

# Enquête. Dans le ventre du Big Data en Chine

mercredi 20 juin 2018, par [JING Wu](#) (Date de rédaction antérieure : 23 mai 2018).

**La multiplication des data centers en Chine ouvre un nouveau marché juteux. Moins performant qu'il n'y paraît, il fait naître cependant des ambitions politiques de contrôle social étendu.**

Nulle part ailleurs qu'en Chine, le Big Data n'est une aussi grande affaire. Outil essentiel à l'exercice du pouvoir et clé de la réforme gouvernementale de la société, il y est promu comme l'instrument d'une stratégie nationale. Le Big Data change la vie de tous, État, entreprise, individu. La Chine est au premier plan mondial en termes de ressources numériques disponibles, et elle caracole en tête dans le domaine de l'analyse et de l'échange de données. Grâce à un fort soutien gouvernemental, à un contrôle assez lâche et au peu d'importance accordé par la population au respect de sa vie privée, le Big Data est en passe d'incarner la volonté de la Chine de devancer le reste du monde.

## Collecte massive de données

Lorsque vous consultez le dentiste, l'emplacement de la moindre de vos caries est désormais enregistré dans la base de données des services de santé. L'usage d'un simple mot-clé suffit aux Affaires civiles pour savoir combien il y a de célibataires dans votre rue. Vous avez acheté deux livres le mois dernier et vous avez mis trois heures à choisir du vernis à ongles : ce sera soigneusement consigné par [Taobao](#) [principal site chinois de commerce en ligne, filiale d'Alibaba]. Votre habitude de vous connecter sur les réseaux sociaux de bon matin ira enrichir l'escarcelle de Tencent [société mère des réseaux sociaux [QQWeibo](#) et [WeChat](#)], tandis que [le moteur de recherches] [Baidu](#) aura retenu qu'avant la Saint-Valentin, vous avez recherché des billets d'avion bon marché. C'est ainsi que des dizaines de milliers de "vos" données sont collectées et assemblées pour former de gigantesques ensembles, que l'on appelle le Big Data, les "mégadonnées" ou "données massives".

La société de conseil McKinsey définit le Big Data comme le produit de la collecte de données dont le volume dépasse largement la capacité habituelle de traitement des logiciels de bases de données. Les mégadonnées se distinguent des données classiques par la diversité de leurs origines, leur variété et la vitesse extrêmement élevée de leur renouvellement. De plus, du fait de leur caractère complexe et surabondant, elles nécessitent d'être "clarifiées" pour prendre de la valeur.

## "Nouvel or noir" du capitalisme numérique ?

Ces données sont bien sûr utiles pour le marketing, secteur d'application par excellence, mais aussi en matière de gestion des risques et de management de précision. Ainsi, des données récoltées par les compagnies d'électricité peuvent permettre au secteur financier de mieux maîtriser les risques : une utilisation anormale d'électricité par une entreprise augmente la probabilité que celle-ci connaisse des problèmes de gestion, ce qui incitera les banques à la prudence dans l'attribution de prêts. Dans le domaine public, les applications sont encore plus larges. Il est ainsi possible de maîtriser les flux humains dans les lieux publics grâce aux données envoyées par les téléphones portables, afin d'éviter tout risque de bousculade ou de mieux ajuster l'offre en moyens de transport.

En surveillant les opinions exprimées sur les réseaux sociaux, on peut éviter des explosions de mécontentement.

Les plus enthousiastes parlent du Big Data comme d'un "nouvel or noir". Ils affirment que nous sommes en train de passer de l'ère du capitalisme financier à celle du capitalisme numérique.

### **À nouveau produit, nouveau marché**

Le marché offre trois grands types de prestations : la fourniture de données brutes, l'analyse de tendances ou de résultats sur la base de données brutes, et un service d'intermédiaire entre deux partenaires commerciaux.

Les données brutes sont dans la plupart des cas collectées puis traitées à partir de sites web. Ainsi, une plateforme a réuni les informations relatives à 500 000 hôtels en Chine et à l'étranger, avec leur adresse, leurs tarifs et des évaluations, et les vend par paquets, à raison de 2 500 yuans [320 euros] le lot. Il arrive aussi que l'accès soit payant à l'unité, comme un site qui vend 0,01 yuan [1 yuan = 0,13 euro] chaque élément de sa base de données sur l'arythmie cardiaque.

Mais la plupart des entreprises n'ont pas la capacité d'analyser ces chiffres ; elles préfèrent s'adresser à des entreprises spécialisées. AliCloud fournit ce genre de service : dans l'immense base de données des utilisateurs d'Alibaba [sa maison mère], son moteur de recherches marketing Open Ad Monitoring permet de filtrer la catégorie de consommateurs visée pour faire du marketing de précision. Si par exemple un laboratoire veut promouvoir de coûteuses gélules de calcium, les filtres devront permettre de sélectionner les personnes d'âge moyen ou avancé dont les dépenses en produits de santé sont relativement élevées.

Certaines sociétés se bornent à établir des ponts. C'est le cas notamment de GBDEX, [le data center de Guiyang](#) [capitale de la province du Guizhou, dans le sud-ouest de la Chine, où Apple a basé son iCloud]. Lorsque les autorités sanitaires locales ont souhaité mettre en vente une partie de leurs données médicales, elles en ont d'abord fixé le prix avec le site de GBDEX qui les a proposées sur le marché. Une société pharmaceutique de la ville a acheté des données, qui, une fois analysées, vont lui permettre de mettre au point des produits mieux ciblés. Le site a pris une commission de 40 % sur la transaction.

*"Pour l'instant, le vendeur dicte le prix",* explique Zhu Guohui, le directeur adjoint de GBDEX. Cette société est la plus grande plateforme de commercialisation de mégadonnées en Chine ; elle a été créée en 2015 avec l'appui de la province du Guizhou et la bénédiction du Premier ministre Li Keqiang.

Ces deux dernières années, les autorités locales et nationales ont pris des mesures de soutien à l'exploitation du Big Data. C'est ainsi que plus d'une vingtaine de centres de données massives ont vu le jour, en particulier dans des villes comme Guiyang, Shanghai, Wuhan ou Harbin.

Les pouvoirs publics sont à la fois vendeurs et acheteurs. Il peut arriver que la direction des transports d'une région "A" achète à son homologue d'une région "B" des données pouvant lui servir de référence. De même, les services de la Sécurité publique se portent parfois acquéreurs d'informations auprès des compagnies de télécommunication.

### **Rafler les ressources foncières**

Les centres de données sont apparus en Chine vers 2010. Au départ, les professionnels étaient convaincus que ce nouveau secteur d'activités pouvait générer des centaines de milliards de chiffre d'affaires, mais ils ont dû déchanter. Selon [une enquête publiée en janvier par le magazine](#)

[économique Caijing](#), les affaires de la plupart de ces centres, même avec l'appui et le financement des pouvoirs publics, ne marchent pas très fort, et les prestataires privés peinent à dégager des bénéfices.

Certaines administrations engloutissent des millions, voire des dizaines de millions de yuans, dans leurs programmes de données massives, alors que les plateformes créées sont incapables de traiter les informations. La construction de certains centres n'a eu pour but que de rafler les ressources foncières et les avantages accordés par les autorités lors de leur création.

En octobre 2017, après plus de deux ans d'activité, GBDEX, largement subventionné par l'État, n'avait réalisé que 120 millions de yuans [15,5 millions d'euros] de transactions. Rares sont les centres d'échange de mégadonnées qui engrangent plus de 10 millions de yuans de revenus annuels, selon un professionnel interrogé par *Caijing*. Même le grand fournisseur de données [Datatang](#) accuse une perte d'exploitation de 4 millions de yuans en six mois. Selon *Caijing*, à eux tous, ces sites officiels ne dégagent pas plus de 5 milliards de yuans de revenus par an.

### **Les entreprises sont peu partageuses**

*“Les entreprises veillent jalousement sur leurs données, car le risque encouru en les cédant est bien supérieur aux gains potentiels, explique Zhu Guohui. Les données vitales qu’une entreprise a amassées au prix de nombreux efforts peuvent être très vite dupliquées dès qu’elles sortent de l’entreprise.”* Le directeur général exécutif de GBDEX, Wang Canshou, reconnaît :

*Nombreux sont ceux qui acquièrent des informations, rares sont ceux qui en fournissent.”*

Soucieuses de protéger leurs secrets commerciaux, beaucoup de sociétés ne souhaitent pas céder leurs données – et il est encore plus rare qu’elles acceptent de le faire gratuitement.

En juin 2017, la célèbre entreprise chinoise de livraison rapide [SF Express](#) a déclaré la guerre à Cainiao, la filiale logistique d'Alibaba. SF Express a annoncé brusquement qu'elle fermait son interface de données avec Cainiao, accusant cette dernière d'avoir voulu l'étouffer pour l'obliger à rapatrier ses données de QCloud à AliCloud. Il a fallu l'intervention du Bureau national de la Poste et des Communications pour que cesse cette guerre, avec la promesse des deux parties de reprendre la transmission mutuelle de données.

Cette bataille a mis en lumière la féroce concurrence que se livrent ces entreprises. Zhao Xiaomin, spécialiste de la livraison express, explique :

*Pour une société (SF Express) valorisée 200 milliards de yuans sur le marché boursier, il est inenvisageable de donner son sang gratuitement à une autre société (Cainiao).”*

### **Des données trop bien gardées ?**

Il est normal qu'une entreprise cherche à protéger ses données, mais de la part des pouvoirs publics qui devraient tenir les informations à la disposition de tous, vouloir tout cacher apparaît difficilement tenable.

En 2016, le Premier ministre Li Keqiang a indiqué lors d'un sommet consacré au Big Data en Chine que plus de 80 % des ressources dans ce domaine étaient détenues par des administrations. En 2015, dans le cadre d'un plan d'action gouvernemental pour le développement des mégadonnées, les autorités ont affirmé leur volonté d'“ouvrir l'accès aux ressources publiques” afin de faciliter le partage des informations entre services administratifs. L'objectif était alors de publier les données détenues par le gouvernement sur un site accessible à tous les services avant la fin de 2018, et d'ouvrir progressivement au public les données administratives dans des dizaines de domaines dont le crédit, les transports et la santé, avant la fin de 2020. Mais cette évolution est beaucoup plus lente que prévu.

En fait, entreprises comme administrations souhaitent accéder à des données, mais sans partager les leurs ; c'est le serpent qui se mord la queue ! Zhu Guohui constate :

*C'est difficile ! À Guiyang nous avons fait beaucoup d'efforts. Les organismes publics cherchent à montrer l'exemple, mais il reste de nombreux problèmes.”*

Comme [l'indiquait début février l'organe du Parti communiste chinois, le Renmin Ribao](#) [le Quotidien du peuple], nombre de données essentielles dorment dans des cartons, car certaines autorités locales ne perçoivent pas l'importance du partage d'informations et vont même jusqu'à considérer ces ressources comme leur “propriété privée”.

Même des plateformes gouvernementales ayant déjà une certaine expérience sont difficilement accessibles. Lors d'une enquête réalisée à Foshan [dans la province du Guangdong, dans le sud du pays] dans le centre de ressources en mégadonnées du district de Nanhai, [des journalistes de l'hebdomadaire cantonais Nanfang Zhoumo ont découvert](#) que seules 22 % des données proposées par la plateforme étaient en accès entièrement libre.

En fait, même les données en accès libre sont difficiles à obtenir. Ainsi, en 2016, Wang Dengke, qui était étudiant en quatrième année d'université, s'est rendu à Pékin pour créer avec des amis une base de données publiques gouvernementales. L'objectif était, grâce à un robot d'indexation, de réorganiser les données en accès libre pour qu'elles puissent être commercialisées. Mais très rapidement, il a constaté que ce ne serait pas faisable. En effet, l'interface des sites gouvernementaux change à tout bout de champ, ce qui empêche souvent d'y accéder. De plus, les données rapatriées à partir de ces sites sont souvent d'un format incorrect qui rend leur exploitation impossible.

## **Des informations souvent redondantes**

Dans le domaine du Big Data, on entend par qualité des informations le fait qu'elles sont concordantes, précises, complètes et toujours valides. Or, une étude réalisée par l'université des Sciences de l'industrie de Harbin [nord-est du pays] montre que les données des assurances sociales sont souvent impossibles à partager, tandis que les systèmes de la Sécurité publique ou des Affaires civiles fournissent des données redondantes ou incohérentes.

Comme nous l'a confié Wang Xuan, contrôleur qualité chez le consultant en Big Data [Primeton](#), dans le cas des banques, on se retrouve souvent avec une base de données comportant des dizaines de milliers de clients dont les informations sont redondantes, une personne pouvant même sembler détenir plus de mille comptes.

À toutes les étapes de la chaîne, la moindre erreur peut entraîner des défauts de qualité. Il peut y avoir une “contamination” par des données inutiles ou fausses pendant le transfert, ou une

destruction involontaire de données au cours de leur traitement. Dans ce dernier cas, il s'agit d'un problème de capacité à traiter les données. *“On observe un grand écart sur le plan technologique en matière de Big Data entre la Chine et l'étranger. Les technologies proviennent toutes de grandes sociétés étrangères comme Google et autres”*, a confié aux médias Li Guanggan, le directeur du Service de recherches sur les informations au Centre d'études et de développement du gouvernement. En fait, le modèle économique de la Chine est en avance par rapport à son niveau technologique.

Si la qualité des données n'est pas garantie, les conclusions que l'on tire de leur exploitation sont-elles bien solides ?

### **Les données personnelles vendues au marché noir**

L'existence d'un marché noir constitue une autre ombre au tableau. À ce jour, la Chine ne dispose pas d'organisme réglementant les échanges de Big Data d'une région ou d'un secteur industriel à l'autre. Les lois et règlements relatifs à la propriété des données restent très vagues.

*“Le manque de normes et de législation dans le domaine du commerce des mégadonnées en restreint les possibilités de développement”*, constate Zhu Guohui, qui a plusieurs fois évoqué la nécessité *“d'éviter les risques”* (sous-entendu, le risque d'atteinte à la confidentialité des données) au cours de notre interview.

En mai 2017, plusieurs responsables de Datatang ont été interpellés par la police après avoir révélé des informations confidentielles de leurs clients ; la valeur en Bourse de la société a alors dégringolé, passant de 2,1 milliards de yuans à 0,7 milliard [de 271 à 90,5 millions d'euros]. L'affaire a soulevé un vent d'inquiétude chez de nombreux autres professionnels du secteur, qui se sont empressés de remettre en ordre leurs services. Quant à GBDEX, s'il a fait évoluer son cœur d'activité de la collecte des données à leur analyse, c'est surtout pour éviter d'être accusé de ne pas respecter ses obligations de confidentialité.

Il n'en reste pas moins que les données brutes sont celles qui présentent le plus d'intérêt aux yeux des acheteurs, car non seulement elles donnent des informations personnelles exactes, mais elles peuvent être utilisées un nombre de fois illimité. C'est ce qui a favorisé l'apparition d'un marché noir. Zhu Guohui explique :

*Les données valent d'autant plus cher qu'elles sont brutes.”*

L'enquête de *Caijing* montre que la plupart des transactions sont réalisées sur le marché noir. Selon des statistiques incomplètes, 5,53 milliards de données personnelles auraient été divulguées, soit quatre en moyenne par Chinois, 80 % provenant de “taupes” au sein du secteur et 20 % seulement étant le fait de hackers. Sur le marché noir, on peut acheter les informations figurant sur la carte d'identité des individus ou sur les crédits qu'ils ont contractés. Le tout à des tarifs très compétitifs : le dixième du prix pour l'achat par des voies officielles de données de reconnaissance faciale (vendues 10 centimes de yuan l'unité).

Les médias disent régulièrement que les données personnelles circulent sur le marché noir, mais personne ne semble trop s'en émouvoir. Pourtant, en août 2016, Xu Yuyu, une étudiante de 18 ans, est décédée d'un arrêt cardiaque après s'être fait dérober 9 900 yuans [1 280 euros] de frais d'inscription à l'université dans une escroquerie par téléphone. L'enquête a montré que le voleur avait pu sévir grâce à l'achat illégal de 100 000 données personnelles sur les étudiants admis.

## Comment définir les informations “personnelles” ?

L'histoire tragique de Xu Yuyu a fait prendre conscience à la population de l'importance de la sécurité des données. Elle pose une autre question encore plus délicate : les informations que l'on génère nous appartiennent-elles ?

Comment garantir la confidentialité des informations personnelles collectées par l'administration et les entreprises ? La Loi sur la sécurité sur Internet, entrée en vigueur en juin 2017, vise à protéger les informations personnelles, en particulier les noms, dates de naissance, numéros de carte d'identité, adresses, numéros de téléphone etc. des citoyens, mais à l'ère du Big Data, son champ d'action semble bien ténu.

Selon Li Yang, professeure assistante à la faculté de droit de l'Université centrale des minorités de Pékin, le droit chinois prévoit la protection de la vie privée, mais à l'ère du Big Data, la définition des informations personnelles est de plus en plus ardue. Les habitudes de consommation d'une personne, ou encore la fréquence avec laquelle elle emprunte telle ou telle ligne de transport, sont-elles ou non des informations personnelles ?

Le 14 avril 2016, le Parlement européen [adoptait à l'issue de quatre années de négociations un Règlement général européen sur la protection des données personnelles](#), qui entrera en vigueur le 25 mai 2018. La Chine, elle, n'a pas encore légiféré. L'article 127 des Dispositions générales du Code civil adopté en mars 2017 indique certes de façon vague : *“La loi prévoit des règlements en matière de protection des données et des biens virtuels, auxquels il convient de se référer”*, mais la notion de “donnée” n'est pas définie, et il n'est pas précisé qui en est propriétaire. Pour certains juristes, les utilisateurs ne sont que des producteurs de données, et ces dernières n'ont aucune valeur sans leur exploitation par les plateformes, auxquelles doit par conséquent revenir la propriété des données. Ces plateformes ont cependant l'obligation de protéger la confidentialité des utilisateurs.

Certains internautes s'insurgent contre une telle manière de voir : même si les données sont traitées, elles doivent rester la propriété d'un individu. Par exemple, si un site marchand se sert de mes données pour mettre au point un nouveau produit, ne serait-il pas logique que je reçoive une part des profits ? Plus largement, n'est-il pas normal d'avoir le droit de savoir comment sont utilisées mes données ?

Bien sûr, dans la Chine actuelle, ni le gouvernement ni les entreprises n'ont l'intention de céder aux internautes le droit de propriété sur leurs propres données numériques.

## Big Data, Big Brother

Dans son roman *1984*, l'écrivain britannique George Orwell décrit une société totalitaire où Big Brother observe les gens en permanence et en tout lieu à travers les “télécrans” qui constituent ses yeux. Du fait de leur haute sensibilité et de leur omniprésence, ces appareils surveillent en permanence vos moindres gestes et vos moindres paroles. Le Big Data entre dans notre vie privée de façon bien plus douce ; il ne nous rappelle pas à tout moment que *“Big Brother nous regarde”*. Il s'intéresse néanmoins à notre style vestimentaire, nos goûts culinaires, notre mode de logement et nos déplacements, et même à nos proches.

Le trajet que vous suivez chaque jour en sortant de chez vous et le moyen de transport emprunté sont des données, votre historique de navigation sur Internet et vos messages publiés en sont aussi, tout comme l'identité de vos amis et le nom des magasins où vous faites vos courses...

## Un premier pas vers la “notation” des comportements ?

Il suffit que vous laissiez faire, et le Big Data pourra dresser un portrait approximatif de vous-même dans le monde virtuel, et anticiper vos comportements futurs. Dès lors, vous n’aurez plus qu’à vous laisser entraîner dans un système de notation et vous deviendrez quelqu’un de très obéissant, sans même l’intervention de Big Brother.

*“Si vous allez tous les jours au boulot en Mercedes, vous aurez moins de points que moi qui y vais en utilisant un vélo en libre-service. Si vous habitez seul dans une grande maison, vous aurez moins de points que moi qui loue un petit logement pour toute ma famille. Au bout de dix ans, j’aurai ainsi accumulé assez de points pour acheter mon propre logement, tandis que vous, vous devrez continuer encore plusieurs années, d’autant plus que vous voudrez acquérir une grande maison. Si vous jetez par la fenêtre la bouteille d’eau que vous avez bue en conduisant, vous perdrez 10 points. Les rues n’étant pas larges, la priorité est donnée aux transports en commun, et si vous prenez votre propre voiture, vous vous heurterez partout à des restrictions. Au printemps, tout le monde va bénévolement planter des arbres ; vous n’y allez jamais et perdez ainsi 100 points par an.”*

Ce n’est pas du cinéma, ni un extrait de roman, mais bel et bien les paroles d’un responsable du comité de gestion de la ville nouvelle de Xiong’an [zone économique spéciale], dans la province du Hebei. Un mois après cette interview réalisée en octobre 2017, le groupe Alibaba a annoncé avoir conclu un accord de coopération stratégique avec la localité, en vertu duquel il s’engage à *“utiliser la puissance des technologies les plus avancées et des ressources innovantes pour faire du nouveau district de Xiong’an un jalon vers la ville du futur et un modèle en Chine.”*

## Un système de bonus-malus

En 2014, le gouvernement a annoncé la réalisation, d’ici 2020, d’un réseau de “crédit social” [déjà expérimenté dans plusieurs localités du pays]. Ce programme prévoit l’attribution de points aussi bien aux particuliers qu’aux entreprises ou aux administrations publiques, assortie d’un mécanisme de bonus-malus selon que les engagements en termes de crédit sont respectés ou non, afin que, comme le souhaite le Premier ministre Li Keqiang :

*les mauvais emprunteurs aient du mal à progresser et les bons emprunteurs aient le champ libre dans leurs démarches”.*

Comment seront attribuées les notes ? En octobre dernier, [la revue Wired a publié un long article présentant le système de notation “Crédit Sésame”](#) [ou *Zhima Credit*], l’une des plus grandes plateformes de prêt aux particuliers en Chine, filiale d’Alibaba. Un internaute est évalué dans cinq domaines : son historique d’emprunteur, son aptitude à tenir ses engagements, ses caractéristiques personnelles, ses préférences et son comportement, ses relations sociales. Tous ces éléments sont collectés la plupart du temps par Alibaba quand vous utilisez les services de ses filiales Taobao, Tmall [site commercial] ou Alipay.

## Révolutionner le contrôle social

Il n’y a bien sûr pas que les géants des hautes technologies qui ont flairé les immenses évolutions que pouvaient entraîner les mégadonnées : le 8 décembre dernier, le Bureau politique du Parti communiste chinois a consacré sa deuxième session plénière à ce sujet. Xi Jinping, que l’on sait animé de grandes ambitions, a exhorté les dirigeants à suivre de près l’évolution du Big Data ainsi que son influence sur le développement économique et social, et à faciliter le partage des ressources en mégadonnées afin d’accélérer l’édification d’une Chine numérique.

Cependant, atteindre les objectifs assignés par Xi Jinping ne sera pas facile. Tant que les obstacles qui entravent l'échange de données par les entreprises et les pouvoirs publics, et même par les différents services de l'État, n'auront pas été démolis, il sera difficile de mettre en place un système de crédit social efficace.

Comment donc abattre ces "murs" ? Le 24 janvier, Ye Zhenzhen, président de [Renminwang](#) [site du *Quotidien du Peuple*], a avancé une solution : *"les données contrôlées par le Parti"*. Jugeant les mégadonnées aussi importantes que *"les fusils"* (l'armée) ou *"la plume"* (la propagande), il estime impossible de mettre en place une grande plateforme de partage de données sans l'intervention du Parti qui contrôlerait les données, ni sans celle de l'administration qui percerait les "murs" actuels. Ces déclarations ont été très remarquées par le Bureau politique. Alors, le "Big Data du Parti" progressera-t-il vite ? Verrons-nous bientôt surgir un tel monstre marin ? Voilà des questions qui ne resteront pas sans réponse dans les années à venir.

[Lire l'article original](#)

---

**Wu Jing**

[Abonnez-vous](#) à la Lettre de nouveautés du site ESSF et recevez chaque lundi par courriel la liste des articles parus, en français ou en anglais, dans la semaine écoulée.

---

**P.-S.**

Courrier International

<https://www.courrierinternational.com/article/enquete-dans-le-ventre-du-big-data-en-chine>