

# Le Bulletin de L'ILLEC

*Ceci n'est pas une newsletter*

## Données, enjeu global et défi marketing – I

### SOMMAIRE

#### BASIQUE

*Éditorial*

page 2

#### ENCORE DU CHEMIN AVANT

##### L'OMNISCIENCE

*Entretien avec Christophe Benavent,  
professeur à l'université Paris Ouest*

page 1

#### “DATA”, PRIORITÉ STRATÉGIQUE

##### CHEZ AUCHAN

*Entretien avec Martine Demaret,  
directrice études et données, Auchan  
France*

page 4

#### ATTENTION, OLIGOPOLE

*Entretien avec David Fayon,  
administrateur des postes et télécoms*

page 5

#### CROISSANCE D'ABORD, RÉGULATION ENSUITE

*Entretien avec Gilles Babinet, chargé  
des enjeux du numérique pour la  
France auprès de la Commission  
européenne*

page 7

### Encore du chemin avant l'omniscience

**L**es ressources techniques et intellectuelles d'une nouvelle ère pour le marketing sont là, autour des données. Mais à l'échelon des entreprises, les ressources humaines de leur exploitation sont trop rares. Notamment dans la distribution.

*Entretien avec Christophe Benavent, professeur à l'université Paris Ouest*

■ *Quand on parle d'« intelligence [du] consommateur », en liaison avec l'idée de « big data », s'agit-il de l'intelligence dont font preuve les consommateurs, ou de celle que le marketing déploie pour les connaître ?*

**Christophe Benavent** : Intelligence du consommateur, oui. Ce qui complique un peu les choses, c'est que le consommateur est idiot et paresseux : idiot, car il n'est pas capable de traiter toutes les informations dont il dispose, et paresseux, car il ne fait pas l'effort de traiter cette information. Remarquons qu'il est raisonnable d'être idiot et paresseux, car si le consommateur voulait vraiment être intelligent, il se trouverait dans la situation de l'âne de Buridan, et il lui faudrait dix heures pour faire ses courses. Grâce aux moteurs de comparaison, le consommateur devient, si ce n'est intelligent, du moins malin. Il est assisté, augmenté.

Une autre catégorie d'intelligence a son origine dans la définition anglaise du mot « intelligence », qui distingue la capacité de compréhension du processus de renseignement. L'intelligence marketing prend son sens dans l'idée de renseignement, qui consiste à avoir l'information qu'il faut, où il faut, pour agir quand il faut. C'est une des vertus du big data. Les neuf mille collaborateurs de Burberry disposent de tablettes nourries par un système CRM<sup>1</sup> dans lesquelles figurent tous les renseignements concernant leurs clients. L'intérêt ici du big data n'est pas de fournir les données qui établissent un modèle général de comportement, mais d'amener, dans un contexte précis, l'information, le renseignement, qui permettra au vendeur de faire la différence.

(suite page 2)

## Basique

**P**as de chiffres, ils seront caducs dans une heure. Big Data est obèse sans remède. Mais elle est trop solide sur ses bases pour se laisser aller, alors qu'une armée de servants se constitue qui lui soumet le monde. Une armée d'élite, de gestionnaires de données qui tiennent les clés de l'avenir, et qui sait avec elles celles des organisations, marchandes ou politiques. Qui pourrait rivaliser avec leur savoir-faire, dans une économie où la seule activité productive véritable, « l'or noir du XXI<sup>e</sup> siècle », est l'information ? L'enchevêtrement des réseaux multiplie les biens comme le faisait, sans rivaux, la terre des physiocrates : chaque donnée en produit d'autres, et ceux qui en maîtrisent la récolte ont le pas.

La prolifération des données, leur disponibilité (« ouverture »), la constitution de bases colossales, semblent renverser l'ordre de l'économie. La superstructure informationnelle (mots, images, sons aussi, par quoi se livrent profils de citoyens et de consommateurs, traits, traces et actions, comportements et opinions) est première, et le reste s'ordonne en fonction. Pas seulement dans la sphère marchande, à l'heure de l'*open data* annoncée par les gouvernements, tandis qu'un récent Livre blanc sur la défense met « l'accent sur la fréquence et l'impact potentiel de la menace que constituent les cyberattaques visant nos systèmes d'information ».

La donnée, c'est la base. Et une base généreuse, pas seulement par son abondance, aussi par son caractère spontané. La donnée est sans façon, elle se donne brut. Au rancart, les épousettes de l'échantillonnage et des critères observables laborieusement tissées ! Avec l'instantanéité de leur production et de leur collecte, les données tendent à se faire exhaustives, explicatives à l'occasion, surtout prédictives avec aplomb. Et la connaissance en sciences humaines voit s'ouvrir une dimension où le « quanti » et le « quali » se fondent.

Le marketing y est déjà engagé. C'est sous cet aspect que ce *Bulletin* et le suivant abordent d'abord la question, avec en filigrane une interrogation sur le risque d'une ivresse des données. Le politique (qui a d'autres sujets d'emballage) accommode ces dimensions nouvelles avec plus de lenteur. Elles lui posent des défis (fiscalité, libertés, souveraineté...) qui seront évoqués aussi – car c'est peu dire que le politique les aborde; il s'en imprègne, et s'en trouvera ébranlé sur ses bases.

François Ehrard

*suite de la page 1*

■ *Toutes les données, directes (effets de l'acte d'achat ou de connexion), indirectes (avis, partages d'expériences...), comportementales, déclaratives, transactionnelles, sont-elles indifféremment exploitables ?*

**C. B. :** Non, bien sûr, vouloir tout exploiter est un vieux fantasme. Les données dépendent principalement de leur mode de production et de leur qualité. Un questionnaire réalisé en sortie de caisses par des enquêteurs professionnels n'aura pas le même contenu ni la même qualité que le même questionnaire adressé sur des interfaces web. Exploiter ce que racontent les consommateurs sur les réseaux sociaux ne donne pas le même résultat qualitatif. Moins de 5 % des Français utilisent Twitter, et parmi eux ceux qui communiquent sur une marque ont des motivations particulières, ils ont en général une attitude favorable ou veulent se faire valoir auprès de leurs amis. Nous ne sommes pas en présence d'un échantillon au sens traditionnel, celui construit par les études. Il est donc très difficile de transformer ces données en informations. Il faut pour cela mettre en place des procédures de redressement et de filtrage. Aujourd'hui, nous avons beaucoup de sources de données, mais elles touchent des populations très différentes. Un vieux rêve des panélistes est d'avoir un seul point de stockage, à partir de sources de données diverses mais qui concernent les mêmes populations. Si la technologie est là, il manque encore une approche raisonnée de l'exploitation des données, et surtout les compétences humaines.

■ *Faut-il s'attendre à une extension sans fin des points de contact, multipliant d'autant les sources de données ?*

**C. B. :** Si les points de contact ne sont pas infinis en nombre, il faut s'attendre à une systématisation de ce qui existe déjà dans la grande consommation. Outre le magasin où de nouveaux points de contact se développent, comme les kiosques, les étiquettes électroniques – testées par Leclerc à Levallois –, les smartphones, les cartes de fidélité, les points de contact sont en aval et en amont. Ajoutons donc les sites comparateurs de prix, les plateformes de réseaux sociaux, les plateformes d'avis, etc. La difficulté ne vient pas du nombre mais des changements de ces plateformes. Par exemple, si en matière de réseaux sociaux leur nombre ne dépasse pas trois ou quatre par personne, avec le temps on constate des désaffections pour certains acteurs, l'apparition de nouveaux, tels que Tumblr, et une rationalisation des usages, particulièrement, aux Etats-Unis et en Grande-Bretagne, chez les jeunes. On semble s'orienter ainsi aujourd'hui vers de micro-messageries telles que Pheed.

■ *Une donnée à T0 peut devenir action à T0 ; la synchronisation, l'instant personnalisé, est-elle le mode opératoire optimal pour un marketeur ?*

**C. B. :** C'est la tendance, et le cœur de la problématique

du big data. Le *trigger marketing* ou marketing activé par les consommateurs est une idée qui remonte aux années 1990. Techniquement, on est en mesure de le mettre en place : les étiquettes de communication en champ proche ou NFC (« *Near Field Communication* ») peuvent communiquer avec des téléphones eux aussi NFC qui disposent d'une fonction – Beam d'Android – permettant le transfert ou la récupération des informations.

Pour autant, aujourd'hui, tous les consommateurs ne sont pas équipés. Une réflexion est encore à mener sur ce type d'interaction, sur ce qu'on peut en faire. Il faudrait en outre modéliser le comportement des acheteurs au regard de ce nouvel outil. Enfin, il faut définir des règles et les introduire dans des moteurs de décision. L'effort ne porte pas tant sur la technologie que sur l'embauche d'analystes nombreux, et sur une connaissance fine des comportements. Seules les grandes entreprises pourront mettre en place de tels outils, du moins dans une première phase, car ils exigent un effort de recherche-développement considérable, qui va bien au-delà de la mise en place des infrastructures techniques.

■ *Le marketing devient-il une science exacte en s'affranchissant du diktat de la moyenne (segmentation, panier moyen...)?*

**C. B. :** Le marketing n'est pas et ne sera jamais une science exacte. C'est une science humaine qui traite du comportement humain, où règne l'incertitude : aucune loi ne détermine précisément le comportement humain, qui est largement probabiliste. Jusqu'à présent, le marketing identifiait des comportements moyens. Le fait d'avoir des données plus fines peut amener ces moyennes à des unités plus petites, par exemple magasin par magasin. On ne raisonnera plus sur une moyenne nationale mais sur des moyennes locales. C'est ce que j'appelle « effet de granularité ». L'avantage du big data n'est pas de connaître mieux mais plus finement, avec un grain très fin. Pour autant, ce grain ne sera jamais l'individu mais celui de niveaux intermédiaires : le rayon, le point de vente... Les bonnes décisions ne sont plus globales mais locales. Pour exploiter ce niveau d'information, il faut des collaborateurs bien formés.

■ *La grande distribution est-elle mieux placée que les industriels pour mener ce travail ?*

**C. B. :** Oui et non. Oui, car une dizaine de centrales d'achat seulement suit, grâce à la caisse du magasin, tout ce qui se vend et s'achète. L'annonceur, lui, est aveugle, il achète donc des données à d'autres acteurs, tels que Symphony IRI. Ce sont ses lunettes. Il peut contourner le problème en donnant des cartes de fidélité à ses clients, mais l'effet

reste limité. Tout le monde n'a pas un égal accès aux données. Mais si les distributeurs disposent d'un avantage, ils ont une culture de la donnée beaucoup moins avancée que celle des industriels. Ils n'ont pas historiquement de culture marketing.

Le métier d'un commerçant n'est pas tant de vendre que de bien acheter et de profiter des effets d'échelle. La concentration du commerce étant forte, l'avantage concurrentiel ne passe plus par la capacité à avoir des

prix moins élevés que les autres, mais par celle de vendre plus cher, ou un peu moins « moins cher ». Donc de développer un véritable savoir-faire et une culture marketing profonde. Dans la distribution, deux modèles se distinguent : le commerce électronique d'Amazon et consort, qui repose sur la personnalisation (et le marketing) avec des techniques telles que le filtrage collaboratif et la capacité de fabriquer une boutique, un site par client, et

le commerce moderne en dur, centré sur le volume des achats. Le secteur des biens culturels, avec Virgin ou la Fnac, est l'exemple éclatant de cette rivalité. Ces enseignes ne peuvent offrir des prix vraiment plus bas, et vendre une grande diversité de biens dans un espace restreint, tandis qu'en individualisant ce qu'il propose Amazon augmente ses chances d'être vu et élargit son espace d'attention.

■ *Le « déluge de données » expose-t-il plus les marques au risque de fausses informations ou de piratage que le marketing traditionnel ?*

**C. B. :** On peut estimer qu'un tiers des données qui circulent sur les réseaux sociaux sont erronées : à cause d'erreurs techniques ou de mensonges délibérés des consommateurs, qui veulent protéger leur vie privée. Mais au-delà, se pose également le problème de leur interprétation, du sens des mots, par exemple dans l'analyse des données textuelles, qui dépend du contexte. C'est le problème général du « TAL » (traitement automatique du langage). Des données, on en a beaucoup, il n'est pas sûr qu'on ait davantage d'informations. C'est sans doute un des enjeux principaux du big data : la fiabilisation des données et la capacité de les transformer en information. Par exemple, un prénom est un ensemble de données de quelques octets qui sous certaines conditions fournit une information sur l'âge. Les photos sont des données aussi, de plusieurs millions d'octets, qui peuvent donner la même information, à condition de disposer d'un algorithme permettant d'inférer l'âge à partir de la reconnaissance de la forme des visages.

■ *Si une marque a le moyen de cibler en continu ses consommateurs potentiels (comme en mode boîte noire), a-t-elle encore besoin d'en conserver la mémoire (« connaissance du consommateur ») ?*

**C. B.** : L'être humain est ce qu'est son histoire. Pour comprendre comment il agit, interpréter son comportement, on a besoin de connaître son passé. Et ce n'est pas suffisant, il faut en plus connaître son environnement. On a donc besoin d'une mémoire, pour historiciser les comportements. Lorsque par exemple les comportements sont saisonniers, et beaucoup dans notre consommation le sont, il faut au moins quatre ans pour le saisir. Cependant, dans la pratique statistique, on constate que beaucoup de phénomènes sont sans mémoire (on les appelle « processus poissonniens ») : il n'est pas nécessaire de savoir ce qui s'est passé hier pour savoir ce qui va se passer aujourd'hui. Il n'y a donc pas de réponse générale à la question de l'utilité de la *mémoire*, c'est affaire de contexte.

■ *Le marketing pourrait-il céder à une ivresse des données ?*

**C. B.** : L'ivresse de la mode des données, oui ! Au reste, une série de mythes se construit autour du discours du big data, qui peuvent conduire certains à surinvestir quand ils ne le devraient pas, quand d'autres sont freinés alors qu'ils devraient investir.

■ *Big data et ses outils marquent-ils la fin du gaspillage de données ?*

**C. B.** : Bien au contraire, c'est la porte ouverte aux gaspillages. C'est même la philosophie du big data, sa

raison d'être : l'idée que la moindre action humaine peut laisser des traces que l'on recueille et qui seront un jour utiles. Toutefois, le stockage des données coûtant de moins en moins cher, c'est de moins en moins un gaspillage. C'est un gaspillage d'effort et non d'argent.

■ *Les outils prédictifs à l'œuvre avec les mégabases de données sont-ils performants au-delà du court terme, ou myopes aux signaux faibles ?*

**C. B.** : Les outils prédictifs utilisés aujourd'hui ont entre trente et cent ans ! Rien de neuf sous le soleil, excepté le vocabulaire. Ces outils, simplement, s'appliquent à un ensemble de données plus vaste. Règle d'or : l'étendue des données n'améliore pas la qualité d'un modèle, comme le démontre la théorie des sondages qu'on apprend en première année de sciences économiques. La précision ne dépend que de la taille des échantillons et elle évolue comme le carré de la taille de l'échantillon : chaque fois qu'on multiplie par quatre l'échantillon, on divise l'imprécision par deux. Pour une précision de plus ou moins 1% il suffit d'un échantillon de 10 000 individus représentatifs. En avoir un million ne sert à rien. Quant aux signaux faibles, ils n'appartiennent pas à la science mais à l'art divinatoire !

*Propos recueillis par Jean Watin-Augouard*

1. *Gestion de la relation avec la clientèle (customer relationship management).*

## “Data”, priorité stratégique chez Auchan

**L**e commerce traditionnel a été jusqu'à présent peu préoccupé de données clients. Devenues numériques et capital précieuses, les données changent la donne.

*Entretien avec Martine Demaret, directrice études et données, Auchan France*

■ *Les données sont-elles devenues seulement beaucoup plus massives, ou plus diverses et nouvelles en qualité ?*

**Martine Demaret** : Elles sont à la fois plus nombreuses et de sources plus variées, notamment les données captées sur les parcours numériques de navigation, internet, mobile. Les données sont moins structurées que celles que nous avions, auparavant, l'habitude de traiter, des données plus formatées comme les comportements d'achat, les tickets de caisse, les commandes. Derrière le big data, il y a un enjeu de massification du volume des données et de capacité à traiter des données très diverses, sur le plan de la forme comme du contenu. Il y a aussi le pari de trouver les bons moyens pour tirer les informations essentielles, pertinentes, des données, et créer de la valeur ajoutée. Il faut donc trouver les bons niveaux d'agrégat, les bons niveaux de rafraîchissement, pour utiliser les données de

façon optimale.

■ *Big data et ses outils marquent-ils la fin du gaspillage de données ?*

**M. D.** : Oui et non. Oui, car les outils ont pour ambition d'aider à mieux qualifier les données, à faire le tri, à analyser, à trouver les bons modèles de corrélation. Non, car face au volume des données, il y a bien sûr un risque d'éparpillement, on peut se perdre dans le flot d'information. Il est donc nécessaire d'avoir toujours des démarches pragmatiques et itératives, pour éviter de s'égarer dans la masse des données comme ce fut le cas, au début des années 1990, quand on a commencé à traiter les données de caisses. On peut se perdre si on n'a pas un objectif de métier précis. Il faut rester vigilant et exprimer clairement la problématique, avant de démarrer toute analyse.

■ *La question du big data est-elle apparue chez Auchan comme un tournant dans sa stratégie, et si oui, quand et comment ?*

**M. D.** : Un tournant, oui, et de façon récente, car le traitement des données n'est pas notre cœur de métier, comme il peut l'être pour les véricistes et le e-commerce, pour lesquels la donnée client est un capital indéniable. Notre capital est plus lié aux surfaces de nos magasins, aux

produits, au trafic que l'on peut gérer sans données clients. Mais il est vrai que la donne change avec les parcours de courses connectés et les diverses nouvelles formes de commerce. Le « data numérique » va devenir hautement stratégique. La réflexion est en cours sur la feuille de route.

■ *Dans quelle mesure les nouveaux outils d'informatique décisionnelle (collecte, consolidation, modélisation des données...) ont-ils amélioré la pertinence et la performance de votre offre ?*

**M. D.** : Nous sommes encore aux prémices. Les nouveaux outils, notamment les outils de *data-mining* (exploration de données) peuvent être utilisés par différents métiers, le marketing, le commercial... Cela permet un décloisonnement des données. Elles ne seront plus uniquement dans les mains du service informatique. Et cela nous permettra de monter en pertinence, en agilité et en efficacité, et d'analyser nos performances commerciales.

■ *Certains panels seraient-ils hypertrophiés ? Y a-t-il trop de données ?*

**M. D.** : Oui, dans l'univers de la grande distribution, le panel reste très utilisé et sans prise en compte d'éléments nouveaux qui pourraient interroger ceux auxquels on est habitué, ou apporter des éléments de réponse complémentaires. Le problème n'est pas qu'il y a trop de données, mais qu'elles ne soient pas utilisées de manière optimale, faute d'organisation, d'outils et de compétence pour le faire. La grande surface est donc toujours tributaire des panels, alors que la « data client » numérique pourrait apporter, elle aussi, une aide à la décision aussi pertinente, voire plus précise dans certains cas.

## Attention, oligopole

**L**a révolution des données peut changer les rapports offre-demande ou les processus de développement économique des nations. Elle peut aussi être accaparée au seul bénéfice de quelques opérateurs ou Etats tentés d'en verrouiller l'usage.

*Entretien avec David Fayon, administrateur des postes et télécoms<sup>1</sup>*

■ *La création d'entreprises en liaison avec le « big data » est-elle dynamique en France ?*

**David Fayon** : Oui, dans la mesure où le big data, traitement de quantités massives de données, nécessite de nouveaux outils dotés d'algorithmes complexes. Les formations telles que Normale Sup, Polytechnique et certaines écoles d'ingénieurs préparent bien aux nouveaux métiers. La création d'entreprises est donc possible en France, mais elle est dynamique à l'étranger aussi, ce qu'atteste l'expatriation

■ *Quels nouveaux métiers voyez-vous émerger chez Auchan, en liaison avec l'économie des données ?*

**M. D.** : Tous les métiers d'analystes, de *data-miners*. Pour l'heure, il est difficile de les recruter, car nous n'avons pas encore de services structurés comme ceux du commerce électronique. Nous sommes donc moins légitimes à leurs yeux. A ce premier type de profil, qui devient une nécessité pour nous, ajoutons un deuxième type, le *data quality manager*, qui saura gérer et gouverner la qualité des données.

■ *Les gestionnaires de données sont-ils les dirigeants de demain ?*

**M. D.** : Non, mais ils auront une place plus importante, et cela va modifier les organisations marketing, qui ne seront plus seulement orientées vers le discours de la marque et la communication, mais devront avoir besoin de fortes compétences techniques et statistiques. Demain, les directions du marketing auront un budget technologique très certainement supérieur à certaines directions informatiques.

■ *La culture et le management, chez Auchan, sont-ils en train de changer du fait de ces évolutions ?*

**M. D.** : C'est encore récent, mais il y a une prise de conscience réelle. L'utilisation de la donnée est de plus en plus systématique. Nous avons aussi décidé, en 2012, de créer une direction regroupement les études marketing et la donnée, signe fort de cette évolution.

*Propos recueillis par J. W.-A.*

de nombreux Français aux Etats-Unis, pour des raisons d'attractivité propre, de reconnaissance ou de moindre pression fiscale. Les entreprises françaises comme Criteo, qui a une antenne à San Francisco, sont des exceptions.

■ *Les outils prédictifs qui scrutent des quantités colossales de données sont-ils performants au-delà du court terme, ou myopes aux signaux faibles ?*

**D. F.** : Il est très difficile de repérer dans l'océan de données, sur le Web en particulier, les signaux faibles et les informations marginales mais porteuses de sens. Ainsi, dans le domaine du marketing, le mécontentement des clients, les suggestions pour améliorer ou concevoir des produits... N'oublions pas que le big data concerne aussi bien Internet que les masses de données accumulées dans les entreprises et autres organisations : intranets et wikis d'entreprises. Le champ est très vaste.

■ *La révolution des données détruit-elle des métiers ? Autant qu'elle en crée ?*

**D. F. :** La destruction de métier n'est pas une spécificité du big data, elle concerne, comme l'a bien analysé Schumpeter et son principe de destruction créatrice, toute révolution, hier agricole, puis industrielle, puis servicielle. Nous entrons dans la quatrième révolution, celle des données, elle aussi destructrice et créatrice d'emplois. Pour autant, il est difficile, aujourd'hui, de quantifier les volumes détruits et induits.

■ *Quelles sont les politiques publiques engagées autour du marché des données, en vue de le réguler ou de le structurer ?*

**D. F. :** Les politiques publiques sont de deux ordres : celui du laisser-faire, où, faute de politique, la France se singularise par une faible attractivité de son territoire et peu de grappes d'entreprises ; et celui d'une politique publique indirecte, où quelques pôles de compétitivité émergent, comme en Ile-de-France Cap Digital ou Systematic, ou le plateau de Saclay, qui réunit de grandes écoles et des viviers pour former les profils de demain en brassant chercheurs, étudiants et industriels. Cette politique indirecte favorise la création d'un terreau où pourront émerger des entreprises innovantes. Pour autant, il faut déplorer l'absence de politique de grande envergure comme le fut le Plan Calcul en 1966.

Il ne faut pas oublier qu'outre le big data il y a l'*open data*. L'*open data* définit toutes les données ouvertes de l'Etat, d'entreprises, d'institutions, d'associations, publiées sur Internet et réutilisables par des tiers afin de développer de nouvelles applications, de nouveaux usages. Par exemple, la géolocalisation des bureaux de postes et leurs horaires d'ouverture, ou la géolocalisation des stations de métro permettront de bâtir des applications. Ces données peuvent être encapsulées dans d'autres applications qui permettent l'optimisation de trajets et offrent des services qui restent à imaginer.

■ *Toutes les données, directes (effet de l'acte d'achat ou de connexion), indirectes (avis, partages d'expériences...), comportementales, déclaratives, transactionnelles, sont-elles à considérer comme des ressources valorisables et... fiscalisables ?*

**D. F. :** Oui, même s'il faut tenir compte du nécessaire retraitement des données ; certaines sont inutiles, d'autres redondantes, d'autres encore fausses. Application typique pour une marque : quels sont les mots utilisés par les internautes pour la décrire, dans un sens positif, négatif ou neutre ; quelle est son image, son « e-réputation » ? On fore aujourd'hui les données comme hier on forait pour le pétrole. Elles sont notre nouvelle énergie.

La question fiscale est liée à celle de l'ouverture : le rapport Collin & Colin<sup>2</sup> préconise une fiscalisation des données, et de moins taxer les entreprises qui rendraient les leurs publiques. Cela permettrait de taxer plus lourdement les Facebook et consort qui verrouillent les données créatrices de valeur. La mission Etalab ([www.etalab.gouv.fr](http://www.etalab.gouv.fr)), qui administre le portail interministériel [Data.gouv.fr](http://Data.gouv.fr), est chargée de réfléchir sur le sujet.

■ *Ce que Collin & Colin appellent « travail gratuit des utilisateurs » doit-il être systématiquement assujéti à l'impôt comme une ressource des acteurs du numérique ?*

**D. F. :** Oui, c'est innovant, mais complexe à mettre en œuvre.

■ *En plus de la collecte et du traitement, va-t-on vers un marché de la protection des données ?*

**D. F. :** Oui, le plus grand supermarché de la donnée, aujourd'hui, c'est Facebook, où on vend son âme au diable en l'échange de la gratuité apparente du service.

■ *Y a-t-il un enjeu de souveraineté ? Un grand opérateur global de l'économie numérique est-il susceptible de différencier sa politique tarifaire pour des motifs politiques ?*

**D. F. :** Un acteur peut très vite atteindre la taille critique, rendre les internautes prisonniers. Cela pourrait, poussé à l'extrême, devenir liberticide, d'autant qu'il n'existe pas de droit à l'oubli sur Internet.

■ *L'appartenance linguistique exerce-t-elle une influence sur l'économie numérique ? (Talend est un leader français, qui travaille dans l'intégration de données, mais l'analyse de données paraît entièrement américaine : question de langue ?)*

**D. F. :** Non, car si l'algorithme est parfait, il peut s'adapter à toutes les langues. Google, Twitter et Facebook ont été traduits en plusieurs langues.

■ *Le phénomène big data et la révolution liée aux données sont-ils une chance historique pour les pays émergents ?*

**D. F. :** Oui, cette révolution leur permet de sauter l'étape de la révolution industrielle, polluante. Ces pays peuvent ainsi se sublimer en passant de la révolution agricole à la révolution des données et rattraper leur retard.

■ *Le big data inaugure-t-il un nouveau type d'économie de la demande, où à la problématique keynésienne de la demande solvable s'ajouterait celle de la demande créatrice (coconception...) ?*

**D. F. :** La valorisation des données change effectivement le rapport traditionnel offre-demande. La demande devient créatrice et la création va au-delà de l'entreprise, pour aller vers des entreprises captant cette intelligence apportée par les données (Facebook, Google, Twitter et les autres, qui savent bâtir un écosystème autour d'eux).

■ *YouTube (Google) sur Facebook, moteur Google par défaut sur Mac... Y a-t-il un risque d'oligopole mondial sur le marché des données ? Faut-il un Sherman Act international ?*

**D. F.** : Oui. La position de Google, Amazon, Facebook et Apple, les « GAFA », y invite. Trois des quatre géants du Web ont des profitabilités à deux chiffres, et la quatrième, Amazon, mise sur la croissance pour détrôner Walmart et devenir à terme le premier supermarché mondial. L'oligopole ne concerne pas seulement les données mais les conditions de leur stockage. Aussi les enjeux du big data et de l'open data sont-ils liés au *cloud computing*. Mais un Sherman Act international... je ne vois pas les Etats-Unis se tirer une balle dans le pied !

■ *La manipulation et l'analyse des big data œuvrent-elles dans le sens d'un retour à la centralisation de la décision (en entreprise avant de l'être en politique ?), les échelons intermédiaires se voyant dessaisir de responsabilités qui peuvent être traitées en bouquet par le sommet ?*

**D. F.** : Même si la Toile est une structure décentralisée, d'égal à égal, certains acteurs comme Google deviennent oligopolistiques, avec une vision centralisatrice des données, donc de la valeur. On entre alors dans des systèmes « bigbrotheriens ».

■ *L'open data des Etats est-il vraiment engagé ?*

**D. F.** : Les initiatives sont encore timides. Pour autant, elles ont du sens, pour créer du service auprès des administrés et d'entreprises partenaires.

■ *Le big data est-il une promesse d'« offshoreleaks » à répétition ?*

**D. F.** : Je ne le crois pas, car les sociétés verrouillent le marché et s'efforcent de ne pas rendre interoperables leurs données, comme le montre l'absence de passerelles entre LinkedIn et Facebook, alors que dans le même temps les contacts de courriels peuvent y être importés pour accroître le parc d'utilisateurs.

■ *Où se situe aujourd'hui le plus grand risque pour la vie privée (d'internautes français) : que les données personnelles soient surexploitées à des fins commerciales, ou que celles qui sont collectées par des moteurs, réseaux sociaux ou sites marchands soient accaparées par des Etats ?*

**D. F.** : Dans un contexte de crise, des données aux mains d'un pouvoir politique ne sont plus neutres et peuvent être utilisées à des fins de marquage, par des dirigeants autoritaires voire totalitaires. La vigilance est de rigueur.

*Propos recueillis par J. W.-A.*

1. *David Fayon est directeur de projets SI à la direction des systèmes d'information des activités courrier de La Poste. Il a publié Géopolitique d'Internet, qui gouverne le monde ? (Economica, 2013) et précédemment Facebook, Twitter et les autres (Pearson, 2012).*

2. [www.economie.gouv.fr/rapport-sur-la-fiscalite-du-secteur-numerique](http://www.economie.gouv.fr/rapport-sur-la-fiscalite-du-secteur-numerique).

## Croissance d'abord, régulation ensuite

**E**n vue d'un impact économique de l'ingénierie des données aussi favorable que possible pour l'économie française, les politiques doivent défendre les interfaces ouvertes.

*Entretien avec Gilles Babinet, chargé des enjeux du numérique pour la France auprès de la Commission européenne*

■ *La révolution « big data » détruit-elle des métiers, autant qu'elle en crée, notamment en évitant des rapports et ou études ?*

**Gilles Babinet** : Il est encore difficile de répondre. Le « big data » doit sans doute avoir un effet comparable au numérique au sens large. Les relocalisations d'emplois et de valeur sont très difficiles à mesurer. D'une façon générale, on considère que le numérique accroît la création de valeur et probablement la répartition des richesses.

■ *Y a-t-il un enjeu de souveraineté ? Un grand opérateur global de l'économie numérique est-il susceptible de différencier sa politique tarifaire pour des motifs politiques ?*

**G. B.** : Non seulement il y a un enjeu de souveraineté, mais il est fondamental. Regardez l'industrie du tourisme.

Expedia, Hotels.com, TripAdvisor et consort ont tout simplement aspiré la marge de ces secteurs. Le départ récent du PDG du groupe Accor illustre une prise de conscience encore limitée des actionnaires du Cac 40 face à cette fragilité.

■ *Y a-t-il un risque de fracture économique entre grands groupes et PME par la capacité à user des données ?*

**G. B.** : Pas vraiment, car ces technologies sont dans le « Cloud », le Nuage. Une petite PME a potentiellement accès aux mêmes technologies que les grands groupes. L'imagination et l'expertise deviennent déterminantes.

■ *L'appartenance linguistique exerce-t-elle une influence sur l'économie numérique ?*

**G. B.** : Oui, par exemple dans le cas des ontologies médicales : le fait d'avoir une grosse masse de données en anglais simplifie considérablement la recherche de « *patterns* » dans les données, et donc la découverte de modèles pathologiques.

■ *La collecte de l'information coûte peu (à Facebook et autres), mais son traitement est coûteux. Est-il en train de se dévaloriser avec les algorithmes adaptés au big data ?*

**G. B.** : C'est tout l'enjeu du big data. Après avoir réduit à zéro les coûts de stockage et de distribution, la technologie réduit également à zéro ou presque les coûts d'algorithmie.

■ *Toutes les données, directes, comportementales, déclaratives, transactionnelles, sont-elles toutes à considérer comme des ressources valorisables et fiscalisables ?*

**G. B.** : Le rapport Colin et Collin ne dit pas cela. Il estime que si les données ne sont pas accessibles à leurs propriétaires, les entreprises, qui créent de la valeur à partir de ces données, doivent être fiscalisées, ce qui est différent.

■ *La fiscalité incitative appliquée à la collecte et au traitement que suggère le rapport Colin et Collin pourrait-elle mettre à contribution les « entreprises globales de l'économie numérique » sans pénaliser la profitabilité des opérateurs nationaux qui acquittent déjà l'IS ?*

**G. B.** : Non, et à ma connaissance ce n'est pas l'idée. L'idée est de créer un écosystème favorable aux interfaces d'application ouvertes (API : *application program interfaces*), donc favorable à la création d'une économie de l'API.

■ *Le rapport Colin et Collin propose, dans le périmètre des entreprises « en amorçage dans l'économie numérique », la fusion du CIR et du statut jeune entreprise innovante (JEI) : quelles seraient les entreprises visées ? Comment traiter celles dont la R&D ne concerne pas seulement le numérique ?*

**G. B.** : Tout cela est à reconsidérer dans le cadre des annonces faites par le Président. En réalité, tous ces mécanismes sont assez complexes et nécessitent un toilettage, a minima.

■ *En plus de la collecte et de traitement, va-t-on vers un marché de la protection des données ?*

**G. B.** : Je parlerai plutôt de management de données. Les entreprises qui ne percevront pas que leurs données

sont des ressources stratégiques se mettront en grand danger. Celles qui chercheront à les cacher ne feront que construire des châteaux de sable face à la marée montante.

■ *Quelles sont les politiques publiques engagées autour du marché des données (en vue de le réguler ou de le structurer) ?*

**G. B.** : En France, la Cnil dit s'emparer du sujet, et exprime l'inquiétude que la synchronisation de jeux de données différents puisse permettre de réidentifier les utilisateurs. Il me semble qu'il est nécessaire de laisser l'innovation agir avant de réguler, plutôt que de prendre le risque de nous mettre dans une situation d'infériorité par rapport à d'autres régions du monde qui n'auraient pas eu cette approche.

■ *Dans la prolifération des bases de données, le droit à l'oubli est-il chimérique ?*

**G. B.** : On parviendra sans doute à prévoir un droit à l'oubli auprès des grands acteurs, mais sur le fond, c'est un leurre ; il suffit de connaître un petit peu le fonctionnement de l'internet pour savoir que sa mémoire est sans limite et qu'il lui est de plus en plus difficile d'oublier. De nombreux acteurs se sont lancés dans le stockage de la mémoire du Web.

■ *Les relations interentreprises sont elles exposées au déluge de données ? Les grands opérateurs (Google, FB et autres) s'intéressent-ils aussi aux données B2B ?*

**G. B.** : Ces acteurs, et surtout ceux qui sont déjà présents dans les entreprises (Salesforce, Omniture), aimeraient beaucoup réexploiter ces données. Pour l'instant, les entreprises se défendent mieux que les particuliers, mais pour combien de temps ?

*Propos recueillis par J. W.-A.*

---

**Bulletin de l'Institut de liaisons et d'études des industries de consommation**

Directeur de la publication : Dominique de Gramont – Editeur : Trademark Ride, 93, rue de la Santé, 75013 Paris (tél. 01 45 89 67 36 - fax 01 45 89 78 74, [jwa@tmride.fr](mailto:jwa@tmride.fr), [www.trademarkride.com](http://www.trademarkride.com)) – Rédacteur en chef : Jean Watin-Augouard – Secrétaire général de rédaction et contact : François Ehrhard (01 45 00 93 88, [francois.ehrhard@ilec.asso.fr](mailto:francois.ehrhard@ilec.asso.fr)) – Maquette et mise en pages : Graph'i Page (06 85 91 40 33, [ividalie@orange.fr](mailto:ividalie@orange.fr))  
Imprimé par : SB Graphic, rue Isaac Newton, ZA de Chanteloup, 93603 Aulnay-sous-Bois ( tél. 01 48 68 05 45) – ISSN : 1271-6200  
Dépôt légal : à parution – Reproduction interdite sauf accord spécial