

Au-delà des défis que pose l'adaptation de la fiscalité internationale aux enjeux de l'économie numérique, le sujet de la fiscalisation des données est désormais clairement identifié.

Les questions qui se posent à cet égard sont multiples et complexes :

- Doit-il y avoir un lien direct entre fiscalité et *data* et si oui pourquoi ? (justification économique, contrainte budgétaire, protection de la concurrence, émergence de champions nationaux, partage de la valeur entre les opérateurs et créateurs).
- Faut-il mettre en place une fiscalité spécifique au numérique en général et aux *data* en particulier ? Si oui, comment taxer les *data* ?
- La fiscalité doit-elle vraiment être un instrument de contre-pouvoir dans le cadre de l'ère numérique ? Ou doit-elle simplement s'adapter aux nouvelles réalités ?

2013, un *momentum* pour la fiscalité de l'économie numérique

Jamais les efforts menés par les États pour restaurer l'équilibre de leurs finances publiques n'ont été aussi intenses, se traduisant par des hausses d'impôts généralisées dans la plupart des pays matures.

L'économie du numérique est particulièrement visée, dans le cadre d'une volonté clairement affichée par les États de réaligner les recettes fiscales avec le lieu de génération du chiffre d'affaires, voire même de destination des services et biens numériques.

En parallèle, les déficits budgétaires réduisent le financement public en faveur de la culture en général, et du développement de la création en particulier.

Le constat est désormais unanime : il est nécessaire d'adapter les règles de la fiscalité internationale aux nouvelles réalités du monde numérique et des nouvelles technologies, principalement en ce qui concerne les règles de territorialité de l'impôt sur les sociétés et les taxes indirectes assises sur le chiffre d'affaires.

En effet, par l'effet combiné de la dématérialisation systématique des services et des biens et de modèles d'affaires bipolaires[2] (fondés principalement sur les incorporels et les technologies de l'information), il n'y a souvent plus de nexus suffisamment caractérisé permettant d'allouer le droit d'imposer aux États sur le territoire desquels les services sont délivrés ou bien le chiffre d'affaires se trouve généré par l'utilisation (payante ou non) d'Internet par des consommateurs toujours plus connectés[3].

2 Modèle d'affaires adossant une activité dite « gratuite » à une activité rémunérée, localisée le plus souvent dans des territoires différents

3 Nicolas Collin et Pierre Colin, *Mission d'expertise sur la fiscalité de l'économie numérique,* janvier 2013

En outre, les différences de régimes fiscaux (taux, bases, régimes fiscaux dits de faveur etc.) entre les pays, parfois au sein de mêmes zones économiques (ex. : Union européenne) créent des disparités fiscales entre les acteurs de l'économie du numérique en fonction de leur taille (multinationales versus PME/ETI) et localisation (États-Unis, Europe, pays émergents). Elles se traduisent également pour les États par des déperditions significatives de recettes dont l'effet est aggravé par une compétition fiscale accrue entre ces mêmes États comme parfois par des pratiques optimisées de la part des opérateurs globalisés.

Les États ont réagi vigoureusement depuis 2012 et ont convenu en 2013 d'un plan d'actions ambitieux à court terme, incluant notamment :

- Le Programme « BEPS » en 15 points de l'OCDE, visant à lutter contre l'érosion de la base d'imposition et le transfert des bénéfices (juin 2013), lequel intègre précisément la prise en compte des défis fiscaux posées par l'économie numérique comme point d'action n° 1.
- L'harmonisation au sein de l'Union européenne des règles de TVA en matière de traitements et services rendus par voie électronique (nouveau régime unifié à partir du 1er janvier 2015, généralisant l'imposition dans l'État de consommation).
- Compte tenu de l'envergure internationale et politique de cette approche, il s'écoulera un certain temps entre le lancement de ces plans d'actions et réformes et leur impact dans l'économie réelle.

De même, l'harmonisation au sein de l'Union européenne des règles de TVA soulève encore certains problèmes d'application et il n'a pas été possible d'accélérer son entrée en vigueur avant 2015, alors même que cette problématique est connue de longue date. Ce sujet de l'économie numérique est récemment venu à l'ordre du jour du Conseil de l'Union européenne, lors d'une réunion qui s'est tenue les 24 et 25 octobre 2013.

Ces difficultés théoriques comme pratiques s'expliquent principalement par la complexité et la spécificité des modèles d'affaires de l'économie numérique qui ne sont toujours pas parfaitement appréhendées par les concepts et outils fiscaux traditionnels.

À cet égard, un mouvement d'opinion de plus en plus fort se fait jour au sein des acteurs et experts de l'économie du numérique pour contester la nécessité d'une fiscalité spécifique au monde numérique, les technologies digitales imprégnant en effet tous les pans de l'économie (industries, services, secteurs marchand et non marchand, privé et public) et nécessitant donc un approche globale, si ce n'est de droit commun.

Consensus autour de la non-taxation des données

Plusieurs acteurs publics ont milité pour la mise en place de mesures fiscales spécifiques au numérique et/ou touchant plus spécifiquement les géants internationaux de l'Internet.

Les buts poursuivis étaient multiples et divers :

- Financement de la culture, pour les pays dotés d'une politique fiscale volontariste en la matière.
- Restauration des finances publiques.
- Équilibrage du taux effectif d'imposition entre les acteurs nationaux et multinationaux.
- Restauration du lien direct entre territorialité de l'impôt et source de la valeur économique à l'ère digitale.

La France est sans doute le laboratoire d'idées le plus dynamique en la matière, comme l'a révélé l'étude des Politiques Fiscales dans le domaine de la Culture menée par EY pour le Forum d'Avignon sur la période 2009-2012, identifiant près de 50 incitations fiscales et 15 taxes spécifiques au domaine culturel.

La question de la fiscalisation des données y avait dès lors suscité un vif débat, qui trouve désormais écho au plan international.

Après les concepts, rapidement abandonnés, de taxe assise sur les recettes publicitaires générées sur Internet[1], de taxation de la bande passante, de taxe « au clic » ou de taxe assise sur les activités de fabrication et ventes de téléphones mobiles[2], c'est le concept de taxe assise sur les données lancé en France qui a été le plus novateur, et sans doute le plus polémique au plan international.

Imaginé par Nicolas Collin et Pierre Colin dans leur rapport, ce concept visait à créer une fiscalité propre liée à l'exploitation des données issues du suivi régulier et systématique de l'activité des utilisateurs sur un territoire donné. Reposant sur le constat de l'importance de la collecte et de l'utilisation des *data* et données dans la chaîne de valeur et le chiffre d'affaires des opérateurs du numérique, ce concept a, sur le papier, le mérite d'un fait générateur simple et garantissant la neutralité du prélèvement.

De façon plus novatrice, cette proposition ne visait pas tant à maximiser le volume des recettes fiscales nouvelles mais, de manière très ambitieuse, à renforcer les libertés individuelles et la concurrence en favorisant les comportements vertueux des opérateurs utilisant les *datas* (*via* l'application de taux d'imposition réduits ou dégressifs selon les comportements observés à l'aune de critère de conformité prédéfini[3]).

Toutefois, bien que séduisant intellectuellement, ce concept de taxation « vertueuse » des données a été largement critiqué au plan international, notamment car il pose, au-delà des débats idéologiques, des difficultés extrêmes et non résolues d'application et de mise en œuvre et pourrait par ailleurs se traduire dans les faits par une augmentation du coût final pour le consommateur.

Il n'a dès lors pas été retenu par le Conseil National du Numérique (septembre 2013) dans son avis n° 2013-3, lequel a privilégié une stratégie de négociation politique pour une réforme internationale des règles de la fiscalité des entreprises.

Le sujet reviendra peut-être sur le devant de la scène, notamment au plan européen, par exemple sur le fondement d'une justification à taxer les exportations de données à partir du territoire de l'Union européenne.

1 Taxe sur les recettes publicitaires proposée par M. Le Sénateur Marini (France) dès 2010

2 Taxes sur les téléphones mobiles intelligents préconisées par le Rapport de la Mission Lescure en 2013 (France), mais non retenues par le gouvernement français

3 Par exemple : niveau d'information de l'utilisateur sur ses droits à la protection des données personnelles, recueil de consentement et facilité de l'exercice de ses droits *via* l'interface, ouverture à la concurrence et nouveaux services, accès des données à des tiers, etc.

Prochaines étapes pour la fiscalité 3.0

Une récente étude EY[1] observait que les problématiques et enjeux liés à l'utilisation croissante de services *via* le *Cloud* dans l'économie numérique étaient encore très largement ignorés ou laissés de côté, notamment :

- Une fiscalité inadaptée ou complexe dans de nombreux pays, règles non uniformes.
- La caractérisation des revenus générés *via* le *Cloud* au regard des règles de retenues à la source (*withholding tax*).

Dans le contexte globalisé et hautement technologique de l'économie numérique, il est plus que jamais nécessaire de privilégier une action coordonnée au plan international, afin d'adapter et d'uniformiser les règles de la fiscalité.

Au plan local, et sans brider l'action des législateurs nationaux, il semble clair que les efforts devront d'abord porter, d'une part, sur la simplification des régimes fiscaux, bien trop complexes à ce jour, et d'autre part, sur une meilleure sécurité des opérateurs confrontés à des changements de législations permanents.

Sans nier la véritable nécessité d'améliorer et harmoniser les régimes fiscaux et de restaurer les finances publiques des États, les opérateurs du numérique, petits comme grands, restent toujours dans l'attente de politiques fiscales claires et lisibles axées sur le long terme et permettant d'assurer le développement des champions de demain.

Plutôt que de créer de nouveaux impôts ou de nouvelles normes, ne faudrait-il pas tout simplement appliquer ceux existants et, si besoin, se contenter de les adapter ou de les améliorer ?

Les pistes de réflexions de l'OCDE devraient désormais privilégier à court terme la mise à jour de la définition fiscale de l'établissement stable et la réforme des standards internationaux en matière de prix de transfert.

À moyen terme, les États souhaiteront sans doute revisiter les principes de territorialité de l'impôt sur les sociétés en matière de services numériques, compte tenu en effet de ce que les États devraient être logiquement tentés, au plan national, de privilégier l'imposition sur le lieu de destination des services.

Dans ce contexte, tout le défi des travaux désormais lancés au niveau de l'OCDE sera de prendre en compte les réalités du numérique pour créer enfin une fiscalité 3.0 adaptée au nouveau monde digital et ce sur la base d'un socle commun et homogène, afin d'éviter des approches disparates au niveau national.

1 *EY Global Survey Report "Tax Considerations in cloud Computing"*, March 2012

Responsabilités citoyennes

Contrôler ses propres données

Face à la montée du contrôle des données émises par les individus ou leurs objets connectés, une prise de conscience s'impose, tout en excluant le renoncement à toute activité numérique, qui ne concerne qu'une partie infime de la population.

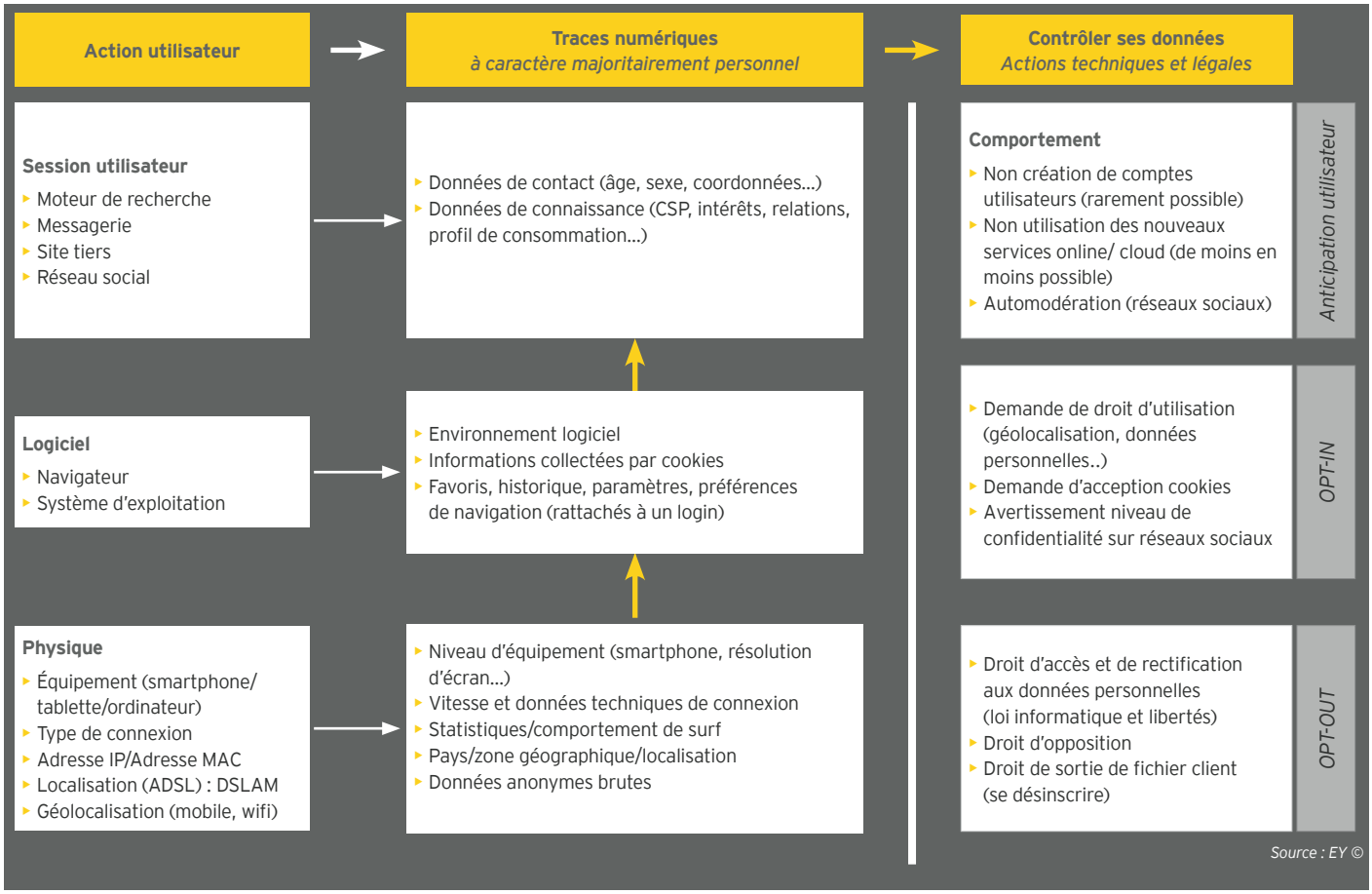
Les recommandations de la Commission européenne en matière de droit à l'oubli, de renforcement des principes de consentement et d'amélioration des voies de recours administratif et judiciaires vont dans le sens d'un contrôle du contrôle.

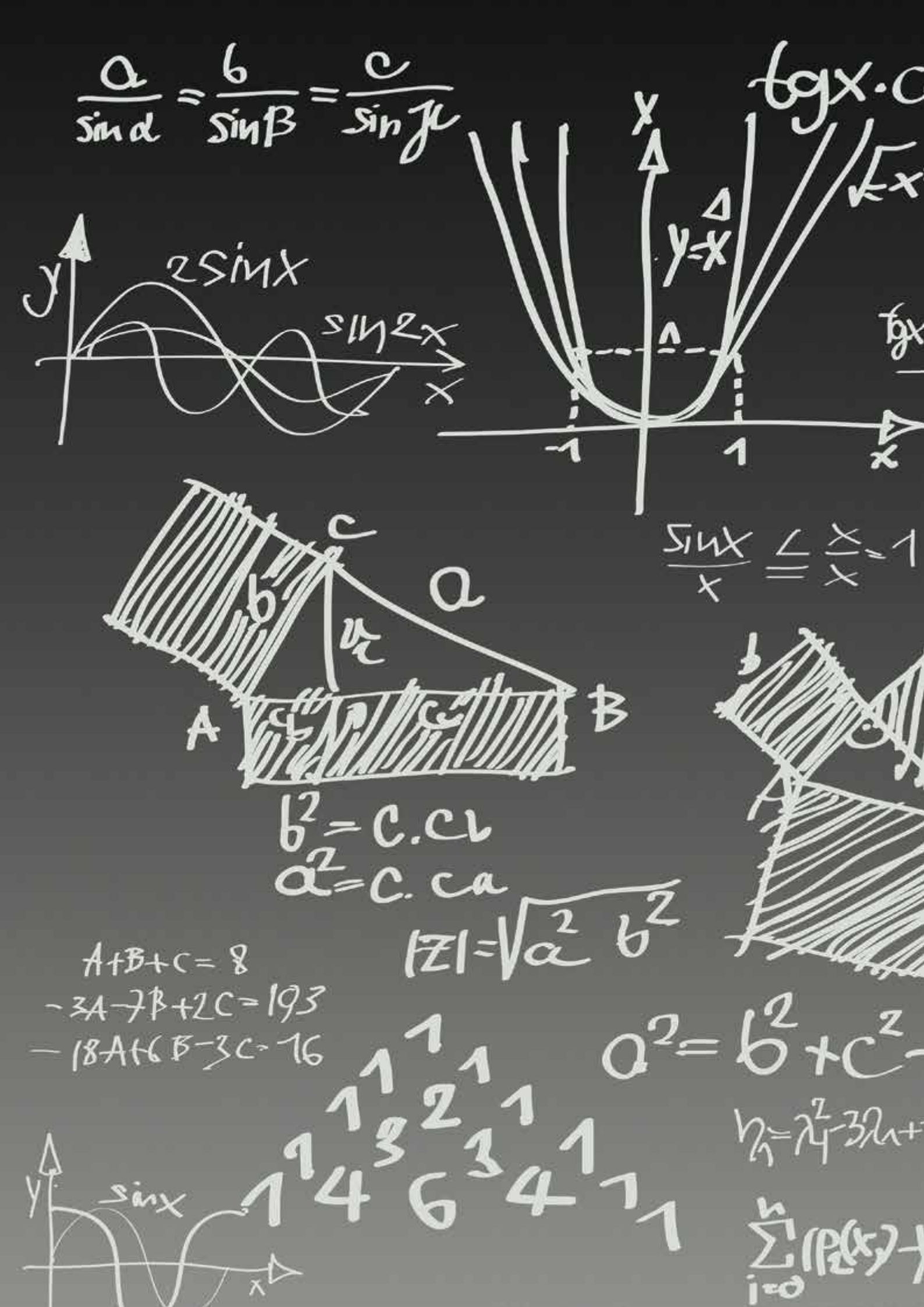
Les recommandations et le renforcement de la notion de responsabilisation *(accountability)* des « opérateurs d'accès/ services Internet » vont de pair avec la confiance et l'e-réputation – actif clé pour l'ensemble des acteurs du marché.

▸ Voir schéma ci-dessous

Les effets conjugués d'une conscience plus aiguë du nécessaire contrôle sur ses propres données d'une part et de la Directive européenne d'autre part, redonneront confiance et pouvoir de contrôle aux citoyens.

Comment contrôler ses propres traces et données numériques *a priori* et *a posteriori* ?




a/sin α = b/sin β = c/sin γ
2sin x
sin 2x
y
x
y=x
-1
1
sin x / x ≤ x/x = 1
c
b
a
A
B
b² = c · c₁
a² = c · c_a
|Z| = √(a² b²)
A+B+C = 8
-3A-7B+2C = 193
-18A+6B-3C = 16
1 1 1 1 2 1 1 3 3 1 1 4 6 4 1 1
a² = b² + c²
sin x

3. Vers un nouveau paradigme économique propice à l'innovation et la création

Opposer le principe de précaution au Big data : un risque réel pour l'innovation

Masse critique d'informations et niveaux d'analyses : l'exploitation économique de la *data* est-elle réellement incompatible avec le respect de la vie privée ?

La thermodynamique

La thermodynamique, qui a permis le développement de l'industrie du XIX[e] siècle en basant ses analyses uniquement sur des indicateurs « macro », vise à comprendre les échanges d'énergie et de chaleur.

Elle est riche en applications pratiques que nous utilisons tous les jours : moteurs, réfrigérateurs, turbines ou encore réacteurs. Les modèles ont besoin de très peu de grandeurs pour décrire le comportement des systèmes et leur évolution, principalement Entropie, Température, Pression, Volume.

L'efficacité de ces prédictions macroscopiques est tout à fait étonnante quand on sait qu'au niveau microscopique, ces comportements sont la résultante des particules composant les gaz et les liquides et que chaque particule est définie *a minima* par son vecteur vitesse (3 données) et sa position (3 données) soit des milliards d'informations nécessaires.

Cette efficacité de l'analyse macroscopique réside dans le fait qu'un grand nombre d'états microscopiques sont possibles pour un même état macroscopique. Aussi, la température est la résultante d'un niveau d'agitation de molécules, qui peuvent se trouver dans des milliards de configurations possibles pour une même mesure de la température.

Ainsi, la thermodynamique nous enseigne que l'on est capable de prendre des décisions sur la base d'une mesure d'un indicateur « macro » comme la température sans pour autant avoir besoin de mesurer toutes les grandeurs au niveau « micro » (positions, mouvements des milliards de molécules à l'origine du niveau de la température). Par exemple, il est possible de pasteuriser du jus de pomme en maintenant une température de 75° . Pour ce faire, il n'est pas nécessaire de modéliser ce qui se passe au niveau de chaque molécule (niveau « micro »).

Par analogie avec le Big data, il est possible de prendre des décisions en analysant les comportements « macro » d'un groupe d'individus sans avoir besoin d'identifier parfaitement le comportement intime de chacun.
A-t-on besoin de tout connaître de M. ou Mme X pour faire des prédictions sur une population de grande taille ? En d'autres termes, est-il nécessaire d'associer la somme d'informations collectées à une personne clairement identifiée (nom, adresse…) afin de lui proposer de façon plus adaptée certains produits ?

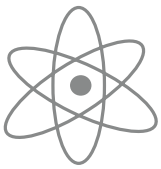
On pourrait imaginer que les grandeurs de mesures sont « macro » et très précises pour permettre une adaptation parfaite à la demande. Cependant la demande, grandeur macroscopique, est la résultante de nombreuses demandes individuelles (microscopiques) qui peuvent rester anonymes. Le retour au niveau « micro » pour l'acte d'achat peut s'analyser de la même façon : une campagne de publicité avec les bons messages envoyés à la population idoine pourrait s'avérer plus efficace qu'un envoi ciblé personnel.

Ce que nous enseigne la science

La thermodynamique

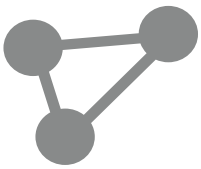
Il est possible de prendre une décision sur la base d'une mesure d'un indicateur « macro » sans avoir besoin de mesurer toutes les grandeurs au niveau « micro ».

La physique quantique



L'observation microscopique modifie le comportement des particules, rendant aléatoire et vaine la précision infinie de la mesure.

La génomique



La combinaison de décryptages au niveau « micro » permet, dans le respect de la confidentialité, d'aboutir à des découvertes sources de progrès.

Applications au Big data :

Big data et respect de la vie privée peuvent être conciliables dans un environnement réglementé et sécurisé.

Source : EY ©

La physique quantique

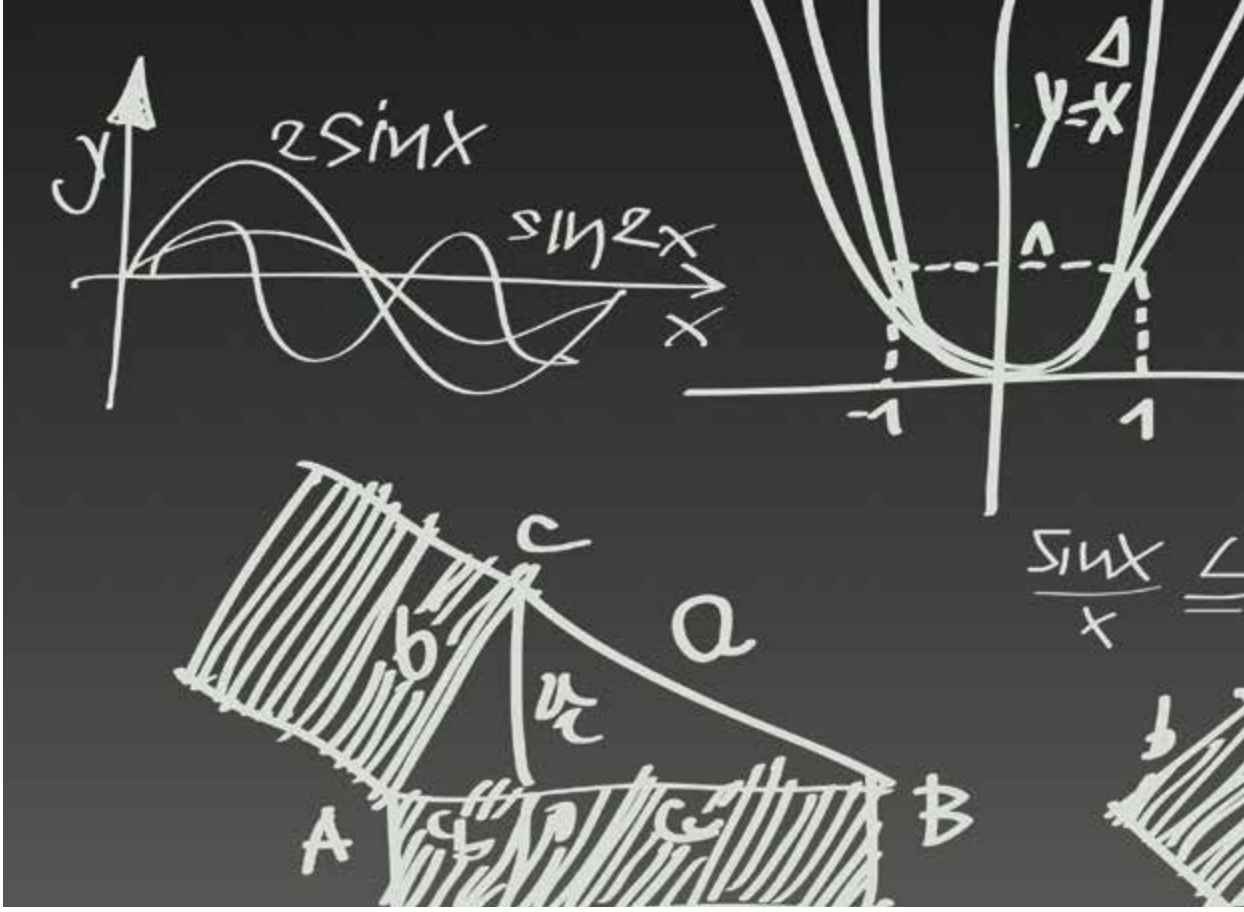
La physique nous enseigne que l'analyse microscopique, domaine de la physique quantique du XX^{e} siècle, est aléatoire par nature et que la précision infinie de la mesure est vaine (on ne peut connaître en même temps la position et la vitesse d'une particule).
Enfin, à ce niveau, l'observation interfère avec l'expérience. En d'autres termes, l'observation modifie le comportement des particules.

Par analogie avec le Big data, il en ressort que l'observation « microscopique » d'un individu modifierait son comportement. **Lorsqu'une personne se sait « observée », elle tend à modifier son comportement, rendant ce dernier moins prédictible.**
D'où l'importance d'un environnement sécurisé, de confiance, afin que les individus se sentent libres et décisionnaires de leurs comportements et ne soient pas réticents à échanger ou communiquer leurs données.

La génomique

Si la cartographie du génome d'un individu ne permet, au niveau « micro », que d'émettre des probabilités de réalisation (maladies, etc.), au niveau « macro », la combinaison de ces probabilités individuelles permettra de révéler, sur une population de grande taille, les risques et opportunités réels, justifiant ainsi des investissements d'ampleur permettant d'améliorer la santé des individus (R&D, vaccins, traitements).

On peut comprendre que si l'on respecte la confidentialité des décryptages individuels permettant, par leur combinaison, d'aboutir à des découvertes et des décisions d'investissement, **l'équilibre entre progrès et respect de la vie privé est atteint.**



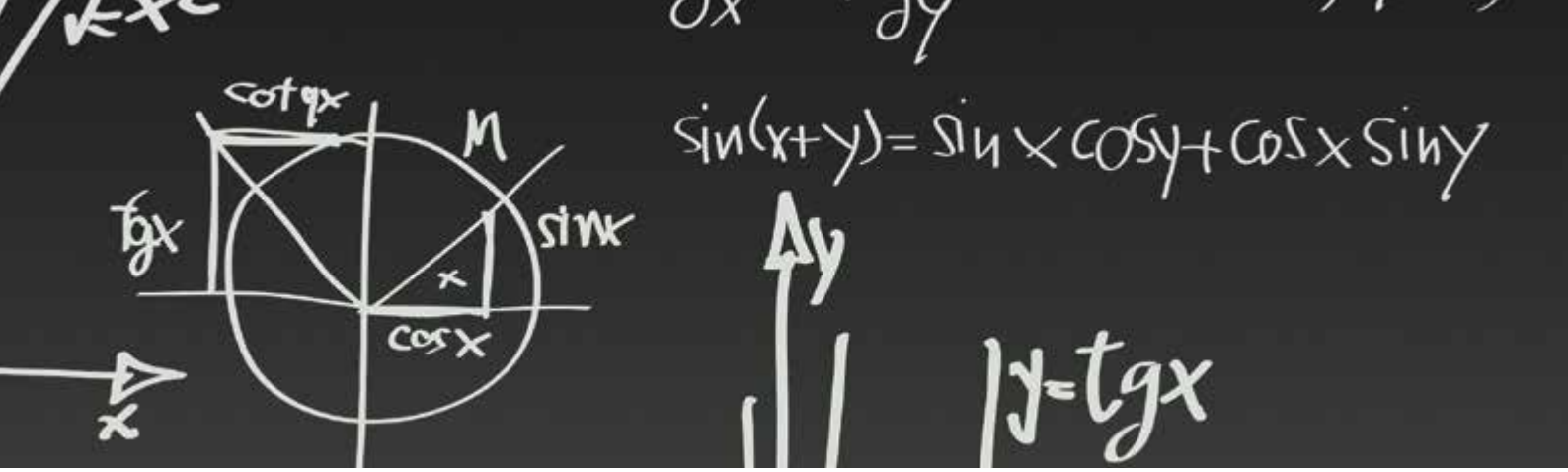
Le temps de l'action

Il serait illusoire d'imaginer pouvoir ignorer le « Big data bang », voire s'en barricader : la vitesse exponentielle avec laquelle la masse de données générées augmente et est stockée, la progression des outils d'analyse et de rapprochement de ces données ainsi que la capacité et l'intérêt que suscitent ces résultats rendent irréaliste un moratoire. Si l'avenir n'est pas totalement clair, il est cependant certain que ceux qui n'auront pas su considérer le défi dès maintenant seront les grands perdants.

Ainsi, il est urgent d'opposer le « principe d'innovation » au « principe de précaution ».

Invoqué face aux dangers, certes réels, que l'exploitation du Big data ferait courir à notre intimité ou à notre identité numérique, le « principe de précaution » n'apparaît pourtant pas comme une réponse adaptée.

Nous devons en passer par une nécessaire phase d'expérimentation *test & learn*, et les erreurs, voire abus, propres à cette phase de transition, devraient éveiller les consciences, en particulier chez ceux qui ont subi l'expérience désagréable d'une mise à nu en ligne de leur vie « privée ». Ces écueils obligeront parfois certains à déployer leur capacité de rebond pour faire peau neuve, en renonçant à leur identité numérique première et aux traces qui pouvaient y être associées.



Entrer dans l'ère du Big data : par où commencer ?

1. Établir un diagnostic

- Dresser une cartographie des données disponibles, qu'elles soient internes (données métiers) ou externes (sites web, réseaux sociaux, *open data*).
- Évaluer la qualité et la pertinence des données au regard des priorités de développement (*ex.* : financement de la création, innovation, services, étude ou segmentation des publics/clients, valorisation des données, efficacité opérationnelle, etc.)

2. Se doter d'une structure décisionnelle et de compétences clés

- Définir l'organisation et la gouvernance adéquates pour l'exploitation de données en pensant l'organisation de façon transverse.
- Renforcer les compétences analytiques, mathématiques, statistiques et sociologiques pour la gestion et l'exploitation des données - que ce soit en croissance organique, par rachat de sociétés digitales ou *via* des partenariats. En d'autres termes, les entreprises devront se doter de *data scientists*, ces experts « capables de traiter le déluge de données et d'en tirer toute la quintessence décisionnelle et managériale[1] ».

3. Se doter d'une stratégie et d'un plan d'action

- Identifier et mettre en œuvre des leviers d'enrichissement de la connaissance client, notamment en :
  - Incitant les clients encore non connus (lecteurs achetant en point de vente, public des salles de spectacle et de cinéma, etc.) à s'identifier *via* un programme de fidélité, en favorisant les actes d'achat sur Internet.
  - Analysant les comportements clients : contenus générés sur les réseaux sociaux, parcours client web, historique des transactions et contenus consultés.
- Identifier le niveau de moyens adaptés à la stratégie : il est indispensable de bien déterminer la donnée utile pour éviter un stockage systématique coûteux et inexploitable, et de privilégier les approches *test & learn* éprouvées par les grands acteurs du digital, consistant à lancer des chantiers tests et mesurer régulièrement et systématiquement leur efficacité afin de les améliorer rapidement, voire à les suspendre si besoin.

4. Développer et insuffler en interne une culture de la *data*

Dans le secteur médiatique et culturel, le niveau de maturité face à la *data* diffère selon les acteurs : ceux qui ont un accès direct à leurs publics, avec des bases de clients ou d'abonnés (groupes de presse, exploitants de salles…) ont déjà une bonne appréhension de ces sujets et une culture du marketing direct. L'enjeu d'une approche Big data est de consolider et d'exploiter des données dans un contexte de croissance en volume, en vitesse et en variété des sources - notamment digitales - difficiles à réconcilier avec les données historiques.

C'est une opportunité, mais aussi un défi de taille à relever, pour des acteurs tels que les diffuseurs audiovisuels et les créateurs de contenus, qui, traditionnellement n'avaient pas de lien direct avec leurs audiences.

Le développement de la culture et des compétences d'exploitation de la donnée client varie également en fonction des services : les métiers marketing ont déjà intégré ces logiques, tandis que les équipes éditoriales et créatives se l'approprient plus difficilement – voire y sont réticentes, pouvant considérer que l'exploitation des données est incompatible avec leurs règles éthiques. Or les outils d'analyse des audiences et d'identification des tendances doivent être envisagés comme des sources complémentaires visant à alimenter leurs contenus et non à se substituer à leur travail de recherche et de hiérarchisation. Et le succès rencontré par les infographies, fortement relayées sur les réseaux sociaux, montre que l'audience est réceptive à l'information quantitative visuelle.

Engager cette mutation interne ne suffira probablement pas à extraire toute la quintessence de la *data*. C'est en nouant des alliances que les acteurs de l'écosystème culturel pourront capitaliser sur les données personnelles culturelles pour innover et créer de nouvelles propositions de valeur.

1 *« Big data – nouveaux défis »*, Telecom ParisTech, N° 169, juillet 2013

Focus : Acquisition de droits

Une société comme Netflix combine deux sources d'information pour définir sa stratégie d'achat de droits pour son offre de vidéo par abonnement : l'analyse des contenus les plus regardés (thématiques, acteurs, formats, etc.) et celle des contenus les plus piratés.

Enrichir la donnée collectée

La gestion de la donnée et son exploitation sur le mode du Big data sont au cœur du modèle économique des grands acteurs globaux du digital.

En dépit de l'intérêt que représente l'enjeu de la valorisation des données personnelles, les acteurs français sont en retard par rapport aux acteurs globaux du numérique, qui se sont positionnés comme des distributeurs ou diffuseurs de référence de la création. Citons par exemple Amazon, qui utilise ainsi les données (pages consultées, historique d'achat, listes de souhaits…) pour recommander des contenus et produits à ses clients : 35 % de ses ventes seraient liées à cette mécanique. Ou encore Apple, leader de la distribution de musique avec iTunes, qui analyse les bibliothèques musicales de ses utilisateurs (composition, titres les plus écoutés…) pour recommander de nouveaux artistes ou disques.

À l'image d'autres industries, les acteurs de l'industrie médiatique et culturelle peuvent imaginer valoriser les données personnelles afin d'affiner la connaissance de leurs publics, à travers par exemple :

- L'analyse rationnelle de la propagation et des contenus des messages (positifs ou négatifs) sur les réseaux sociaux, à l'image des travaux réalisés dans les transports (Quantas Airline, RATP) ou dans le secteur bancaire *(Bank of America)* ; L'identification d'opportunités de diffuser des créations en fonction des centres d'intérêt et des goûts avérés des utilisateurs ou des communautés.
- L'anticipation des comportements, comme l'analyse de la fréquentation d'une zone touristique ou d'une salle de concert.
- L'analyse croisée de données externes (réseaux sociaux, etc.) et internes (données de navigation, historiques de consultation/achat, centres d'intérêt, etc.) pour mieux connaître les attentes et les intégrer dans la conception des contenus et services proposés.

C'est en adoptant une « approche *data* », comme levier d'optimisation de la valeur client, que les acteurs de l'industrie culturelle pourront relever les trois défis que sont la relation avec les publics et audiences, la diffusion des créations et contenus ainsi que le financement de la création.

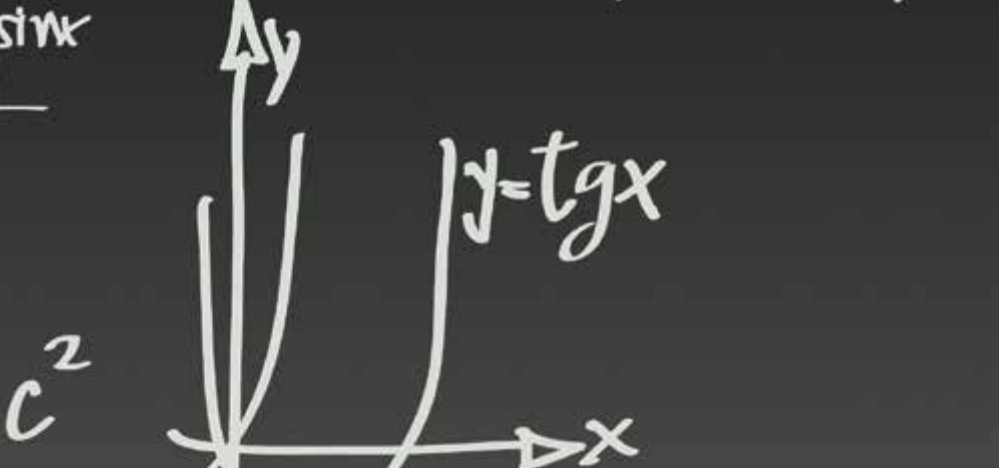
1. Assurer la diffusion et l'interaction des œuvres avec leur public

Dans le secteur médiatique et culturel, le développement de la « connaissance client » permet de répondre à un défi majeur qui consiste à instaurer une relation privilégiée et pérenne avec les publics. C'est en particulier le cas pour la production d'événements, de contenus et de créations, afin de créer des projets et contenus au plus près des publics visés.

Dans les médias, des outils d'analyse des tendances comme Trendsboard utilisent les données issues des réseaux sociaux et blogs pour identifier les sujets qui commencent à faire l'actualité, permettant ainsi aux rédactions d'anticiper des « buzz » ou de couvrir certains événements, jusqu'alors imprévisibles.

Proposer une **interaction personnalisée** en fonction des données analysées apparaît alors comme un levier clé pour créer ou renforcer la relation de proximité avec son public :

- L'exploitation des données est un levier de promotion de la diversité culturelle, aussi bien dans les lieux que *via* les supports numériques. La proposition de contenus sur la base de l'analyse des centres d'intérêts, sur le modèle des moteurs de recommandation, suscite ainsi la rencontre entre les œuvres et leur public.
- La collecte et l'exploitation de données personnelles sur les publics permet de prolonger la relation au-delà d'un événement donné (spectacle vivant, visite d'un musée, visionnage d'un film…), à travers l'animation de communautés et la proposition de services et contenus complémentaires.
- Dans l'industrie audiovisuelle, les offres proposées sur les supports numériques permettent de développer et de valoriser une connaissance individualisée des audiences.



2. Obtenir une vision globale du parcours des publics

Les outils utilisés pour la collecte et l'exploitation de données sont souvent inadaptés et/ou fragmentés du fait de logiques de fonctionnement d'activités en silos. Ainsi, un établissement public comme la Rmn-GP[1] ne peut, à ce jour, savoir si une personne qui a acheté un livre donné dans une de ses 40 librairies ou boutiques pourrait être intéressée par une exposition au Grand Palais, au Musée du Luxembourg ou tout autre service, comme un abonnement à une newsletter ou un parcours à thème autour d'un artiste précis. En effet, les informations qui permettraient d'établir de tels liens ou prédictions ne sont actuellement pas disponibles, soit parce qu'elles ne sont pas collectées, soit parce qu'elles sont dispersées au sein de différentes bases de données gérées par divers systèmes d'information, non encore interopérables à ce jour.

L'enjeu est donc double : il faut non seulement investir pour se doter des compétences nécessaires en CRM mais aussi assurer la transversalité entre les services impliqués dans la collecte et le traitement de données.

Cette vision globale est indispensable pour proposer une expérience plus personnalisée et contextualisée et espérer ainsi la prolonger dans le cadre d'une relation d'engagement. De même les lieux culturels (galeries, sites touristiques, musées…) deviendraient plus accessibles grâce à l'utilisation de données sur le trafic ou la fréquentation. Si les projets sont nombreux dans le domaine des transports publics, il reste encore beaucoup à faire sur les lieux culturels. L'analyse de la fréquentation permet d'informer en temps réel les potentiels visiteurs, par exemple, sur la durée d'attente à l'entrée d'une exposition, et elle pourrait également permettre de définir des tarifications fines en fonction du remplissage.

La construction d'une vision globale du parcours des publics pourra également s'appuyer sur l'ouverture et la réutilisation des données publiques culturelles. L'*open data* culturelle doit s'inscrire en complément des actions que les acteurs des industries culturelles et créatives mèneront pour utiliser et valoriser leurs données. Offrir la possibilité à des acteurs tiers, et notamment des start-ups, d'utiliser les données des acteurs de la culture élargit considérablement le champ des possibles, notamment pour créer des services pour les usagers ou clients.

3. Assurer le financement de la création

Les données peuvent être mises à la disposition d'autres acteurs de la culture, de la distribution, des réseaux sociaux, etc. pour qu'ils puissent les utiliser en l'état ou les intégrer dans d'autres combinaisons de données, dans une approche d'ouverture des données *(open data)* ou dans une logique commerciale.
Une utilisation commerciale des données permet ainsi de créer des offres de contenus gratuites ou partiellement gratuites grâce à la publicité. Pour les annonceurs, les audiences ont d'autant plus de valeur qu'elles sont qualifiées et permettent un ciblage efficace, d'où l'importance de la donnée récente, même anonyme.

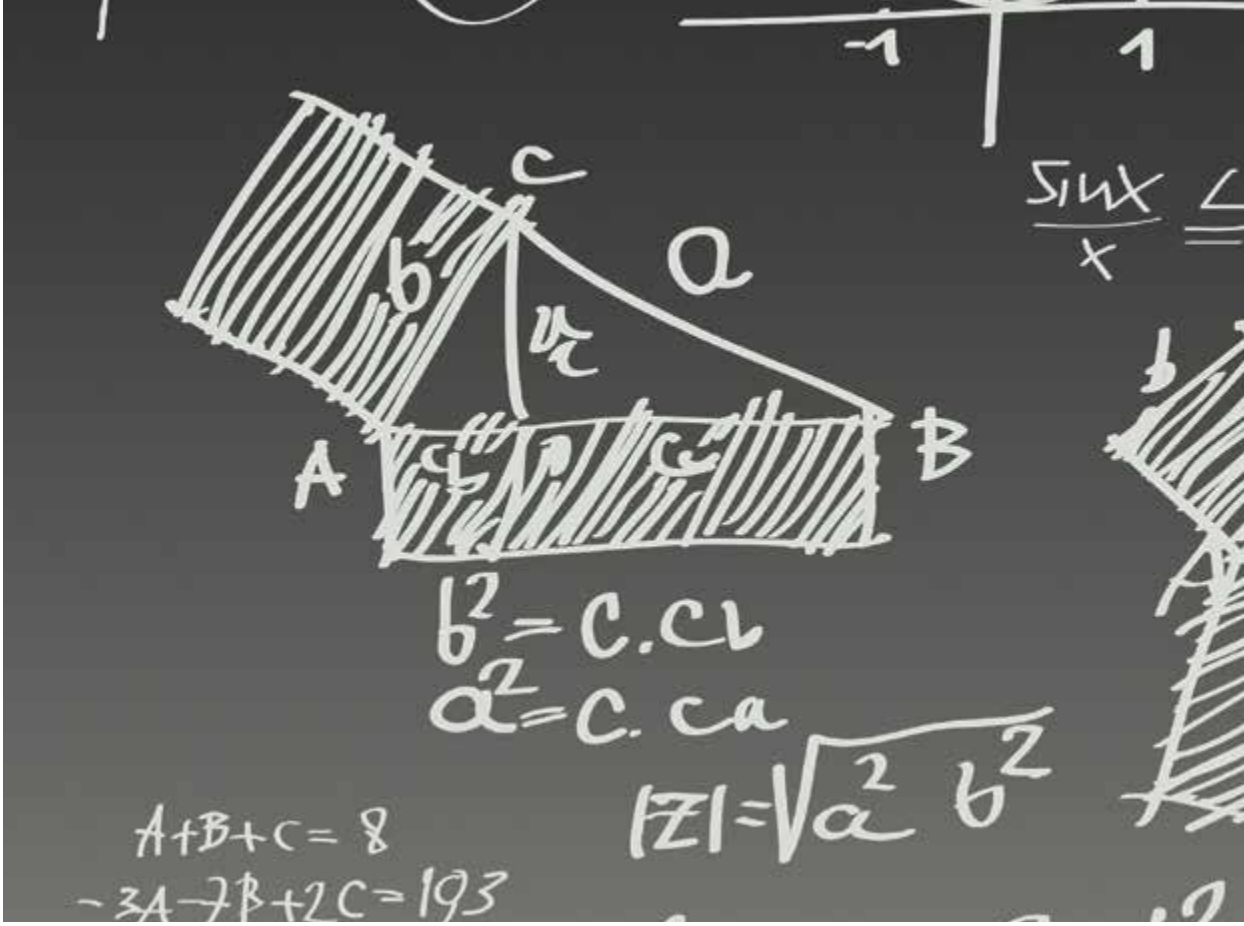
Les données personnelles culturelles ont, dès lors, un rôle majeur à jouer dans un contexte où le secteur culturel cherche et invente de nouveaux modèles de financement.

Dès aujourd'hui, ces données sont au cœur du financement participatif, dont le modèle repose sur la création de communautés identifiées, d'individus connectés qui s'engagent personnellement pour financer un projet créatif.

Le *crowdfunding* est aussi générateur de données (d'un genre comparable aux réseaux sociaux) et, si les sites ne monétisent pas cette donnée sur un mode publicitaire, ils peuvent s'en servir pour animer le réseau et faire des recommandations et contribuer à accroître les financements.

Demain, les données personnelles culturelles seront sans nul doute un nouveau relais de financement, dans un contexte marqué par la raréfaction des financements publics. Un projet culturel pourrait demain valoriser, au moment de son financement, sa capacité à générer des données pour le distributeur, au même titre qu'il peut générer des ventes.
Les plans de financement de projets cinématographiques ou discographiques pourraient, par exemple, intégrer la valorisation des données nouvelles collectées : un producteur exécutif céderait à un co-producteur le droit d'administrer la communauté de fans de l'œuvre créée, et les revenus publicitaires éventuellement générés.

1 *Réunion des musées nationaux - Grand Palais*



En France, plusieurs entreprises et établissements publics ont déjà rendu publiques leurs données culturelles

Depuis l'ouverture en 2011 du portail *data.gouv.fr,* le ministère de la Culture et de la Communication s'attache à recenser et mettre à disposition des développeurs différents jeux de données tels que la liste des établissements publics culturels géolocalisés et de leur offre éditoriale, les données de la médiathèque de l'architecture et du patrimoine (liste des immeubles protégés au titre des monuments historiques, liste des objets mobiliers propriété publique classés au titre des monuments historiques), ou encore une trentaine de jeux de données du Centre national du cinéma et de l'image animée (CNC), comme par exemple les recettes à l'international des films français ou la liste des établissements cinématographiques. De nombreuses villes et régions ont quant à elles initié l'ouverture des données publiques, qui peuvent être des statistiques démographiques (anonymisées), aussi bien que la liste des lieux de tournage de film et les statistiques de consultations d'ouvrages en bibliothèques.

L'ouverture de ces données, croisées avec les centres d'intérêt d'individus ou de communautés, représente non seulement l'opportunité de développer de nouveaux services et usages, mais aussi de faciliter la promotion et l'accès à la culture.

De nombreuses applications s'appuyant sur les données culturelles ouvertes existent déjà :

- Les cartes interactives, comme le module développé par le ministère de la Culture et de la Communication représentant l'ensemble des lieux culturels français sous forme de galaxie, ou Cartographone, carte regroupant les lieux de tournage de film à Paris.
- Les interfaces de visualisation de données, avec par exemple un article du journal *Le Monde* contenant un module permettant de visualiser de façon interactive les statistiques de consultation des ouvrages dans les bibliothèques parisiennes.
- Les applications utilitaires, permettant par exemple de trouver la bibliothèque la plus proche, des informations sur les musées de la ville (comme Musambule à Marseille), ou encore des agendas culturels multi-éditeurs (tels que Cibul en région PACA).
- La visite augmentée : comme Culture Clic, proposant non seulement des informations pratiques sur les musées français mais aussi un catalogue de 900 œuvres visualisables en réalité augmentée.

La nouvelle chaîne de valeur de la donnée personnelle culturelle

Nouveaux acteurs

Le marché de la donnée personnelle culturelle est en pleine structuration : tout au long de la chaîne de valeur de la *data* se déplacent et se positionnent producteurs, agrégateurs et utilisateurs de données personnelles culturelles.

Un marché qui devrait continuer à évoluer et faire émerger de nouveaux acteurs et de nouveaux métiers : des *pure players* qui se spécialiseront dans la production/*sourcing* de *data*, d'autres dans l'agrégation et l'analyse de *data* (croiser les données des sourceurs entre elles et avec d'autres données, afin de les contextualiser et de fournir l'analyse nécessaire aux utilisateurs) et les utilisateurs de données personnelles qui vont développer des services et applications ciblés et à plus forte valeur ajoutée. Tous ces services auront besoin de l'infrastructure Big data, faisant intervenir une autre catégorie d'acteurs.

Le bon fonctionnement de l'ensemble de l'écosystème de la donnée personnelle culturelle dépend de ce que nous appellerons le « filtre de confiance ».

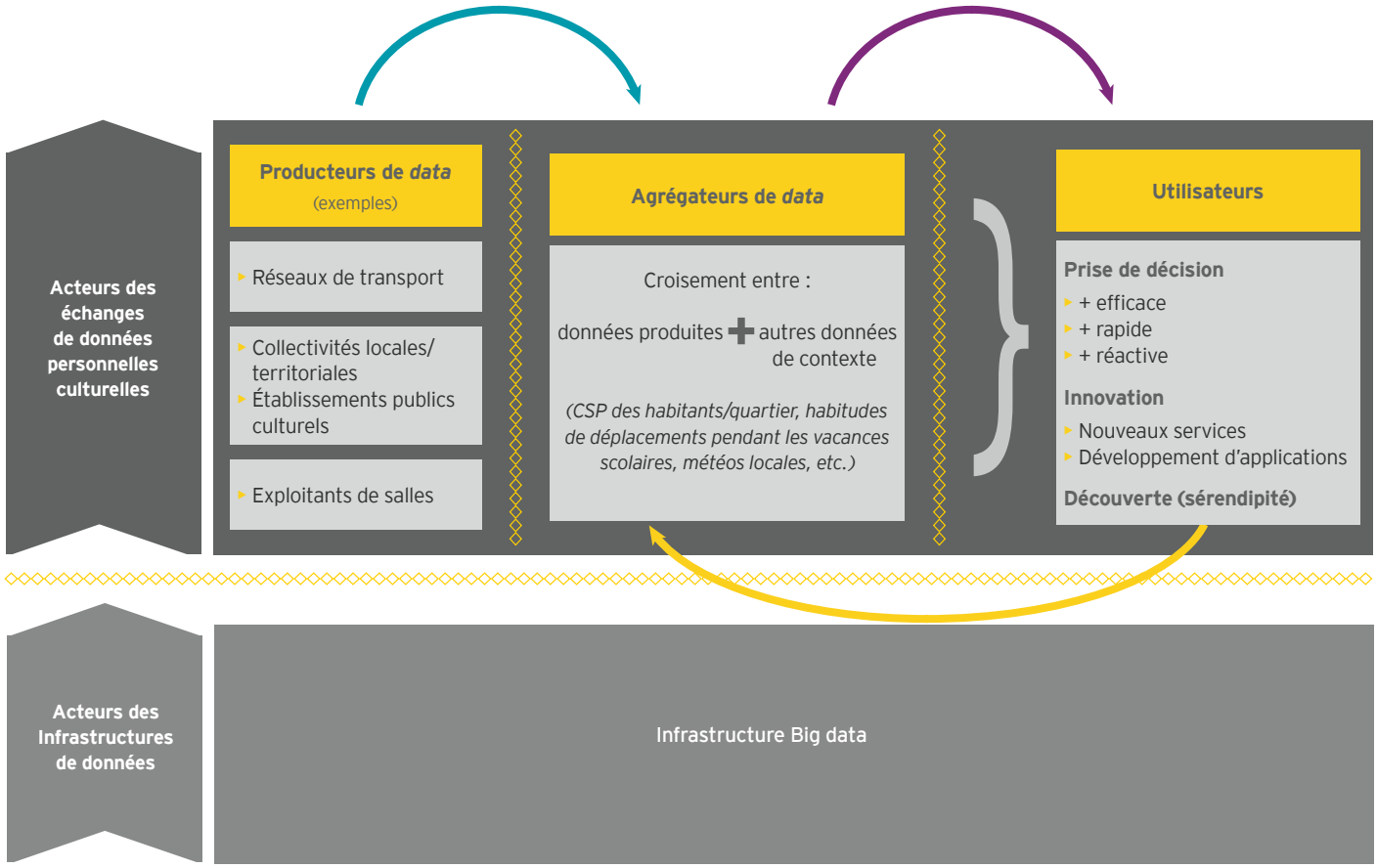
Le marché devrait donc évoluer vers une séparation entre la *data* et son utilisation, une évolution qui aurait le triple avantage d'apporter une réponse à un marché qui évoluerait vers un monopole naturel, à l'enjeu des données nominatives collectées/échangées et de libérer tout le potentiel de la donnée en matière d'innovation. ▸ Voir schéma p. 48

Les acteurs de la nouvelle chaîne de valeur de la *data*

- **Les producteurs de données :** ils constitueraient de véritables bases de données fiables qui, grâce à leur capital confiance, pourraient être revendues à des tiers.
- **Les agrégateurs de données :** ils agrégeraient, croiseraient et contextualiseraient ces données à des fins d'analyse.
- **Les utilisateurs de données :** ils développeraient, à partir des analyses obtenues, des nouveaux services et nouvelles applications. Ces utilisateurs peuvent être des entreprises de média et des *start-up* développant des applications spécifiques et ciblées pour ces groupes de média ou des établissements culturels (application dédiée en marge de l'exposition Hopper au Grand Palais).
- **Les opérateurs d'infrastructures du Big data :** nouveaux acteurs du numériques, opérateurs de télécommunications…
- **Les gestionnaires d'infrastructures :** ils pourront louer, en plus de leurs capacités de stockage, des capacités de calcul à des sociétés ayant des besoins ponctuels en matière de traitement de Big data.
- **Des acteurs de confiance** s'assurant du maintien de l'intégrité/qualité des données, depuis leur *sourcing* jusqu'à leur utilisation, rassurant ainsi utilisateur et consommateur finaux. Cette phase est indispensable pour garantir, par exemple, que la version électronique d'une œuvre, téléchargée en ligne (e-book) corresponde à l'œuvre originale.

Les acteurs de la nouvelle chaîne de valeur de la *data*

Source : EY ©



Où et comment se positionner

Les champions du numérique et les entreprises culturelles vont chercher à étendre leur présence au sein de la chaîne de valeur de la data, horizontalement (de la production à l'utilisation de la data) et/ou verticalement (de l'infrastructure Big data aux services destinés aux utilisateurs et clients finaux). Ils procéderont par croissance interne ou externe, par des alliances avec des *pure players*, etc.

Des *business models (pure players)* et univers de services innovants vont émerger/se développer :

- Le ***Sourcing*** (producteur de données fiables).
- La **spécialisation** dans le développement de services et d'applications mobiles ultra-ciblés (exemple du marché du jeu vidéo mobile).
- **La certification et l'audit liés à la sécurisation des données tout au long de la chaîne de valeur,** pour répondre aux exigences de transparence et de confiance des consommateurs et utilisateurs, mais aussi prévenir le risque de réputation.

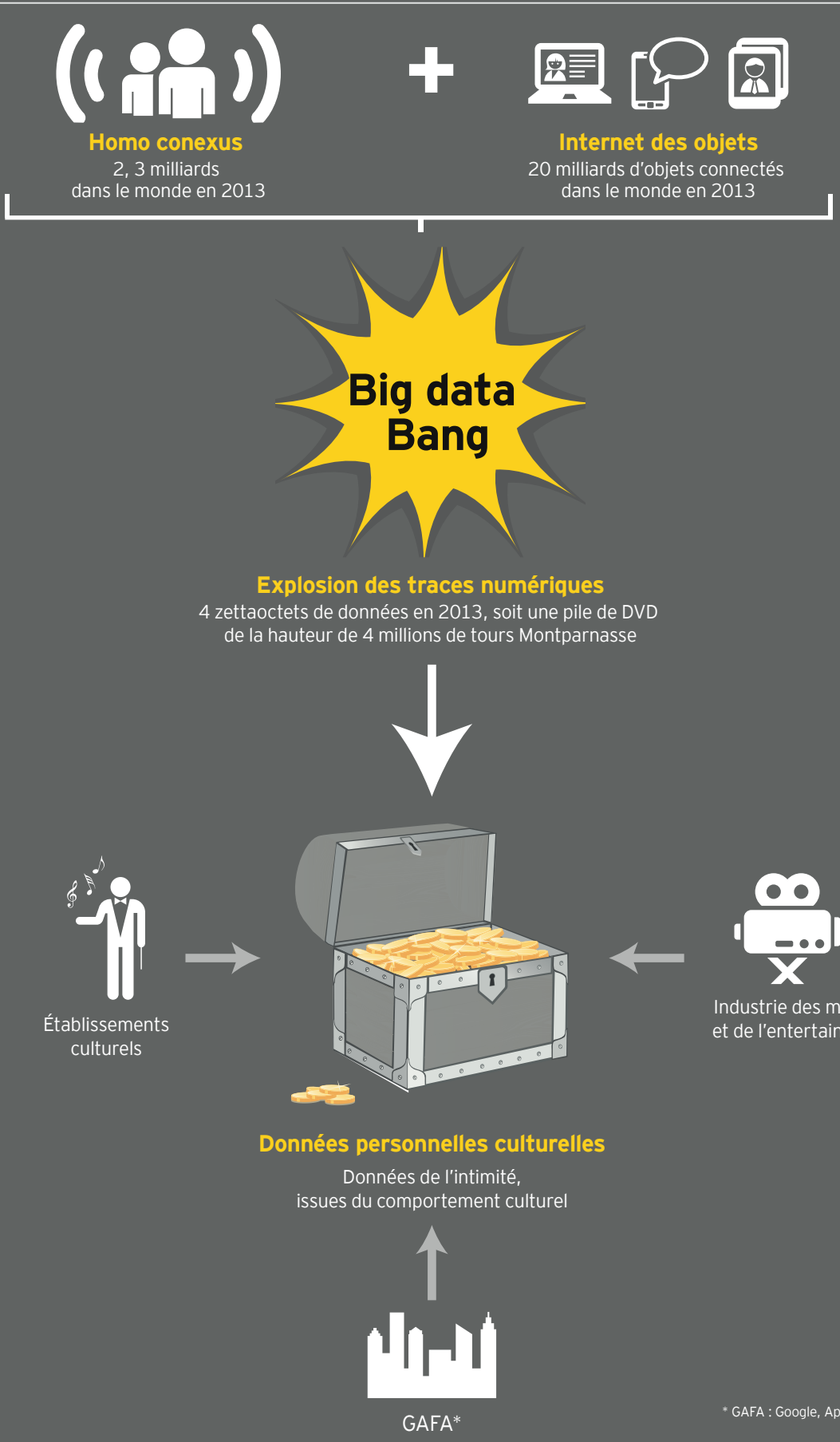
Infographie

La donnée personnelle culturelle au cœur du Big data

La donnée personnelle culturelle, une *data* qui vaut de l'or

Un marché qui se structure autour de la donnée personnelle culturelle

Source : EY ©

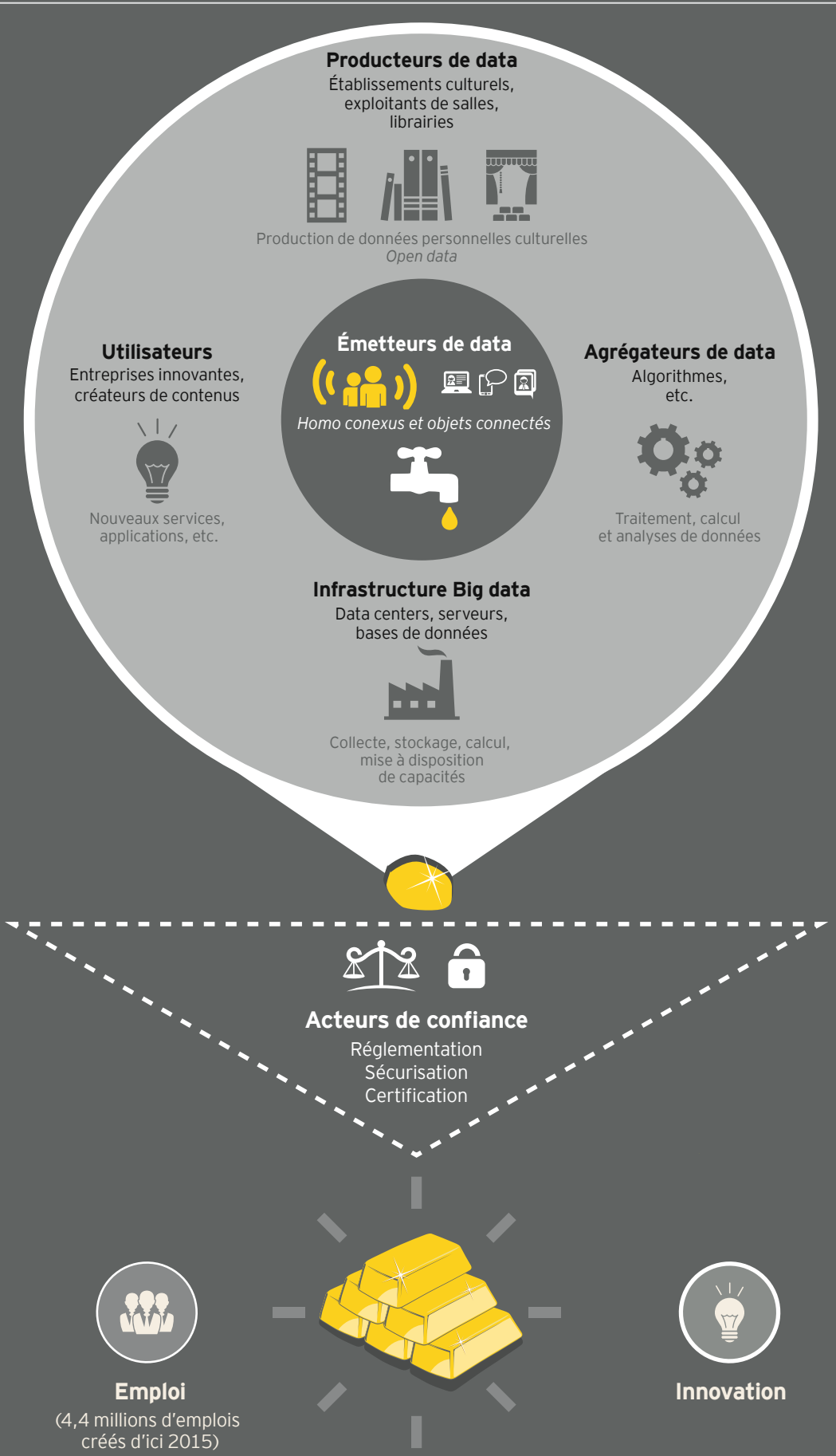


* GAFA : Google, Apple, Facebook, Amazon

Du Big data à la Big value

Vers un écosystème basé sur la confiance

Source : EY ©



EY | Audit | Conseil | Fiscalité & Droit | Transactions

EY est un des leaders mondiaux de l'audit, du conseil, de la fiscalité et du droit, des transactions. Partout dans le monde, notre expertise et la qualité de nos services contribuent à créer les conditions de la confiance dans l'économie et les marchés financiers. Nous faisons grandir les talents afin qu'ensemble, ils accompagnent les organisations vers une croissance pérenne. C'est ainsi que nous jouons un rôle actif dans la construction d'un monde plus juste et plus équilibré pour nos équipes, nos clients et la société dans son ensemble.

EY désigne l'organisation mondiale et peut faire référence à l'un ou plusieurs des membres d'Ernst & Young Global Limited, dont chacun est une entité juridique distincte. Ernst & Young Global Limited, société britannique à responsabilité limitée par garantie, ne fournit pas de prestations aux clients. Retrouvez plus d'informations sur notre organisation sur www.ey.com.

© 2013 Ernst & Young et Associés
Tous droits réservés.

Studio EY France - 1309SG178
SCORE France N° 13-032
Photos : © Fotolia - © Gettyimages - © EY

Document imprimé conformément à l'engagement d'EY de réduire son empreinte sur l'environnement.

Cette publication a valeur d'information générale et ne saurait se substituer à un conseil professionnel en matière comptable, fiscale ou autre. Pour toute question spécifique, vous devez vous adresser à vos conseillers.

ey.com/fr

Contacts

Bruno Perrin
Associé, Ernst & Young et Associés
Responsable du secteur Technologies, Médias et Télécoms en France
Tél. : + 33 1 46 93 65 43
E-mail : bruno.perrin@fr.ey.com

Fabrice Naftalski
Associé, Ernst & Young Société d'Avocats
Tél. : + 33 1 55 61 10 05
E-mail : fabrice.naftalski@ey-avocats.com

Éric Fourel
Associé, Ernst & Young Société d'Avocats
Tél. : + 33 1 55 61 11 22
E-mail : eric.fourel@ey-avocats.com

Solenne Blanc
Directrice Associée, Ernst & Young Advisory
Tél. : + 33 1 46 93 59 17
E-mail : solenne.blanc@fr.ey.com

Louisa Melbouci
Responsable Marketing Technologies, Médias et Télécoms
Tél. : + 33 1 46 93 76 47
E-mail : louisa.melbouci@fr.ey.com

Contact Presse EY

Velislava Chichkova
Tél. : + 33 1 46 93 45 06
E-mail : velislava.chichkova@fr.ey.com

Cette étude a été réalisée par EY, sous la direction de Bruno Perrin, Fabrice Naftalski et Régis Houriez, avec la participation de Marie-Pierre Bonnet-Desplan, Solenne Blanc, Vincent Placer, Louisa Melbouci, Pierrick Vaudour, Guillaume Marcerou, Sébastien Bardou, Aurèle Tabuchi, Colin Garnier, France de Roquemaurel à la rédaction et Sandrine da Cunha au graphisme.