

Le Big Data: Big Brother?

Ou comment trouver un équilibre stable avec la protection de la vie privée

Stéphane Hurtaud
Partner
Governance, Risk
& Compliance
Deloitte

Laureline Senequier
Senior Manager
Governance, Risk
& Compliance
Deloitte

Aurelia Schwander
Consultant
Governance, Risk
& Compliance
Deloitte

“La disponibilité et l’utilisation des Big Data sont essentielles au maintien de la compétitivité de l’UE. Pour l’heure, il est difficile de savoir clairement à qui appartiennent ces données industrielles et de quelle manière elles peuvent être utilisées ou non. Cette situation n’encourage pas le développement de services d’analyse de données”¹, c’est en ces termes que le Commissaire en charge du portefeuille Economie numérique et Société numérique, Günther Oettinger, décrit la situation du Big Data au niveau européen.

¹ Discours à la Foire d'Hannovre: “Europe's future is digital” https://ec.europa.eu/commission/2014-2019/oettinger/announcements/speech-hannover-messe-europes-future-digital_en



Même s'il peut être source de nouvelles opportunités, le Big Data pose un dilemme pour les acteurs du marché dans la mesure où la législation actuelle ne fournit pas d'indications claires pour en encadrer le développement. Comme point de départ pour résoudre ce dilemme, il importe de comprendre quelles sont les bonnes pratiques actuelles et quelle est la tendance à venir.

Bienvenue dans l'ère du Big data

Ces dernières années ont été marquées par une véritable explosion du volume des « données numériques en temps réel »² grâce aux innovations technologiques et à la démocratisation des appareils numériques. Selon une étude publiée par l'initiative Global Pulse des Nations Unies, le volume de données numériques devrait être multiplié par 44 entre 2007 et 2020 et devrait donc doubler tous les 20 mois. Dans ce contexte, le Big Data répond au besoin des acteurs du marché de réduire l'imprévu en créant des informations pertinentes à partir de données constamment produites.

Nous pouvons définir le Big Data comme la capture de données provenant de sources différentes combinée en une seule et même source.³ La première nouveauté qu'apporte le Big Data est qu'il permet la collecte de données de nature et de format différents tels que

les journaux d'appels téléphoniques, les transactions bancaires mobiles, les posts de blogs, les tweets, les recherches en ligne, les images satellite ou encore les vidéos. Grâce à cela, le Big Data dépasse de loin les possibilités qu'offrent les bases de données traditionnelles dont le fonctionnement implique le traitement d'informations structurées. Le Big Data permet de faire émerger des modèles prédictifs, des corrélations ou des phénomènes depuis un ensemble de données non structurées.

La deuxième nouveauté du Big Data est qu'il implique le traitement de volumes de données d'une ampleur telle que des outils et techniques de calcul spécifiques sont nécessaires pour extraire les données pertinentes et obtenir un résultat exploitable en un laps de temps réduit. Des méthodes telles que l'analyse de grappes (regroupement d'éléments ou d'individus similaires) ou l'analyse associative (co-occurrence d'éléments) permettent de recouper les données et d'obtenir des résultats exploitables. Enfin, le Big Data permet de générer de l'information pertinente à partir d'un flux continu de données. Dès lors, le Big Data représente, tant pour le secteur privé que pour le secteur public, une opportunité en termes de compétitivité ainsi qu'une occasion unique de tirer profit de l'économie numérique.

² *Big Data for Development: Challenges & Opportunities*, UN Global Pulse <http://unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf>

³ *Big data: opportunities and privacy concerns*, Piotr Bąkowski, Service de recherche pour les députés, Service de recherche du Parlement européen [http://www.europarl.europa.eu/RegData/bibliotheque/briefing/2014/140771/LDM_BRI\(2014\)140771_REV1_EN.pdf](http://www.europarl.europa.eu/RegData/bibliotheque/briefing/2014/140771/LDM_BRI(2014)140771_REV1_EN.pdf)

Le Big Data, entre promesses et opportunités

Le traitement de tels volumes de données devrait déboucher sur l'amélioration de la productivité, de la connaissance du besoin du client et, par conséquent, de l'offre de services.

Dans le secteur public, le Big Data pourra contribuer à affiner le processus décisionnel dans des domaines tels que la santé publique, la sécurité nationale, l'urbanisme, la sécurité alimentaire, l'efficacité énergétique ou les systèmes de transport intelligents. Il pourrait par exemple, en complément des sondages et statistiques officiels, « permettre de suivre les progrès en matière de développement, d'améliorer la protection sociale et de comprendre où il convient d'ajuster les politiques et programmes existants »⁴. A cet effet, le retour que fournit l'analyse presque en temps réel par le Big Data sur l'efficacité des politiques publiques devrait réduire le décalage existant en termes de temps et de connaissances entre chaque études, permettant ainsi aux dirigeants politiques d'adopter une approche ciblée et adaptative opportune.

A titre d'exemple, le gouvernement britannique soutient actuellement une initiative Big Data visant, d'une part, à mieux contrôler les services et infrastructures de transports publics et, d'autre part, à mieux comprendre les besoins des utilisateurs. En effet, l'analyse des données relatives aux transports publics permet de mieux comprendre les habitudes des voyageurs et ainsi d'améliorer la planification des services de transport⁵.

Un autre exemple pratique de l'emploi du Big Data dans le secteur public est la tendance des forces de polices à adopter des stratégies ayant recours aux informations produites par les nouvelles technologies, telles que les détecteurs de coup de feu, la vidéosurveillance, les médias sociaux etc. pour la prévention, ou tout du moins, pour une action plus rapide et plus efficace face aux activités criminelles.

Dans le secteur privé, le processus décisionnel pourra aussi bénéficier de l'utilisation du Big Data car, contrairement à la veille stratégique qui vise à expliquer

les tendances actuelles, le Big Data cherche à anticiper le comportement des consommateurs et à prévoir les tendances du marché grâce à des modèles prédictifs permettant une approche proactive et donc la mise en avant de l'offre au moment le plus approprié.

En résumé, le Big data permet de tirer parti du volume croissant de données numériques que nous produisons chaque jour. Cependant, le manque de législation encadrant Big data et, en particulier l'incertitude juridique relative à la protection des données, suscitent des questionnements quant à l'utilisation de Big Data.

La protection de la vie privée à l'ère du yottaoctet : un défi permanent

Chaque jour, « les individus, tels les vers à soie déposant leur soie, laissent des traces permanentes dans le cyberspace qui sont ensuite exploitées à différentes fins par les organismes publics et privés ». Il existe un risque accru de détention d'informations personnelles ou sensibles à laquelle les personnes concernées n'auraient pas consenti dans la mesure où des données à caractère personnel telles que des photos, adresses e-mail, coordonnées bancaires, informations médicales ou adresses IP puissent être collectées à des fins d'analyse ultérieure dans le cadre de Big Data. Il s'agit là d'une préoccupation majeure car le « droit des individus au respect de leur vie privée et familiale, de leur domicile et de leurs communications » est un droit fondamental bénéficiant d'un régime spécifique de protection au niveau européen et constituant une obligation légale pour tous les acteurs du marché.

Par ailleurs, le Big Data est capable, de par sa nature, de relier diverses données qui, prises isolément, ne seraient pas considérées comme des données à caractère personnel mais, qui une fois recoupées, permettent d'identifier des individus. Il s'agit là d'une caractéristique propre au Big Data particulièrement complexe à gérer pour qui l'utilise, car les méthodes classiques de dépersonnalisation ou d'anonymisation complète des données trouvent leurs limites et peuvent ainsi rendre l'utilisation du Big Data illégale.

⁴ *Big and Open Data in Transport*, Chambres du Parlement britannique, Juillet 2014, <http://researchbriefings.files.parliament.uk/documents/POST-PN-472/POST-PN-472.pdf>

⁵ *Rapport du workshop sur la vie privée, les consommateurs, la concurrence et le Big data du 2 juin*, https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Big%20data/14-07_11_EDPS_Report_Workshop_Big_data_EN.pdf

Le Big Data représente un risque pour la protection de la vie privée à cause d'usages impropres ou d'abus volontaires tels que le suivi du comportement, les systèmes de profilage ou de discrimination ou encore l'utilisation de données et d'algorithmes à des fins d'exclusion ou d'évaluation des individus. La possibilité d'analyser et de croiser un certain volume de données afin d'identifier des modèles peut également déboucher sur des débordements tels que la surveillance de masse, à l'instar du programme PRISM de l'Agence pour la sécurité nationale américaine.

Enfin, le Big Data peut entraîner la divulgation d'informations confidentielles à cause du partage de base de données avec des tiers ou bien du manque de protection adéquate des données. Compte tenu de sa nature, le Big Data constitue un défi en termes de protection de la vie privée. En effet, les principes classiques de protection des données ne peuvent pas lui être appliqués en tant que tels. Par exemple, l'article 4 sur la qualité des données de la loi luxembourgeoise sur la protection des données, qui pose les principes de nécessité et de proportionnalité, est incompatible avec la nature même du Big Data.

Big data et protection de la vie privée : une réconciliation possible

Malgré les zones grises évoquées précédemment, le Big Data fait son apparition dans la législation européenne qui devrait être un « facilitateur des services de Big data en Europe »⁶. Cette tendance se concrétise dans certains instruments juridiques tels que la Directive concernant la réutilisation des informations du secteur public qui promeut la disponibilité des informations produites, collectées et commandées par le secteur public au sein de l'UE, également appelée politique d'ouverture des données.

De plus, la récente réforme européenne du cadre législatif en matière de protection des données pose que ce droit fondamental ne devrait pas empêcher l'innovation que représente le Big Data mais devrait, au contraire, être considéré comme un facilitateur, un avantage concurrentiel.

La question qui se pose alors est de savoir comment initier des projets dans le domaine du Big Data tout en respectant la législation actuelle pour le moins inadaptée.

Le Big Data répond au besoin des acteurs du marché de réduire l'imprévu en créant des informations pertinentes à partir de données constamment produites.

Pour l'heure, et compte tenu du cadre juridique actuel, il est possible d'implémenter les mesures suivantes :

- Procéder à une analyse d'impact sur la protection des données (PIA, Privacy Impact Assessment) au moment de la définition du traitement des données, en mettant l'accent sur les finalités spécifiques du recours au Big Data
- Implémenter les principes de protection des données tout au long du cycle de vie des données et donc aux moments de la collecte et de l'accès aux données, de la conservation et de l'agrégation des données, ainsi qu'au moment de leur analyse et de leur présentation
- Faire du principe de transparence, par le biais d'une information claire au citoyen ou au consommateur sur le recours au Big Data, un principe clé assurant ainsi la qualité et l'exactitude des données traitées. Il s'agira dès lors d'indiquer comment les données à caractère personnel seront traitées et sous la responsabilité de qui. Il pourra d'ailleurs s'avérer utile de solliciter, dès le début du projet, l'avis de l'autorité nationale chargée de la protection des données
- Organiser la sécurité des données conformément à l'analyse d'impact sur la protection des données. A ce sujet, il faudra définir une classification des données à caractère personnel mais aussi assurer l'audit et la traçabilité de ces données dès leur origine
- Garantir l'intégrité des données et limiter l'accès et la modification de celles-ci par des moyens tels que le contrôle d'accès granulaire et le chiffrement des données au repos et en transit dans l'infrastructure

⁶ The EU Data Protection Reform and Big Data – Factsheet – Avril 2015 http://ec.europa.eu/justice/data-protection/files/data-protection-big-data_factsheet_web_en.pdf

