
Data driven governance : hypothèse d'application à l'état des modèles big data déployés par le secteur financier

Auteur : Diez, Sylvain

Promoteur(s) : Blavier, André

Faculté : HEC-Ecole de gestion de l'ULg

Diplôme : Master en sciences de gestion, à finalité spécialisée en Financial Analysis and Audit

Année académique : 2015-2016

URI/URL : <http://hdl.handle.net/2268.2/1250>

Avertissement à l'attention des usagers :

Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.

Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.

DATA DRIVEN GOVERNANCE :
HYPOTHESE D'APPLICATION A L'ETAT
DES MODELES BIG DATA DEPLOYES PAR
LE SECTEUR FINANCIER

Jury :
Promoteur :
André BLAVIER
Lecteurs :
Ashwin ITTOO
Wilfried NIESSEN

Mémoire présenté par
Sylvain DIEZ
En vue de l'obtention du diplôme de
Master en Sciences de Gestion
à finalité spécialisée
Audit et Analyse Financière

Année académique 2015/2016

Je tiens à remercier, pour leur contribution à la réalisation de mon mémoire :

Monsieur André Blavier, mon promoteur, pour son implication, son accompagnement et ses précieux conseils,

Messieurs Ashwin Ittoo et Wilfried Niessen, mes lecteurs, pour le temps consacré,

Xavier et Christiane, mes parents, Damien, mon frère, et Sarah, ma fiancée, pour leur soutien et leurs nombreux encouragements,

Xavier, mon père, Sylvain Coudray et Patricia Blanchard pour leur relecture et leurs commentaires constructifs,

Micaël Dekleyn, pour avoir contribué aux annexes,

Toutes les autres personnes qui, de près ou de loin, ont rendu possible cette réalisation.

Sommaire

Introduction	1
PARTIE 1 : Qu'est-ce que le Big Data ?	5
PARTIE 2 : Le Big Data dans le secteur financier	25
PARTIE 3 : Une Data Driven Governance	49
Conclusion.....	69
Annexes.....	75
Références bibliographiques	89

Introduction

En 1984, Internet reliait 1000 universités et laboratoires d'entreprises. En 1998, nous sommes passés à 50 millions d'utilisateurs. En 2015, Internet permet à 3,2 milliards de personnes d'être connectées, soit 44 % de la population mondiale (Institut des Actuairens en Belgique, 2015).

En 2000, 800 000 pétaoctets (1 pétaoctet étant égal à 10^{15} octets) de données étaient stockés dans le monde. On s'attend à ce que cette quantité atteigne 35 zettaoctets (1 zettaoctet étant égal à 10^{21} octets) d'ici 2020 (Zikopoulos, Eaton, deRoos, Deutsch, & Lapis, 2012).

En 2011, plus de 5 milliards de téléphones mobiles étaient utilisés dans le monde, 30 milliards de publications de contenus étaient partagées sur Facebook chaque mois, et 235 téraoctets (1 téraoctet étant égal à 10^{12} octets) de données ont été collectés par la bibliothèque nationale américaine (Bibliothèque du Congrès). Cette même année, aux Etats-Unis, sur les 17 secteurs que compte l'économie, 15 contenaient plus de données stockées par entreprise que la Bibliothèque du Congrès (McKinsey Global Institute [MGI], 2011).

En 2013, on dénombrait chaque minute plus de 98 000 tweets, 695 000 mises à jour de statut sur Facebook, 11 millions de messages instantanés, 698 445 recherches sur Google, plus de 168 millions d'e-mails envoyés, 1820 téraoctets de données créées et 217 nouveaux utilisateurs du Web mobile (Hewlett-Packard Development Company [HP], 2013). Par jour, cela représentait plus de 7 téraoctets de données générés par Twitter et 10 téraoctets par Facebook. Certaines autres entreprises généraient des téraoctets de données chaque heure (Zikopoulos et al., 2012).

Aujourd'hui, nous générons chaque jour environ 2,5 trillions d'octets de données, soit 2.500.000.000.000.000 octets d'informations par jour ! Par ailleurs, 90 % de l'ensemble des données du monde ont été créées ces deux dernières années (Brasseur, 2013, p. 30 cité par Hamel & Marguerit, 2013 ; Desouza, 2014 ; PricewaterhouseCoopers [PwC], 2013b).

Ces quelques chiffres sont impressionnants et nous montrent à quel point notre société contemporaine est envahie par les données. On entend parfois dire que ces données sont l'« or noir » de XXI^e siècle et que ceux qui les contrôlent ont le pouvoir (GovLab Research, 2013). Ces données concernent tous les domaines de notre vie quotidienne : ce sont des informations financières, environnementales, médicales, comportementales, démographiques,

psychographiques sur les consommateurs, de surveillance, etc. Elles proviennent, entre autres, de capteurs, d'appareils mobiles, d'appareils techniques, des réseaux sociaux et des blogs (PwC, 2013b ; Zikopoulos et al., 2012).

Dans le début des années 2000, un terme est né pour nommer cette quantité constamment grandissante de données : le « Big Data » (Desouza, 2014). Aujourd'hui, le Big Data est au centre des recherches de nombreux scientifiques et directeurs d'entreprise qui souhaitent en retirer de la valeur pour leur organisation. D'ailleurs, selon une étude de PwC (2013a), 62 % des directeurs généraux et technologiques interrogés pensent que le Big Data a un potentiel significatif pour créer un avantage concurrentiel dans le chef de leur organisation.

En parlant de création de valeur et d'avantage concurrentiel pour les organisations, le McKinsey Global Institute, dans son rapport sur le Big Data publié en 2011, donne quelques chiffres sur la valeur potentielle de l'utilisation Big Data, tant au niveau du gouvernement qu'au niveau de l'entreprise et du consommateur final :

- une valeur potentielle annuelle de 300 milliards de dollars pour le service de la santé américain, notamment par la réduction des dépenses nationales de soins de santé ;
- une valeur annuelle potentielle de 250 milliards d'euros pour l'administration publique européenne, dont une centaine seulement par l'amélioration de l'efficacité opérationnelle en interne ;
- un surplus annuel potentiel de 600 milliards de dollars pour le consommateur en utilisant les données de géolocalisation personnelles au niveau mondial ;
- ou encore une hausse potentielle de 60 % de la marge opérationnelle des détaillants.

Ces chiffres faramineux exposés par McKinsey, un des plus gros acteurs mondiaux de la consultance technologique, montrent pourquoi le Big Data suscite tellement d'intérêt auprès des multinationales, mais aussi auprès de toutes les organisations qui sont amenées à posséder des données en masse. Attention, « *tout l'intérêt des "données massives" ne réside pas dans leur quantité, le volume à partir duquel il est possible de parler de Big Data ne faisant d'ailleurs pas l'unanimité, mais plutôt dans le traitement qu'il est possible d'en faire* » (Hamel & Marguerit, 2013, p. 438), à l'aide des « Big Data analytics » (Chen, Chiang, & Storey, 2012 ; Schenk & Berman, 2015).

Gilles Babinet, dans son ouvrage « *L'ère numérique, un nouvel âge pour l'humanité* », expose que « *la révolution numérique pourrait compter parmi les trois grandes inflexions de l'humanité, au même titre que l'invention de l'écriture par la civilisation sumérienne – qui*

entraîna le développement des villes, routes et remises à grains – et l'invention de l'imprimerie par Gutenberg qui rendit possible l'émergence des Lumières » (Babinet, 2014, p. 1).

De son côté, Satya Nadella, le CEO (Chief Executive Officer) de Microsoft, dans une keynote intitulée « *Les données sont la nouvelle électricité* » (en anglais, « *Data is the new electricity* »), parle de la transformation digitale de la société, dont le Big Data fait partie, comme la quatrième révolution industrielle, après celle de la vapeur, de l'électricité et de l'informatique. Selon lui, cette transformation digitale va bouleverser la manière de faire du business, et est rendue possible par trois caractéristiques : les connexions (entre humains, entre objets et entre humains et objets), les données qui sont générées par ces connexions et les technologies informatiques qui permettent de traiter ces données (Microsoft, 2016).

Par ailleurs, deux des plus grands experts du monde en Big Data reconnus, Viktor Mayer-Schönberger (Professeur de la Gouvernance et de la Régulation d'Internet à l'Université d'Oxford) et Kenneth Cukier (« Editeur data » de The Economist et penseur éminent des développements dans le Big Data), ont publié, en 2013, un ouvrage référence dans le domaine en question : « *Big Data : A revolution that will transform how we live, work and think* ». De nos jours, plus personne ne doute que le Big Data est synonyme d'une « *révolution qui va transformer la manière dont nous vivons, travaillons et pensons* » (Mayer-Schönberger, & Cukier, 2013 ; voir aussi Boyer, 2015).

Enfin, pour confirmer à quel point chaque individu est concerné par la révolution du Big Data, Zikopoulos et al., de l'International Business Machines Corporation (IBM), ont intitulé le premier chapitre de leur ouvrage (2012, p. 3, traduction personnelle de l'anglais) : « *Qu'est-ce que le Big Data ? Indice : vous en faites partie chaque jour* ».

C'est dans ce contexte de révolution de notre monde par le Big Data que nous avons souhaité réaliser un exposé qualitatif sur le sujet et étudier la piste d'une « *Data Driven Governance* ». Nous entendons par là un Etat qui utilise le Big Data pour l'amélioration de sa gouvernance et de ses services publics, à l'image de la banque et de la compagnie d'assurances pour générer du profit.

Pourquoi les secteurs de la banque et de l'assurance ? D'une part, le phénomène Big Data concerne des données en masse. Il est donc logique que les secteurs de l'économie qui détiennent le plus de données soient les plus impactés par la révolution du Big Data. Les secteurs de la banque et de l'assurance sont de tels secteurs. D'autre part, la banque et la

compagnie d'assurances sont deux acteurs financiers dont le consommateur ne peut se passer, ayant besoin de pouvoir déposer son argent en lieu sûr, emprunter de l'argent et être assuré pour les divers risques quotidiens.

Pourquoi l'Etat ? La relation entre la banque ou la compagnie d'assurances et son client peut être apparentée à la relation entre l'Etat et son citoyen. En effet, de la même manière que le consommateur ne peut subsister sans un banquier et un assureur, le citoyen ne peut se passer de son Etat. Ensuite, tout comme le client est amené à confier à son banquier ou à son assureur des informations sur ses revenus et son patrimoine en échange d'un service financier, le citoyen est aussi amené à communiquer à l'Etat un certain nombre d'informations privées afin d'être considéré comme « citoyen » et de pouvoir bénéficier des services publics. Dès lors, l'Etat dispose, comme la banque et la compagnie d'assurances, de volumes importants de données confidentielles au sujet des citoyens.

Les données en masse ouvrent la porte à de nombreuses opportunités de création de valeur pour la banque et la compagnie d'assurances, et elles font de même pour l'Etat. Par ailleurs, de la même manière que le Big Data implique des challenges et des risques pour le banquier et l'assureur, ces challenges et risques sont aussi valables pour l'Etat.

Ainsi, puisque les bienfaits et les risques de l'utilisation du Big Data dans les secteurs de la banque et de l'assurance sont transposables au domaine de la gouvernance publique, nous désirons étudier la possibilité d'une Data Driven Governance en Belgique, à l'image de l'impact du Big Data dans le secteur financier.

Notre exposé est composé de trois parties. Dans la première partie, nous tenterons de comprendre ce qu'est le Big Data et explorerons les principaux changements que la révolution du Big Data engendre. Dans la deuxième partie, nous étudierons l'impact du Big Data dans les secteurs de la banque et de l'assurance et nous en synthétiserons les principales tendances. Dans la troisième partie, nous transposerons ces tendances d'impact du Big Data dans le secteur financier au domaine de la gouvernance et nous étudierons de quelle manière elles peuvent s'appliquer au secteur public. Enfin, nous évaluerons la possibilité d'une Data Driven Governance en Belgique sur base de ce que nous aurons développé et en faisant preuve d'esprit critique.

PARTIE 1 : Qu'est-ce que le Big Data ?

Dans cette première partie, nous tentons de comprendre ce qu'est le Big Data et pourquoi il est en voie de changer radicalement notre mode de vie et notre manière de faire du business.

D'abord, nous parcourrons les différentes définitions que les experts associent à ce concept. Ensuite, nous décomposerons le Big Data et verrons ce qui le constitue et quelles en sont les caractéristiques principales. Par après, nous exposerons les différentes sources du phénomène ainsi que les outils et les techniques qui permettent de manipuler ces données en masse. Nous évoquerons par ailleurs l'impact que le Big Data a sur plusieurs secteurs de l'économie. Nous parlerons aussi des risques engendrés par la révolution des données pour notre civilisation. Enfin, suite à ces premiers éléments d'analyse, nous proposerons notre propre définition du Big Data.

1. Une définition du Big Data

Tout le monde est assez d'accord pour dire qu'il n'existe pas de définition officielle et unanimement acceptée par les experts dans le domaine du Big Data (Mayer-Schönberger, & Cukier, 2013). Une des raisons de ce problème est sans doute qu'à l'image de l'illustration présentée en annexe 1, tirée d'un « information paper » de l'Institut des Actuaire en Belgique (2015), les experts ne connaissent qu'une minime partie de ce qu'est vraiment le Big Data. Ils en sont donc encore au stade d'essayer de comprendre et de définir ce phénomène qui dépasse même les plus grands acteurs de la technologie.

Selon Press (2014), le terme « Big Data » est utilisé pour la première fois par des scientifiques de la NASA en décrivant le problème auquel ils font face au niveau de la visualisation d'un très grand nombre de données sur leurs systèmes informatiques. Ils en concluent qu'ils ont un « problème de Big Data » quand les supports informatiques habituels ne suffisent plus pour contenir la quantité de données présente. En 2008, dans un article intitulé « *Big-Data Computing: Creating revolutionary breakthroughs in commerce, science, and society* », les scientifiques universitaires Bryant, Katz et Lazowska évoquent le phénomène du « *Big-Data computing* », sans pour autant le définir, mais popularisent le terme « Big Data » (Press, 2014).

En 2011, le McKinsey Global Institute propose la définition suivante : « *Le Big Data se réfère à un ensemble de données dont la taille va au-delà de la capacité des logiciels de bases de*

données classiques à capturer, stocker, gérer et analyser » (MGI, 2011, p. 11, traduction personnelle de l'anglais). Selon les chercheurs de McKinsey, cette définition est intentionnellement subjective, dans le sens où tout est relatif. Ils définissent donc le Big Data comme ce qui sera toujours au-delà de la puissance de nos technologies.

La même année, Gantz et Reinsel de l'International Data Corporation (IDC) proposent la définition suivante : « *Les technologies Big Data décrivent une nouvelle génération de technologies et d'architectures, conçues pour extraire économiquement de la valeur à partir de très grands volumes d'une large variété de données, en permettant une capture, une découverte et/ou une analyse à très grande vitesse* » (2011, p. 6, traduction personnelle de l'anglais).

En 2013, le terme « Big Data » entre officiellement dans l'Oxford English Dictionary. Il est défini comme « *des données d'une très grande taille, dans la mesure où leur manipulation et leur gestion entraînent d'importants challenges logistiques* » (Oxford English Dictionary, 2016, traduction personnelle de l'anglais).

Une autre source dont on ne peut se passer lorsqu'on parle de Big Data est l'ouvrage de Mayer-Schönberger et Cukier, dans lequel les auteurs exposent le problème de la manière suivante : « *Le Big Data se réfère aux choses que nous pouvons faire à une large échelle et qui ne peuvent pas être faites à une échelle moindre, pour extraire de nouvelles connaissances ou créer de nouvelles formes de valeur, de façon à transformer les marchés, les organisations, les relations entre les citoyens et les gouvernements, et encore d'autres* » (Mayer-Schönberger & Cukier, 2013, p. 8, traduction personnelle de l'anglais). Les auteurs ajoutent que, selon eux, le Big Data est amené à modifier notre manière de vivre et d'interagir avec le monde.

Gartner, un des géants du conseil et de la recherche dans le domaine des techniques avancées, définit le Big Data comme « *des actifs d'information de grand volume, de haute vitesse et de grande variété qui requièrent des outils innovants et rentables de traitement de données et qui permettent une amélioration des connaissances, de la prise de décision et de l'automatisation des processus* » (Gartner, 2016a). Nous remarquons une grande ressemblance entre cette définition de Gartner et celle donnée plus haut de Gantz et Reinsel de IDC : la présence des termes « volume », « variété » et « vitesse ». Nous verrons dans la section suivante pourquoi ces termes sont si importants lorsqu'on parle de Big Data.

Enfin, la TechAmerica Foundation nous propose la définition suivante : « *Big Data est un terme qui décrit des grands volumes de données à haute vitesse, complexes et variables qui requièrent des techniques et des technologies avancées pour permettre la capture, le stockage, la distribution, la gestion et l'analyse de l'information* » (TechAmerica Foundation's Federal Big Data Commission, s.d., p. 10, traduction personnelle de l'anglais).

Nous remarquons que certains auteurs définissent le Big Data en expliquant ce qu'il est par nature, alors que d'autres le définissent en développant ce qu'il permet de faire. Malgré les différences entre les définitions citées, nous pouvons constater un point commun entre toutes : le Big Data se réfère à une grande quantité de données qui est difficile à maîtriser mais qui a un potentiel énorme de création de valeur pour améliorer la vie de l'Homme et pour augmenter la performance de son entreprise. Le Big Data est donc un immense volume de données, mais il a aussi d'autres caractéristiques très spécifiques, que nous exposons dans la section suivante.

2. Les « V » du Big Data

A côté des définitions qui ont été citées dans la section précédente, la plupart des auteurs et acteurs du domaine de la technologie préfèrent définir le Big Data en exposant ses différentes caractéristiques. Ceux qui ont déjà entendu parler du Big Data ont plus que probablement entendu parler des fameux « V ». A vrai dire, tout le monde est assez d'accord pour dire que le Big Data se caractérise par des termes qui commencent par la lettre « V ». Cependant, tous ne sont pas d'accord sur le nombre de « V » qu'il faut considérer, si ce n'est qu'il y en a au moins trois.

2.1. Les trois « V » de base

Les trois « V » qui reviennent systématiquement sont Volume, Vitesse et Variété. C'est par exemple ce qui ressort de la définition des chercheurs d'IDC, citée plus haut (Gantz & Reinsel, 2011). Zikopoulos et al. (2012), IBM (2012) ainsi que Schenk et Berman (2015), se contentent aussi de ces trois « V » pour définir le Big Data. A ce propos, il est intéressant de remarquer qu'en 2001, alors que le terme « Big Data » était encore peu connu et utilisé, Laney, de META Group, a publié un article intitulé « *3D Data management : Controlling Data volume, velocity and variety* ». Nous constatons qu'au début des années 2000, le Big Data, sous la forme de ses trois principales caractéristiques, était déjà dans l'esprit des experts de la technologie.

Nous présentons donc ici les trois « V » de base, ou les trois caractéristiques de base du Big Data.

2.1.1. Le Volume du Big Data

Comme présenté dans l'introduction, c'est avant tout l'immense quantité de données qui existent déjà et qui sont encore générées chaque seconde qui caractérise le Big Data. Alors que l'on parlait, il y a quelques années, de gigaoctets (10^9 octets), on parle aujourd'hui plutôt d'exaoctets (10^{18} octets) et de zettaoctets (10^{21} octets), 1,8 zettaoctet représentant toutes les informations enregistrées en 2011. Etant donné que le volume mondial de données devrait atteindre 35 zettaoctets en 2020, il serait nécessaire de multiplier le nombre de serveurs physiques par dix d'ici dix ans (Hamel & Marguerit, 2013).

2.1.2. La Variété du Big Data

Pour que l'on puisse parler de Big Data, il faut que les données soient variées. Des données sont variées à partir du moment où elles proviennent de multiples sources, telles que des capteurs, des objets intelligents, Internet et les technologies de collaboration sociale. Les données n'incluent plus seulement des données relationnelles traditionnelles, mais aussi des données brutes, semi-structurées ou non structurées provenant de pages web, de fichiers en ligne, d'index de recherche, de forums de médias sociaux, d'e-mails, de tweets, de statuts Facebook, de blogs, de vidéos et de capteurs entre autres (PwC, 2013b ; Zikopoulos et al., 2012). Puisque les données ont des origines différentes, elles ont aussi des formats différents (texte, image, audio, vidéo, etc.).

La TechAmerica Foundation's Federal Big Data Commission (s.d.) prétend que 85 % des données sont non structurées, 90 % selon Atos (2014). Les organisations font donc face à un véritable défi de tri des données afin de pouvoir en retirer des informations utiles, susceptibles de produire de la valeur. Desouza (2014) estime que cette dimension du Big Data est le défi le plus important pour les entreprises. En effet, celles-ci ont développé des systèmes de gestion de données dans le but de manipuler des informations de catégories spécifiques. Par exemple, un logiciel du département des ressources humaines est fait pour gérer des données sur le personnel, alors qu'un logiciel de GRC (Gestion de la Relation Client) détient des données sur les clients.

Aujourd'hui, les organisations ne peuvent plus se contenter de gérer seulement l'une ou l'autre catégorie de données, mais sont contraintes de devoir jongler avec des bases de données hétérogènes, dans lesquelles les données sont différentes autant par leur origine que

par leur nature. Par exemple, le défi d'une entreprise sera de devoir mélanger des données provenant de médias sociaux avec les données client d'un programme de GRC. Ensuite, elle devra analyser le mélange et en tirer des conclusions qui lui permettront de passer à un marketing personnalisé.

2.1.3. La Vitesse du Big Data

Par vitesse, on entend à la fois la vitesse à laquelle les données sont produites et transformées et la vitesse avec laquelle ces données doivent être collectées, comprises et traitées par les organisations (TechAmerica Foundation's Federal Big Data Commission, s.d.). On peut donc parler de Big Data à partir du moment où les données sont créées en temps réel ou presque réel (Kitchin, 2014). En ce qui concerne la vitesse de traitement des données, l'augmentation du nombre de capteurs RFID (Radio Fréquence Identification) et autres outils de transfert d'informations ont mené à un flux continu de données dont le rythme est devenu impossible à gérer pour les systèmes de gestion de données traditionnels. C'est ce que Zikopoulos et al. appellent les « *données en mouvement* » (2012, p. 8, traduction personnelle de l'anglais).

De plus, dans un monde où une entreprise peut devancer ses concurrents en identifiant une tendance, un problème ou une opportunité seulement quelques secondes avant les autres, nous réalisons à quel point la vitesse d'analyse des données devient un enjeu important. Enfin, il faut aussi noter que de plus en plus de données produites aujourd'hui sont disponibles pendant un laps de temps très court. Il est donc primordial pour les organisations de pouvoir analyser les données presque en temps réel afin de pouvoir en tirer des informations valorisables pour leur performance économique (Zikopoulos et al., 2012).

Pour illustrer notre propos, en annexe 2 se trouve une infographie qui montre comment ces trois dimensions du Big Data se concrétisent dans le monde d'Internet :

- Le Volume est représenté par les chiffres, soit des quantités ahurissantes de données.
- La Variété est symbolisée par les 13 sources de données qui sont listées dans l'infographie et qui génèrent des données sous des formats très variés (textes, photos, vidéos, requêtes, connexions, actions, etc.).
- La Vitesse est représentée par le peu de temps qu'il faut pour générer toutes ces données variées sur Internet, soit 60 secondes.

2.2. Toujours plus de « V »

Au fil du temps, le Big Data s'est montré toujours plus complexe et les experts mondiaux de la technologie ont relevé de nouvelles caractéristiques du Big Data, commençant aussi par la lettre « V ».

2.2.1. La Véracité du Big Data

IBM (2012), PwC (2013a) et InnovaTech (s.d.) définissent le Big Data en citant les trois premières dimensions que nous venons d'exposer et en y ajoutant un quatrième « V » : la Véracité des données. C'est aussi le cas des data scientists (littéralement, « scientifiques des données ») d'IBM, comme cela est illustré dans l'annexe 3. Les scientifiques exposent notamment qu'un business leader sur trois ne fait pas confiance à l'information qu'il utilise pour prendre ses décisions, ou encore que la faible qualité des données coûte à l'économie américaine 3 100 milliards de dollars par an. Ces chiffres montrent le challenge de la qualité des données que les organisations traitent.

La TechAmerica Foundation's Federal Big Data Commission (s.d.), qui utilise aussi quatre « V » pour définir le Big Data, explique que la qualité de l'information peut être biaisée en raison de l'incohérence des données, de leur caractère incomplet, à cause d'ambiguïtés, de latences, de tromperies ou encore des approximations d'un modèle. C'est pourquoi la véracité des données est un enjeu et les décisions qui sont prises sur base des données doivent être solidement justifiées. Enfin, avec la prolifération des réseaux sociaux comme Facebook et Twitter, les informations provenant de ces nouvelles sources de Big Data doivent être examinées minutieusement quant à leur véracité avant de s'en inspirer pour prendre des décisions (Desouza, 2014 ; Hamel & Marguerit, 2013).

2.2.2. La Valeur du Big Data

Boyer (2015), Hamel et Marguerit (2013) ainsi que Vesset et al. (2012) ajoutent un cinquième « V » pour décrire la dimension Valeur du Big Data. Il est donc question de la valeur que les données peuvent générer dans le chef d'une organisation, de « *l'intelligence* » que les données peuvent produire (Groupement français de l'industrie de l'information [GFII], 2012, p. 12 cité par Hamel & Marguerit, 2013). Pour les organisations, cette valeur se manifeste autant par l'amélioration des services existants que par la création de nouveaux services dans le but de trouver sans cesse de nouvelles sources de revenus (Hamel & Marguerit, 2013).

Le McKinsey Global Institute (2011) confirme l'importance de la dimension Valeur lorsqu'on parle du Big Data. Comme énoncé dans l'introduction, l'entreprise donne des chiffres clés qui

quantifient le potentiel de création de valeur du Big Data pour les organisations de tous les secteurs de l'économie. La recherche de McKinsey montre par ailleurs que « *les données peuvent créer une valeur significative pour l'économie mondiale, améliorant la productivité et la compétitivité des entreprises et du secteur public, et créant un surplus économique substantiel pour les consommateurs* » (MGI, 2011, p. 1-2, traduction personnelle de l'anglais).

2.2.3. La Visualisation du Big Data

L'Institut des Actuairens en Belgique (2015) propose une définition du Big Data contenant six « V ». En plus des cinq dimensions déjà développées, celle de la Visualisation du Big Data est ajoutée. Les connaissances résultant de l'analyse des données doivent être partagées de manière efficiente et compréhensible, notamment aux actionnaires de l'entreprise. Au vu de la quantité absolument inimaginable de données qui existe déjà et qui est encore générée chaque jour, il est évident que pouvoir présenter l'information de manière claire et lisible constitue un autre défi du Big Data.

2.2.4. La Viscosité, Variabilité et Volatilité du Big Data

De son côté, Desouza (2014) choisit de définir le Big Data à l'aide de sept dimensions. En plus des dimensions de Volume, Variété, Vitesse et Véracité déjà abordées, il propose trois autres « V » :

- La Viscosité, qui mesure la résistance dégagée par le Big Data lorsqu'il s'agit de naviguer à travers les données. Cette résistance provient du fait que les données ont pour origine différentes sources, qu'il existe différents débits de flux de données ou encore que les processus nécessaires à la transformation des données en informations utiles sont très complexes (Wang, 2012).
- La Variabilité, qui tient compte des changements de taux de débit de données.
- La Volatilité, qui indique le laps de temps au cours duquel les données sont valides et pour combien de temps elles doivent être stockées.

2.2.5. La Vulnérabilité du Big Data

Enfin, notons que HP (2013), un autre géant de la technologie, parle du Big Data en développant 4 « V ». Les trois premiers sont les dimensions de base : le Volume, la Variété et la Vitesse. Le quatrième « V » est une dimension qui n'est pas apparue dans les ouvrages des autres auteurs cités : la Vulnérabilité. Les scientifiques de HP soulèvent ici un aspect important de la problématique des données. En effet, parmi toutes les données qui sont

générées chaque seconde, certaines peuvent être rendues publiques, mais d'autres revêtent un caractère confidentiel qui est crucial pour l'organisation qui les détient.

HP pose alors la question suivante : « *Comment sécurise-t-on un volume toujours plus grand d'informations sensibles ?* » (2013, p. 5, traduction personnelle de l'anglais). Effectivement, il est essentiel que ces données, collectées et analysées afin de détecter des menaces potentielles, soient particulièrement bien protégées. De plus, à partir du moment où les plateformes qui contiennent ces données sensibles se déplacent dans des environnements technologiques hybrides, en gérer la sécurité et la disponibilité devient un autre challenge du Big Data, qui exige des diagnostics et une surveillance permanente (HP, 2013).

Nous avons à présent un bel aperçu de toutes les dimensions qui caractérisent le Big Data. Comme illustré en annexe 4, certaines de ces caractéristiques définissent ce qu'est le Big Data en nature (Volume, Variété, Vitesse) et ce qu'il peut produire (Valeur), alors que d'autres en traduisent les défis (Véracité, Visualisation, Viscosité, Variabilité, Volatilité, Vulnérabilité).

Avant de continuer, une remarque importante doit être faite : il faut bien garder à l'esprit que « *le Big Data en lui-même ne crée pas de valeur* » (Turner, Schroeck, & Shockley, 2013, p. 6, traduction personnelle de l'anglais ; voir aussi GovLab Research, 2013 ; Microsoft, 2016), comme nous l'expliquons dans la section suivante.

3. Les Big Data analytics

En octobre 2015, Gartner (2015a) publiait un article intitulé « *Gartner dit que ce n'est pas juste une question de Big Data ; il s'agit de savoir ce qu'on en fait : Bienvenue dans l'économie algorithmique* » (traduction personnelle de l'anglais). En effet, au-delà du fait d'avoir accès à des grands volumes de données de différentes sortes, il est primordial, pour les organisations qui veulent créer de la valeur à partir du Big Data, de posséder des capacités puissantes d'analyse de données. Ces capacités englobent des outils et des techniques, mais aussi et surtout les compétences pour les utiliser (Turner et al., 2013 ; voir aussi Microsoft, 2016). C'est ici qu'apparaît un terme qui va de pair avec la révolution du Big Data : les « Big Data analytics ».

Chen et al. (2012) définissent les « Big Data analytics » comme les techniques analytiques qui sont utilisées sur de grands et complexes volumes de données et qui requièrent des technologies avancées de stockage, de gestion, d'analyse et de visualisation des données. Ces capacités analytiques peuvent prendre plusieurs formes :

- rapports sur base d'analyses de requêtes,
- data mining (littéralement, « exploration de données »),
- visualisations de données,
- modélisations prédictives,
- optimisations,
- simulations,
- analyses de textes à l'état naturel (comme des transcriptions de conversations de call centers),
- analyses de données géospatiales,
- analyses de données en streaming (littéralement, en « flux »),
- analyses de vidéos et de voix.

(Turner et al., 2013)

Précisons que la matière première des Big Data analytics est l'algorithme (Schenk & Berman, 2015). Selon Cardon, « *ce terme d'informatique a une signification bien plus large qu'on ne le croit. Comme la recette de cuisine, un algorithme est une série d'instructions permettant d'obtenir un résultat. À très grande vitesse, il opère un ensemble de calculs à partir de gigantesques masses de données (les « big data »). Il hiérarchise l'information, devine ce qui nous intéresse, sélectionne les biens que nous préférons et s'efforce de nous suppléer dans de nombreuses tâches* » (2015, p.7).

D'un point de vue plus scientifique, un algorithme est une « *suite finie, séquentielle de règles que l'on applique à un nombre fini de données, permettant de résoudre des classes de problèmes semblables. [C'est un] ensemble de règles opératoires propres à un calcul ou à un traitement informatique* » (Petit Robert, 2010, p. 65). Les algorithmes peuvent être très simples comme très complexes. C'est eux qui permettent, à partir d'une base de données, d'établir des corrélations et de produire des résultats face à des questions ou des problèmes posés. Par conséquent, le but des Big Data analytics, de cet « art » ou de cette science qui consiste à analyser des données en masse, est de pouvoir en tirer des informations et des conclusions cohérentes et pertinentes par rapport à un problème ou une question (Schenk & Berman, 2015).

A présent, nous savons que le Big Data en lui-même ne crée pas de valeur. En effet, il s'agit avant tout de savoir ce que l'on fait de ces données en masse. Il est clair que, sans des technologies et des techniques avancées de manipulation et d'analyse de données, il n'est pas

possible de produire de la valeur à partir du Big Data. Voyons maintenant les différentes sources du Big Data qui existent et qui permettent que des immenses volumes de données puissent être générés chaque jour.

4. Les sources du Big Data

Le Big Data est composé d'informations structurées et non structurées. Les premières proviennent des bases de données traditionnelles, et constituent environ 10 % du Big Data. Les autres 90 % sont des informations non structurées et proviennent :

- d'une part, de sources « humaines », via Internet (e-mails, photos, vidéos, tweets, publications Facebook, « clics » sur les sites web, requêtes sur des moteurs de recherche, achats et transactions en ligne) et via les réseaux de télécommunications (appels téléphoniques et conversations de call centers) ;
- d'autre part, de « machines » qui génèrent des données automatiquement, comme les téléphones mobiles (données de géolocalisation et données historiques d'utilisation), les capteurs (utilisés par exemple pour collecter des informations climatiques ou sur des voitures) et les objets connectés en général (comme des montres ou des habits « intelligents », connectés à l'aide de codes barres ou de puces RFID).

(Atos, 2014 ; Chen et al., 2012 ; IBM, s.d.a)

Parmi toutes ces sources du Big Data, nous pouvons distinguer des données qui sont générées à cause d'une action humaine volontaire, et d'autres qui sont produites automatiquement par des machines, ou qui découlent d'une action humaine sans but de fournir des données.

4.1. Trois catégories de sources

Selon Kitchin (2014), les sources du Big Data peuvent généralement être réparties en trois catégories, décrites ci-dessous.

4.1.1. Les sources dirigées

Elles génèrent des données grâce notamment à des technologies de surveillance traditionnelles, qui sont dirigées vers une personne ou un lieu par un opérateur humain. C'est par exemple le cas des systèmes de contrôle de passeports. Ils collectent et vérifient des informations sur des passagers qui sont stockées dans des bases de données en temps réel. A partir ces dernières, des nouvelles données peuvent être générées comme des photos d'identité, des empreintes digitales ou encore des empreintes rétiniennes. C'est aussi le cas

des capteurs environnementaux électromagnétiques ou thermiques qui permettent des cartographies 2D et 3D mobiles et en temps réel (Kitchin, 2014).

4.1.2. Les sources automatisées

Les données sont générées par des appareils ou systèmes avec des fonctions automatiques qui ne nécessitent pas d'intervention humaine. Ces sources incluent :

- les téléphones mobiles, qui enregistrent et communiquent automatiquement les informations historiques de leur utilisation ;
- les appareils mobiles en général, qui peuvent transmettre des données de géolocalisation ;
- les interactions et transactions sur les réseaux digitaux, qui non seulement transfèrent des informations, mais génèrent aussi des données à propos des interactions et des transactions elles-mêmes (comme les virements bancaires ou les e-mails) ;
- les données de flux de « clics », qui enregistrent comment les gens naviguent à travers un site web ou une application mobile (données historiques) ;
- les données générées par les objets connectés et les capteurs ;
- les données audio provenant de call centers ou d'écoutes téléphoniques dans le cas de services de renseignements gouvernementaux.

(Kitchin, 2014 ; Turner et al., 2013)

Beaucoup de ces données automatisées ont pour origine des objets physiques. Ce sont des appareils mobiles et des capteurs qui sont capables de « sentir » des caractéristiques de l'environnement, telles que la géolocalisation, la température ou le mouvement. Dès lors que ces objets communiquent des informations à d'autres objets ou à une base de données, on parle « d'objets connectés ». A l'échelle mondiale, le réseau formé par ces objets connectés est appelé l'« Internet des Objets » (Gartner, 2016b).

L'Internet des Objets est une tendance complémentaire à la tendance Big Data et va donc aussi révolutionner le monde actuel. Aujourd'hui apparaissent déjà des « montres connectées », des « voitures connectées », des « vêtements connectés » et peut-être qu'un jour notre corps humain sera aussi connecté. Ainsi, l'Internet des Objets constitue une des sources principales du Big Data.

4.1.3. Les sources volontaires

Les données sont littéralement « données » par les utilisateurs. Celles-ci proviennent principalement des interactions à travers les réseaux sociaux, comme les publications de contenus et les téléchargements (uploads) de fichiers (images et vidéos). Par ailleurs, le crowdsourcing (littéralement, « contribution par la foule ») sur des plateformes « open source », où les utilisateurs contribuent eux-mêmes au contenu du web (comme Wikipédia), constitue aussi une source volontaire du Big Data (Institut des Actuaire en Belgique, 2015 ; Kitchin, 2014 ; Turner et al., 2013).

Cette multitude de sources d'informations différentes traduit la dimension Variété. Celle-ci engendre une des forces majeures du Big Data : rassembler ces données provenant d'origines différentes et les croiser afin d'en faire ressortir des corrélations que l'on ne pourrait découvrir autrement.

4.2. L'Open Data

Une autre source importante du Big Data est l'Open Data (littéralement, « les données ouvertes »). Gartner définit l'Open Data comme « *l'information ou le contenu mis gratuitement à disposition pour l'utilisation et la redistribution, et soumis à la seule exigence de l'attribuer à la source. Le terme peut aussi être utilisé plus généralement pour décrire toutes les données qui sont partagées à l'extérieur de l'organisation et au-delà de son utilisation initiale voulue, par exemple avec des partenaires commerciaux, des clients ou des associations industrielles* » (2016c, traduction personnelle de l'anglais).

Pour donner une idée du contexte, l'Open Data, en 2013, c'était plus de 40 pays avec des plateformes gouvernementales d'Open Data, plus de 90 000 bases de données sur Data.gov (plateforme Open Data du gouvernement des Etats-Unis), 1,4 millions de vues de page sur le site Open Data de Royaume-Uni, 102 villes qui ont participé à l'International Open Data Hackathon Day, ou encore plus d'un million de bases de données rendues disponibles par les gouvernements au niveau mondial (McKinsey Global Institute, McKinsey Center for Government, & McKinsey Business Technology Office [MGI, MCG & BTO], 2013).

Selon McKinsey, les bases de « données ouvertes » revêtent les caractéristiques suivantes :

- L'accessibilité : un nombre très important d'utilisateurs sont autorisés à accéder à l'Open Data.
- La lisibilité automatique : les données peuvent être traitées automatiquement.
- Le coût : l'accès à l'Open Data est gratuit ou se fait à un coût négligeable.

- Les droits : il existe très peu de limites quant à l'utilisation, la transformation ou la distribution des données.

Ce sont les raisons pour lesquelles McKinsey qualifie l'Open Data d'« *information liquide* » (MGI, MCG & BTO, 2013, p. 1, traduction personnelle de l'anglais).

Par ailleurs, il est essentiel de ne pas confondre Big Data et Open Data. L'Open Data est une source du Big Data, mais l'entièreté de l'Open Data ne se trouve pas dans le Big Data, comme illustré en annexe 5. L'infographie du McKinsey Global Institute montre que l'Open Data constitue une grande partie du Big Data, mais la majeure partie du Big Data provient des autres sources de données, qui a priori ne sont pas « ouvertes » au grand public.

L'Open Data est donc une des sources du Big Data, et il revêtira un rôle important lorsque nous aborderons l'impact du Big Data sur les gouvernements et les services publics, dans la troisième partie.

5. Les outils pour manier le Big Data

Manier le Big Data n'est certainement pas chose aisée. Comme cela a été dit dans les précédentes sections, le Big Data est par définition un volume de données tellement important qu'il ne peut pas être manipulé avec les outils de gestion de données traditionnels. C'est pourquoi les entreprises qui ont percé dans le Big Data ont dû avoir à leur disposition des outils informatiques puissants. Nous exposons ici les principales technologies qui ont soutenu et soutiennent encore l'industrie du Big Data.

5.1. La plateforme Apache Hadoop

Un des outils les plus importants quand on parle de Big Data est la plateforme Apache Hadoop, lancée en décembre 2011. C'est un système open source, écrit en Java (langage de programmation informatique) libre, qui permet le stockage et le traitement de très gros volumes de données, cela au travers d'un groupement d'ordinateurs utilisant des algorithmes simples.

Pour illustrer à quel point cette plateforme est devenue incontournable pour les leaders de l'industrie du Big Data, il faut savoir que Hadoop est utilisé, pour la recherche et la production de données, par des entreprises comme Amazon, Facebook, Google, Twitter et Yahoo! entre autres (Boyer, 2015 ; Institut des Actuairens en Belgique, 2015 ; MGI, 2011).

5.2. Les bases de données non-relationnelles

Les bases de données relationnelles organisent et présentent les données selon un système de relations (tableaux à deux dimensions) et utilisent le langage de programmation informatique SQL (Structured Query Language). Au contraire, les bases de données non-relationnelles, rebaptisées « NoSQL » (Not only SQL) en 2009, sont des outils qui ne sont plus fondés sur l'architecture classique des bases relationnelles et qui permettent de faire face à deux des grands challenges du Big Data : son volume et sa variété.

La plupart de ces bases de données sont incorporées dans des sites web comme Google, Yahoo!, Amazon et Facebook (IBM, s.d.b ; Institut des Actuairens en Belgique, 2015 ; MGI, 2011 ; Techopedia, 2016).

5.3. Le cloud computing

Littéralement « informatique dans le nuage », cette technologie permet des capacités importantes de stockage de données mais aussi de ressources informatiques, disponibles en ligne, à des prix fixes et avec de hauts niveaux de flexibilité et de fiabilité. Grâce au cloud computing, les entreprises actives dans l'industrie du Big Data disposent « en quelques clics » d'une quantité infinie de ressources informatiques de qualité (Boyer, 2015 ; Institut des Actuairens en Belgique, 2015 ; MGI, 2011).

5.4. Les systèmes de data mining

Ces systèmes d'exploration de données constituent un ensemble de méthodes, d'algorithmes et de techniques qui permettent de manipuler le Big Data en apportant des solutions efficaces face aux challenges du volume, de la variété et de la vitesse des données. PageRank et MapReduce sont des modèles informatiques de ce type qui sont utilisés par Google et qui sont les symboles d'une avancée importante dans le domaine de la manipulation et de l'analyse des grands volumes de données (Institut des Actuairens en Belgique, 2015 ; MGI, 2011).

L'exposé de ces quelques outils et techniques de maniement du Big Data nous montre la complexité informatique et technologique qui se cache derrière les géants de l'industrie des données. A présent, nous allons parcourir quelques applications concrètes du Big Data dans le monde de l'entreprise et dans notre vie quotidienne.

6. Tour d'horizon de l'impact du Big Data

Le Big Data va avoir un impact majeur sur notre civilisation en bouleversant notre manière de vivre, de travailler et de penser. Plus aucun secteur de l'économie ni aucun domaine de notre

vie n'est à l'abri de la révolution du Big Data (Mayer-Schönberger & Cukier, 2013 ; MGI, 2011 ; Microsoft, 2016). C'est ce que nous montrons dans cette section en donnant, pour plusieurs secteurs de l'économie et domaines de la société, des applications du Big Data et des Big Data analytics, ainsi que des exemples concrets.

6.1. Secteur de la santé

De manière générale, les données médicales sont, selon IBM, non structurées à hauteur de 80 %, et elles sont pourtant toutes pertinentes. Le problème est qu'elles sont éparpillées dans tous les sens, car elles comprennent notamment :

- les dossiers médicaux des individus,
- les données des laboratoires et des systèmes d'imagerie,
- les rapports des médecins traitants,
- les données sur la santé du patient collectées en temps réel par les appareils médicaux.

Les technologies Big Data permettent aux organisations de soins de santé de rassembler toutes ces informations dans un seul endroit afin de les analyser et d'en tirer des conclusions. Cela résulterait en une meilleure coordination des soins de santé, des modèles de remboursement basés sur ces conclusions, une augmentation de l'implication des patients, de meilleurs programmes d'aide sociale et aussi des avancées importantes dans les traitements médicaux à l'aide d'analyses d'expériences cliniques (Institut des Actuaire en Belgique, 2015 ; PwC, 2013a).

6.2. Secteur de la vente au détail

Pour les grands détaillants, le Big Data est une magnifique opportunité d'acquérir des connaissances toujours plus approfondies sur leur clientèle, incluant ses intérêts et ses habitudes d'achats. Ainsi, les données permettent :

- de personnaliser le marketing,
- d'augmenter l'efficacité des campagnes de publicité,
- d'optimiser les décisions de présentation de l'offre de détail,
- de développer de nouveaux produits,
- de supprimer toutes les inefficiences dans la distribution et les opérations.

Le résultat est évidemment une augmentation des ventes et des marges bénéficiaires.

Un exemple concret est, comme le font certains grands détaillants, l'incorporation des flux de données provenant de Twitter dans leurs analyses des données de programmes de fidélité. Un

avantage considérable de cette méthode est de pouvoir anticiper les besoins et les pics de demande ou encore de proposer des promotions toujours plus appropriées aux clients (Institut des Actuairens en Belgique, 2015; PwC, 2013a ; voir aussi Brown, Chui, & Manyika, 2011).

6.3. Domaine de la politique et des services publics

L'avènement du Web 2.0, avec la naissance des contenus web générés par les utilisateurs eux-mêmes, des plateformes de médias sociaux et des systèmes de crowdsourcing, a suscité une vague d'engouement pour réinventer les gouvernements et leur fonctionnement. En 2008, les élections présidentielles aux Etats-Unis ont montré les premiers signes de succès des campagnes électorales en ligne et du concept de la participation politique.

Ce mouvement, surnommé la « *politique 2.0* » (Chen et al., 2012, p. 1170, traduction personnelle de l'anglais), a vu des politiciens se mettre à utiliser des plateformes web multimédias et participatives. Cela a abouti à une réussite, que ce soit au niveau des débats politiques, des campagnes de publicité, de la mobilisation des électeurs, des annonces d'événement ou des dons en ligne (Chen et al., 2012).

Un cas très connu de l'utilisation du Big Data en politique a eu lieu en 2012 lorsque Barack Obama a conduit sa campagne présidentielle à l'aide des Big Data analytics. Mais le Big Data va encore plus loin et va permettre une amélioration de l'efficacité de nombreuses responsabilités gouvernementales, comme nous le verrons dans la troisième partie. Par exemple, nous pouvons déjà citer la prévention et la prédiction des menaces et de la criminalité, ou aussi la détection de la fraude fiscale ou sociale (Institut des Actuairens en Belgique, 2015).

6.4. Prévention des cyber-risques

Le cyber-terrorisme est devenu une menace pour tout le monde. Aujourd'hui, il est devenu primordial pour toutes les organisations de se prémunir contre les cyber-menaces dans le but de protéger notamment leurs infrastructures informatiques et leurs actifs intellectuels. Elles ont maintenant la capacité d'analyser les trafics de données sur leurs réseaux informatiques en temps réel, ce qui leur permet de détecter des anomalies, indicateurs de potentielles cyber-attaques.

Gartner estime qu'en 2016, plus de 25 % des organisations internationales auront adopté les Big Data analytics pour au moins une utilisation de sécurité ou de détection de fraude, alors

qu'il n'y en n'avait que 8 % en 2014 (Chen et al., 2012 ; Institut des Actuairens en Belgique, 2015).

6.5. Secteur de la vente en ligne

Les plateformes de vente en ligne comme Amazon et eBay ont sans doute fait partie des pionniers de l'utilisation des Big Data analytics. Avec leur concept particulièrement innovant de « plateforme e-commerce », ces entreprises ont littéralement bouleversé les modèles économiques traditionnels. Une de leurs forces majeures a certainement été le déploiement des recommandations de produit.

Aujourd'hui, notamment avec ces systèmes de recommandations, Amazon mais aussi Google et Facebook continuent à mener le marché de la vente en ligne dans le développement des méthodes d'analyses de données web, du cloud computing et des plateformes de médias sociaux (Chen et al., 2012).

6.6. Sciences et technologie

Beaucoup de domaines des sciences et de la technologie ont récolté les bénéfices des capteurs et autres instruments générant des données à haut débit. C'est notamment le cas de l'astrophysique, de l'océanographie, de la génomique ou de la recherche environnementale. Ainsi, plusieurs disciplines scientifiques ont déjà commencé leur voyage vers le monde des Big Data analytics.

C'est par exemple le cas de la biologie qui a vu naître la plateforme iPlant Collaborative, dont le site web arbore le slogan « *Transformer la science à travers une découverte conduite par les données* » (iPlant Collaborative, s.d., traduction personnelle de l'anglais). Cette plateforme utilise une cyber-infrastructure pour soutenir une communauté de chercheurs, éducateurs et étudiants travaillant dans les sciences botaniques. Elle contient non seulement des données de toutes sortes (données canoniques, de référence, expérimentales, d'observation, de simulations et de modèles entre autres), mais aussi des données en open source et des outils analytiques (Chen et al., 2012).

6.7. Secteur des entreprises B2C

Pour les entreprises B2C (Business to Consumer), les données provenant des réseaux sociaux constituent une arme puissante de sondage des sentiments de leurs clients. Grâce à l'analyse du Big Data généré par les réseaux sociaux, ces entreprises ont l'opportunité de savoir ce que leurs clients disent à leur propos. En outre, elles peuvent aussi déterminer comment les

sentiments exprimés par les clients sur les réseaux sociaux impactent leurs ventes, l'efficacité et la réceptivité de leurs campagnes de marketing, ou encore l'optimisation de leur « *marketing mix* » (produit, prix, promotion et placement) (Zikopoulos et al., 2012, p. 24 ; voir aussi Quoistiaux, 2016).

Ce petit tour d'horizon nous donne une idée de l'ampleur de la révolution du Big Data, qui impacte déjà de nombreux secteurs de l'économie et domaines de notre vie quotidienne. A présent, après avoir exploré les avantages de la révolution en question, voyons quels risques peuvent être associés au Big Data.

7. Les risques du Big Data

Nous avons vu que la révolution du Big Data implique de nombreux avantages et bienfaits : d'une part, les entreprises connaissent mieux leurs clients et génèrent plus de bénéfices, d'autre part, les consommateurs reçoivent des produits et services toujours plus adaptés à leurs besoins. Nous avons aussi compris que le Big Data implique des défis pour les organisations qui souhaitent entrer dans la tendance des données. Effectivement, utiliser le Big Data pour son business, c'est bien, mais techniquement, ce n'est pas si facile. Maintenant, nous abordons la question suivante : le Big Data présente-t-il des risques ? Si oui, qui a du souci à se faire : l'entreprise, le consommateur, ou l'Etat ?

Mayer-Schönberger et Cukier, dans leur ouvrage (2013), écrivent que nous vivons dans un monde où « *jamais autant de données n'ont été collectées et stockées à propos de chacun de nous. Nous sommes constamment sous surveillance, que ce soit quand nous utilisons nos cartes de crédits pour payer, nos téléphones pour communiquer ou nos numéros de sécurité sociale pour nous identifier. En 2007, les médias anglais ont soulevé l'ironie du fait qu'il y avait plus de 30 caméras de surveillance dans un rayon de 200 mètres autour de l'appartement de Londres où George Orwell a écrit "1984"* ».

Avant même l'avènement d'Internet, des firmes spécialisées comme Equifax, Experian et Acxiom collectaient, mettaient sous forme de tableaux et fournissaient l'accès à des informations personnelles pour des centaines de millions de personnes dans le monde entier. Internet a rendu le pistage plus facile, moins cher et plus utile. Et les agences gouvernementales clandestines à trois lettres ne sont pas les seules à nous espionner. Amazon surveille nos préférences d'achats et Google nos habitudes de navigation, pendant que Twitter sait ce qui occupe nos esprits. Il semblerait que Facebook parviennent aussi à saisir

toutes ces informations, en plus de nos relations sociales. Les opérateurs mobiles ne savent pas seulement à qui nous parlons, mais aussi qui est tout près de nous.

Avec le Big Data qui promet des connaissances valorisantes à ceux qui l'analysent, tous les signes semblent pointer vers une montée d'autres rassemblements, stockages et réutilisations de nos données personnelles. La taille et l'échelle des collections de données vont continuer à augmenter à pas de géant au fur et à mesure que les coûts de stockage continuent à chuter et que les outils analytiques deviennent plus puissants que jamais. Si l'ère d'Internet a menacé notre vie privée, le Big Data la met-il encore plus en danger ? Est-ce cela le côté sombre du Big Data ? » (Mayer-Schönberger & Cukier, 2013, p. 150-151, traduction personnelle de l'anglais ; voir aussi Brown et al., 2011).

Voilà des questions qui font plutôt froid dans le dos... C'était le premier risque du Big Data pour les libertés individuelles, soulevé par Mayer-Schönberger et Cukier : le Big Data pourrait mettre en danger le respect de la vie privée. Selon eux, il existe deux autres risques du Big Data pour les libertés individuelles :

- « *La tendance à punir et à juger avant même que les individus n'aient agi, sur la base de la simple présomption de ce qu'ils auraient pu faire, et enfin,*
- *La dictature des données, soit, pour ceux qui les exploitent, le fait de n'envisager que les promesses qu'offrent ces données sans tenir compte de leurs limites.*

Les entreprises sont visées mais pas seulement. Les Etats sont eux aussi susceptibles de commettre de tels abus » (Chamaret, 2014, p. 96).

Nous pouvons en conclure que, face à ces risques, il y a vraisemblablement deux camps : d'un côté, les potentielles « victimes », et de l'autre, les potentiels « criminels ». Ces potentielles « victimes », manifestement, ce sont les citoyens, les consommateurs, et personne d'autre. Les potentiels « criminels », il semblerait donc que ce soit les entreprises comme les Etats. Quand nous lisons à quel point Mayer-Schönberger et Cukier (2013) font un constat alarmant de la révolution du Big Data, il est facile de s'inquiéter à propos de la potentielle menace qui pèse sur notre société et les droits de l'Homme les plus fondamentaux.

Ces craintes sont-elles réalistes ou disproportionnées ? L'Humanité est-elle en danger ? Quelles limites doivent être imposées au Big Data pour ne pas que cela dégénère ? Est-il vraiment possible de maîtriser la révolution du Big Data ? Ce sont là des questions importantes qui se posent et que nous aborderons aussi dans notre analyse de l'impact du Big

Data sur le secteur financier (cf. PARTIE 2 : Le Big Data dans le secteur financier) ainsi que dans le domaine de la gouvernance publique (cf. PARTIE 3 : Une Data Driven Governance).

8. Notre définition du Big Data

Sur base de tout ce qui vient d'être dit, nous désirons faire une synthèse en donnant notre propre définition du Big Data.

Selon nous, le Big Data, c'est un ensemble de données qui :

- d'une part, par son volume, sa variété, et sa vitesse, possède un potentiel énorme de création de valeur, tant au bénéfice de l'organisation que de l'individu ;
- d'autre part, implique des défis de grande taille qui s'articulent au niveau de la véracité des données, de leur visualisation, leur viscosité, leur variabilité, leur volatilité et leur vulnérabilité.

Les Big Data analytics, c'est-à-dire la manipulation et l'analyse de cet ensemble de données par des algorithmes, permet de générer des connaissances et des techniques qui vont bouleverser la manière dont le monde actuel fonctionne. De quelle manière ? En y apportant d'innombrables nouvelles opportunités pour améliorer la vie quotidienne de l'Homme et créer de la valeur pour son entreprise. Toutefois, la révolution du Big Data fait aussi apparaître de nouveaux risques qui, s'ils ne sont surveillés, pourront mettre en danger certains principes fondamentaux de notre civilisation.

PARTIE 2 : Le Big Data dans le secteur financier

Dans le cadre de notre analyse de l'impact du Big Data sur notre société et notre économie, nous nous penchons à présent sur le secteur financier. En réalité, la révolution des données semble engendrer de grands changements dans deux sous-secteurs de la finance en particulier : celui de la banque et celui de l'assurance.

D'abord, nous verrons les points communs qui caractérisent ces deux secteurs et nous tenterons de comprendre pourquoi le Big Data a un impact important sur ceux-ci. Ensuite, nous exposerons les grands changements qui sont à l'œuvre dans chaque secteur ainsi que les perspectives d'avenir et les risques du Big Data dans le secteur financier. Nous concluons cette deuxième partie en synthétisant les grandes tendances du Big Data qui ressortent de notre exposé.

Nous cherchons donc à comprendre pourquoi et comment le Big Data impacte les secteurs de la banque et de l'assurance. Toutefois, une autre question peut d'abord venir à l'esprit : pourquoi se concentrer sur ces deux secteurs et pas sur d'autres ? Premièrement, parce qu'à titre personnel et dans le cadre des mes études en sciences de gestion, ce sont des secteurs pour lesquels j'éprouve un intérêt plus important. Deuxièmement, parce que la banque et la compagnie d'assurances se ressemblent sur plusieurs points, comme nous le verrons dans la section 1. Troisièmement, parce que ce sont des secteurs dans lesquels le Big Data semble provoquer des changements particulièrement importants, ce que nous explorerons dans les sections 2 et 3 de cette partie.

1. Les points communs entre les deux secteurs

La banque et la compagnie d'assurances figurent incontestablement parmi les acteurs de l'économie dont le consommateur ne peut se passer. En effet, nous avons besoin d'argent pour vivre, et cet argent, nous ne pouvons pas nous permettre de le cacher sous forme de liasses en dessous de nos matelas. Le commun des mortels a donc besoin d'un endroit où pouvoir laisser son argent en sécurité, ce qui lui permet en même temps d'en retirer quelques intérêts.

Ensuite, pour la grande majorité de la population mondiale, il est sans doute difficile d'acheter une voiture et encore moins une maison « en cash », donc sans emprunter d'argent. Etant donné que ces besoins d'avoir un toit et un moyen de locomotion pour aller travailler et

gagner sa vie sont sans aucun doute des besoins primaires, pouvoir faire crédit devient aussi un besoin de première nécessité. Nous avons donc ici une certitude : tout le monde a besoin d'une banque, que ce soit pour y laisser son argent ou pour y faire un emprunt.

Par ailleurs, nous vivons dans un monde de risques. Lorsque nous avons une maison, il y a notamment un risque d'incendie. Quand nous prenons notre voiture pour aller au travail, il y a non seulement un risque d'accident et de dégâts matériels pour la voiture, mais aussi un risque de dégâts physiques qui pèse sur le conducteur et ses éventuels passagers, et potentiellement un risque de décès. Nous pourrions citer encore d'autres risques auxquels nous sommes soumis dans notre vie quotidienne, mais ces deux exemples suffisent à comprendre pourquoi il est important d'être assuré.

Bien qu'il semble insensé de ne pas vouloir s'assurer contre de tels risques, la loi belge a tout de même dû obliger l'individu à être assuré dans un certain nombre de cas, notamment pour ce qui touche aux voitures et à la responsabilité civile des conducteurs vis-à-vis des autres utilisateurs de la route. Nous comprenons donc pourquoi la compagnie d'assurances est aussi devenue un partenaire indispensable pour chaque individu.

Ceci nous montre un premier point commun entre la banque et la compagnie d'assurances : ce sont des entités dont personne ne peut se passer. Par conséquent, si le Big Data impacte considérablement ces deux secteurs, alors nous sommes, en tant qu'individus, aussi touchés par la révolution des données.

En outre, le fait que chaque individu soit en relation avec au moins une banque et une compagnie d'assurances engendre une autre caractéristique commune : mis à part les mineurs d'âge et peut-être les étudiants qui ne sont pas encore dans la vie active, le marché cible de l'organisation est égal à l'ensemble de la population. Cela implique évidemment un très grand nombre de clients, et donc un très grand nombre de données.

1.1. Les données

Selon Chamaret, « *trois types d'entreprises peuvent être identifiés tout au long de la chaîne de valeur des Big Data* :

- *celles qui possèdent les données : leur activité leur permet de collecter un nombre très important de données (mais elles n'ont pas nécessairement l'envie ou la volonté de les utiliser) ;*

- *celles qui disposent des compétences nécessaires au traitement des données pour les faire parler et en tirer de l'information (pas forcément de manière innovante) ;*
- *celles qui possèdent les idées : ces entreprises conçoivent de nouvelles formes de valeur dans l'utilisation des données.*

Google est un des rares acteurs à être présent tout au long de cette chaîne de valeur, il donne même à d'autres la possibilité de réutiliser ses données au travers d'interfaces de programmation » (2014, p. 95).

De par la nature de leurs activités et de leurs offres de service (Boyer, 2015), les banques et les compagnies d'assurances font sans aucun doute partie de la première catégorie d'entreprises : *« celles qui possèdent les données »*. En effet, que ce soit pour contracter un crédit ou une assurance, le client doit fournir un certain nombre de données personnelles à son banquier ou à son assureur.

Quelles données personnelles exactement ? Xavier Diez, directeur d'agences bancaires et courtier en assurances dans la marque Fintro du groupe BNP Paribas Fortis (communication personnelle, 28 Mars, 2016), décrit les informations qui doivent être demandées au client.

Dans le cas de la banque, que ce soit pour un crédit hypothécaire ou à la consommation, voici les informations requises :

- Informations personnelles de base : nom, date de naissance, état civil, adresse du domicile, ancienneté du domicile, nombres d'enfants à charge, etc.
- Informations relatives à l'activité professionnelle : nom et adresse de l'employeur, type de contrat de travail, ancienneté dans l'entreprise, etc.
- Informations financières relatives à la capacité de remboursement : revenu professionnel net mensuel, prime de fin d'année, chèques-repas, voiture société (avec ou sans carte carburant), certificats verts, revenus locatifs, revenus mobiliers, etc.
- Informations financières relatives à la solvabilité : propriétés foncières et avoirs mobiliers.

Pour certaines de ces informations, le client a l'obligation de fournir des documents justificatifs. L'objectif de toutes ces informations est de mesurer la capacité de remboursement et la solvabilité du client.

Dans le cas de la compagnie d'assurances, voici les informations qui sont demandées, en fonction du type d'assurances :

- Assurances habitation : valeur, ancienneté, superficie, nombres de chambres, propriété ou en location, etc.
- Assurances auto : marque, valeur, conducteur, puissance du moteur, etc.
- Assurances hospitalisation : informations sur la santé de la personne.

La véracité de ces informations repose sur la bonne foi du client qui va être assuré (pas toujours d'obligation de fournir des justificatifs). S'il s'avère que des informations qui ont été données sont fausses ou incomplètes, la compagnie d'assurances peut se retourner contre l'assuré. L'objectif de toutes ces informations est de mesurer le risque assuré.

Il est important de noter que ces données, même en masse, ne sont pas du Big Data. En effet, le Big Data ne se caractérise pas seulement par le volume et la vitesse des données, mais aussi par leur variété. Or, ces données traditionnelles, bien qu'elles puissent fournir des indications intéressantes à l'organisation sur son client, ne sont pas considérées comme variées au sens où nous l'entendons lorsque nous parlons du Big Data (cf. PARTIE 1 : Qu'est-ce que le Big Data ?). Si nous évoquons ces données traditionnelles, c'est pour donner une idée du volume d'informations qui est déjà généré depuis longtemps dans les secteurs de la banque et de l'assurance.

En effet, les données, ce n'est pas nouveau dans ces secteurs. « *Depuis toujours, les banques ont collecté des données pour évaluer la solvabilité des demandeurs de crédits. [Et depuis toujours], les compagnies d'assurances ont utilisé les données pour développer des modèles actuariels sophistiqués afin de déterminer les primes d'assurance* » (HP, 2013, p. 3, traduction personnelle de l'anglais). Toutefois, nous allons voir que le Big Data, en tant qu'ensemble de données volumineuses mais aussi et surtout rapides et variées, va ouvrir la porte à de nouvelles opportunités de création de valeur, autant pour les banques que pour les compagnies d'assurances.

Par ailleurs, ces organisations sont probablement en voie de faire aussi partie de la deuxième catégorie d'entreprises dont nous avons parlé plus haut : « *celles qui disposent des compétences nécessaires au traitement des données pour les faire parler et en tirer de l'information* ». En effet, les dirigeants dans ces secteurs ont compris que toutes ces données à leur disposition constituaient une opportunité incroyable de génération de valeur pour leur organisation. Afin de ne pas se laisser dépasser par les concurrents qui ont sauté dans la tendance du Big Data, chaque organisation a dû ou doit très rapidement mettre les moyens en œuvre afin de ne pas rester à la traîne.

1.2. La relation avec le client

La nature des activités de la banque et de la compagnie d'assurances implique que la relation entre l'organisation et le client est une relation basée sur la confiance et la fidélité. C'est le dernier point commun que nous relevons entre les deux secteurs analysés. La relation qu'un client entretient avec sa banque est assez semblable à la relation entre le client et sa compagnie d'assurances : en échange d'un service financier, le client transmet à l'organisation des données personnelles qu'il veut évidemment garder confidentielles, et l'organisation s'engage à respecter ce désir et à garder ces informations pour elle. C'est le secret professionnel.

Cette caractéristique commune va jouer un rôle important au fur et à mesure que les banques et les compagnies d'assurances vont accumuler des données client et vont être tentées de les utiliser de toutes sortes de manières afin d'en retirer un maximum de valeur.

A présent que nous avons énuméré quelques caractéristiques communes des secteurs de la banque et de l'assurance, nous allons analyser séparément l'impact du Big Data sur chacun de ces secteurs.

2. Le Big Data et la banque

Avant d'expliquer comment le Big Data affecte le secteur bancaire, nous allons tenter de comprendre le pourquoi. Nous avons dit que ce secteur était un secteur intéressant à analyser à la lumière du Big Data car il avait accès à une masse de données considérable, provenant principalement des clients qui contractent des crédits ou qui déposent à la banque leurs avoirs.

Les géants mondiaux de la consultance technologique confirment cette affirmation. Le McKinsey Global Institute avance que « *certaines secteurs exposent des niveaux d'intensité de données beaucoup plus élevés que d'autres, impliquant qu'ils ont un meilleur potentiel à court terme de saisir de la valeur à partir du Big Data. Les secteurs des services financiers, incluant les services de titres et d'investissements ainsi que les services bancaires, ont le plus de données digitales stockées par entité en moyenne* » (2011, p. 19, traduction personnelle de l'anglais). La recherche de McKinsey montre même qu'en 2009, les banques étaient le secteur qui avait le plus de données stockées (619 pétaoctets, soit 619×10^{15} octets) après les secteurs de la manufacture, du gouvernement et de la communication (MGI, 2011, Exhibit 7), comme le montre le graphique en annexe 6.

Ces affirmations sont bien confirmées par Atos (2014), un autre géant dans le domaine du numérique, qui parle même, dans son analyse du secteur des services financiers, de « *lac de*

données » (2014, p. 2, traduction personnelle de l'anglais). La firme HP n'a pas de doute non plus : « *Les données sont définitivement l'élément vital de l'industrie des services financiers. [...] Les marchés et les firmes des services financiers génèrent probablement le plus gros volume de données de toutes les industries* » (2013, p. 3, traduction personnelle de l'anglais). Enfin, PwC ajoute que « *les institutions financières ne se posent plus la question du besoin du Big Data* » (2013b, p. 7, traduction personnelle de l'anglais).

Nous sommes donc certains de deux choses. Premièrement, plus un secteur possède de données, plus il est enclin à utiliser les techniques d'analyse du Big Data (les Big Data analytics) pour générer de la valeur. Deuxièmement, le secteur des services financiers, et en particulier de la banque, est un secteur de ce type.

Nous comprenons donc que, si le secteur bancaire est fortement touché par la révolution du Big Data, c'est parce qu'il a à sa disposition des masses de données considérables. Maintenant, voyons comment le Big Data impacte ce secteur.

2.1. Les grands changements à l'œuvre

Selon les scientifiques Turner et al. (IBM), « *aujourd'hui, [le Big Data] est un impératif du business et apporte des solutions aux challenges économiques les plus anciens pour les banques et les entreprises des marchés financiers dans le monde entier. Les firmes de services financiers tirent profit du Big Data pour transformer leurs processus, leur organisation et bientôt, l'industrie entière* » (2013, p. 1, traduction personnelle de l'anglais). Nous constatons que le Big Data ne provoque pas des petits changements dans l'amélioration des services ou de la performance, mais engendre plutôt des changements radicaux par rapport aux modèles économiques existants.

Atos (2014) a estimé pour 2015 une croissance du marché du Big Data de 6,4 milliards de dollars dans le secteur des services financiers, soit 22 % de croissance annuelle. Selon l'étude de la même firme, les organisations qui utilisent le mieux les Big Data analytics ont une probabilité plus importante d'atteindre des performances financières élevées (x2), de prendre des décisions beaucoup plus rapidement que leurs concurrents (x5), de mettre en œuvre ces décisions comme désiré (x3) et d'utiliser toujours plus souvent le Big Data pour prendre de nouvelles décisions (x2).

Les banques doivent comprendre les enjeux du Big Data et examiner les données à leur disposition afin de voir comment celles-ci peuvent être utilisées dans leur stratégie. Les deux

domaines de la banque qui sont les plus impactés et améliorés par les Big Data analytics sont, d'une part, la gestion financière et des risques et, d'autre part, la relation client (Atos, 2014).

2.2. L'amélioration de la gestion financière et des risques

Le premier domaine de changement, la gestion financière et des risques, caractérise une problématique qui est interne à la banque. Les données qui seront générées et analysées dans le cadre de cette problématique seront principalement d'ordre financier. Nous listons ici six exemples d'application des Big Data analytics, évoqués notamment par Turner et al. (2013), pour une meilleure gestion financière et des risques dans le chef de la banque.

2.2.1. Optimisation de l'allocation entre actifs et passifs

Depuis la crise financière de 2008, qui a touché tout le secteur bancaire, le Comité de Bâle pour la Supervision bancaire a renforcé ses exigences en matière de fonds propres et de liquidités. Depuis lors, les banques sont contraintes de respecter une série de règles très strictes. Une non-conformité face à ces régulations peut entraîner des rapports d'audit à opinion défavorable, des amendes ou des actions disciplinaires. Dans ce contexte, il existe trois types de données générées : des données de gouvernance financière, de risque et de conformité.

Les technologies du Big Data permettent de rassembler toutes ces données provenant de sources différentes et de les analyser en temps réel. L'objectif est que l'institution financière puisse s'assurer d'être toujours en conformité avec les exigences légales d'allocation du capital et de liquidités. Ainsi, la banque peut veiller à détenir constamment la juste quantité de capital nécessaire et pas un surplus qui serait inutile et synonyme d'inefficience.

Un autre avantage de ces techniques se trouve dans les simulations que la banque peut effectuer. Elles lui permettent de mesurer l'impact des potentiels changements de sa politique financière sur son bilan et ses indicateurs financiers, en fonction de conditions politiques et économiques variables (Atos, 2014 ; Bank for International Settlements, s.d. ; Groll, O'Halloran, Maskey, McAllister, & Park, 2015 ; HP, 2013).

2.2.2. Meilleure gestion du risque

Les Big Data analytics rendent possible une analyse des actifs et des activités risqués plus rapide et de meilleure qualité, ce qui permet à la banque de retirer les risques sur des milliards de dollars. En effet, aujourd'hui, les « data analysts » (littéralement, « analystes des données ») des banques construisent des modèles de données qui leur permettent de prédire

les crédits pour lesquels il existe une probabilité de défaut. Le but est de pouvoir agir de manière proactive avant que le défaut soit effectif.

Par exemple, la banque peut créer une cartographie qui identifie les régions où les cas de défaut sont courants, détaillée jusqu'au niveau des codes postaux. La banque peut aussi analyser les données historiques de transactions de ses clients, identifier des tendances de dépenses au niveau des ménages, monter des schémas de dépenses et ainsi prédire les capacités probables de remboursement des ménages (Atos, 2014 ; Henno, 2016 ; Oracle, 2015 ; PwC, 2013a).

2.2.3. Prévention plus efficace de la fraude

Les analyses de masses de données par les banques permettent une amélioration dans l'efficacité de la détection et de la prévention de la fraude, comme le blanchissement d'argent, les activités boursières illégales ou les activités de financement du terrorisme. En effet, les Big Data analytics permettent de relever des corrélations entre des données de différentes sources et ainsi d'identifier plus facilement et rapidement les cas de fraude.

Par exemple, alors que des événements (comme des envois d'e-mails ou des coups de téléphone) seuls ne peuvent pas être des signaux de cas de fraude, ces mêmes événements, mis en relation avec des données de transaction ou de géolocalisation, peuvent faire ressortir des activités suspectes. Ces techniques d'utilisation du Big Data permettent d'étudier et de pister les comportements des clients, ce qui entraîne des améliorations significatives dans la prévention de la fraude par les banques (HP, 2013 ; Oracle, 2015 ; PwC, 2013b).

2.2.4. Identification des menaces de cyber-attaques

Comme nous l'avons évoqué précédemment, les banques détiennent beaucoup de données confidentielles. Par conséquent, une question importante se pose pour elles : comment neutraliser les cyber-attaques avant que celles-ci ne les menacent, elles et leurs clients ? Les techniques du Big Data permettent à la banque non seulement de visualiser et d'enregistrer tout ce qu'il se passe à travers son système interne, mais aussi d'analyser et de comprendre comment les anomalies apparaissent. De cette manière, la banque est alertée plus rapidement en cas de menace imminente (HP, 2013).

2.2.5. Optimisation des prix

Les technologies du Big Data peuvent conduire les banques à analyser les données pour une optimisation des solutions de prix. Y a-t-il vraiment une valeur ajoutée à cela ? Oui. Par exemple, une optimisation de la hauteur d'un seul point de base (égal à 0,1 %) sur un

portefeuille de centaines de milliards de dollars aura un impact positif non négligeable sur le résultat de la banque. Il y a donc un réel intérêt à utiliser les Big Data analytics pour optimiser les prix (HP, 2013).

2.2.6. Amélioration de l'efficacité opérationnelle

Les banques utilisent les Big Data analytics afin d'améliorer l'efficacité opérationnelle entre les personnes, les processus et les actifs physiques. Des processus appropriés doivent être mis en place dans le but d'assurer une productivité maximale parmi les employés et ainsi permettre une réduction des coûts (HP, 2013).

En conclusion, nous constatons que si les bonnes techniques d'analyse sont appliquées sur les bonnes données, les Big Data analytics peuvent apporter une réelle plus-value à la banque au niveau interne. Voyons maintenant comment la banque peut aussi se transformer au niveau externe.

2.3. L'amélioration de la relation client

Pour ce qui concerne l'amélioration de la « customer intimacy » (littéralement, « intimité avec le client »), les Big Data analytics font passer la relation client d'une relation traditionnelle du type « même traitement pour tout le monde » à une relation du type « customer engagement » (littéralement, « engagement du client »). A travers ce nouveau type de relation, la banque trouve des manières plus efficaces et réactives de comprendre son client et d'en établir un profil qui permette un service individualisé et de meilleure qualité (Atos, 2014 ; Oracle, 2015).

La relation client n'est plus une problématique interne à la banque, comme la gestion financière et des risques, mais externe, entre la banque et le consommateur. Par conséquent, la banque n'est plus l'unique bénéficiaire de l'usage de l'analyse des données, mais le client aussi. L'intérêt pour la banque est évidemment d'augmenter la satisfaction du client, mais aussi de proposer des offres de services qui correspondent tellement bien aux besoins du client que ce dernier ne peut refuser. Le client en ressort aussi gagnant puisqu'il bénéficie d'un service toujours plus personnalisé.

Nous détaillons ci-dessous les différentes manières dont la banque peut améliorer la relation client grâce aux Big Data analytics.

2.3.1. Une analyse à grande échelle pour un marketing et un service personnalisé

Chaque jour, les consommateurs fournissent de nombreuses informations valorisables à leur banque. Ce sont des quantités faramineuses de données structurées (comme des données historiques et de transactions) et non structurées (comme des données audio de call centers ou des données relatives à l'utilisation des services bancaires sur des appareils mobiles via les données cellulaires) à propos des transactions du client et de ses interactions avec la banque. Une approche « multicanale » permet d'analyser les informations de l'ensemble de la clientèle à grande échelle, et ensuite de générer des indications précises sur les besoins, les comportements et les préférences des clients individuellement.

Une fois que la banque connaît bien ses clients, elle peut alors mener des actions marketing ciblées et leur proposer des produits ou des services personnalisés (comme un contrat de crédit sur-mesure). L'analyse des données peut aussi conduire la banque à inventer toujours plus de nouveaux produits et services afin de répondre aux besoins des clients dans un monde en évolution permanente. Etant donné la concurrence rude du secteur bancaire, tout ceci est primordial pour une banque si elle veut se distinguer de ses concurrents et tenter d'augmenter sa part de marché (Atos, 2014 ; Gartner, 2015b ; HP, 2013 ; Oracle, 2015; PwC, 2013b ; Turner et al., 2013).

2.3.2. Un besoin anticipé et une offre au bon moment

Ce deuxième point est lié de près au premier. Si la banque connaît son client, elle peut, en plus de lui fournir un service personnalisé, aussi prévoir ses besoins. En effet, un « mapping » (littéralement, une « cartographie ») à grande échelle des moments clés dans la vie des gens permet d'anticiper les besoins des clients individuellement et d'intervenir auprès d'eux plus efficacement au moment où ils en ont le plus besoin.

Concrètement, voilà comment cela pourrait se traduire : la banque, grâce à l'analyse des données de ses milliers de clients, serait capable d'anticiper, pour un client en particulier, un besoin de crédit. Elle lui proposerait alors une offre de crédit qui correspond à son besoin prochain avant même que celui-ci ne se présente (Atos, 2014 ; PwC, 2013b). En résumé, les Big Data analytics permettent à la banque de mieux connaître son client afin de lui proposer « *le bon produit, au bon moment, et au bon prix* » (Oracle, 2015, p. 6, traduction personnelle de l'anglais).

2.3.3. Un service qui tient mieux compte du feedback client

La tendance Big Data a encouragé les banques à développer des systèmes interactifs à travers lesquels elles sont à l'écoute des avis, réactions et préférences des clients (Atos, 2014 ; PwC, 2013b). Il est clair que si les clients se sentent écoutés et compris dans leurs besoins, la valeur ajoutée pour la banque est une meilleure satisfaction client.

2.3.4. Un contrôle du risque de réputation

Dans le but de protéger la réputation de sa marque, la banque peut effectuer une surveillance des données provenant du web, et surtout des réseaux sociaux. De cette manière, l'entité peut savoir quels sont les sentiments exprimés par les clients au sujet ses produits, de ses employés ou encore de sa direction (Oracle, 2015).

En parcourant les différentes manières dont les Big Data analytics peuvent être utilisés pour générer de la valeur dans le secteur bancaire, nous comprenons comment fonctionnent les Big Data analytics dans la pratique. Le principe est d'analyser une très grande quantité de données (la population) et d'y identifier des corrélations, des liens de causalité et des schémas d'occurrence récurrents qui permettent de mieux saisir comment se comporte la population. Ensuite, l'objectif est d'en tirer des conclusions applicables à un seul élément de la population, cet élément étant un client, un ménage, un crédit ou encore un cas de fraude.

Rappelons que, le Big Data, ce sont des données non seulement volumineuses et collectées presque en temps réel, mais aussi et surtout très variées. Dans ce sens, l'atout majeur des Big Data analytics est le fait de mettre en relation des données qui a priori n'ont pas de lien entre elles (une transaction avec un e-mail, un crédit avec une zone géographique ou bien même un tweet avec la performance financière de la banque) et d'en faire ressortir des corrélations qui permettent de mieux comprendre les risques et les comportements.

2.4. La naissance de la banque digitale

Quand Atos (2014) évoque les changements engendrés par le Big Data dans le secteur bancaire, il parle de transformation digitale des modèles traditionnels. Il est certain qu'« *alors que les consommateurs européens déménagent sur Internet, les banques commerciales vont devoir faire de même* » (Olanrewaju, 2014, p. 1, traduction personnelle de l'anglais). Il est ici question du concept de la transformation de la banque traditionnelle en « banque digitale ». La « banque digitale », c'est donc le déménagement de la banque physique sur Internet. Pourquoi ce changement est-il indispensable ?

La digitalisation de la banque traditionnelle est indispensable parce que l'émergence des Big Data analytics fait peser sur elle trois types de menace (Metge, 2015, p. 96-97) :

- *« D'abord une menace de contournement des barrières à l'entrée, de nouveaux acteurs se fondant uniquement sur des maillons spécifiques de la chaîne de valeur (l'entrée en relation, la comparaison, etc.), sans avoir à réaliser les investissements nécessaires pour devenir un acteur au sens propre du secteur (appliquer la réglementation bancaire, concevoir des produits, construire des usines de production, etc.) ;*
- *Ensuite, une menace liée à la réduction de l'asymétrie d'information. La détention et le partage des données, autrefois administrés par la banque dans des conditions extrêmement restrictives, deviennent l'apanage du client, qui décide souverainement qui peut avoir accès ou non à tout ou partie de ses données personnelles. Dans une économie de la donnée, la fin de cette asymétrie d'information représenterait pour le secteur bancaire la perte d'un avantage compétitif ;*
- *Enfin, une menace liée au développement d'usages à caractère addictif : chercher une information (Google), développer ses relations sociales (Facebook, Twitter) ou professionnelles (LinkedIn), consommer (Amazon) sont autant d'usages sur lesquels des acteurs se sont focalisés au point d'acquérir des positions monopolistiques à l'échelle globale. »*

Pour ces raisons, si la banque ne veut pas se laisser distancer par ses concurrents, autant financiers que non financiers, elle doit absolument entamer sa digitalisation. Certaines banques ont déjà bien entamé leur digitalisation, comme BNP Paribas Fortis et sa banque en ligne Hellobank! (Hellobank!, s.d.). Mais nous avons aussi évoqué les concurrents non financiers de la banque, ce qui pourrait paraître étrange.

Pourtant, c'est bien une réalité : des acteurs non financiers de l'économie, comme des opérateurs téléphoniques, maîtrisant les technologies mobiles, commencent à se lancer dans la banque digitale. C'est le cas de l'opérateur mobile Orange, en France, qui envisage sérieusement sa propre banque en ligne et que le secteur bancaire ne peut ignorer, comme le montrent ces titres de presse en ligne : *« Orange veut racheter Groupama Banque pour lancer sa banque mobile »* (Belouezzane & Grandin de l'Eprevier, 2016) et *« Pourquoi Orange peut faire trembler les banques »* (Ninon, 2016) de janvier 2016.

Après cette analyse, nous réalisons que le Big Data est bien en voie de bouleverser le secteur bancaire. Voyons à présent ce qu'il en est dans le secteur de l'assurance.

3. Le Big Data et la compagnie d'assurances

Le secteur de l'assurance fait aussi partie du secteur général des services financiers. Par conséquent, les conclusions que nous avons tirées à propos du secteur financier et du secteur bancaire, sur le pourquoi de l'impact du Big Data, sont aussi valables pour le secteur de l'assurance. Bien que l'étude du McKinsey Global Institute montre que ce dernier possède moins de données que le secteur bancaire (MGI, 2011, Exhibit 7 ; voir annexe 6), personne ne conteste que les compagnies d'assurances font aussi partie des organisations qui détiennent des quantités considérables de données et où les Big Data analytics peuvent apporter une réelle valeur ajoutée (Atos, 2014 ; HP, 2013 ; MGI, 2011 ; PwC, 2013a, 2013b).

En particulier, dans le secteur de l'assurance, « *le champ d'opportunités techniques du Big Data est immense en ce qui concerne la tarification (et le marketing qui lui est lié), même si le Big Data et le digital modifient l'ensemble de la chaîne de valeurs, de la désintermédiation à l'accélération de la gestion* » (Boyer, 2015, p. 82).

3.1. Les analyses prédictives

Une autre raison qui explique pourquoi les Big Data analytics affectent de manière significative le secteur de l'assurance, et qui est plus propre à ce secteur en particulier, est que l'assurance repose principalement sur l'évaluation d'un risque futur. En effet, tout le challenge des compagnies d'assurances est de calculer la probabilité qu'un événement se produise à l'avenir. Elles sont donc amenées à mettre des moyens en œuvre pour prédire l'avenir le plus précisément possible. Les Big Data analytics, qui permettent des analyses prédictives de plus en plus performantes, deviennent donc une magnifique opportunité pour les compagnies d'assurances.

L'analyse prédictive n'est pas un concept nouveau. Depuis les premiers jours de l'assurance, c'est ce que les actuaires tentent de faire en analysant, par exemple, des statistiques de mortalité afin de déterminer les prix des produits d'assurances vie. Ce concept de modélisation prédictive est défini comme un ensemble de techniques dans lequel un modèle mathématique est créé ou choisi. Ce modèle analyse de grandes masses de données historiques pour y identifier des corrélations et des interactions, dans le but de prédire au

mieux la probabilité d'occurrence d'événements futurs (Institut des Actuairens en Belgique, 2015 ; MGI, 2011 ; PwC, 2013a).

Aujourd'hui, comme cela a déjà été largement développé dans les sections précédentes, le Big Data mêle ces données historiques traditionnelles structurées avec des données nouvelles non structurées, qui proviennent principalement d'Internet et des réseaux sociaux, mais aussi de toutes sortes de capteurs et d'appareils mobiles. De plus en plus de ces données sont collectées en temps réel, et nous retrouvons ainsi les trois caractéristiques du Big Data : volume, vitesse et variété.

Bien que, techniquement et mathématiquement, ce soit très complexe, le principe des analyses prédictives est assez simple : le but est d'« *utiliser des algorithmes pour trouver dans les données des schémas qui pourraient prédire des résultats similaires dans le futur* » (Gualtieri, 2013, p. 2, traduction personnelle de l'anglais ; voir aussi Schenk & Berman, 2015).

Avec le Big Data, les analyses prédictives passent à un autre niveau de performance. D'abord, les techniques traditionnelles de modélisation prédictive doivent évoluer afin d'être capables de manipuler et analyser le Big Data. Ensuite, lorsqu'on applique ces techniques sur du Big Data, les analystes en arrivent à des prédictions de comportements et d'événements qui battent tous les records. Globalement, les analyses prédictives de Big Data ont un impact considérable sur le business modèle de l'entreprise : il en résulte une réduction des risques, une meilleure prise de décision et une amélioration de l'expérience client (Gualtieri, 2013 ; IBM, 2012 ; Schenk & Berman, 2015 ; voir aussi Henno, 2016).

Comme nous le voyons dans les prochaines sections, les analyses prédictives sont devenues une arme redoutable de création de valeur pour les compagnies d'assurances, qui sont amenées à évaluer chaque jour des risques futurs. Nous distinguons trois principaux domaines d'impact des Big Data analytics, et surtout des analyses prédictives, dans le secteur de l'assurance : premièrement, le marketing et le service rendu au client, deuxièmement, l'évaluation des risques et le calcul des primes d'assurance, et troisièmement, l'identification et la prévention de la fraude.

3.2. L'amélioration du marketing et du service pour le client

« *Les technologies du Big Data peuvent aider [l'assureur] à innover et à apporter un meilleur service au client en lui permettant d'explorer des données de beaucoup de sources différentes, dans le but de gagner des connaissances et de l'inspiration pour la création de services toujours plus intéressants et pertinents* » (HP, 2013, p. 8, traduction personnelle de

l'anglais). En effet, les Big Data analytics permettent des actions marketing mieux ciblées et un service de meilleure qualité et plus innovant pour l'assuré.

3.2.1. Une meilleure connaissance des besoins du client

Certains des grands changements engendrés par les Big Data analytics dans le secteur bancaire, dont il a été question plus haut, sont aussi d'actualité pour le secteur de l'assurance. C'est le cas de l'analyse « multicanale » des données structurées comme non structurées (croisement des données traditionnelles avec les données provenant des réseaux sociaux, de capteurs, d'appareils mobiles, etc.) qui permet une connaissance approfondie du client, de ses besoins et de ses préférences (Atos, 2014 ; Boyer, 2015 ; Capgemini Consulting, 2015 ; IBM, 2012 ; MGI, 2011 ; PwC, 2013b). En conséquence, les analyses de données rendent possible un marketing ciblé plus performant que jamais.

3.2.2. Un marketing ciblé sur l'anticipation des besoins

L'idée d'un tel marketing est de proposer au client des offres de produits et de services en fonction « d'où il en est dans sa vie ». L'objectif pour la compagnie d'assurances est d'arriver à un stade où elle a accumulé tellement de données sur ses clients que les Big Data analytics rendent possibles des analyses prédictives de haute performance. Ainsi, elle serait capable d'anticiper les besoins de ses clients et de leur proposer des produits et services qui conviennent à leur profil, avant même que ceux-ci ne se rendent compte qu'ils en ont besoin, ou avant qu'ils n'aillent voir chez un concurrent.

Très concrètement, voilà comment cela peut se passer. Par exemple, la compagnie d'assurances peut utiliser les technologies d'analyse du Big Data afin de pouvoir détecter qu'un client étudiant va être diplômé sous peu. Celui-ci va alors commencer à travailler, à recevoir un revenu, et il va avoir incontestablement besoin de nouvelles assurances. De même pour une personne qui partirait à la retraite et qui nécessiterait une rente viagère plutôt qu'un programme d'épargne-pension : l'assureur doit anticiper l'opportunité et venir à l'encontre de son assuré de manière proactive (Boyer, 2015 ; Gartner, 2015b ; IBM, 2012 ; MGI, 2011 ; PwC, 2013b ; voir aussi Quoistiaux, 2016).

Dans ce type de marketing, en plus du Volume, c'est la dimension Variété qui amène la différence. La valeur dégagée de l'utilisation du Big Data provient surtout du croisement de données de sources différentes et de types différents afin de prédire la prochaine étape dans la vie de l'assuré.

3.2.3. Un marketing ciblé sur la géolocalisation

L'idée d'un tel marketing est de proposer au client des offres de produits et de services en fonction « d'où il se trouve à un moment précis ». La source première de ce type de marketing est naturellement la donnée de géolocalisation récoltée en temps réel. Par exemple, dans le domaine du voyage, en fonction de la ville où le client atterrit, son assureur pourrait lui envoyer des offres d'assurances qui sont appropriées à son nouvel environnement, comme des propositions d'assistance d'urgence dans une zone qui serait considérée comme risquée par les Big Data analytics (HP, 2013 ; PwC, 2013b).

Dans ce type de marketing, en plus du Volume, c'est la dimension Vitesse qui amène la différence. La valeur dégagée de l'utilisation du Big Data provient surtout de la capacité à savoir rapidement où le client se trouve et à réagir de manière appropriée au plus tôt.

3.2.4. Un service innovant et de meilleure qualité

Une meilleure connaissance des besoins des clients peut permettre à l'assureur d'imaginer et de développer des solutions et des services innovants. Des compagnies d'assurances qui auraient accès aux données personnelles de géolocalisation de leurs clients pourraient entreprendre, grâce aux Big Data analytics, le développement de nouvelles technologies et de nouveaux services pour les accompagner et les aider dans leur vie quotidienne. Par exemple, dans le but d'encourager une conduite plus sûre, l'assureur pourrait offrir un service d'alertes en temps réel à propos de la circulation, des conditions météo, des endroits de parking à risque élevé, ou des changements de limitations de vitesse (Gartner, 2015b ; MGI, 2011).

Dans la même idée, une compagnie d'assurances pourrait fournir à son client, sous forme d'une plateforme en ligne, un service d'aide à la compréhension des risques relatifs à la propriété d'un bien immobilier. C'est ce que propose la firme FM Global avec RiskMark, un outil qui inspecte chaque propriété de client en collectant de nombreuses données comme des informations photographiques, géographiques et des paramètres de construction. Non seulement l'assureur peut analyser ces données et fournir à l'assuré une meilleure évaluation des risques existants, mais ce service permettrait aussi à l'assuré d'analyser lui-même les risques. Par exemple, ce dernier pourrait utiliser la plateforme pour comparer les risques associés à des propriétés qui se trouvent à des localisations différentes.

Les Big Data analytics permettent aussi à la compagnie d'assurances de fournir un conseil de meilleure qualité au client, en mettant à la disposition des agents de conseil une interface performante de présentation des données du client. C'est ce qu'a fait l'assureur MetLife en

créant MetLife Wall, une plateforme « conduite par le Big Data » donnant accès à toutes les données de l'assuré en un seul lieu. Ainsi, l'agent de conseil ne doit pas naviguer à travers de multiples sources de données pour tenter d'apporter au client quelques informations pertinentes sur les produits qui pourraient l'intéresser. Au contraire, l'agent peut lui apporter un conseil approprié sur base de la vue à 360 degrés que donne l'application sur l'ensemble des données du client. Le résultat est une amélioration nette de la qualité et de la pertinence du conseil en assurances (Capgemini Consulting, 2015).

3.2.5. Une meilleure gestion de la réputation

Les réseaux sociaux sont devenus un espace de puisage de données incroyablement utiles pour les compagnies d'assurances. Au-delà du fait d'analyser les données provenant d'Internet pour analyser les comportements des individus, ces réseaux sont aussi intéressants pour analyser les sentiments des clients à propos de la marque de l'assureur. Les Big Data analytics permettent alors un scannage à grande échelle de « ce que les internautes disent d'une marque ou d'un produit ». Ainsi, la compagnie d'assurances peut entreprendre les bonnes démarches afin de gérer au mieux sa réputation et son image de marque.

De plus, les Big Data analytics deviennent aussi un moyen puissant de « tâter le terrain » avant de lancer un nouveau produit sur le marché. Les mêmes analyses de données sur les réseaux sociaux permettent à la compagnie d'assurances de prédire comment le marché va accueillir un nouveau produit et de faire des prévisions de ventes plus précises que jamais (IBM, 2012 ; PwC, 2013b).

3.3. Des primes d'assurance qui reflètent les risques réels

Les méthodes traditionnelles de calcul des primes d'assurance sont basées sur le profil de risque d'un groupe de clients et ne tiennent pas compte des profils de risque individuels. De cette manière, étant donné que, dans ce groupe de clients, chacun d'eux paye la même prime, ceux qui sont en réalité moins risqués compensent ceux qui sont réellement plus risqués.

Aujourd'hui, les Big Data analytics permettent une évaluation des risques qui est non seulement plus réelle, mais aussi plus précise et individualisée. Pourquoi ? Parce que les compagnies d'assurances peuvent analyser les informations personnelles des clients en les mettant en relation avec des informations provenant des réseaux sociaux ainsi que des données télématiques et géospatiales. Les données sont volumineuses, car des millions de clients sont concernés, elles sont rapides, car collectées presque en temps réel, et elles sont

variées, car elles proviennent de sources différentes et sont de formats différents. Nous avons donc bien affaire à du Big Data.

En croisant et en analysant ces différents types de données, les compagnies d'assurances parviennent à faire ressortir des corrélations entre les données et à mieux discerner les risques relatifs à chaque client. En conséquence, l'assureur a la capacité de calculer des primes d'assurance sur base des profils de risque individuels, primes qui reflètent plus que jamais les véritables risques associés à chaque assuré (Atos, 2014 ; Boyer, 2015 ; Capgemini Consulting, 2015 ; HP, 2013 ; IBM, 2012 ; Institut des Actuairens en Belgique, 2015 ; PwC, 2013b ; voir aussi Brown et al., 2011 ; Quoistiaux, 2016).

3.3.1. Des contrats d'assurance basés sur les comportements

« Pour les assureurs, l'utilisation de données vraies et en temps réel à la place d'algorithmes actuariels ouvre la possibilité à des contrats basés sur les comportements et à une tarification minute par minute : c'est un facteur vital qui va transformer le marché vers un meilleur modèle » (Atos, 2014, p. 5, traduction personnelle de l'anglais). Les données de type télématique, données en temps réel qui combinent *« les moyens de l'informatique avec ceux des télécommunications »* (Petit Robert, 2010, p. 2520), sont un des moyens pour rendre possible cette nouvelle tendance d'assurance sur-mesure.

En effet, l'analyse des données télématiques permet des contrats d'assurance qui conviennent parfaitement aux profils de risque des clients. Un exemple concret est celui d'une prime d'assurance voiture basée sur le véritable comportement au volant du conducteur, plutôt que sur des informations classiques comme l'âge et le sexe du conducteur ou la puissance de sa voiture. Pour ce faire, un capteur (du type accéléromètre ou gyromètre) est placé dans la voiture de l'assuré et « mesure » son style de conduite. Ces informations sont ensuite transmises à la base de données de la compagnie d'assurances, qui associe à cet assuré un profil spécifique de conduite (plus ou moins agressif). Enfin, l'assureur est en mesure de calculer la prime d'assurance de son client sur base de ce profil spécifique.

C'est une petite révolution puisque les primes d'assurance ne seraient plus basées sur des données comme l'expérience du chauffeur, qui s'apparentent à des préjugés, mais sur des données réelles et objectives de style de conduite et de risques potentiels encourus par l'assureur. Ainsi, un conducteur agressif, quelque soit son âge, son sexe et la puissance de sa voiture, même s'il a beaucoup d'années d'expérience sur la route, devrait payer une prime

d'assurance plus élevée (Atos, 2014 ; HP, 2013 ; IBM, 2012 ; Institut des Actuares en Belgique, 2015; MGI, 2011).

3.3.2. Réduction des coûts liés aux déclarations de sinistre

C'est un point directement lié au précédent, dans le cas de l'assurance voiture. En effet, si l'assuré se sent surveillé dans sa manière de conduire, alors forcément il va adopter une conduite plus prudente. Or, une conduite plus prudente implique un risque d'accident moins élevé. À grande échelle, si tous les clients d'une compagnie d'assurances roulaient plus prudemment, il pourrait (car cela doit encore être prouvé) s'ensuivre une diminution du nombre d'accidents et donc une baisse des déclarations de sinistre et des coûts de remboursements liés à ces sinistres (MGI, 2011).

3.3.3. Récompenser le client pour une bonne attitude

Dans cette nouvelle tendance de contrats d'assurance basés sur les comportements, est né un principe de récompenser le client pour sa « bonne attitude », à travers des réductions sur ses primes d'assurance. Par exemple, le programme d'assurance auto « Pay as you drive » (littéralement, « Paye au fur et à mesure que tu roules ») octroyait des réductions de primes si l'assuré parcourait moins de kilomètres qu'un seuil fixé. Aujourd'hui, grâce aux opportunités offertes par les Big Data analytics et les données télématiques, ce type de contrat a fait place à des programmes d'assurance « Pay how you drive » (littéralement, « Paye en fonction de comment tu conduis »).

Dans le cadre d'un tel programme, la compagnie d'assurances peut analyser le style de conduite de chaque assuré, lui associer un profil de risque individuel, lui attribuer une prime d'assurance basée sur ce profil et ensuite lui octroyer des réductions sur ses primes en cas de conduite prudente. Par exemple, l'assureur Progressive Insurance, très avancé dans l'utilisation des technologies Big Data, a développé un programme du type « Pay how you drive », appelé Snapshot (Boyer, 2015 ; Capgemini Consulting, 2015 ; IBM, 2012 ; Institut des Actuares en Belgique, 2015 ; MGI, 2011).

3.3.4. Une meilleure évaluation des risques liés à l'immobilier

Toujours en combinant l'analyse de données traditionnelles structurées avec des données « nouvelle génération » non structurées, une compagnie d'assurances peut voir son évaluation des risques liés à l'immobilier passer à un niveau supérieur.

Par exemple, elle pourrait découvrir que des données de détection à distance (images de satellites) croisées avec d'autres types de données font ressortir des corrélations intéressantes

à propos de risques associés à des propriétés dans une certaine zone géographique. De cette manière, la prime calculée pour l'assurance d'une propriété reflèterait mieux les risques réels (MGI, 2011).

3.4. Identification et prévention des risques de fraude

En combinant les données de déclarations de sinistre avec les données GRC (Gestion de la Relation Client), les données provenant des réseaux sociaux et les données de géolocalisation (grâce aux appareils mobiles et aux capteurs), les Big Data analytics deviennent une arme redoutable pour combattre les déclarations de sinistre frauduleuses.

Par exemple, les informations des réseaux sociaux permettraient de voir si l'assuré connaît la personne avec qui un accident de voiture a été commis. De plus, les données de géolocalisation, transmises par le téléphone mobile ou un capteur dans la voiture, pourraient même confirmer ou infirmer le lieu de l'accident indiqué par l'assuré (Boyer, 2015 ; Gartner, 2015b ; IBM, 2012 ; MGI, 2011).

4. Les perspectives d'avenir

Nous venons d'explorer les raisons de l'impact du Big Data dans les secteurs de la banque et de l'assurance, et de voir comment les Big Data analytics transforment les business modèles dans ces secteurs. Parmi toutes les possibilités d'usage du Big Data qui ont été citées, certaines sont déjà très répandues au sein des entreprises, d'autres sont en voie de développement. L'avenir nous réserve encore sans doute de nombreuses surprises et le Big Data n'a probablement pas fini de défier notre imagination. Mais qu'est-ce que les scientifiques des technologies présagent encore pour les prochaines années dans ces secteurs ?

Dans le secteur de l'assurance, les capteurs biométriques pour prédire les risques de pathologies feront leur entrée tôt ou tard, même si la législation en vigueur devra être adaptée en conséquence. Ce sont des montres et bracelets « intelligents », des lentilles de contact électroniques, des capteurs intracomprimés ou encore des séquenceurs ADN de poche (Boyer, 2015).

Dans le secteur bancaire, les particuliers n'auront bientôt plus besoin de leur banquier pour contracter un crédit. Selon Benoît Legrand (ancien président d'ING France et actuellement responsable des FinTechs chez ING), « *le Big Data et le crédit en ligne constituent, après la banque en ligne à la fin des années 1990, une deuxième rupture technologique sur laquelle nous misons pour rebattre les cartes dans le secteur bancaire* ». Cette citation provient d'un

article de Jacques Henno, paru dans Les Echos en mars 2016, intitulé « *Comment le Big Data va bousculer le crédit* ». Le journaliste y décrit une nouvelle tendance qui arrive des Etats-Unis et d'Asie. Celle-ci consiste à « *utiliser les données publiques, les réseaux sociaux et les algorithmes prédictifs pour évaluer, en quelques minutes, les capacités de remboursement des individus ou des petites entreprises* ».

Sommes-nous bientôt à l'ère du crédit sans banquier (Henno, 2016) ? Il en est fort probable oui, et il n'y a pas que l'activité d'octroi de crédit qui est concernée. Selon Faljaoui (2016a), le conseil en placement aussi, c'est pourquoi « *la révolution numérique n'épargnera même pas les banquiers... ce n'est qu'une question de temps* ». Il semblerait que le métier tout entier du banquier soit menacé.

Jusqu'à présent, l'impact du Big Data a été analysé dans le secteur de la banque d'une part, et dans celui de l'assurance d'autre part. Cependant, la banque et l'assurance sont des activités qui peuvent se retrouver au sein d'une même entité. Dès lors, le Big Data offre encore de nouvelles opportunités aux « bancassureurs ». Nous avons vu que la force majeure des Big Data analytics est de mettre en relation des données qui n'ont a priori aucun lien entre elles, et d'en déduire des corrélations afin de mieux comprendre les comportements.

Dans cette optique, la force des bancassureurs va être de mettre en relation des données bancaires avec des données d'assurances. Par exemple, il est probable qu'un jour un bancassureur établisse, pour un jeune conducteur, une corrélation entre un retrait d'argent par carte bancaire un samedi soir (donnée bancaire) et sa couverture d'assurance auto (donnée d'assurance), qui traduirait un risque potentiel d'accident de voiture (Boyer, 2015).

Dans le thème de la bancassurance et du Big Data, nous en profitons pour parler de Belfius qui, les 28 et 29 avril 2016, à Londres, dans les locaux du Level39 Technology Accelerator (Incubateur technologique), a dévoilé ses « *ambitions 2020* ». Lors de cette présentation, Marc Raisière, CEO de Belfius, a exposé les priorités d'investissements du bancassureur belge pour les années à venir : les « Smart Cities », le « Mobile Banking » et les « FinTechs », en collaboration avec Accenture (Trends-Tendances, 2016).

Nous découvrons ici trois concepts-clés, résultats de la révolution du Big Data dans le monde du business et de la gouvernance. Les « FinTechs » se réfèrent aux technologies financières ou aux entreprises qui investissent dans ce domaine, le « Mobile Banking » est directement lié à notion de « banque digitale » que nous avons abordée plus haut, et les « Smart Cities » sont les « villes intelligentes » dont nous parlerons brièvement dans la troisième partie. Alors que

les « FinTechs » constituent une concrétisation du Big Data dans le secteur de la finance, les « Smart Cities » seront un exemple de concrétisation du Big Data dans le domaine de la gouvernance publique.

5. Les risques du Big Data dans le secteur financier

Face à tout cela, Boyer pose une question intéressante : « *Sommes-nous dans la prévention (comme les détecteurs de fumée) ou dans la surveillance (avec le risque d'une société "Big brother") ?* » (2015, p. 85). En effet, à côté de toutes ces opportunités de création de valeur pour les entreprises et d'amélioration de services pour les consommateurs, il reste malgré tout une question majeure : l'usage des données personnelles.

« Ces nouvelles approches et outils nous posent, collectivement, des questions dépassant le cadre d'un simple usage par une entreprise. Lorsqu'un nouvel acteur financier [ne] proposera des prêts ou des assurances qu'aux contreparties avec les meilleurs risques, aidé en cela par l'analyse de données, il remettra en cause le principe de la mutualisation du risque. La question transcende la banque, l'assureur, elle touche à nos choix et modèles de sociétés » (Metge, 2015, p. 100).

Les risques de l'usage du Big Data en général, soulevés par Mayer-Schönberger et Cukier (2013), que nous avons abordés dans la première partie, sont transposables aux secteurs de la banque et de l'assurance. Le premier risque concerne le respect de la vie privée, et un des challenges du Big Data est de « *gérer le paradoxe en termes de protection et d'exposition de la vie privée : on va beaucoup plus sur Internet, mais on souhaite plus de confidentialité de ses données* » (Boyer, 2015, p. 85).

Cette question du respect de la vie privée va avoir un impact important sur la législation, qui d'une manière ou d'une autre, va devoir progressivement adapter les lois sur la vie privée au fur et à mesure que le Big Data se répand dans le monde de l'entreprise. Toutefois, le pouvoir législatif aura aussi un rôle important à jouer lorsqu'il s'agira d'encadrer la révolution du Big Data et d'en limiter les excès (Boyer, 2015 ; Institut des Actuairens en Belgique, 2015 ; Metge, 2015).

A côté de cette question de respect de la vie privée se pose une autre question : comment les banques et compagnies d'assurances peuvent-elles assurer la sécurité et la confidentialité de toutes ces données privées ? En effet, si un client accepte de donner accès à certaines de ses données personnelles, il s'attend évidemment à ce que ses données soient protégées et que

personne d'autre ne puisse y avoir accès que son banquier ou son assureur. A ce niveau-là, il existe un risque non négligeable de cyber-sécurité. Ce risque constitue un défi primordial dans le chef de l'entité qui détient les données, dans notre cas la banque ou la compagnie d'assurances (Boyer, 2015 ; Metge, 2015 ; voir aussi Brown et al., 2011).

Le deuxième risque évoqué par Mayer-Schönberger et Cukier (2013) est le fait de punir ou juger sur base de ce que « disent » les données, avant même que les individus n'aient agi. Dans les secteurs de la banque et de l'assurance, ce risque est très clair. Imaginons le cas extrême d'un citoyen auquel les Big Data analytics associeraient un profil tellement risqué que plus aucune banque ne voudrait lui octroyer un crédit et que plus aucune compagnie d'assurances ne voudrait l'assurer. Que pourrait-il faire ? Est-il possible d'en arriver à un stade où un individu serait rejeté de la société à cause des données et de ce qu'elles « disent » de lui (Boyer, 2015 ; Metge, 2015 ; voir aussi Henno, 2016) ?

Enfin, le troisième risque développé par Mayer-Schönberger et Cukier est le risque de la « *dictature des données* » (2013, p. 151, traduction personnelle de l'anglais). Pour les secteurs en question, un autre risque majeur pour la banque et la compagnie d'assurances serait de considérer uniquement les Big Data analytics et les informations apportées par les données, sans tenir compte de leurs limites.

Le Big Data comporte donc de nombreux avantages, autant pour l'individu que pour l'entreprise, mais il implique aussi des risques importants qui ne peuvent être ignorés.

6. Les grandes tendances

Après cette analyse des avantages, des opportunités et des risques de l'usage du Big Data dans les secteurs de la banque et de l'assurance, nous voulons dresser une synthèse des principales tendances remarquées. Cela va nous permettre de faire la transition entre cette partie et la suivante, qui aborde l'impact du Big Data dans le monde des gouvernements.

6.1. La personnalisation

Le Big Data est un ensemble de données volumineuses, variées et collectées presque en temps réel. Globalement, les informations qui sont mises en relation sont des données traditionnelles, des données provenant d'Internet et des données générées par des capteurs. L'analyse de ces données croisées (les Big Data analytics) permet à la banque et à la compagnie d'assurances de mieux connaître les besoins et les préférences de leurs clients et ainsi de leur fournir un service personnalisé et de meilleure qualité.

6.2. La prédictibilité

Lorsqu'une entité a à sa disposition un ensemble de données qui peut être qualifié de Big Data, au-delà d'utiliser ces données pour mieux connaître ses clients, elle peut aussi y appliquer des analyses prédictives. Le but de ces analyses est d'identifier dans les données, des corrélations, des interactions et des schémas récurrents afin de prédire des comportements ou des événements futurs. Une fois que la banque ou la compagnie d'assurances peut prédire, elle peut agir en conséquence :

- Si ces comportements ou événements futurs constituent des opportunités de revenus, alors l'organisation peut déployer les moyens nécessaires au bon moment.
- Si ces comportements ou événements futurs constituent des menaces de pertes de revenus (fraude) ou des menaces pour la sécurité (cyber-terrorisme), alors l'organisation peut mettre les moyens en place pour prévenir les menaces.

6.3. L'optimisation

L'analyse des données en interne permet une augmentation de l'efficacité des processus internes et une diminution des coûts. Cette optimisation se manifeste aussi dans les calculs toujours plus performants de prix qui permettent une rentabilité maximale.

6.4. L'objectivation

L'analyse des données a une autre force majeure : elle rend la prise de décision et le jugement objectifs. En effet, les Big Data analytics permettent à l'organisation une prise de décision et un jugement qui ne sont plus basés sur des intuitions, des suppositions, des règles traditionnelles ou des hypothèses raisonnables, mais plutôt sur des analyses de données qui sont objectives, puisque réalisées par des algorithmes et non plus par des humains.

A présent, nous allons transposer ces tendances au domaine de la gouvernance publique et explorer dans quelle mesure le Big Data impacte les gouvernements et la manière dont ils fonctionnent.

PARTIE 3 : Une Data Driven Governance

Dans la première partie, nous avons décrit le contexte de la révolution du Big Data dans notre société. Dans la deuxième partie, nous avons analysé l'impact du Big Data sur deux secteurs où les données ont une importance particulière : les secteurs de la banque et de l'assurance. Nous avons conclu cette deuxième partie en relevant quatre grandes tendances de l'usage du Big Data dans les secteurs étudiés : la personnalisation, la prédictibilité, l'optimisation et l'objectivation. Nous avons vu qu'au travers de ces quatre tendances, la banque et la compagnie d'assurances parviennent à créer de la valeur autant en interne qu'en externe, dans la relation avec leur client.

A présent, nous nous intéressons à l'influence du Big Data sur les gouvernements, sur l'Etat de manière générale. Comme les secteurs de la banque et de l'assurance, l'Etat est une organisation qui, de par son rôle de gouverner des citoyens, est amenée à manipuler de grandes quantités de données. En 2009, aux Etats-Unis, comme illustré dans l'annexe 6, le gouvernement était le deuxième secteur avec le plus de données stockées, après le secteur de la fabrication. Si l'Etat possède aussi des données en masse, alors il détient aussi un énorme potentiel de création de valeur (MGI, 2011).

Le but de cette troisième partie est d'étudier la possibilité d'une Data Driven Governance, ou d'une gouvernance « conduite » par les données. Nous désirons envisager un Etat qui utilise le Big Data à l'image de la banque et de la compagnie d'assurances, de manière à créer de la valeur autant en interne qu'en externe, pour le citoyen. Pour cela, nous nous posons deux questions :

- Premièrement, comment l'Etat peut-il utiliser le Big Data pour créer de la valeur ?
- Deuxièmement, quels sont les challenges et les risques auxquels l'Etat doit faire face s'il veut entrer dans la tendance Big Data ?

Puisque nous parlons d'Etat et de Big Data, il est évidemment intéressant de savoir où en est la Belgique par rapport aux autres pays en matière de stratégie numérique. Le département des affaires économiques et sociales de l'Organisation des Nations unies réalise tous les deux ans une enquête sur l'état des e-gouvernements à travers le monde (United Nations, 2014). L'enquête évalue les gouvernements sur base de trois aspects clés de l'administration

électronique : les services en ligne, les infrastructures de télécommunications et le capital humain.

Comme illustré en annexe 7, le résultat de l'enquête est notamment un tableau qui représente le classement des leaders mondiaux en matière d'e-gouvernement. Nous constatons qu'en 2014, la Belgique se classait à la dernière place de ce classement de 25 pays, dont le podium était occupé par la République de Corée, l'Australie et Singapour. Ce tableau montre que notre pays était surclassé par 15 pays européens, 5 pays asiatiques, l'Amérique du Nord et l'Océanie. La Belgique peut sans doute mieux faire, nous allons voir comment.

Dans la première section, nous explorons les pistes principales d'utilisation du Big Data dans plusieurs domaines publics tels que l'enseignement, la sécurité, la circulation routière et le respect de la loi. Nous pensons que l'Etat belge pourrait notamment s'inspirer de ces quelques pistes pour l'amélioration globale de ses services publics.

Avant de continuer, une remarque préalable doit être faite. Dans cette partie, nous parlons donc d'« Etat » et de « gouvernement ». Or, l'Etat belge est composé d'un Etat fédéral et d'entités fédérées, ce qui peut créer une confusion quant à l'entité visée par notre exposé. Dès lors, nous souhaitons préciser que, par « Etat » ou « gouvernement », nous entendons le pouvoir politique et les services publics au sens large du terme. Les différentes applications du Big Data qui vont être développées s'adressent donc, dans le cadre de la Belgique, autant à l'Etat fédéral qu'aux entités fédérées, en fonction des compétences qui leur sont assignées.

1. Comment l'Etat peut-il utiliser le Big Data ?

Afin de répondre à cette question, nous étudions comment le Big Data peut révolutionner la manière dont le gouvernement fonctionne à la lumière des quatre grandes tendances que nous avons découvertes dans la deuxième partie. En calquant ces tendances au domaine de la gouvernance publique, voici les quatre principales possibilités d'utilisation du Big Data :

- La personnalisation du service au citoyen ;
- La prédictibilité : prévenir la fraude et les menaces ;
- L'optimisation interne ;
- L'objectivation de la prise de décision et du jugement.

L'émergence de ces quatre tendances du Big Data, dans notre société et plus particulièrement dans le domaine de la gouvernance, nous a été confirmée par deux experts. Le premier provient du monde de l'entreprise et souhaite rester anonyme, nous le surnommerons donc

« Monsieur X ». Il occupe le poste de « Head of Customer Analytics » dans une grande banque belge, et nous avons eu l'occasion de l'interviewer par téléphone le 15 avril 2016. Le second expert est issu du milieu académique et s'appelle Jean-Marc Van Gyseghem, directeur de l'unité « Libertés et société de l'information » au Centre de Recherche Information, Droit et Société (CRIDS), basé à l'Université de Namur. Nous avons rencontré M. Van Gyseghem à Namur le 19 avril 2016.

À travers l'analyse de ces tendances, nous allons voir que, grâce au Big Data, l'Etat est en mesure de proposer au citoyen un nouveau modèle de gouvernance, un modèle de plus en plus objectif, « conduit » par les données.

1.1. La personnalisation du service au citoyen

Dans les secteurs de la banque et de l'assurance, nous avons vu que les Big Data analytics permettent à l'organisation d'améliorer la satisfaction du client en lui offrant un service toujours plus personnalisé. De plus, ils permettent aussi d'augmenter les ventes en proposant au client des services et produits qui répondent tellement bien à ses besoins qu'il peut difficilement refuser. A présent, nous transposons cette dimension de personnalisation des services au domaine du gouvernement : de quelle manière l'Etat peut-il créer de la valeur en offrant un service personnalisé à son citoyen ?

1.1.1. De la standardisation à la personnalisation

La segmentation du marché pour personnaliser les actions marketing est une tendance qui a déjà été acceptée depuis longtemps dans le secteur privé. Le secteur public, quant à lui, a fonctionné pendant trop longtemps avec un système standardisé de « même service pour tout le monde ». Il faisait l'hypothèse que les économies d'échelles rendraient le système moins coûteux qu'un programme de services publics personnalisés. Cependant, il a été démontré que la segmentation de la population et la personnalisation des services publics augmentent l'efficacité des services, réduisent les dépenses qui y sont liées et augmentent la satisfaction du citoyen.

Un service public personnalisé, c'est « *un service public unique dans ses finalités, mais différencié dans ses modalités en fonction des besoins de chaque usager* » (Grosdhomme Lulin, 2015, p. 19). Cette différenciation est rendue possible lorsque l'Etat, qui possède des données en masse sur ses citoyens, donne du sens à ces données en y appliquant des techniques d'analyse avancées qui permettent des traitements sophistiqués et individualisés.

Dans le chef du gouvernement, il y a donc bien une réelle opportunité de création de valeur au niveau organisationnel, mais cette personnalisation des services publics profite surtout au citoyen, qui ressort avec un service de meilleure qualité et plus approprié à ses besoins (Grosdhomme Lulin, 2015 ; MGI, 2011 ; Schenk & Berman, 2015 ; voir aussi Centre d'analyse stratégique de la République française, 2013 ; Chui, Farrell, & Jackson, s.d. ; Desouza, 2014 ; Hamel & Marguerit, 2013 ; Yiu, 2012).

Ici, nous pouvons noter une première différence par rapport au secteur financier, étudié dans la deuxième partie. Pour ce secteur, la personnalisation des offres de services permet, au-delà d'une meilleure satisfaction des clients, une augmentation des ventes pour l'entreprise. C'est donc un « win-win » : tout le monde y gagne. En revanche, dans le domaine de la gouvernance, le but de l'Etat n'est pas de vendre des services, mais de remplir son devoir de services publics envers le citoyen. Dès lors, c'est surtout le citoyen qui va bénéficier de cette personnalisation. C'est toujours un « win-win », mais le citoyen a sans doute plus à y gagner que l'Etat.

1.1.2. Un exemple dans l'enseignement

Nous illustrons comment la personnalisation du service au citoyen peut se concrétiser en prenant l'exemple de l'enseignement. Aujourd'hui, la Belgique se caractérise par un enseignement principalement standardisé : un professeur donne le même cours, les mêmes exercices, et évalue selon les mêmes critères tous les élèves de sa classe, alors que chacun d'eux possède un niveau différent. C'est le principe de la diversité.

En s'inspirant des premiers progrès de la stratégie numérique de l'enseignement français, nous avons un aperçu de ce qui nous attend à l'avenir avec la révolution du Big Data. La première étape consiste, à l'aide des technologies informatiques les plus basiques (interaction, expérimentation, essai-erreur, outils de « e-learning »), à proposer à chaque élève un parcours d'apprentissage individualisé. Ce parcours permettrait à l'élève de travailler les matières où il a plus de difficulté et de passer moins de temps sur les matières qu'il maîtrise.

La seconde étape, c'est le professeur qui aurait accès à un « journal de travail numérique », qui lui permettrait d'avoir une meilleure vision du niveau de ses élèves, autant au niveau individuel qu'au niveau de la classe. Le professeur pourrait voir le travail effectivement accompli par ses élèves, repérer les endroits où ils ont eu plus de mal et ainsi les évaluer de manière plus pertinente.

Pour la troisième étape, nous sommes au niveau national, dans le cadre d'étudiants qui arrivent dans l'enseignement supérieur. Avec les outils informatiques nécessaires, le ministère de l'Education nationale pourrait collecter et rassembler toutes sortes de données relatives à l'éducation et développer des algorithmes d'aide à la décision. Ainsi, à l'image de la banque et de la compagnie d'assurances qui peuvent faire des suggestions à leurs clients en fonction de leurs achats précédents de services, l'Etat pourrait recommander à un étudiant tel ou tel cursus en fonction de son parcours et de ses points forts.

Cela pourrait mener assez rapidement à la création d'un portail numérique qui donnerait accès à de nombreuses données utiles pour les étudiants, comme par exemple des statistiques de réussite dans une telle formation en fonction des formations déjà suivies. Si on va encore plus loin, ce portail numérique pourrait être mis en relation avec les réseaux sociaux, dans le but de créer toujours plus d'interactions entre les étudiants du pays, mais aussi entre les étudiants et les établissements d'enseignement supérieur. Grâce à des techniques de data mining (exploration de grandes quantités de données variées à très grande vitesse pour détecter des corrélations) appliquées sur ce Big Data naissant, ces établissements seraient même capables de faire des offres de cursus à des profils précis d'étudiant.

Nous voyons que mettre en œuvre une stratégie Big Data peut permettre à l'Etat la création d'une véritable valeur ajoutée au niveau de la personnalisation du service rendu au citoyen, dans ce cas-ci d'un jeune en quête de formation pour se préparer au marché du travail. Au-delà de l'enseignement, l'Etat pourrait aussi, grâce à l'utilisation du Big Data, apporter au citoyen un service personnalisé à haute valeur ajoutée dans les domaines :

- de l'emploi, avec un meilleur accompagnement des demandeurs d'emploi (une étude de cas sur le sujet se trouve en annexe 8) ;
- de la justice, avec l'individualisation des peines ;
- des transports publics, notamment par l'agrégation dynamique des offres de différents types de moyen de transport.

Ainsi, cette nouvelle philosophie du Big Data permettrait à l'Etat de mieux répondre aux besoins de ses citoyens, en faisant usage des données en masse mais aussi en les rendant disponibles au grand public (Grosdhomme Lulin, 2015 ; voir aussi TechAmerica Foundation's Federal Big Data Commission, s.d.). En d'autres termes, l'Etat peut utiliser les technologies du Big Data pour offrir à son citoyen un service public plus « *smart* » (Damon, Messatfa, & Chabut, 2011).

1.2. La prédictibilité : prévenir la fraude et les menaces

Nous retrouvons ici la deuxième tendance qui découle de l'utilisation des Big Data analytics dans les secteurs de la banque et de l'assurance : la prédictibilité. Comme nous l'avons vu, les analyses prédictives de données deviennent de plus en plus tendance dans ces secteurs, ce qui permet à la banque et à la compagnie d'assurances de créer de la valeur en prédisant plus efficacement les risques et les comportements. Pour l'Etat, cette prédictibilité des risques et des comportements offre aussi des opportunités uniques de création de valeur, qui se traduisent par une identification et une prévention plus efficaces de la fraude et des menaces. Ces dernières concernent principalement la sécurité publique (criminalité et menaces terroristes), mais pas seulement.

Pour prévenir la fraude et améliorer la sécurité publique, la stratégie Big Data de l'Etat consiste d'abord à collecter en continu et à rassembler toutes sortes de données comportementales qui décrivent la réalité de la vie de la population. Ensuite, à l'aide d'algorithmes toujours plus performants, les Big Data analytics, et en particulier les techniques de data mining, permettent une identification plus efficace des anomalies, des erreurs et des cas de fraude (Desouza, 2014 ; Grosdhomme Lulin, 2015 ; Kim, Trimi, & Chung, 2014 ; MGI, 2011).

Aujourd'hui, les Big Data analytics et les analyses prédictives se « *rapprochent peu à peu de l'intelligence artificielle et permettent de passer du rétrospectif au prospectif, du constat à l'anticipation, de l'analyse à l'action* » (Grosdhomme Lulin, 2015, p. 5 ; voir aussi Microsoft, 2016 ; Yiu, 2012).

Ainsi, l'analyse des données permet deux opportunités de création de valeur :

- Une création de valeur d'ordre financier, à partir du moment où les Big Data analytics deviennent des instruments puissants de lutte contre la fraude fiscale et sociale ;
- Une création de valeur qui se traduit par l'amélioration de la sécurité publique, en ceci que les Big Data analytics permettent au gouvernement de faire baisser la criminalité et d'être plus efficace dans l'identification et la prévention des menaces terroristes.

1.2.1. Prévention de la fraude

Zikopoulos et al. (2012) avancent qu'en général, là où il y a une sorte de transaction financière, il y a aussi forcément un risque d'abus ou de fraude. Dans ce contexte, ils affirment que, si on exploite le Big Data, on a clairement l'opportunité d'identifier la fraude

et, mieux encore, de l'empêcher. Pour les gouvernements et leurs administrations fiscale et sociale, c'est une possibilité d'utilisation du Big Data à forte valeur ajoutée.

Dans la lutte contre la fraude fiscale, mettre en place des algorithmes qui vont être capables d'analyser rapidement de grandes quantités de données variées (techniques de data mining) peut permettre au gouvernement d'identifier des corrélations et des activités suspectes. Par exemple, l'Etat pourrait :

- relever qu'une transaction spécifique a eu lieu au même moment qu'un changement d'adresse ;
- découvrir qu'un citoyen est actif dans deux pays à la fois et effectue régulièrement des schémas de transactions qui s'apparentent à de la fraude ;
- détecter un réseau de fraude et en identifier les acteurs principaux.

Dans ce cadre de lutte contre la fraude fiscale, le gouvernement belge utilise déjà ces techniques de data mining.

Ces outils d'analyses de données en masse sont aussi utilisés pour combattre la fraude sociale. Par exemple, à l'aide du Big Data, un Etat pourrait détecter qu'une personne perçoit des allocations de chômage auprès d'une institution et en même temps rentre un dossier d'accident professionnel auprès d'un autre organisme. Par ailleurs, aux Etats-Unis, un système de versement des allocations sociales via des cartes de paiement électroniques prépayées, permet au gouvernement de contrôler l'usage qui est fait de ces allocations. Concrètement, cette carte de paiement peut être configurée dans le but d'interdire de dépenser ces montants d'allocations dans certains types de commerce, notamment les commerces de vente d'alcool ou les salles de jeux.

Ces quelques exemples montrent que, lorsqu'ils sont utilisés par l'Etat, les Big Data analytics, et en particulier les techniques de data mining, sont devenus des armes redoutables de lutte contre la fraude fiscale mais aussi sociale.

Au niveau fiscal, de cette meilleure prévention de la fraude découle évidemment pour l'Etat une augmentation des recettes. En effet, les Big Data analytics engendrent une collecte plus efficace des taxes et par conséquent une augmentation, ou en tout cas une récupération, des recettes fiscales. Le Big Data constitue donc un moyen puissant pour les administrations fiscales de réduire la différence entre les recettes fiscales potentielles et effectives. Les administrations qui performant le mieux utilisent le Big Data pour segmenter les

contribuables afin de mieux organiser le prélèvement des impôts. Pour ce faire, celles-ci rassemblent des données géographiques, financières, démographiques, historiques, de fidélité au paiement de l'impôt, de risque de défaut et de difficulté de prélèvement des contribuables. Elles catégorisent ensuite ces derniers afin de mieux appréhender les risques et maximiser les recettes fiscales.

Concrètement, aux Etats-Unis, grâce à l'utilisation des Big Data analytics :

- en deux ans, le gouvernement de l'Etat de Géorgie a empêché des cas de fraude fiscales pour plus de 30 millions de dollars ;
- le gouvernement de l'Etat de Floride a réussi à augmenter le prélèvement de ses recettes fiscales de 66 millions de dollars au cours de l'année d'imposition 2013.

Ainsi, un gouvernement qui déploie une stratégie Big Data se crée une réelle opportunité de création de valeur au niveau de ses finances (Grosdhomme Lulin, 2015 ; MGI, 2011 ; Schenk & Berman, 2015 ; voir aussi Desouza, 2014 ; Hamel & Marguerit, 2013 ; Kim et al., 2014 ; TechAmerica Foundation's Federal Big Data Commission, s.d.).

1.2.2. Baisse de la criminalité

A coté de l'amélioration des finances publiques par une meilleure prévention de la fraude, les Big Data analytics offre aussi une opportunité de création de valeur au niveau la sécurité publique. Dans ce domaine, le véritable enjeu se trouve dans le fait de prédire la criminalité afin de « prévenir plutôt que guérir ». Puisque nous parlons de « prédire », nous avons besoin des analyses prédictives du Big Data.

Dans ce domaine, Grosdhomme Lulin parle de la « *police prédictive* » (2015, p. 27), soit une police qui croise des données sur les circonstances et les modalités de faits de criminalité passés avec des données actuelles de géolocalisation collectées en temps réel. Ainsi, les forces de l'ordre deviennent capables de :

- prédire le lieu et le moment des futurs crimes ;
- prédire et identifier quels individus vont probablement commettre un crime ;
- constituer des profils types de criminels en fonction des crimes passés ;
- prédire même les victimes des futurs crimes.

Une fois que la « *police prédictive* » a établi ces prédictions d'occurrence de faits criminels, elle peut anticiper et mettre les moyens en œuvre pour « prévenir la criminalité plutôt que la guérir ». Concrètement, elle pourra par exemple être amenée à prépositionner des patrouilles

de police à des endroits stratégiques où les risques de criminalité sont plus élevés (Grosdhomme Lulin, 2015 ; Schenk & Berman, 2015 ; voir aussi Desouza, 2014 ; Neyfakh, 2015).

1.2.3. Détection des menaces terroristes

En ce début de XXI^e siècle, notre société contemporaine est marquée par une guerre qui est une première dans l'histoire de l'Humanité : la guerre contre le terrorisme. Les premiers éclats de cette guerre ont eu lieu le 11 septembre 2001 et les récents attentats de Paris le 13 novembre 2015 et de Bruxelles le 22 mars 2016 ont montré que, malheureusement, cette guerre est encore loin d'être terminée. Si le terrorisme est un ennemi difficile à combattre et à vaincre, c'est sans aucun doute parce qu'il se cache parmi les innocents. Il est donc clair que la sécurité est devenue un sujet de préoccupation majeur, pour les Etats mais également pour toutes les organisations en général. En effet, tout le monde est concerné. De plus, comme nous l'avons déjà abordé, le cyber-terrorisme existe aussi.

Dans ce contexte hostile, de même que pour la criminalité, les Big Data analytics constituent pour l'Etat une magnifique opportunité d'amélioration de la sécurité publique. C'est la raison pour laquelle les agences de renseignement et de sécurité publique se mobilisent pour rassembler de grandes quantités de données provenant de sources multiples afin d'identifier les potentielles menaces terroristes et de cyber-sécurité. C'est dans ce cadre que les Big Data analytics peuvent jouer un rôle primordial d'identification et de prévention des risques et des menaces pour améliorer la sécurité publique. En outre, des systèmes sophistiqués d'algorithmes automatisés sont maintenant capables d'analyser de larges ensembles de données afin de détecter plus efficacement et contrer plus rapidement les menaces terroristes.

Concrètement, nous pouvons prendre l'exemple du Passenger Name Records (PNR), qui a fait parler de lui après les récents attentats en France et en Belgique. Le PNR, ce sont les données sur les passagers qui sont détenues par les compagnies aériennes. Si les forces de l'ordre ont accès à ces données, elles peuvent alors les croiser avec leurs propres données de renseignement. Ainsi, des Big Data analytics appliqués à cette masse d'informations croisées permettraient de détecter les comportements suspects au niveau des transports aériens (Chen et al., 2012 ; Grosdhomme Lulin, 2015 ; MGI, 2011 ; voir aussi Kim et al., 2014 ; TechAmerica Foundation's Federal Big Data Commission, s.d.).

En découvrant que le Big Data pourrait améliorer drastiquement la sécurité publique, une question délicate peut aisément nous venir à l'esprit : les événements tragiques du 22 mars

2016 à Bruxelles auraient-ils pu être évités si l'Etat belge avait déployé au préalable une vraie stratégie Big Data ?

Enfin, comme nous l'avons dit en introduction de cette section sur la prédictibilité, les menaces qui pèsent sur la société concernent la sécurité publique, mais pas seulement. Nous avons parlé de la criminalité et des menaces terroristes, mais les maladies et les épidémies à l'échelle nationale et internationale peuvent aussi constituer des menaces pour la société. Il n'est donc plus question de sécurité publique mais de santé publique. Tout comme le Big Data et les analyses prédictives peuvent apporter des solutions concrètes pour améliorer la sécurité publique, ils le peuvent aussi pour le domaine de la santé publique. En annexe 9 se trouve une étude de cas sur le sujet.

1.3. L'optimisation interne

L'optimisation est la troisième tendance que nous avons relevée quant à l'impact du Big Data dans les secteurs de la banque et de l'assurance. Nous avons vu que l'analyse des données en interne permet une augmentation de l'efficacité des processus internes et une diminution des coûts. En étudiant l'impact du Big Data sur l'Etat à la lumière de cette tendance, nous distinguons deux types d'optimisation interne.

Le premier type est l'optimisation des processus internes au sein des administrations gouvernementales qui, par le déploiement de technologies Big Data, crée de la valeur, notamment par une baisse des coûts. Le second type est l'optimisation du fonctionnement du pays, qui crée de la valeur pour tous ses usagers (citoyens et étrangers de passage).

1.3.1. Optimisation des processus internes

De manière générale, développer des moyens qui permettent de collecter, manipuler et analyser le Big Data requiert une optimisation des outils et des processus technologiques. Cette optimisation technologique est évidemment indispensable pour une organisation qui souhaite utiliser le Big Data pour créer de la valeur, que ce soit à travers la tendance de personnalisation ou de prédictibilité, que nous venons d'étudier. Dans cette optique, l'optimisation des processus internes a un impact principalement externe : le citoyen bénéficie d'un service personnalisé et le pays est plus sûr. Cependant, cette optimisation peut aussi avoir un impact interne et permettre à l'Etat de diminuer les coûts à travers ses différentes administrations. C'est ce que nous voyons dans cette section.

En 2011, le McKinsey Global Institute estimait qu'en utilisant le Big Data, l'administration gouvernementale des économies développées d'Europe pouvait économiser plus de 100

milliards d'euros, juste par l'amélioration de l'efficacité opérationnelle. En outre, le géant de la consultance technologique avance que l'augmentation de la transparence des données et l'application des Big Data analytics peut engendrer une hausse de la productivité et de l'efficacité interne du gouvernement, notamment en réduisant les coûts des activités administratives jusqu'à 20 %.

Cette optimisation générale de la productivité grâce au Big Data se manifeste donc, entre autres, par une baisse des coûts au sein des administrations. Cette baisse est due principalement à la diminution des erreurs dans les procédures administratives et à l'accélération de celles-ci grâce à des échanges de données plus efficaces et plus rapides (réduction des temps de recherche des données) (MGI, 2011 ; Schenk & Berman, 2015 ; voir aussi Centre d'analyse stratégique de la République française, 2013 ; Desouza, 2014 ; Hamel & Marguerit, 2013 ; Yiu, 2012).

Pour donner un exemple concret d'optimisation des processus internes, nous rapportons un projet qui nous a été présenté par M. Van Gyseghem. Lors de notre rencontre, il nous a fait part du « Décret source authentique », un projet qui vise à utiliser les nouvelles technologies pour aider le gouvernement wallon à simplifier son administration en général. C'est un projet de collaboration entre la Fédération Wallonie-Bruxelles et la Région wallonne qui a pour but d'optimiser les échanges de données entre les différentes administrations.

Concrètement, chaque administration possède des données et l'objectif est multiple : les mettre en commun, veiller à leur fiabilité, garantir leur sécurité et enfin s'assurer que les administrations aient accès uniquement aux données dont elles ont besoin et pas à d'autres (pour éviter les abus). Une fois concrétisé, ce projet apportera une réelle valeur ajoutée en simplifiant de nombreuses tâches administratives au sein des administrations et en réduisant les erreurs (gain en temps et en efficacité).

Le citoyen, lui, y gagnera aussi, puisque de nombreuses tâches lui seront simplifiées, notamment grâce à un portail numérique sur le site de la Région wallonne, qui permettra par exemple qu'un formulaire de demande de subside soit pré-rempli via une interface du type « Mon dossier » (Van Gyseghem, 2016 ; voir aussi Desouza, 2014 ; Hamel & Marguerit, 2013 ; Kim et al., 2014 ; MGI, 2011 ; Yiu, 2012). Cette dernière application rejoint la tendance de personnalisation des services publics, abordée précédemment.

A grande échelle, cela donne un système intra-connecté de données qui sont à la fois volumineuses (toute la population du pays est concernée), collectées et échangées toujours

plus rapidement entre les différents acteurs, et variées (mélange d'informations humaines, légales, financières, etc.). Cette transformation de l'administration publique s'inscrit donc bien dans la tendance Big Data.

1.3.2. Optimisation du fonctionnement du pays

En prenant de la hauteur, nous réalisons que le pays peut aussi, grâce à l'utilisation des Big Data analytics, être optimisé au niveau de son fonctionnement global, à l'image de l'administration publique. C'est ce que pense Monsieur X, selon qui les Big Data analytics pourraient permettre un basculement vers une société plus harmonieuse, où la vie pour le citoyen y serait plus confortable. Dans cette optique, il prend l'exemple de la circulation sur les routes belges.

A ce sujet, il est bien connu que la Belgique, étant un véritable carrefour au milieu de l'Europe et ayant sur son territoire la capitale de l'Union européenne, est un pays où le trafic routier est un fléau. Selon une étude de l'INRIX (2015), le belge a passé, en 2015, en moyenne 44 heures de son temps dans les embouteillages, ce qui fait de la Belgique le deuxième pays le plus congestionné du monde, après les Etats-Unis. Par ailleurs, l'étude a aussi établi un classement des villes européennes les plus congestionnées, et le top 5 comprend deux villes belges : Anvers en troisième position et Bruxelles en cinquième. Ces chiffres ne mettent évidemment pas la Belgique à l'honneur, mais là où il y a une faiblesse, il y a forcément une possibilité d'amélioration.

Dans ce sens, Monsieur X affirme que les Big Data analytics constituent pour l'Etat une réelle opportunité de création de valeur au niveau du trafic routier. En effet, la congestion a de multiples impacts négatifs. Ces longues heures perdues sur la route affectent le bien-être du citoyen, contribuent à la pollution atmosphérique et coûtent chaque année des milliards d'euros à l'économie toute entière (Commission européenne, 2015 ; Organisation de Coopération et de Développement Economiques [OCDE] Belgique, 2015). Si l'utilisation des technologies Big Data permet de diminuer la congestion nationale, il en résultera une création de valeur à la fois en faveur du bien-être du citoyen, de la préservation de l'environnement et de la compétitivité économique du pays. La question à présent est : comment ?

Aujourd'hui, les données de géolocalisation, collectées via les appareils portables des citoyens ou des capteurs dans les voitures, offrent une occasion unique pour le gouvernement de mieux gérer son réseau routier et donc d'y diminuer considérablement les embouteillages. Monsieur X avance que des techniques de Big Data analytics, en mettant en relation les

données de géolocalisation avec les systèmes de régulation du trafic (feux rouges), permettraient aux autorités publiques de rendre la circulation plus fluide dans les villes congestionnées.

Optimod est un exemple de projet de prévention de la congestion urbaine, développé par la ville de Lyon, en France. De même, les chercheurs de l'Institut Royal de Technologie de Suède ont déployé un logiciel de « streams analytics » (littéralement, « analyses de flux ») pour fluidifier le trafic dans les métropoles et réduire le nombre d'accidents dus à la congestion (Grosdhomme Lulin, 2015 ; TechAmerica Foundation's Federal Big Data Commission, s.d.).

Cette optimisation de la circulation dans les villes fait notamment partie des objectifs du projet « *Data driven: Connecting cars to smarter cities* », déployé par l'INRIX. La vision de cette organisation est de profiter de la tendance du Big Data et de l'Internet des Objets pour optimiser le fonctionnement global des villes en les rendant plus « intelligentes » (INRIX, s.d. ; voir aussi Desouza, 2014).

1.4. L'objectivation de la prise de décision et du jugement

La quatrième tendance que nous avons relevée pour les secteurs de la banque et de l'assurance est l'objectivation. Le Big Data permet en effet une prise de décision, un jugement et une évaluation qui sont basées sur des analyses de données objectives plutôt que sur des intuitions et des raisonnements humains. Dans le cadre de l'Etat, les analyses de données en masse permettent « *la possibilité de prendre des décisions publiques de manière automatique, sans intervention humaine, avec effet immédiat, sur la base des données ainsi collectées et traitées* » (Grosdhomme Lulin, 2015, p. 6). Avec l'expansion du Big Data à travers les différents départements du gouvernement, c'est un nouveau modèle de gouvernance qui émerge : un modèle où la prise de décision devient de plus en plus objective.

Aujourd'hui, il devient possible pour le gouvernement de remplacer ou de soutenir la prise de décision humaine avec des algorithmes automatisés. Comme pour le secteur financier, le but est de baser le processus décisionnel sur des données objectives plutôt que sur des intuitions et sur un jugement humain. Ainsi, les Big Data analytics permettent au gouvernement de prendre de meilleures décisions et d'apporter au pays des solutions plus pertinentes face aux problèmes sociétaux existants (Chui et al, s.d. ; Grosdhomme Lulin, 2015 ; MGI, 2011 ; Schenk & Berman, 2015 ; Yiu, 2012).

Nous abordons cette opportunité d'objectivation par l'analyse des données sous deux angles. D'une part, l'objectivation et les meilleures décisions qui en découlent peuvent être partagées par le gouvernement en augmentant la transparence des données : c'est l'Open Data. D'autre part, dans le cadre du respect de la loi, l'objectivation engendrée par le Big Data peut permettre une sorte de « respect automatique » en programmant les objets dans ce sens.

1.4.1. Transparence des données et Open Data

Un autre bénéfice de l'utilisation des Big Data analytics au sein du gouvernement est l'augmentation de la transparence des données (Schenk & Berman, 2015). Augmenter la transparence des données, c'est donner l'accès à l'Open Data. En effet, comme nous l'avons vu dans la première partie, l'Open Data se réfère aux données que l'Etat met à la disposition du grand public.

L'Open Data est vu comme un potentiel important de valeur ajoutée pour le domaine public. A ce sujet, Zuiderwijk et Janssen écrivent que « *les gouvernements et politiciens à travers le monde entier ont déclaré soutenir la publication des données gouvernementales "ouvertes" dans le but d'accomplir une grande variété de valeurs publiques* » (2015, p. 1, traduction personnelle de l'anglais). Selon eux, des valeurs publiques peuvent se traduire par de meilleurs services publics, une relation de confiance améliorée entre l'Etat et son citoyen ainsi que l'émergence d'un capital social (en sens sociologique du terme : « *ensemble de ressources actuelles ou potentielles qui sont liées à la possession d'un réseau durable de relations plus ou moins institutionnalisées d'interconnaissance et d'inter-reconnaissance* » (Bourdieu, 1980)).

L'Agence wallonne des Télécommunications (aujourd'hui rebaptisée Agence du Numérique), citée par InnovaTech (s.d., p. 3), confirme cette tendance et déclare que « *la mise à disposition gratuite, mais sous licence, de ces données publiques dans des formats adaptés, rend possible la création de nouvelles applications permettant de conférer une réelle valeur ajoutée à ces données (applications composites, tableaux de bord et infovisualisations, transports publics, géolocalisations, cartographies, applications d'intérêt public, etc.). Les entreprises privées et autres porteurs de projets numériques trouveraient là le moyen d'opérationnaliser rapidement leurs idées innovantes* ».

En effet, avec l'accès aux larges bases de données du secteur public, ce sont tous les acteurs de l'économie, autant les individus que les organisations, qui peuvent augmenter l'efficacité de leurs activités. Ainsi, tous ces acteurs deviennent plus performants économiquement, ce

qui est évidemment favorable à la croissance économique du pays (Chui et al, s.d. ; MGI, 2011 ; Yiu, 2012).

Rendre les données à disposition de l'Etat plus transparentes afin que les citoyens et les entreprises puissent en profiter, c'est notamment la vision de Toon Vanagt et Eric Rodriguez, fondateurs des SPRL Data.be et Open Lex. Dans cette tendance de transparence des données, les deux Belges sont à l'origine de deux initiatives Open Data. La première, le site « data.be », créé en 2012, est une plateforme qui met à disposition du grand public les informations financières et légales des sociétés belges. La seconde initiative, déployée en 2016, est le site « lex.be », une interface unique qui permet aux juristes, citoyens et entreprises de rechercher facilement et gratuitement des textes légaux et des décisions de justice dans leur version intégrale (Data.be SPRL, 2016 ; Malengreau, 2016 ; Open Lex SPRL, 2016).

Ces deux projets d'Open Data, et en particulier le second, d'ailleurs soutenu par la Commission européenne, montrent que la transparence des données et le partage des connaissances publiques sont devenus de nouvelles opportunités de création de valeur, autant pour les citoyens que pour les entreprises.

1.4.2. Mieux faire respecter la loi

La voiture « qui roule toute seule » est désormais une réalité, comme le montrent les premiers prototypes de Google (Google, 2016). Pour que ces voitures puissent fonctionner, il est évident qu'elles doivent être munies de toutes sortes de capteurs qui analysent l'environnement du véhicule afin de rouler sans faire d'accident. Or, qui dit capteur dit aussi données. En effet, ces capteurs transmettent les données environnementales à un ordinateur de bord qui les analyse et ensuite tire des conclusions sur la manière dont la voiture doit rouler à un endroit précis.

A partir du moment où une grande quantité de voitures de ce type sont sur la route, cela signifie que des données en masse sont produites et transmises. Etant donné que ces informations sont variées (mélange de données de géolocalisation, d'environnement, de vitesse, de météo, etc.) et peuvent être collectées en temps réel, les dimensions de volume, variété et vitesse font naître du Big Data.

Comment le gouvernement peut-il faire de cette situation une opportunité de mieux faire respecter la loi ? Par exemple, en équipant ces voitures « intelligentes » de capteurs qui seraient capables, grâce à des systèmes de reconnaissance d'image, de sonars et de radars longue portée, de ralentir automatiquement à la vue d'un panneau de limitation de vitesse. Ces

technologies ne sont pas des fictions, mais caractérisent le dernier véhicule construit par la société Tesla, la « Tesla D ». Ainsi, le respect de la loi n'est plus laissé à la subjectivité des conducteurs, mais est garanti par l'objectivité des données.

De la même manière, une carte de paiement pourrait empêcher son utilisateur, directement et sans contestation possible, d'effectuer une dépense dans un domaine non autorisé. Ces deux exemples symbolisent une transformation de l'Etat vers un modèle objectif de gouvernance qui, dans certains domaines, permet un meilleur respect de la loi par le citoyen. Si la loi est respectée, ce n'est donc pas nécessairement parce que le citoyen le veut, mais parce que l'objet qu'il utilise est configuré pour. En résumé, le Big Data permet un processus automatique de décision qui débouche directement sur une action (Grosdhomme Lulin, 2015).

L'augmentation de la transparence des données publiques et les Big Data analytics pour faire respecter la loi montrent le potentiel de création de valeur entre les mains de l'Etat qui utilise efficacement les données qu'il a à sa disposition. La révolution du Big Data fait apparaître un nouveau modèle de gouvernance, un modèle plus objectif et plus efficient dans lequel les processus de prise de décision et de jugement sont « conduits » par l'analyse des données.

Nous venons de transposer les tendances du Big Data observées dans le secteur financier au domaine de la gouvernance publique, et nous réalisons qu'il est en effet possible d'envisager un Etat utilisant les données à l'image de la banque et de la compagnie d'assurances. Certes, cette liste d'applications du Big Data n'est pas exhaustive, mais elle nous donne déjà un bel aperçu sur la possibilité d'une Data Driven Governance.

Effectivement, l'Etat qui parvient à utiliser le Big Data de manière pragmatique peut déployer de nombreuses opportunités de création de valeur. Dorénavant, il a la possibilité d'améliorer la productivité au sein de ses administrations, de rendre le pays plus harmonieux et agréable à vivre pour les citoyens et les entreprises, de leur fournir des services plus personnalisés et enfin de basculer vers un modèle objectif de gouvernance, plus performant pour s'attaquer aux problèmes sociétaux actuels.

1.5. Un Etat à l'image de la « Smart City »

A côté du Big Data, la « Smart City » est aussi un concept fort en vogue ces dernières années. Après le « Smart Phone », la « Smart Watch » et encore d'autres termes anglais du genre, on parle aujourd'hui de plus en plus de « Smart City ». Tout comme le Big Data, ce terme, qui en français signifie littéralement « ville intelligente », désigne un concept nouveau, et il n'y a

donc pas encore de définition unanimement acceptée par les experts de la technologie. Néanmoins, deux choses sont certaines par rapport à la Smart City :

- Elle est la conséquence d'une population de plus en plus urbanisée. En effet, cette urbanisation en hausse au niveau mondial a fait naître pour les villes des challenges sans précédent dans les domaines du développement durable, de l'éducation, de l'énergie, de l'environnement, de la sécurité et des services publics entre autres.
- Les technologies de l'information et de la communication ainsi que les données sont considérées comme la solution pour faire face à ces nouveaux défis économiques, sociaux et environnementaux.

Très rapidement, nous comprenons que la Smart City est intimement liée au Big Data. Par ailleurs, bien qu'il n'existe pas de définition officielle, les scientifiques semblent assez d'accord pour caractériser la Smart City par six dimensions : Smart Economy, Smart Mobility, Smart Environment, Smart People, Smart Living et Smart Governance.

Ces dimensions rappellent des thèmes que nous avons abordés dans le cadre de l'utilisation du Big Data par l'Etat, respectivement : l'augmentation de la productivité, la gestion du trafic, moins de congestion pour moins de pollution, des citoyens plus informés grâce à l'Open Data, l'amélioration de la sécurité publique et enfin un modèle de gouvernance plus « smart » car « conduit » par les données.

L'objet de ce mémoire n'étant pas la Smart City, nous n'approfondirons pas plus le thème, mais nous retenons que la Smart City est une sorte de « Smart State » miniature. Etant donné que de nombreuses villes dans le monde ont déjà entrepris des initiatives « smart », Liège en faisant partie, nous sommes confortés dans le potentiel de l'hypothèse d'une Data Driven Governance à l'échelle du pays, soit d'un Etat à l'image de la Smart City (European Parliament, 2014 ; Kitchin, 2014 ; Rodríguez-Bolívar, 2015 ; Smart City Institute HEC-ULg, 2015 ; voir aussi Schenk & Berman, 2015).

2. Quels sont les challenges et les risques ?

A présent que nous savons de quelle manière l'Etat peut créer de la valeur en déployant une stratégie Big Data, nous nous posons la question des challenges et des risques. Premièrement, nous nous demandons quelles sont les conditions que l'Etat doit remplir pour utiliser le Big Data de manière efficace. Deuxièmement, nous étudions dans quelle mesure l'utilisation du Big Data par l'Etat peut comporter des risques pour le citoyen.

2.1. Les challenges pour l'Etat

Comme nous l'avons développé à travers les tendances de personnalisation, de prédictibilité, d'optimisation et d'objectivation, le Big Data constitue définitivement une opportunité de création de valeur pour l'Etat. Mais quels sont les challenges auxquels l'Etat doit faire face afin de rendre possible cette création de valeur ? Quelles sont les conditions pour que cela fonctionne ?

Selon le MGI, « *utiliser le Big Data intelligemment a un potentiel de génération de valeur significative pour le secteur public, à condition que des actions soient mises en œuvre pour surmonter les barrières technologiques, pour recruter et reformer des personnes avec les compétences appropriées, et pour gérer les changements nécessaires dans les organisations afin d'adopter et de tirer profit du Big Data* » (2011, p. 63, traduction personnelle de l'anglais). Cette citation fait ressortir trois challenges pour l'Etat : un challenge technique, un challenge de compétences et un challenge de mentalité (voir aussi Eggers & Bellman, 2015).

2.1.1. Un challenge technique

Les investissements dans la mise en place de technologies appropriées afin d'être capable de rassembler, intégrer, stocker et assurer la sécurité et la confidentialité des données constituent la première condition. Etant donné le coût de ces investissements et les finances publiques limitées, le challenge technique est la première barrière vers une Data Driven Governance. Si l'Etat veut bénéficier des bienfaits du Big Data, il doit se donner les moyens d'y arriver en faisant les investissements nécessaires (Grosdhomme Lulin, 2015 ; MGI, 2011 ; Schenk & Berman, 2015).

2.1.2. Un challenge de compétences

Une Data Driven Governance ne sera possible que si l'Etat parvient à recruter et à former des personnes compétentes, capables de dégager de la valeur du Big Data. Comme Monsieur X et Jean-Marc Van Gyseghem le confirment, le recrutement et la conservation des talents constituent un des principaux points faibles du secteur public. L'Etat doit donc absolument y remédier s'il veut tirer le meilleur profit de toutes les données qu'il détient (Grosdhomme Lulin, 2015 ; MGI, 2011 ; Schenk & Berman, 2015 ; voir aussi Brown et al., 2011).

2.1.3. Un challenge de mentalité

Les mentalités peuvent parfois s'avérer être un frein plus important que les deux premiers challenges. La gestion du changement depuis la culture traditionnelle vers un nouveau modèle de gouvernance « conduit » par les données est le deuxième défi auquel l'Etat doit faire face

(Grosdhomme Lulin, 2015 ; MGI, 2011). Selon Monsieur X, l'administration est trop rigide et doit basculer vers une mentalité qui promeut le dynamisme, la flexibilité et l'innovation.

En annexe 10 se trouve une infographie de Deloitte University Press (Eggers & Bellman, 2015, Figure 2) qui illustre ces trois barrières que l'Etat doit surmonter pour opérer sa transition digitale vers une Data Driven Governance. Dans cette illustration, les pôles « people », « processes » et « preparedness » représentent respectivement les challenges de compétences, d'ordre technique et de mentalité. Voyons à présent les risques du Big Data qui pèsent sur le citoyen.

2.2. Les risques pour le citoyen

Comme nous l'avons vu, les Big Data analytics présentent de nombreux avantages, autant pour l'Etat que pour le citoyen. Il semblerait que le Big Data puisse nous mener vers un monde meilleur, une société plus harmonieuse. Toutefois, comme cela a été développé dans les deux premières parties, il y a aussi le revers de la médaille. En effet, les risques du Big Data qui ont été développés précédemment sont tout à fait transposables au domaine de la gouvernance publique.

Ainsi, deux risques principaux, confirmés par Monsieur X, menacent le système : un risque de sécurité des données et un risque pour les libertés individuelles. Ces deux risques ont deux points communs : premièrement, seul l'Etat est en mesure de les gérer, et deuxièmement, c'est le citoyen qui est la victime potentielle.

2.2.1. Un risque de sécurité des données

La sécurité et la confidentialité des données est un risque majeur, et l'Etat doit mettre tous les moyens en œuvre pour les garantir. Nous vivons dans un monde de plus en plus numérique et aucun individu ni aucune organisation n'est à l'abri d'un piratage informatique. De plus, nous avons parlé du terrorisme, mais le cyber-terrorisme existe aussi.

Si l'Etat déploie une stratégie Big Data, il est clair qu'il aura à sa disposition des volumes importants de données sensibles à propos de ses citoyens. Il ne peut donc se permettre qu'une personne ou une organisation mal intentionnée ait accès à ces données (MGI, 2011 ; Schenk & Berman, 2015 ; voir aussi Bertot, Gorham, Jaeger, Sarin, & Choi, 2014 ; Chui et al, s.d.). Dans cette situation, la partie vulnérable est le citoyen, l'Etat a la responsabilité d'être gardien des données et l'ennemi est une tierce personne ou organisation.

2.2.2. Un risque pour les libertés individuelles

En ce qui concerne le risque pour les libertés individuelles, la partie vulnérable reste le citoyen, mais à présent l'Etat est à la fois celui qui peut commettre les abus et celui qui doit protéger le citoyen de ces abus. Comme nous l'avons exposé dans les deux premières parties, certains droits de l'Homme les plus fondamentaux peuvent être menacés par la révolution du Big Data, en particulier le droit à la vie privée et le droit d'être différent (discrimination).

Par conséquent, afin d'encadrer l'usage du Big Data au sein de ses administrations et d'éviter les abus, l'Etat doit mettre en place de nouvelles lois et réglementations. Il doit aussi faire preuve de transparence (politique d'utilisation des données, avantages et risques) et faire en sorte que le citoyen « *reste propriétaire de ses données et puisse les valoriser à son bénéfice* » (Grosdhomme Lulin, 2015, p. 8 ; voir aussi Bertot et al., 2014 ; Chui et al, s.d. ; Janssen & van den Hoven, 2015 ; MGI, 2011 ; Washington, 2014).

Pour résumer cette troisième partie, nous pouvons dire que l'Etat bénéficie, avec le Big Data, des mêmes opportunités de création de valeur que le secteur financier. En effet, les tendances de personnalisation, de prédictibilité, d'optimisation et d'objectivation sont tout à fait transposables au domaine des services publics. Néanmoins, pour que le déploiement d'une stratégie Big Data apporte une réelle valeur ajoutée au citoyen et à l'administration publique en interne, plusieurs barrières doivent être outrepassées : l'Etat doit faire face à un challenge technique, un challenge de compétences et un challenge de mentalité. Enfin, si les bienfaits du Big Data peuvent être transposés du secteur financier au secteur public, les risques qu'il provoque le sont tout autant. C'est pourquoi l'Etat et ses citoyens sont menacés par un risque de sécurité des données et un risque pour les libertés individuelles.

En annexe 11 se trouve une infographie réalisée par la TechAmerica Foundation et le géant de l'informatique SAP qui résume les grandes lignes de l'exposé de cette troisième partie. Elle présente, de manière visuelle et à l'aide de quelques chiffres, les domaines où le Big Data a un véritable potentiel de valeur ajoutée et dans lesquels l'Etat devrait investir. De plus, l'infographie rappelle aussi les défis qui se trouvent sur la route d'une Data Driven Governance.

Conclusion

Avant de conclure, rappelons notre problématique de départ. Dans le cadre de la révolution du Big Data (Mayer-Schönberger & Cukier, 2013), nous envisageons, en Belgique, la possibilité d'une Data Driven Governance à l'image du secteur financier. Nous nous posons donc la question suivante : est-il possible de concevoir un Etat belge qui utilise le Big Data pour améliorer sa gouvernance, comme le font la banque et la compagnie d'assurances pour générer du profit ?

Notre exposé était composé de trois parties. Dans la première partie, nous avons présenté le contexte de la révolution du Big Data. Dans la deuxième, nous avons étudié l'impact de l'utilisation du Big Data dans les secteurs de la banque et de l'assurance, dans lesquels quatre tendances ont été relevées : la personnalisation, la prédictibilité, l'optimisation et l'objectivation. Enfin, dans la troisième partie, nous avons transposé ces tendances au domaine de la gouvernance publique et exploré comment l'Etat peut, grâce au Big Data, créer de la valeur, autant pour lui que pour ses citoyens. A présent, nous sommes en mesure de répondre à notre question de recherche.

A cette question de la possibilité d'une Data Driven Governance en Belgique, nous répondons : oui, mais... Oui, un Etat belge qui utilise le Big Data à l'image du secteur financier est concevable, mais plusieurs questions se posent...

Tout d'abord, « *de la même manière que les entreprises utilisent le Big Data pour générer des profits, les gouvernements l'utilisent en faveur du bien public* » (Kim et al., 2014, p. 1, traduction personnelle de l'anglais). Oui, l'Etat bénéficie bien des mêmes opportunités de création de valeur que le monde de l'entreprise. Néanmoins, il faut souligner que, si les modalités d'utilisation sont les mêmes (cf. les quatre tendances), la finalité reste différente : contrairement à l'entreprise, l'Etat n'a pas un but de lucre. Comme dirait Jean-Marc Van Gyseghem, « *l'Etat est un genre d'ASBL* ». En effet, le gouvernement existe pour servir le citoyen, non pour faire du profit ; il a des compétences et il est important qu'il ne les dépasse pas.

Bien que le pouvoir législatif belge vote des lois pour encadrer l'utilisation des données détenues par le secteur public, comme le projet de loi adopté par la Chambre des représentants le 14 avril 2016 (Chambre des représentants de Belgique, 2016), il est primordial que des

organes indépendants soient constitués afin de surveiller le pouvoir politique et empêcher les abus. Aujourd'hui, en Belgique, de tels organes existent déjà, comme la Commission de Protection de la Vie Privée (CPVP, 2016). Toutefois, selon M. Van Gyseghem, la question de l'indépendance est encore parfois douteuse. Nous nous posons donc une première question : comment nous assurer que l'Etat continue à agir pour le bien public ? Qui va le surveiller et l'empêcher d'agir pour son « propre intérêt » ?

Ensuite, nous nous demandons si l'Etat peut réussir à déployer une véritable stratégie Big Data. Nous avons vu que, pour atteindre une utilisation efficace du Big Data, les gouvernements doivent faire face à des challenges d'ordre technique, de compétences et de mentalité. Au niveau du challenge technique, nous doutons que l'Etat belge soit capable de faire les bons investissements. En effet, à ce niveau, le gouvernement belge souffre d'une réputation en dégradation. Les récents faits d'actualité, comme la crise des tunnels et les attentats du 22 mars 2016 à Bruxelles, ont mis notre gouvernement à l'épreuve.

Malheureusement pour nos politiques, ces situations ont fait ressortir des faiblesses dans la gestion de certaines administrations et nous avons beaucoup entendu parler de « manque de moyens », en l'occurrence pour la mobilité et la sécurité publique. Justement, nous avons vu que le Big Data peut créer de réelles opportunités de création de valeur dans ces deux domaines. En outre, nous savons aussi que les finances publiques de notre pays sont souvent en déficit et que l'Etat est sans cesse en recherche de nouvelles sources de revenus pour compenser la hausse de ses dépenses.

Au sujet du challenge de compétences, nous avons l'impression qu'en Belgique, la formation est en retard sur son temps. Nous croyons que l'enseignement supérieur belge forme encore trop d'étudiants pour des professions du passé. Dans ce sens, Monsieur X et Jean-Marc Van Gyseghem affirment que le marché du travail manque cruellement de data scientists et autres métiers du monde numérique. L'Etat risque donc de manquer de compétences pour déployer une stratégie Big Data opérationnelle. Il est temps de former pour l'avenir et d'investir dans des « data formations » plutôt que de former à des professions en voie de disparition. Enfin, au-delà de ces deux premiers challenges, nous craignons que les mentalités traditionnelles du monde politique constituent des entraves à l'innovation technologique et à une transition vers une « Data Driven Culture ».

Dès lors, nous nous demandons si le pouvoir public belge est financièrement, professionnellement et psychologiquement capable de prendre le train du Big Data en marche.

Certes, la Belgique a tout de même des points forts dans le domaine du numérique, notamment avec sa Banque Carrefour de la Sécurité Sociale (Hamel & Marguerit, 2013) et son service fiscal Tax-on-web (Service Public Fédéral Finances, s.d.), mais cela suffira-t-il ?

Dans la continuité de cette réflexion, nous nous posons une autre question : si l'Etat belge ne parvient pas à rentrer efficacement dans la tendance du Big Data, quelqu'un d'autre pourrait-il le faire à sa place ? Notre société connaît un phénomène très important de désintermédiation (GovLab Research, 2013), et l'économie collaborative est en plein essor.

Uber et Lyft (transport privé), BlaBlaCar (covoiturage), Airbnb et Booking.com (logement), ListMinut et TaskRabbit (services quotidiens divers), Peerby (emprunts d'objets), Cookening et Menu Next Door (cuisine), Deliveroo et Take Eat Easy (livraisons de repas) ainsi que Kickstarter et KissKissBankBank (crowdfunding ou financement participatif) sont quelques exemples célèbres d'une économie nouvelle basée sur la collaboration et les nouvelles technologies. Le nouveau modèle économique proposé par ces entreprises consiste en une plateforme numérique qui met en relation des particuliers entre eux, les uns recherchant un service et les autres proposant un service.

Ce phénomène de désintermédiation de la société implique que les intermédiaires traditionnels peuvent être contournés. Par exemple, le transport collaboratif permet aux particuliers de se passer des taxis et le logement collaboratif leur permet de se passer des hôtels. Bientôt, grâce à des algorithmes qui produisent des profils de risque individuels, une plateforme mettra en relation des particuliers qui pourront se prêter de l'argent entre eux au lieu d'aller contracter un emprunt à la banque (Henno, 2016). Pourrait-il arriver la même chose à l'Etat ? Pourrait-il se faire « ubériser » ? Un jour, les citoyens pourront-ils se passer du pouvoir public en se rendant des services mutuellement ?

Dans cette même optique, nous pouvons aussi considérer l'économie du Big Data. Nous entendons par là les organisations dont le business modèle est basé sur les données, comme les géants Google, Amazon et Facebook. Avec l'économie collaborative, nous nous demandons si les citoyens eux-mêmes peuvent prendre la place de l'Etat. Maintenant, la question suivante se pose : des entreprises privées pourraient-elles prendre la place de l'Etat ? Des fonctions régaliennes pourraient-elles être privatisées ?

M. Van Gyseghem rappelle qu'en janvier 2015, le Premier Ministre belge Charles Michel a demandé les services de Google pour surveiller Internet, afin de repérer d'éventuelles menaces terroristes (Datanews, 2015). Quels sont les risques si l'Etat commence à déléguer

certaines de ses fonctions à des entreprises privées dont le but premier est de faire du profit ? Quel doit être le rôle du pouvoir public s'il s'avère que ces entreprises privées sont plus compétentes que l'Etat dans certains domaines ? L'Etat « *doit-il continuer à investir ses ressources dans des tâches qui sont accomplies aussi bien, voire mieux, à l'extérieur* » (Grosdhomme Lulin, 2015, p. 52) ?

Tant que nous sommes dans le débat des fonctions de l'Etat et de sa manière de gouverner, il est intéressant d'évoquer aussi les tendances Big Data d'optimisation et d'objectivation que nous avons développées. La plupart des experts aspirent à un modèle de gouvernance objectif, « conduit » par les données. Alors que Peter De Keyzer affirme qu'il est essentiel « *que la politique se base enfin sur des chiffres et non sur des sentiments* » (2016), d'autres justifient ce nouveau modèle par le fait qu'il y a plus de vérité sur Internet que nulle part ailleurs (Henno, 2016). Ils prônent un modèle de gouvernance dans lequel les décisions seraient prises sur base d'algorithmes qui analysent des données objectives plutôt que sur des intuitions et raisonnements humains.

Selon tous ces experts, une telle Data Driven Governance prendrait alors de meilleures décisions que la gouvernance traditionnelle, car les données objectives permettraient de mieux appréhender la réalité de la situation. Cependant, les données sont-elles vraiment toujours objectives ? Ne peut-on pas faire dire aux données ce que l'on veut ? Est-ce vraiment la solution d'écarter toute intuition et raisonnement humain de la prise de décision ? A quoi ressemblerait notre société si les algorithmes gouvernaient ?

Nous avons vu que les algorithmes pouvaient même devenir un moyen efficace pour le gouvernement de faire respecter la loi, en configurant les objets que les citoyens utilisent pour qu'ils respectent la loi automatiquement (cf. la voiture qui ralentit automatiquement en passant devant un panneau de limitation de vitesse). Seulement, une question fondamentale se pose : « *quel arbitrage voulons-nous établir entre l'efficacité de l'action publique et la liberté de comportement des individus* » (Grosdhomme Lulin, 2015, p. 18) ?

Par ailleurs, nous souhaitons réagir à ce qui a été développé au sujet de la tendance Big Data de personnalisation. Dans le domaine public, il est clair que le Big Data permet au citoyen de bénéficier d'un service à haute valeur ajoutée, au point qu'il lui sera probablement difficile de s'en passer. En outre, dans un avenir proche, il sera possible de payer avec un bracelet, avec notre voiture, via une reconnaissance d'identité par les battements de notre cœur ou un selfie (reconnaissance faciale), ou juste en énonçant « je paye avec Google » (Mobistar, 2016). Il est

intéressant de constater que ces nouveaux moyens de paiement riment clairement avec « *disparition de l'argent liquide* » (Bourguinat, 2016).

Intrinsèquement, le « cash » ne peut pas être contrôlé. Il est donc un symbole de liberté. La disparition de l'argent liquide est synonyme de contrôle de tous les flux d'argent par le pouvoir public, permettant la prévention de nombreuses activités illégales (évasion fiscale, blanchiment d'argent, drogue, prostitution, commerce des armes et travail au noir). Pour l'Etat, c'est évidemment le « bonheur » ; mais qu'implique ce contrôle absolu de l'argent des citoyens ? A l'avenir, le gouvernement sera-t-il capable de bloquer les comptes bancaires d'un individu en un seul clic ? Donc oui, un service personnalisé présente un intérêt indéniable, mais quelles en sont les limites ?

Si nous ajoutons à la discussion la tendance Big Data de prédictibilité, nous pouvons aller encore plus loin. L'analyse prédictive est une réalité. Aujourd'hui, une application appelée « HeroQuizz » (HeroQuizz, 2016) promeut, en collaboration avec Facebook, le slogan suivant : « *Découvre le statut que tu posteras dans 10 ans !* ». Pour ce faire, l'application a besoin d'accéder aux informations Facebook de l'utilisateur : profil public, liste d'amis, adresse électronique, publications du journal, anniversaire et photos. Si l'utilisateur l'autorise à analyser ces données, HeroQuizz est capable de formuler un statut que l'individu pourrait probablement publier dans dix ans. Ayant fait personnellement le test, il faut avouer que le résultat obtenu peut être impressionnant quant à la pertinence de la prédiction. Il semblerait donc que Facebook soit capable de prédire notre avenir en analysant les données personnelles que nous lui confions.

Qu'en est-il de l'Etat s'il en devient aussi capable ? Comme nous en avons déjà parlé, avec le Big Data et les analyses prédictives, la sécurité publique s'en trouvera nettement améliorée, mais à quel prix ? Celui des libertés individuelles ? Après les attentats du 22 mars 2016 à Bruxelles, nous avons vu le gouvernement belge se permettre des atteintes aux droits fondamentaux des citoyens sous le couvert d'une nécessité d'ordre public. Quelles doivent être les limites pour trouver un juste équilibre entre sécurité publique et respect de la vie privée ?

Un jour, en tant qu'individu, nous serons probablement amenés à faire un choix entre, d'une part, garder notre vie privée et vivre dans une société telle qu'elle est aujourd'hui, avec les limites et les risques qu'elle présente ou, d'autre part, opter pour une société « améliorée », beaucoup plus sûre que la société actuelle et où tout devient possible, mais à condition de

donner toujours plus accès à nos données personnelles. C'est un peu faire le choix entre une société risquée avec une vie privée ou une sécurité publique garantie, sans vie privée.

En réalité, on fait croire au citoyen qu'en échange de l'accès à ses données personnelles, l'Etat sera capable de contrôler tout le monde et ainsi de garantir la sécurité publique. Mais est-il vraiment possible pour l'Etat de tout contrôler ? Un expert des implications et enjeux du Big Data nous a à ce propos confié ceci, sous couvert d'anonymat : « *On ne pourra jamais contrôler tout le monde. Faire croire qu'on va pouvoir tout contrôler est un mensonge !* ».

Le débat devient alors philosophique. Dans le contexte de menace terroriste dans lequel nous vivons actuellement, nous nous demandons évidemment si le Big Data peut faire disparaître le terrorisme. Si oui, jusqu'où est-on prêt à aller pour arriver à cette fin ? Tandis que l'expert susnommé affirme que « *faire croire que le terrorisme peut disparaître grâce au Big Data est un leurre* », la question se pose alors sur les véritables intentions d'un Etat qui a envie de tout contrôler.

Pour terminer, il est donc sujet d'Etat, de pouvoir public, de gouvernance, de Big Data, de contrôle, de bien public, de libertés individuelles et de droits fondamentaux entre autres, et ces termes nous rappellent le roman « 1984 », de Georges Orwell (Faljaoui, 2016b). L'auteur y met en scène le fameux « Big Brother », une sorte de dictateur qui surveille les moindres faits, gestes et paroles de chaque individu. Est-ce un monde comme celui décrit par Orwell qui nous attend ? Est-ce que « Big Data » peut devenir « Big Brother » ? Sommes-nous déjà à l'ère de Big Brother ?

En conclusion de ce mémoire, nous disons oui à une Data Driven Governance, mais y a-t-il un risque que « Data Driven *Governance* » devienne « Data Driven *Dictatorship* » ?

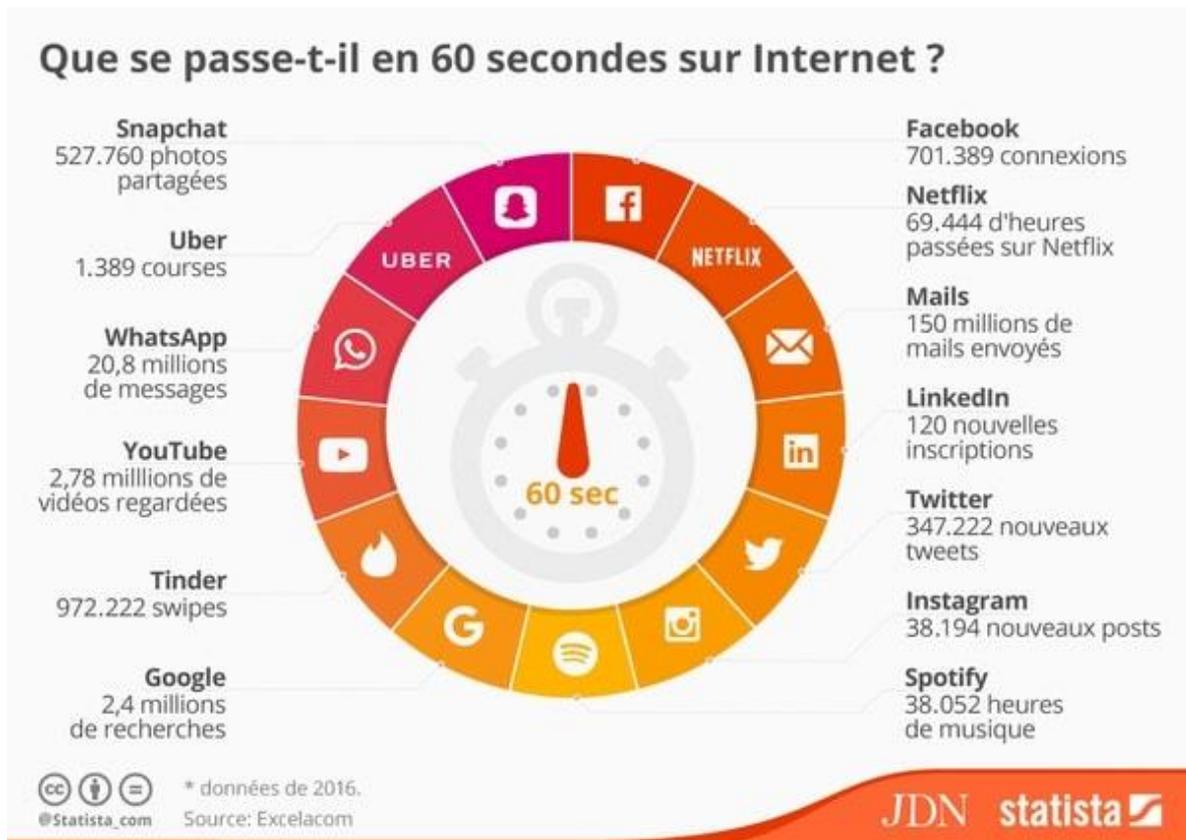
Annexes

Annexe 1 : Que savons-nous vraiment du Big Data ?



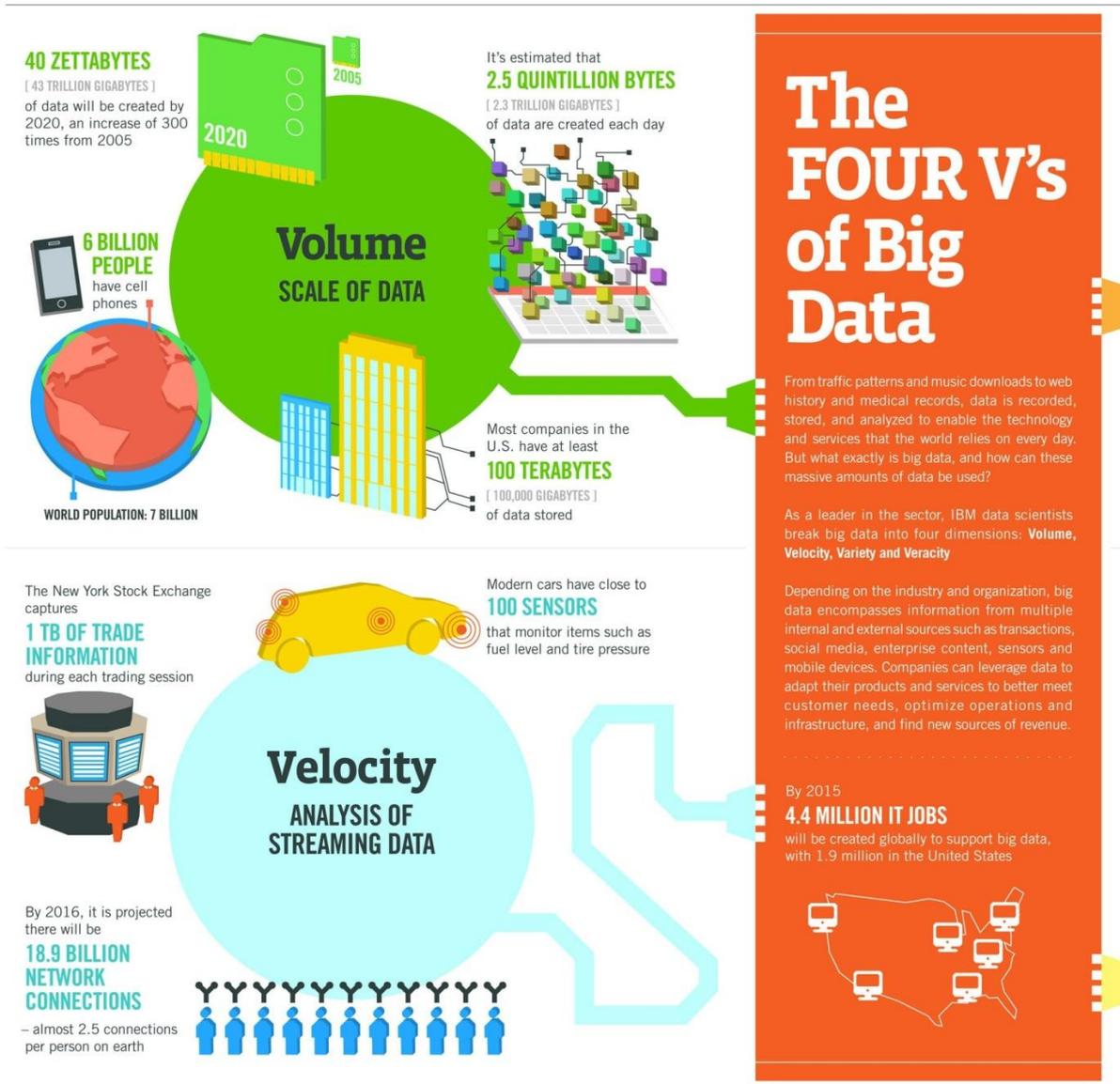
Source : Institut des Actuairens en Belgique. (2015). *Big Data : An actuarial perspective [Information paper]*. Consulté en ligne le 12/04/2016 : http://www.iabe.be/sites/default/files/bijlagen/Big_Data_paper_full_v009.pdf

Annexe 2 : Que se passe-t-il en 60 secondes sur Internet ?



Source : Consulté en ligne le 22/04/2016 : <http://www.journaldunet.com/ebusiness/le-net/1177195-que-se-passe-t-il-en-60-secondes-sur-internet-statista/>

Annexe 3 : les 4 « V » du Big Data selon IBM



Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTec, QAS

The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015
4.4 MILLION IT JOBS
 will be created globally to support big data, with 1.9 million in the United States



As of 2011, the global size of data in healthcare was estimated to be

150 EXABYTES
 [161 BILLION GIGABYTES]

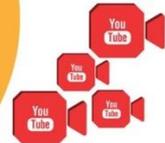


30 BILLION PIECES OF CONTENT
 are shared on Facebook every month



By 2014, it's anticipated there will be
420 MILLION WEARABLE, WIRELESS HEALTH MONITORS

4 BILLION+ HOURS OF VIDEO
 are watched on YouTube each month



400 MILLION TWEETS
 are sent per day by about 200 million monthly active users



Variety
 DIFFERENT FORMS OF DATA

1 IN 3 BUSINESS LEADERS

don't trust the information they use to make decisions



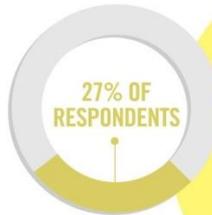
Poor data quality costs the US economy around

\$3.1 TRILLION A YEAR



27% OF RESPONDENTS

in one survey were unsure of how much of their data was inaccurate

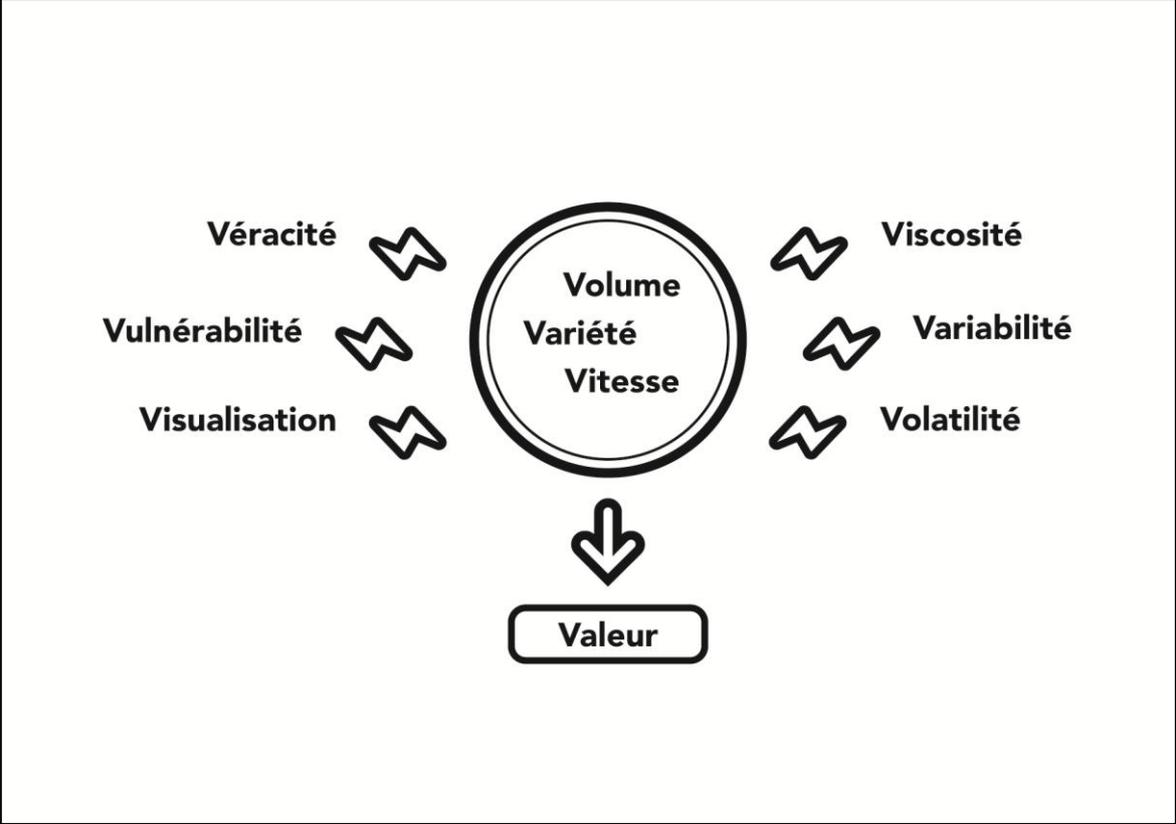


Veracity
 UNCERTAINTY OF DATA



Source : Consulté en ligne le 07/05/2016 : <http://www.ibmbigdatahub.com/infographic/four-vs-Big-Data>

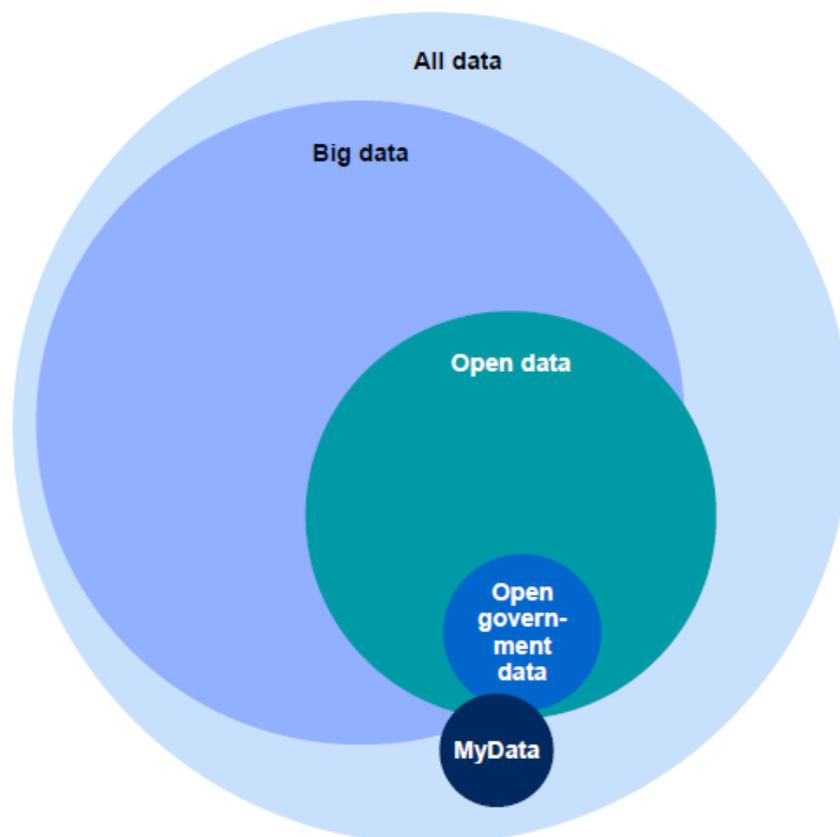
Annexe 4 : Notre schéma des « V » du Big Data



Annexe 5 : Les différents types de « data »

How open data relates to other types of data

ILLUSTRATIVE



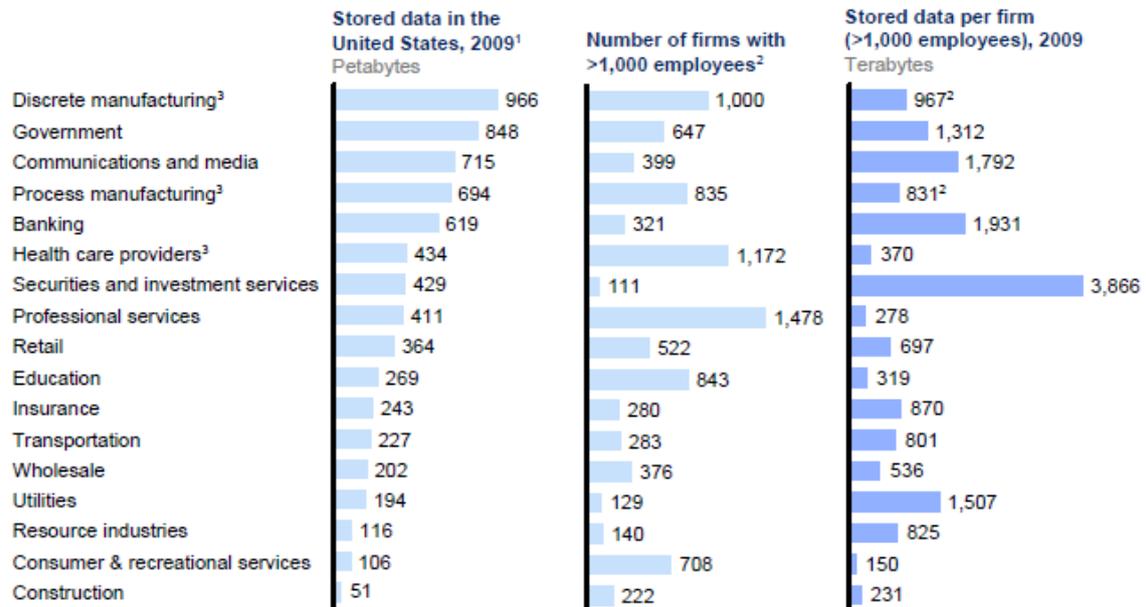
SOURCE: McKinsey Global Institute analysis

Source: McKinsey Global Institute, McKinsey Center for Government, & McKinsey Business Technology Office [MGI, MCG & BTO]. (2013). *Open Data : Unlocking innovation and performance with liquid information*. McKinsey&Company. Consulté en ligne le 12/04/2016 : <http://www.mckinsey.com/business-functions/business-technology/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information>

Annexe 6 : Données stockées par secteur aux Etats-Unis en 2009

Exhibit 7

Companies in all sectors have at least 100 terabytes of stored data in the United States; many have more than 1 petabyte



1 Storage data by sector derived from IDC.

2 Firm data split into sectors, when needed, using employment

3 The particularly large number of firms in manufacturing and health care provider sectors make the available storage per company much smaller.

SOURCE: IDC; US Bureau of Labor Statistics; McKinsey Global Institute analysis

Source: McKinsey Global Institute [MGI]. (2011). *Big Data : The next frontier for innovation, competition, and productivity*. McKinsey&Company. Consulté en ligne le 12/04/2016 : <http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation>

Annexe 7 : Les e-gouvernements leaders mondiaux en 2014

Table 1.1. World e-government leaders (Very High EGDI) in 2014

Country	Region	2014 EGDI	2014 Rank	2012 Rank	Change in Rank (2012–2014)
Republic of Korea	Asia	0.9462	1	1	-
Australia	Oceania	0.9103	2	12	↑ 10
Singapore	Asia	0.9076	3	10	↑ 7
France	Europe	0.8938	4	6	↑ 2
Netherlands	Europe	0.8897	5	2	↓ 3
Japan	Asia	0.8874	6	18	↑ 12
United States of America	Americas	0.8748	7	5	↓ 2
United Kingdom	Europe	0.8695	8	3	↓ 5
New Zealand	Oceania	0.8644	9	13	↑ 4
Finland	Europe	0.8449	10	9	↓ 1
Canada	Americas	0.8418	11	11	-
Spain	Europe	0.8410	12	23	↑ 11
Norway	Europe	0.8357	13	8	↓ 5
Sweden	Europe	0.8225	14	7	↓ 7
Estonia	Europe	0.8180	15	20	↑ 5
Denmark	Europe	0.8162	16	4	↓ 12
Israel	Asia	0.8162	17	16	↓ 1
Bahrain	Asia	0.8089	18	36	↑ 18
Iceland	Europe	0.7970	19	22	↑ 3
Austria	Europe	0.7912	20	21	↑ 1
Germany	Europe	0.7864	21	17	↓ 4
Ireland	Europe	0.7810	22	34	↑ 12
Italy	Europe	0.7593	23	32	↑ 9
Luxembourg	Europe	0.7591	24	19	↓ 5
Belgium	Europe	0.7564	25	24	↓ 1
Very High EGDI Average		0.8368			
World Average		0.4712			

Source : United Nations. (2014). *United Nations e-government survey 2014*. Consulté en ligne le 12/04/2016 : http://unpan3.un.org/egovkb/portals/egovkb/documents/un/2014-survey/e-gov_complete_survey-2014.pdf

Annexe 8 : Etude de cas – Big Data et service public de l'emploi

Aujourd'hui, le chômage est sans aucun doute un des plus gros défis auxquels les gouvernements occidentaux doivent faire face. Alors qu'aider les demandeurs d'emploi à trouver du travail est devenu une priorité pour les politiques, certains voient la solution dans le Big Data. Nous présentons ici un exemple de déploiement du Big Data dans le service public de l'emploi français, tiré de l'ouvrage de Grosdhomme Lulin (2015).

Jobijoba est une « *start-up bordelaise qui propose un moteur intelligent de recherche d'emploi* » (p. 50). Le fondateur de Jobijoba est parti du constat qu'en France, un individu qui recherche un emploi doit mener ses recherches sur de nombreux sites web différents comme Monster, cadreemploi.fr, pole-emploi.fr et d'autres sites plus spécialisés. Jobijoba vient alors combler un vide et proposer une plateforme unique qui non seulement agrège les offres des sites traditionnels, mais aussi présente ces offres de manière intelligente grâce à des algorithmes avancés. Jobijoba offre trois innovations principales :

- Les métiers ne sont plus référencés selon les noms de fonction renseignés par les recruteurs, mais ils sont déduits des descriptions de fonction. Lorsqu'une description de fonction est publiée par un recruteur, les algorithmes de Jobijoba identifient les mots-clés de la description et les associent à des mots-clés déjà enregistrés. Une certaine combinaison de mots-clés correspondra à un métier, l'algorithme fera le lien et associera alors à la description de fonction un métier sur base de son contenu plutôt que sur base du nom théorique de la fonction. Ainsi, les noms d'offre correspondent vraiment au contenu des métiers, ce qui donne des résultats de recherche plus pertinents.

- Les algorithmes de Jobijoba exploitent l'ensemble des données afin de rendre possible un système de recommandations, autant pour les employeurs que pour les demandeurs d'emploi. D'une part, les employeurs recevront des suggestions en fonction de ce que les autres employeurs font (par exemple : compléter une description de fonction avec des compétences que la majorité des recruteurs requiert pour un poste semblable). D'autre part, des recommandations seront faites au demandeur d'emploi pour augmenter ses chances de trouver un travail (par exemple : signaler au demandeur une compétence que les recruteurs exigent souvent pour la fonction recherchée et qu'il ne possède pas encore).

- Les offres et les demandes de travail ne sont plus mises en relation selon la méthode traditionnelle qui considère les métiers et les secteurs. Les algorithmes rapprochent les offres et les demandes plutôt en fonction des compétences et des talents. Cela permet à un demandeur d'emploi de tomber sur des offres sur lesquelles il ne serait pas tombé avec le système traditionnel. Il s'agit alors de rechercher des métiers sur base des compétences plutôt que sur base des noms de fonction que l'on croit correspondre à certaines compétences.

« *Aujourd'hui, Jobijoba travaille avec neuf cents sites partenaires, référence 6 millions d'offres d'emploi dans onze pays et reçoit 4 millions de visiteurs uniques par mois. En France, ce sont 600 000 offres d'emplois issues de quatre cents sites partenaires* » (p. 51). Avec sa « *mentalité Big Data* », le fondateur de Jobijoba a créé un outil qui apporte une réelle plus-value pour le service public de l'emploi.

Et la plateforme n'a pas fini de se développer. De nouveaux services sont en voie de développement, tels qu'un concept de conseiller virtuel pour « coacher » les demandeurs d'emploi et un service d'analyse du marché de l'emploi, tant pour l'offre que pour la demande. Bien qu'il ne soit pas encore prouvé que le taux de chômage en France ait diminué suite à cette initiative, l'étude de cas montre le potentiel du Big Data pour créer de la valeur en offrant au citoyen un service personnalisé qui répond plus efficacement à ses besoins.

Annexe 9 : Etude de cas – Big Data et santé publique

Le Big Data, à travers sa tendance de prédictibilité, offre à l'Etat une opportunité unique de préserver la santé publique en lui permettant de mieux anticiper les éventuelles menaces de maladies et d'épidémies à l'échelle nationale et internationale. Nous présentons ici une étude de cas qui illustre notre propos, issue de l'ouvrage de Mayer-Schönberger et Cukier (2013).

En 2009, un nouveau type de virus a été découvert : la grippe H1N1. La maladie s'étendant rapidement, les agences de santé publique à travers le monde ont crié l'état d'urgence en quelques semaines. Alors que certains spécialistes redoutaient une épidémie à l'échelle de la grippe espagnole de 1918, qui avait touché des dizaines de millions de personnes, le plus alarmant était qu'aucun vaccin n'avait encore été trouvé. Dès lors, la seule chose que les autorités de santé publique pouvaient faire était de ralentir la propagation du virus tant bien que mal. Pour cela, il fallait encore savoir où elle avait déjà commencé.

Pour ce faire, aux Etats-Unis, les Centres de Contrôle et de Prévention des Maladies (CCPM) ont commencé à collecter un maximum d'informations sur l'épidémie en requérant des médecins des rapports sur leurs patients. Le problème était que les patients ne se rendaient pas directement chez leur médecin dès les premiers signes de maladie, mais attendaient que les symptômes s'aggravent. Par conséquent, le temps qu'un médecin se rende compte qu'un patient était atteint du virus H1N1 et qu'il le communique à un centre de renseignement, la maladie s'était déjà fortement propagée.

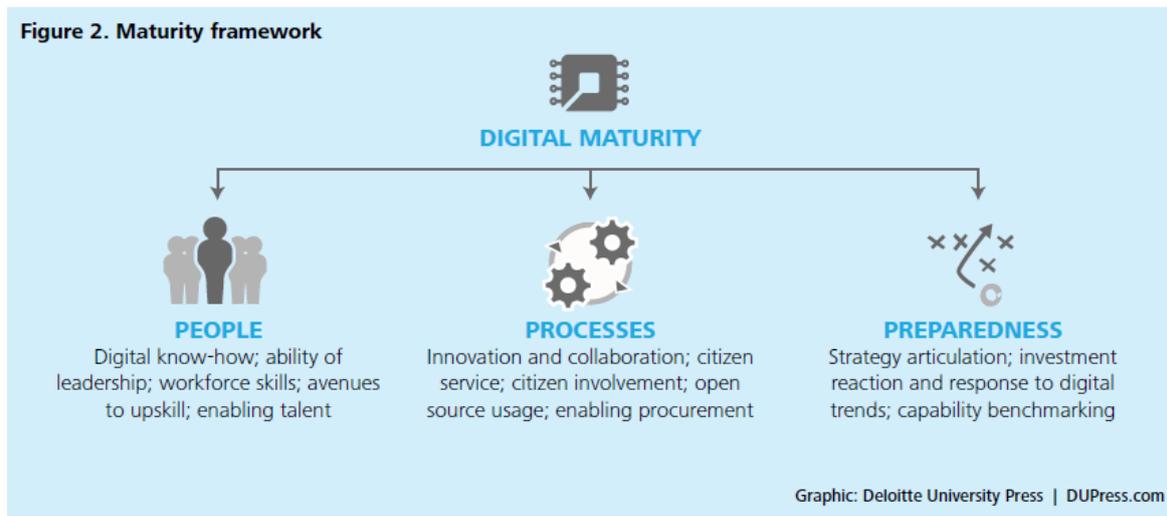
Il se trouve que, quelques semaines avant les premiers éclats de l'épidémie dans l'actualité, des ingénieurs de Google avaient publié un article scientifique qui expliquait comment Google parvenait à « prédire » la propagation de la grippe hivernale au niveau national mais aussi au niveau local. Ce sont les plus de 3 milliards de requêtes exécutées chaque jour sur le moteur de recherche et enregistrées par Google qui permettaient au géant d'Internet d'arriver à ses fins. En effet, en comparant les 50 millions de termes les plus utilisés par les américains avec les données des CCPM de propagation de la grippe hivernale entre 2003 et 2008, les analystes de Google ont réussi à établir des corrélations entre les zones affectées par le virus et ce que les gens dans ces zones recherchent sur Internet.

Après de nombreuses recherches d'un modèle mathématique dont les prédictions de propagation de la grippe hivernale en 2007 et 2008 seraient comparables aux chiffres officiels de la propagation effective du virus pendant ces deux années, les ingénieurs de Google ont fini par atteindre leur objectif. Ils ont trouvé un modèle mathématique qui, en combinant 45 termes de recherche, produisait des prédictions fortement corrélées avec les statistiques officielles au niveau national. Les CCPM étaient déjà capables de dire comment le virus se propageait, mais Google le faisait presque en temps réel et pas une ou deux semaines plus tard.

Lorsque l'épidémie du virus H1N1 a frappé en 2009, les algorithmes de Google se sont avérés plus utiles et efficaces que les systèmes statistiques gouvernementaux pour anticiper la propagation de la maladie et mettre en œuvre les solutions appropriées pour la ralentir. Pourquoi Google s'est-il montré plus performant que l'Etat pour faire face à la menace de santé publique ? Parce que nul autre que Google n'avait autant de données à disposition ni les outils nécessaires pour les analyser.

Lorsque les ingénieurs de Google ont mis en relation les requêtes du moteur de recherche avec les données officielles de propagation de la maladie, on a assisté à un « phénomène Big Data ». Les données étaient déjà volumineuses (milliards de requêtes) et collectées presque en temps réel mais, une fois les requêtes croisées avec les statistiques officielles de santé, les données devenaient aussi variées. C'est donc le Big Data qui a fait la différence.

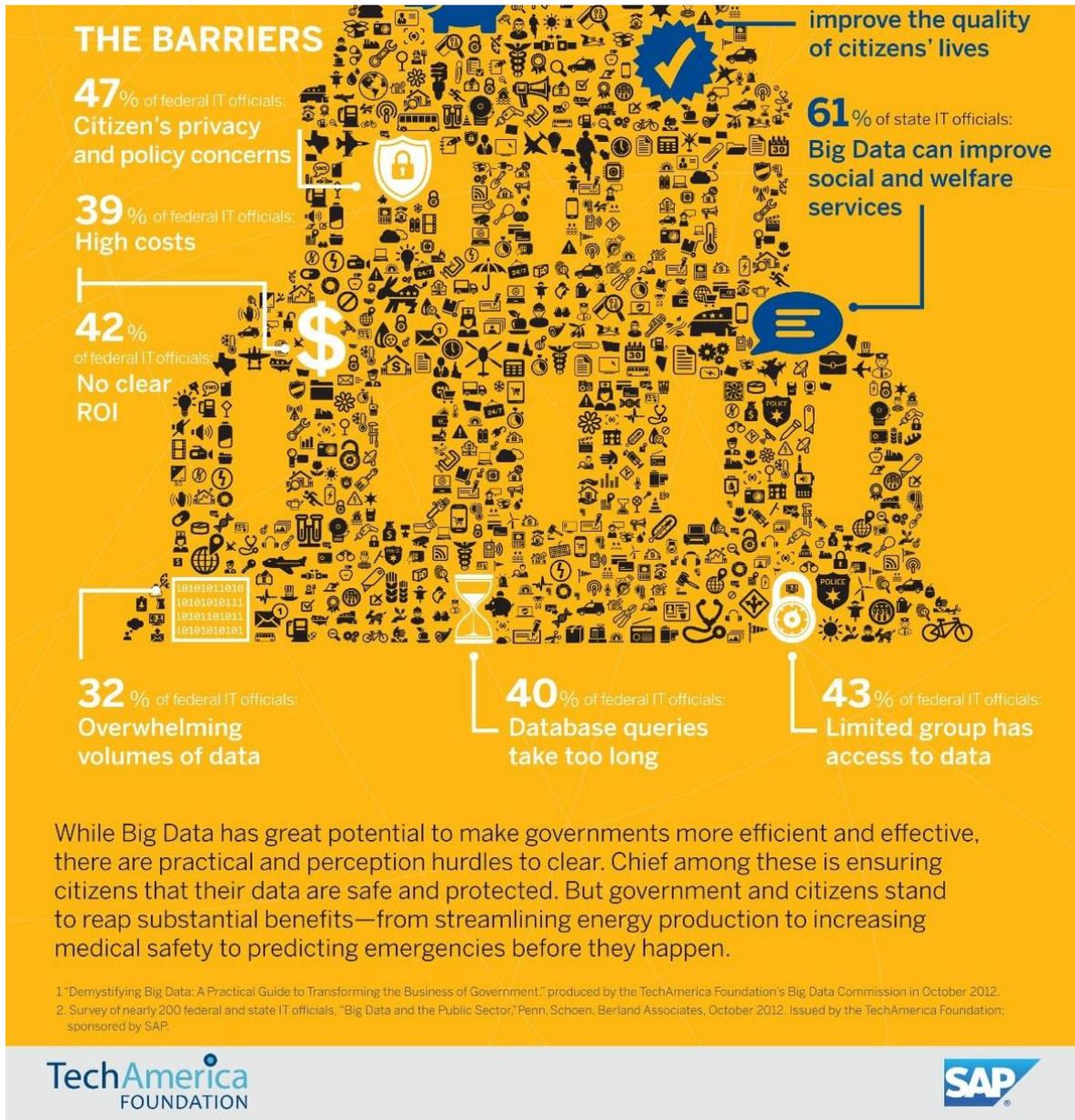
Annexe 10 : Modèle de maturité digitale du gouvernement



Source : Eggers, W. D., & Bellman, J. (2015, Octobre 2). The journey to government's digital transformation [A Deloitte Digital global survey]. *Deloitte University Press*. Consulté en ligne le 08/05/2016 : http://d27n20517rookf.cloudfront.net/wp-content/uploads/2015/10/DUP_1081_Journey-to-govt-digital-future_MASTER.pdf

Annexe 11 : Big Data, gouvernement et Data Driven Governance





Source : Consulté en ligne le 11/05/2016 : <http://www.govtech.com/technology/Can-Big-Data-Save-Lives.html>

Références bibliographiques

- Atos. (2014). *Powering Big Data analytics everywhere in financial services*. Consulté en ligne le 12/04/2016 : <https://atos.net/content/dam/global/documents/your-business/atos-big-data-fs-brochure.pdf>
- Babinet, G. (2014). *L'ère numérique, un nouvel âge pour l'humanité [Résumé de l'ouvrage]*. Paris, France : Le Passeur. Consulté en ligne le 12/04/2016 : <http://www.gqmagazine.fr/uploads/freepages/Ere-numerique.pdf>
- Bank for International Settlements. (s.d.). *Basel Committee on Banking Supervision*. Consulté en ligne le 29/03/2016 : https://www.bis.org/list/bcbs/lang_fr/index.htm
- Belouezzane, S., & Grandin de l'Eprevier, J. (2016, Janvier 4). Orange veut racheter Groupama Banque pour lancer sa banque mobile. *Le Monde*. Consulté en ligne le 30/03/2016 : http://www.lemonde.fr/entreprises/article/2016/01/04/orange-veut-racheter-groupama-banque-pour-lancer-sa-banque-mobile_4841575_1656994.html
- Bertot, J. C., Gorham, U., Jaeger, P. T., Sarin, L. C., & Choi, H. (2014). Big data, open government and e-government: Issues, policies and recommendations. *IOS Press : Information Polity*, 19, 5–16. doi: 10.3233/IP-140328
- Bourdieu, P. (1980, Janvier). Le capital social. *Actes de la Recherche en Sciences Sociales*, 31, 2-3. Consulté en ligne le 07/05/2016 : http://www.persee.fr/doc/arss_0335-5322_1980_num_31_1_2069
- Bourguinat, H. (2016, Mars 2). Vers « la disparition de l'argent liquide » ? *Le Monde Economie*. Consulté en ligne le 30/04/2016 : http://www.lemonde.fr/idees/article/2016/03/02/vers-la-disparition-de-l-argent-liquide_4875004_3232.html
- Boyer, J.-M. (2015). La tarification et le Big Data : quelles opportunités ? *Revue d'Economie Financière (2015/4)*, 120, 81-92. Consulté en ligne le 12/04/2016 : <http://www.cairn.info/revue-d-economie-financiere-2015-4-page-81.htm>

- Brasseur, C. (2013). *Enjeux et usages du Big Data. Technologies, méthodes et mises en œuvre*. Paris, France : Lavoisier.
- Brown, B., Chui, M., & Manyika, J. (2011). *Are you ready for the era of “Big Data” ?* McKinsey Global Institute : McKinsey Quarterly. Consulté en ligne le 12/04/2016 : <http://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/are-you-ready-for-the-era-of-big-data>
- Bryant, R.E., Katz, R.H., & Lazowska, E.D. (2008). *Big-Data computing : Creating revolutionary breakthroughs in commerce, science, and society [White paper]*. Computing Community Consortium (Computing Research Association). Consulté en ligne le 06/04/2016 : <http://cra.org/ccc/resources/ccc-led-whitepapers/>
- Capgemini Consulting. (2015). *Fixing the insurance industry : How Big Data can transform customer satisfaction*. Consulté en ligne le 12/04/2016 : <https://www.capgemini-consulting.com/fixing-the-insurance-industry>
- Cardon, D. (2015). *A quoi rêvent les algorithmes : Nos vies à l'heure des Big Data [Extrait]*. Paris, France : Seuil-La République des Idées. Consulté en ligne le 12/04/2016 : <http://www.seuil.com/livre-9782021279962.htm>
- Centre d'analyse stratégique de la République française. (2013). *Comment utiliser les technologies numériques pour poursuivre l'amélioration des relations entre l'administration et ses usagers [Note d'analyse n°317] ?* Consulté en ligne le 24/04/2016 : <http://archives.strategie.gouv.fr/cas/system/files/2013-01-23-tic-administration-usagers-na317.pdf>
- Chamaret, C. (2014). La révolution Big Data. *Annales des Mines - Gérer et comprendre* (2014/2), 116, 94-96. doi: 10.3917/geco.116.0094
- Chambre des représentants de Belgique. (2016, Avril 14). *Projet de loi relatif à la réutilisation des informations du secteur public [DOC 54 1619/004]*. Texte adopté en séance plénière et soumis à la sanction royale.
- Chen, H., Chiang, R.H.L., & Storey, V.C. (2012, Décembre). Business intelligence and analytics : From Big Data to big impact. *MIS Quarterly*, 36(4), 1165-1188. Consulté en

ligne le 12/04/2016 : <http://www.misq.org/misq/downloads/download/editorial/567/%3FSID%3Djau9e1kjh4ba7i79d4vad02om3>

Chui, M., Farrell, D., & Jackson, K. (s.d.). *How government can promote Open Data and help unleash over \$3 trillion in economic value*. McKinsey&Company. Consulté en ligne le 23/04/2016 :

[http://www.mckinsey.com/~media/mckinsey/industries/public%20sector/our%20insights/how%20government%20can%20promote%20open%20data/how_govt_can_promote_open_data_and_help_unleash_over_\\$3_trillion_in_economic_value.ashx](http://www.mckinsey.com/~media/mckinsey/industries/public%20sector/our%20insights/how%20government%20can%20promote%20open%20data/how_govt_can_promote_open_data_and_help_unleash_over_$3_trillion_in_economic_value.ashx)

Commission de la protection de la vie privée [CPVP]. (2016). *La Commission vie privée veille à ce que les données à caractère personnel soient utilisées et sécurisées soigneusement, et que votre vie privée future soit également garantie [Site web]*. Consulté en ligne le 01/05/2016 : <https://www.privacycommission.be/>

Commission européenne. (2015). *Rapport 2015 pour la Belgique contenant un bilan approfondi sur la prévention et la correction des déséquilibres macroéconomiques*. Consulté en ligne le 21/04/2016 : http://ec.europa.eu/europe2020/pdf/csr2015/cr2015_belgium_fr.pdf

Damon, J., Messatfa, H., & Chabut, D. (2011). L'innovation technologique au service d'une protection sociale plus « smart ». *Regards (revue de l'En3s)*, 40, 35-57. Consulté en ligne le 24/04/2016 : <http://eclairs.fr/wp-content/uploads/2011/01/articlejdhammouvers3-1.pdf>

Data.be SPRL. (2016). *Data.be business info*. Consulté en ligne le 22/04/2016 : <https://data.be>

Datanews. (2015, Janvier 22). Charles Michel demande l'aide de Google pour repérer les terroristes. *Le Vif*. Consulté en ligne le 30/04/2016 : <http://datanews.levif.be/ict/actualite/charles-michel-demande-l-aide-de-google-pour-reperer-les-terroristes/article-normal-362791.html>

De Keyser, P. (2016, Avril 2). Hercule Poirot et le syndrome belge. *L'Echo*. Consulté en ligne le 23/04/2016 : http://www.lecho.be/actualite/archive/Hercule_Poirot_et_le_syndrome_belge.9750650-1802.art

- Desouza, K.C. (2014). *Realizing the promise of Big Data : Implementing Big Data projects*. IBM Center for the Business of Government. Consulté en ligne le 12/04/2016 : <http://www.businessofgovernment.org/sites/default/files/Realizing%20the%20Promise%20of%20Big%20Data.pdf>
- Eggers, W. D., & Bellman, J. (2015, Octobre 2). The journey to government's digital transformation [A Deloitte Digital global survey]. *Deloitte University Press*. Consulté en ligne le 08/05/2016 : http://d27n20517rookf.cloudfront.net/wp-content/uploads/2015/10/DUP_1081_Journey-to-govt-digital-future_MASTER.pdf
- European Parliament. (2014). *Mapping Smart Cities in the EU [Study]*. Consulté en ligne le 25/04/2016 : <http://www.smartcities.at/assets/Publikationen/Weitere-Publikationen-zum-Thema/mappingsmartcities.pdf>
- Faljaoui, A. (2016a, Avril 26). Efficaces et moins chers, les robots s'appêtent à braquer les banquiers. *Trends-Tendances*. Consulté en ligne le 27/04/2016 : http://trends.levif.be/economie/banque-et-finance/efficaces-et-moins-chers-les-robots-s-appretent-a-braquer-les-banquiers/article-opinion-494035.html?utm_source=Newsletter-27/04/2016&utm_medium=Email&utm_campaign=Newsletter-RNBTECZ
- Faljaoui, A. (2016b, Avril 27). La nouvelle arme de Big Brother pour nous surveiller. *Trends-Tendances*. Consulté en ligne le 27/04/2016 : http://trends.levif.be/economie/high-tech/la-nouvelle-arme-de-big-brother-pour-nous-surveiller/article-opinion-494499.html?utm_source=Newsletter-27/04/2016&utm_medium=Email&utm_campaign=Newsletter-RNBTECZ
- Gantz, J., & Reinsel, D. (2011). *Extracting value from chaos*. IDC iView. Consulté en ligne le 12/04/2016 : <https://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf>
- Gartner. (2015a). *Gartner says it's not just about Big Data; it's what you do with it: Welcome to the algorithmic economy [Press release]*. Consulté en ligne le 30/03/2016 : <http://www.gartner.com/newsroom/id/3142917>
- Gartner. (2015b). *Innovating with analytics : 40 real world examples in 40 minutes [Presentation]*. Las Vegas, NV : Gartner Business Intelligence & Analytics Summit

- (30 March – 1 April 2015). Consulté en ligne le 12/04/2016 :
http://www.gartner.com//it/content/3163300/3163318/innovating_with_analytics_lane_y.pdf?userId=86196850
- Gartner. (2016a). *IT glossary : Big Data*. Consulté en ligne le 24/03/2016 :
<http://www.gartner.com/it-glossary/big-data/>
- Gartner. (2016b). *IT glossary : Internet of Things*. Consulté en ligne le 09/05/2016 :
<http://www.gartner.com/it-glossary/internet-of-things>
- Gartner. (2016c). *IT glossary : Open Data*. Consulté en ligne le 26/03/2016 :
<http://www.gartner.com/it-glossary/open-data>
- Google. (2016). *Google self-driving car project [Website]*. Consulté en ligne le 01/05/2016 :
<https://www.google.com/selfdrivingcar/>
- GovLab Research. (2013). *Toward reimagining governance : Mapping the pathway toward more effective and engaged governance*. Consulté en ligne le 12/04/2016 :
<http://thegovlab.org/wp-content/uploads/2013/06/GovLabMapDocument.pdf>
- Groll, T., O'Halloran, S., Maskey, S., McAllister, G., & Park, D. (2015, Décembre). Big Data and the regulation of banking and financial services. *Banking & Financial Services Policy Report*, 34(12). Consulté en ligne le 12/04/2016 :
<http://web.a.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=68969e8d-9310-40a1-96b2-0863e9cdf21f%40sessionmgr4004&vid=4&hid=4106>
- Grosdhomme Lulin, E. (2015). *Gouverner à l'ère du Big Data : Promesses et périls de l'action publique algorithmique*. Paris, France : Institut de l'Entreprise. Consulté en ligne le 12/04/2016 : http://www.paradigmes.com/wp-content/uploads/Gouverner_Ere_Big_Data_IDEP.pdf
- Groupement français de l'industrie de l'information [GFII]. (2012). *Dossier de synthèse de la journée d'étude du GFII « Big Data : exploiter de grands volumes de données : quels enjeux pour les acteurs du marché de l'information et de la connaissance ? »*. Paris, France : Maison de l'Europe. Consulté en ligne le 12/04/2016 :
http://www.gfii.fr/uploads/docs/BigData_synth%C3%A8seVF.pdf

- Gualtieri, M. (2013). *The Forrester WaveTM : Big Data predictive analytics solutions, Q1 2013*. Cambridge, MA : Forrester Research Inc. Consulté en ligne le 12/04/2016 : <http://www.sas.com/resources/asset/Forrester85601-LR8KBD.pdf>
- Hamel, M.-P., & Marguerit, D. (2013). Quelles possibilités offertes par l'analyse des Big Data pour améliorer les téléservices publics ? *Revue Française d'Administration Publique* (2013/2), 146, 437-448. doi: 10.3917/rfap.146.0437
- Hellobank! (s.d.). *Une banque et plus que ça [Site web]*. Consulté en ligne le 01/05/2016 : <https://www.hellobank.be>
- Henno, J. (2016, Mars 8). Comment le Big Data va bousculer le crédit. *Les Echos*. Consulté en ligne le 23/04/2016 : <http://www.lesechos.fr/idees-debats/sciences-prospective/021739541126-comment-le-big-data-va-bousculer-le-credit-1205427.php#>
- HeroQuizz. (2016). *Découvre le statut que tu posteras dans 10 ans ! [Site web]*. Consulté en ligne le 13/04/2016 : <http://fr.heroquizz.com/s/p64k5fwvg0>
- Hewlett-Packard Development Company [HP]. (2013). *Capitalize on Big Data in financial services [Business white paper]*. Consulté en ligne le 12/04/2016 : <http://www8.hp.com/h20195/V2/getpdf.aspx/4AA4-8165ENW.pdf?ver=1.0>
- InnovaTech. (s.d.). *Big Data : D'incroyables opportunités pour nos entreprises et nos territoires*. Les dossiers d'InnovaTech. Consulté en ligne le 12/04/2016 : http://www.innovatech.be/upload/espace_innovation/articles_interviews/Innoflash_Big_Data.pdf
- INRIX. (2015). *Le panorama du trafic : 2015*. Consulté en ligne le 21/04/2016 : http://inrix.com/wp-content/uploads/2016/03/INRIX-2015-Traffic-Scorecard_FR.pdf
- INRIX. (s.d.). *Data driven : Connecting cars to smarter cities*. Consulté en ligne le 21/04/2016 : <http://inrix.com/wp-content/uploads/2014/10/Smart-Cities-Connect-Car-Brochure.pdf>

- Institut des Actuairens en Belgique. (2015). *Big Data : An actuarial perspective [Information paper]*. Consulté en ligne le 12/04/2016 : http://www.iabe.be/sites/default/files/bijlagen/Big_Data_paper_full_v009.pdf
- International Business Machines [IBM]. (2012). *Big Data brings big opportunities for insurers*. Consulté en ligne le 12/04/2016 : http://www-935.ibm.com/services/uk/en/attachments/pdf/IBM_BAO_Big_Data_Insurance_WEB.pdf
- International Business Machines [IBM]. (s.d.a). *Bringing Big Data to the enterprise : What is Big Data ?* Consulté en ligne le 25/03/2016 : <http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>
- International Business Machines [IBM]. (s.d.b). *Dictionary of IBM and computing terminology*. Consulté en ligne le 26/03/2016 : <https://www-03.ibm.com/ibm/history/documents/pdf/glossary.pdf>
- iPlant Collaborative. (s.d.). *Transforming science through data-driven discovery [Site web]*. Consulté en ligne le 01/05/2016 : <http://www.iplantcollaborative.org>
- Janssen, M., & van den Hoven, J. (2015). Big and Open Linked Data (BOLD) in government: A challenge to transparency and privacy? *Government Information Quarterly*, 32, 363–368. doi: 10.1016/j.giq.2015.11.007
- Kim, G.-H., Trimi, S., & Chung, J.-H. (2014, Mars). Big-Data applications in the government sector. *Communications of the Association for the Computing Machinery*, 57(3). doi: 10.1145/2500873
- Kitchin, R. (2014). The real-time city? Big Data and smart urbanism. *GeoJournal*, 79, 1–14. doi: 10.1007/s10708-013-9516-8
- Laney, D. (2001, Février 6). *3D Data management : Controlling Data volume, velocity and variety [File 949]*. META Group Application Delivery Strategies. Consulté en ligne le 12/04/2016 : <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>

- Malengreau, D. (2016, Mars 9). Un « Google de la législation belge » voit le jour. *L'Echo*. Consulté en ligne le 21/04/2016 : http://www.lecho.be/dossier/startups/Un_Google_de_la_legislation_belge_voit_le_jour.9742092-8019.art?ckc=1
- Mayer-Schönberger, V., Cukier, K. (2013). *Big Data : A revolution that will transform how we live, work, and think*. Londres, Royaume-Uni : John Murray.
- McKinsey Global Institute [MGI]. (2011). *Big Data : The next frontier for innovation, competition, and productivity*. McKinsey&Company. Consulté en ligne le 12/04/2016 : <http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation>
- McKinsey Global Institute, McKinsey Center for Government, & McKinsey Business Technology Office [MGI, MCG & BTO]. (2013). *Open Data : Unlocking innovation and performance with liquid information*. McKinsey&Company. Consulté en ligne le 12/04/2016 : <http://www.mckinsey.com/business-functions/business-technology/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information>
- Metge, P. (2015). Le Big Data et la banque. *Revue d'Economie Financière* (2015/2), 118, 93-101. doi: 10.3917/ecofi.118.0093
- Microsoft. (2016). *Data Driven Keynote 1 : Data is the new electricity [Video]*. Consulté en ligne le 14/04/2016 : <https://www.microsoft.com/en-us/server-cloud/data-driven.aspx>
- Mobistar. (2016, Avril). Payer avec un selfie, son cœur ou sans les mains. *Mobistar Newsletter*. Consulté en ligne le 30/04/2016 : https://www.imgnm.be/dw/hello/2016/avril16/post/mobistar_avril16_fr_art2.html?utm_source=Newsletter&utm_medium=email&utm_term=hello&utm_campaign=avril2016
- Neyfakh, L. (2015, Janvier/Février). The Future of getting arrested. *The Atlantic*. Consulté en ligne le 13/04/2016 : <http://www.theatlantic.com/magazine/print/2015/01/the-future-of-getting-arrested/383507/>
- Ninon, R. (2016, Janvier 5). Pourquoi Orange peut faire trembler les banques. *Les Echos*. Consulté en ligne le 30/03/2016 : <http://www.lesechos.fr/finance-marches/banque->

assurances/021598147412-pourquoi-orange-peut-faire-trembler-les-banques-1189734.php

Olanrewaju, T. (2014). *The rise of the digital bank*. McKinsey&Company. Consulté en ligne le 12/04/2016 : <http://www.mckinsey.com/business-functions/business-technology/our-insights/the-rise-of-the-digital-bank>

Open Lex SPRL. (2016). *Lex.be*. Consulté en ligne le 22/04/2016 : <https://lex.be>

Oracle. (2015). *Big Data in financial services and banking : Architect's guide and reference architecture introduction [White paper]*. Consulté en ligne le 12/04/2016 : <http://www.oracle.com/us/technologies/big-data/big-data-in-financial-services-wp-2415760.pdf>

Organisation de Coopération et de Développement Economiques [OCDE] Belgique. (2015). *Etudes économiques de l'OCDE Belgique : Février 2015*. Consulté en ligne le 21/04/2016 : http://www.oecd.org/fr/economie/etudes/Overview_Belgique_2015_Fr.pdf

Oxford English Dictionary. (2016). *Oxford English Dictionary : Big Data*. Consulté en ligne le 23/03/2016 : <http://www.oed.com/view/Entry/18833#eid301162177>

Petit Robert. (2010). *Le nouveau Petit Robert : Dictionnaire alphabétique et analogique de la langue française*. Paris, France : Le Robert.

Press, G. (2014). 12 Big Data definitions : What's yours? *Forbes*. Consulté en ligne le 23/03/2016 : <http://www.forbes.com/sites/gilpress/2014/09/03/12-big-data-definitions-whats-yours/#3d651ef621a9>

PricewaterhouseCoopers [PwC]. (2013a). *Capitalizing on the promise of Big Data : How a buzzword morphed into a lasting trend that will transform the way you do business*. Consulté en ligne le 12/04/2016 : <https://www.pwc.com/us/en/increasing-it-effectiveness/assets/capitalizing-on-the-promise-of-big-data.pdf>

PricewaterhouseCoopers [PwC]. (2013b). *Where have you been all my life ? How the financial industry services can unlock the value in Big Data*. Consulté en ligne le

12/04/2016 : <https://www.pwc.com/us/en/financial-services/publications/viewpoints/assets/pwc-unlocking-big-data-value.pdf>

Quoistiaux, G. (2016, Mars 25). TIMi, le logiciel belge qui prédit l'avenir. *Trends-Tendances*. Consulté en ligne le 23/04/2016 : <http://trends.levif.be/economie/high-tech/numerik/timi-le-logiciel-belge-qui-predit-l-avenir/article-normal-482865.html>

Rodríguez-Bolívar, M. P. (2015). *Transforming city governments for successful Smart Cities*. Springer International Publishing Switzerland. doi: 10.1007/978-3-319-03167-5

Schenk, T., Jr., & Berman, B. (2015). Big Data and analytics : Driving performance and improving decision-making [Research report]. *Public CIO : The Center for Digital Government (2015/Q3)*. Consulté en ligne le 12/04/2016 : https://www.vion.com/assets/site_18/files/vion%20collateral/pcio15_special%20report%20q3_v.pdf

Service Public Fédéral Finances. (s.d.). *Tax-on-web [Site web]*. Consulté en ligne le 01/05/2016 : www.taxonweb.be

Smart City Institute HEC-ULg. (2015). *Smart Cities in Belgium: Qualitative analysis of 11 projects [Executive summary]*. Consulté en ligne le 25/04/2016 : <http://labos.ulg.ac.be/smart-city/wp-content/uploads/sites/12/2015/10/Executive-Summary-EN.pdf>

TechAmerica Foundation's Federal Big Data Commission. (s.d.). *Demystifying Big Data : A practical guide to transforming the business of government*. TechAmerica Foundation. Consulté en ligne le 12/04/2016 : <https://www-304.ibm.com/industries/publicsector/files/serve?contentid=239170>

Techopedia. (2016). *Dictionary : Non-relational database*. Consulté en ligne le 26/03/2016 : <https://www.techopedia.com/definition/25218/non-relational-Database>

Trends-Tendances. (2016, Avril 29). *Belfius dévoile sa stratégie et ses objectifs d'ici à 2020 (vidéo)*. Consulté en ligne le 29/04/2016 : <http://trends.levif.be/economie/banque-et-finance/belfius-devoile-sa-strategie-et-ses-objectifs-d-ici-a-2020-video/article-normal->

495229.html?utm_campaign=Newsletter-
RNBTECZ&utm_medium=Email&utm_source=Newsletter-29%2F04%2F2016

Turner, D., Schroeck, M., & Shockley, R. (2013). *Analytics : The real-world use of Big Data in financial services : How innovative banking and financial markets organizations extract value from uncertain Data [Executive report]*. IBM Corporation : IBM Institute for Business Value. Consulté en ligne le 12/04/2016 : http://www-935.ibm.com/services/multimedia/Analytics_The_real_world_use_of_big_data_in_Financial_services_Mai_2013.pdf

United Nations. (2014). *United Nations e-government survey 2014*. Consulté en ligne le 12/04/2016 : http://unpan3.un.org/egovkb/portals/egovkb/documents/un/2014-survey/e-gov_complete_survey-2014.pdf

Van Gyseghem, J.-M. (2016). *L'e-gouvernement*. Centre de Recherche Information, Droit et Société (CRIDS) – Université de Namur. Consulté en ligne le 22/04/2016 : http://lola.hec.ulg.ac.be/pluginfile.php/40646/mod_resource/content/1/Jean%20Marc%20Van%20Gyseghem%201101.pdf

Vesset, D., Morris, H.D., Little, G., Borovick, L., Feldman, S., Eastwood, M., ... Yezhkova, N. (2012). *Market analysis : Worldwide Big Data technology and services (2012-2015 forecast)*. IDC. Consulté en ligne le 12/04/2016 : http://ec.europa.eu/information_society/newsroom/cf/dae/document.cfm?doc_id=6242

Wang, R. R. (2012, Février 27). *Monday's musings: Beyond the three V's of Big Data—viscosity and virality*. Consulté en ligne le 24/03/2016 : <http://blog.softwareinsider.org/2012/02/27/mondays-musings-beyond-the-three-vs-of-big-data-viscosity-and-virality/>

Washington, A. L. (2014, Juillet). Government information policy in the era of Big Data. *Review of Policy Research*, 31, 319-325. doi: 10.1111/ropr.12081

Yiu, C. (2012). *The Big Data opportunity : Making government faster, smarter and more personal*. Londres, Royaume-Uni : Policy Exchange. Consulté en ligne le 24/04/2016 : <http://www.policyexchange.org.uk/images/publications/the%20big%20data%20opportunity.pdf>

Zikopoulos, P.C., Eaton, C., deRoos, D., Deutsch, T., & Lapis, G. (2012). *Understanding Big Data: analytics for enterprise class Hadoop and streaming data*. McGraw-Hill.

Consulté en ligne le 12/04/2016 :

<http://public.dhe.ibm.com/common/ssi/ecm/im/en/im114296usen/IML14296USEN.PDF>

Zuiderwijk, A., & Janssen, M. (2015). Towards decision support for disclosing Data : Closed or Open Data ? *IOS Press : Information Polity*, 20, 103–117. doi: 10.3233/IP-150358

Table des matières

Introduction	1
PARTIE 1 : Qu'est-ce que le Big Data ?	5
1. Une définition du Big Data	5
2. Les « V » du Big Data.....	7
2.1. Les trois « V » de base	7
2.1.1. Le Volume du Big Data.....	8
2.1.2. La Variété du Big Data	8
2.1.3. La Vitesse du Big Data.....	9
2.2. Toujours plus de « V ».....	10
2.2.1. La Véracité du Big Data	10
2.2.2. La Valeur du Big Data.....	10
2.2.3. La Visualisation du Big Data.....	11
2.2.4. La Viscosité, Variabilité et Volatilité du Big Data.....	11
2.2.5. La Vulnérabilité du Big Data.....	11
3. Les Big Data analytics.....	12
4. Les sources du Big Data.....	14
4.1. Trois catégories de sources	14
4.1.1. Les sources dirigées.....	14
4.1.2. Les sources automatisées.....	15
4.1.3. Les sources volontaires.....	16
4.2. L'Open Data.....	16
5. Les outils pour manier le Big Data.....	17
5.1. La plateforme Apache Hadoop	17
5.2. Les bases de données non-relationnelles	18
5.3. Le cloud computing	18

5.4.	Les systèmes de data mining.....	18
6.	Tour d’horizon de l’impact du Big Data	18
6.1.	Secteur de la santé.....	19
6.2.	Secteur de la vente au détail.....	19
6.3.	Domaine de la politique et des services publics	20
6.4.	Prévention des cyber-risques	20
6.5.	Secteur de la vente en ligne.....	21
6.6.	Sciences et technologie	21
6.7.	Secteur des entreprises B2C.....	21
7.	Les risques du Big Data	22
8.	Notre définition du Big Data.....	24
PARTIE 2 : Le Big Data dans le secteur financier		25
1.	Les points communs entre les deux secteurs.....	25
1.1.	Les données.....	26
1.2.	La relation avec le client	29
2.	Le Big Data et la banque	29
2.1.	Les grands changements à l’œuvre	30
2.2.	L’amélioration de la gestion financière et des risques.....	31
2.2.1.	Optimisation de l’allocation entre actifs et passifs.....	31
2.2.2.	Meilleure gestion du risque	31
2.2.3.	Prévention plus efficace de la fraude.....	32
2.2.4.	Identification des menaces de cyber-attaques	32
2.2.5.	Optimisation des prix.....	32
2.2.6.	Amélioration de l’efficacité opérationnelle	33
2.3.	L’amélioration de la relation client.....	33
2.3.1.	Une analyse à grande échelle pour un marketing et un service personnalisé. 34	
2.3.2.	Un besoin anticipé et une offre au bon moment.....	34

2.3.3.	Un service qui tient mieux compte du feedback client.....	35
2.3.4.	Un contrôle du risque de réputation	35
2.4.	La naissance de la banque digitale.....	35
3.	Le Big Data et la compagnie d'assurances.....	37
3.1.	Les analyses prédictives.....	37
3.2.	L'amélioration du marketing et du service pour le client	38
3.2.1.	Une meilleure connaissance des besoins du client	39
3.2.2.	Un marketing ciblé sur l'anticipation des besoins.....	39
3.2.3.	Un marketing ciblé sur la géolocalisation	40
3.2.4.	Un service innovant et de meilleure qualité	40
3.2.5.	Une meilleure gestion de la réputation	41
3.3.	Des primes d'assurance qui reflètent les risques réels.....	41
3.3.1.	Des contrats d'assurance basés sur les comportements.....	42
3.3.2.	Réduction des coûts liés aux déclarations de sinistre	43
3.3.3.	Récompenser le client pour une bonne attitude.....	43
3.3.4.	Une meilleure évaluation des risques liés à l'immobilier.....	43
3.4.	Identification et prévention des risques de fraude	44
4.	Les perspectives d'avenir	44
5.	Les risques du Big Data dans le secteur financier.....	46
6.	Les grandes tendances	47
6.1.	La personnalisation	47
6.2.	La prédictibilité.....	48
6.3.	L'optimisation.....	48
6.4.	L'objectivation	48
PARTIE 3 : Une Data Driven Governance		49
1.	Comment l'Etat peut-il utiliser le Big Data ?.....	50
1.1.	La personnalisation du service au citoyen	51

1.1.1.	De la standardisation à la personnalisation.....	51
1.1.2.	Un exemple dans l'enseignement	52
1.2.	La prédictibilité : prévenir la fraude et les menaces	54
1.2.1.	Prévention de la fraude	54
1.2.2.	Baisse de la criminalité.....	56
1.2.3.	Détection des menaces terroristes.....	57
1.3.	L'optimisation interne.....	58
1.3.1.	Optimisation des processus internes.....	58
1.3.2.	Optimisation du fonctionnement du pays.....	60
1.4.	L'objectivation de la prise de décision et du jugement.....	61
1.4.1.	Transparence des données et Open Data	62
1.4.2.	Mieux faire respecter la loi	63
1.5.	Un Etat à l'image de la « Smart City »	64
2.	Quels sont les challenges et les risques ?	65
2.1.	Les challenges pour l'Etat.....	66
2.1.1.	Un challenge technique	66
2.1.2.	Un challenge de compétences.....	66
2.1.3.	Un challenge de mentalité	66
2.2.	Les risques pour le citoyen.....	67
2.2.1.	Un risque de sécurité des données	67
2.2.2.	Un risque pour les libertés individuelles	68
	Conclusion.....	69
	Annexes.....	75
	Annexe 1 : Que savons-nous vraiment du Big Data ?.....	75
	Annexe 2 : Que se passe-t-il en 60 secondes sur Internet ?	76
	Annexe 3 : les 4 « V » du Big Data selon IBM.....	77
	Annexe 4 : Notre schéma des « V » du Big Data.....	79

Annexe 5 : Les différents types de « data »	80
Annexe 6 : Données stockées par secteur aux Etats-Unis en 2009	81
Annexe 7 : Les e-gouvernements leaders mondiaux en 2014.....	82
Annexe 8 : Etude de cas – Big Data et service public de l’emploi	83
Annexe 9 : Etude de cas – Big Data et santé publique.....	84
Annexe 10 : Modèle de maturité digitale du gouvernement	85
Annexe 11 : Big Data, gouvernement et Data Driven Governance	86
Références bibliographiques	89

Executive summary

We live in a world overwhelmed by data. The increasing quantity of data which is impacting the way we live and do business is called “Big Data”. Because of the amount of data they possess, the Big Data revolution is a great value-generating opportunity for at least three sectors: banking, insurance and public governance.

In this context, we study the possibility of a Data Driven Governance. We wonder whether the State would be able to use Big Data to improve its governance and public services in the same way the bank and the insurance company use Big Data to generate profits. To answer this question, we divide our work in three parts.

Firstly, we define Big Data. Although this concept is new, scientists and technology leaders agree that Big Data is characterised by at least three “Vs”: Volume, Variety and Velocity. They also agree on the fact that using Big Data analytics can create a lot of value. However, that implies important technical challenges for companies willing to manipulate Big Data. Furthermore, after studying its sources, its tools and its overall impact on the economy and society, we underline that Big Data also presents risks which should not be neglected.

Secondly, we study the impact of Big Data on the financial sector, more precisely on the banking and insurance sectors. After analysing each sector individually, we summarise the Big Data revolution in the bank and the insurance company through four trends: personalisation, predictability, optimisation and objectification.

Thirdly, we transpose these four trends to the public governance sector in order to study the impact of Big Data on governments. This allows us to evaluate the possibility of a Data Driven Governance, i.e. a State using Big Data to personalise its services towards the citizen, to predict and prevent fraud and threats, to optimise the way the administration and the country work as well as to make decision-making process and judgment more objective.

To conclude, our answer to the question of a Data Driven Governance is: yes, but... Yes, the State can use Big Data to improve its governance in the same way the financial sector uses it to generate profits. Yet, some important questions arise, the more crucial one being: can “Big Data” become “Big Brother” and “Data Driven Governance” become “Data Driven Dictatorship”?