



Le Big Data en santé
Note complémentaire n° 1 du
rapport " État des lieux de
l'innovation en santé numérique "

Professeur Hervé Dumez
Professeur Etienne Minvielle
Madame Laurie Marraud

novembre 2015

En partenariat avec



EHESP



Le Big Data en Santé

« État des lieux de l'Innovation en Santé Numérique » : Note complémentaire n° 1

Laurie Marraud, Hervé Dumez, Etienne Minvielle

29 Mai 2015

Table des matières

I. Big Data : Définition et caractéristiques.....	3
Le volume	3
La vélocité.....	3
La variété	4
Les 3V, un changement de paradigme dans le traitement des données.....	4
<i>Le volume, une exhaustivité qui casse les échantillons</i>	<i>4</i>
<i>La vélocité, une perspective longitudinale méconnue jusque-là.....</i>	<i>5</i>
<i>La variété, une fiabilité des données</i>	<i>5</i>
<i>Le bouleversement des forces traditionnelles de la concurrence.....</i>	<i>6</i>
II. Des pistes de développement pour la Mutualité française	6
Les exploitations possibles du Big Data dans le domaine de la santé	7
<i>Soutien à l'intégration de l'offre de soin</i>	<i>7</i>
<i>Service de prévention personnalisée, Evaluation Comportementale et Prédiction.....</i>	<i>7</i>
<i>Le patient autonome : Une nouvelle figure à accompagner</i>	<i>8</i>
<i>Des alliances stratégiques</i>	<i>8</i>
Formulation pour une stratégie de développement.....	9
<i>Aspects Technologiques : Les infrastructures technologiques de l'information et les</i> <i>ressources humaines dans le traitement des données.....</i>	<i>10</i>
<i>Aspects économiques.....</i>	<i>11</i>
<i>Aspects organisationnels: La confiance numérique</i>	<i>11</i>

La notion de Big Data, comme son application en santé, restent encore assez confuses. Cette note vise dans un premier temps à donner des éléments de définition, et des repères sur l'évolution du marché relatif à son essor. Dans un second temps, des pistes sont proposées quant à une stratégie de développement de la part de la Mutualité Française dans ce domaine. L'ensemble des informations réunies sont fondées sur une revue de la littérature ainsi que sur une série d'entretiens. Les pistes avancées n'engagent que les auteurs, et répondent à la demande exprimée par la Fondation de l'Avenir.

I. Big Data : Définition et caractéristiques

C'est au tout début des années 2000 que furent énoncées les principales caractéristiques du Big Data qui font à ce jour toujours référence (Laney, 2001). On ne parlait d'ailleurs pas encore à l'époque de Big Data, mais simplement de « *Data* » ou de « *Data management* ». Ainsi, la production de données nouvelles permises par les outils connectés et Internet allaient se définir suivant trois caractéristiques principales, aussi appelées les 3V: le volume, la vitesse et la variété (Laney, 2001; McAfee & Brynjolfsson, 2012; Roski, Bo-Linn, & Andrews, 2014).

Le volume

Dès les années 2000, l'importance des données disponibles sur Internet est soulignée. Dix ans plus tard, c'est environ 2,5 exaoctets de données qui sont créés chaque jour, et ce nombre double tous les 40 mois ou plus. Aujourd'hui les entreprises travaillent sur des pétaoctets de données (c'est à dire 10^{15} ou un million de milliards de données, soit l'équivalent en texte d'environ 20 millions d'armoires de classeurs) (McAfee & Brynjolfsson, 2012). Par exemple, il est estimé que Walmart (entreprise américaine multinationale spécialisée dans la grande distribution) recueille plus de 2,5 pétaoctets de données toutes les heures à partir de ses opérations avec la clientèle.

Ces quantités massives de données mettent à rude épreuve la capacité et les moyens de stockage traditionnels, ainsi que la gestion et les systèmes de récupération tels que les serveurs de données. Le Big Data nécessite des solutions de stockage et de gestion de données flexibles et facilement extensibles (Roski et al., 2014).

La vitesse

La vitesse doit être comprise comme la capacité de réactivité offerte par le Big Data dans le traitement des données. Elle a rapidement été perçue dans le e-commerce comme un facteur de différenciation concurrentielle, jouant sur la réactivité d'un site Internet, l'analyse de la disponibilité des stocks, l'exécution d'une transaction, ou le suivi de commande.

Aujourd'hui, la donnée en "*temps réel*" offre un autre avantage concurrentiel fort dû au fait que nous sommes équipés d'outils mobiles connectés (McAfee & Brynjolfsson, 2012). Par exemple, les GPS fonctionnent sur ce principe d'analyse des données en temps réel. Dans le domaine de la santé publique, cette caractéristique peut aussi s'avérer particulièrement utile pour des détections précoces de symptômes, ou pour

orienter les populations en cas de catastrophe naturelle vers les lieux de soin pouvant les accueillir.

La variété

Dans les années 2000, la variété des données générait des problèmes d'interopérabilité. Si des efforts conséquents ont été fournis, beaucoup reste à faire aujourd'hui, notamment dans le domaine de la santé, pour trouver des solutions : les textes capturés par les dossiers électroniques, les données de l'imagerie, et les données en *streaming* de médias sociaux et applications mobiles représentent une source d'information précieuse. Avec le Big Data, les flux de données qui arrivent en quantités très importantes sont aussi très variés et émanent de sources protéiformes : données internes (SI interne, etc.), données externes (médias sociaux, mails, terminaux mobiles, etc.), données structurées (documents, images, etc.) ou non structurées (tweets, données GPS, capteurs, etc.). Ces données restent difficilement manipulables à travers les outils classiques de traitement (Karoui, Devauchelle & Duzert 2014).

Le phénomène appelle des capacités de partage des données d'un nouveau type, les serveurs rassemblant ces données dans un effort de triangulation pouvant être extérieurs à l'entreprise (ce qu'il est commun d'appeler le « *cloud computing* »).

C'est sur cette base des « 3V » que peuvent se comprendre les enjeux du Big Data.

Les 3V, un changement de paradigme dans le traitement des données

Les modèles de traitement des données évoluent en effet avec le Big Data. D'une manière générale, la tendance peut se résumer de la manière suivante : l'analyse des événements passés correspond à un modèle de reporting, l'analyse des événements en cours correspond à un modèle de monitoring, et l'analyse des événements à venir correspond à un modèle prédictif : or, avec le Big Data, les deux derniers modèles se transforment (Karoui, Devauchelle & Duzert, 2014). Plus précisément chacune des qualités du 3V introduit de nouvelles formes dans le traitement des données que certains considèrent comme un changement de paradigme.

Le volume, une exhaustivité qui casse les échantillons

Le volume du Big Data permet de s'affranchir d'un certain nombre de méthodes de calcul statistique traditionnel, en ce qu'il permet l'analyse de statistiques descriptives, c'est-à-dire de toutes les données réelles, et non pas uniquement de statistiques inférentielles, basées sur un échantillon. L'échantillonnage est comme la photographie imprimée : à distance, il donne l'illusion du détail mais, plus on s'approche et plus l'image est floue (Mayer-Schönberger & Kenneth, 2013).

A contrario, avec le Big Data, il s'agit d'observer la complexité au lieu de privilégier des faits particuliers d'une part (en postulant des hypothèses à vérifier), et de détecter un signal dans le « bruit de fond » d'autre part (dans les statistiques classiques, des signaux faibles peuvent échapper alors qu'ils peuvent constituer des événements annonciateurs de tendances lourdes). Le Big Data permet ainsi d'établir des corrélations inédites. Cependant celles-ci ne font pas toujours sens à la première lecture et, même si elles décrivent une réalité, ne doivent pas être considérées comme un matériau déjà interprété. Par exemple, on constate que les patients atteints de cancer qui combinent aspirine et jus d'orange guérissent mieux que les

autres. Un réflexe simple pourrait être d'en induire une causalité. Or celle-ci n'a pas grand sens. *"Big Data is about what, not why"* (Mayer-Schönberger & Cukier, 2013). Les hypothèses que génère le Big Data se fondent sur des corrélations réelles mais doivent être supervisées le plus souvent par des experts du domaine en question pour avoir la garantie de faire sens. Le développement d'algorithmes de plus en plus « *intelligents* », capitalisant sur les connaissances précédentes, ouvre pourtant à une voie encore plus ambitieuse dit du traitement non supervisé, au sens où l'intelligence artificielle se substitue à l'expertise humaine dans l'interprétation des résultats.

La vélocité, une perspective longitudinale méconnue jusque-là

La vélocité des données du Big Data offre des possibilités d'analyse en temps réel. Par exemple, en 2011, l'arrivée de l'ouragan Irene illustre bien la manière dont des données ouvertes et recueillies en temps réel ont pu faciliter les interventions d'urgence. A New York, les maisons de soins infirmiers déclaraient leur capacité en lits et les disponibilités hebdomadaire, *via* des données rafraîchies sur le site de données de santé de la ville. Alors que l'ouragan Irene approchait de New York City, les administrateurs de soins infirmiers à domicile ont utilisé ce recensement en temps réel pour identifier les emplacements où ils pouvaient évacuer les patients résidant dans la zone d'inondation. Cet exemple montre que la donnée ouverte et accessible en temps réel peut être perçue comme un « *bien commun* ». Son partage et son utilisation peuvent devenir ainsi plus légitimes au regard du bénéfice apporté à la communauté des patients (ou des citoyens).

Ces remontées d'information en temps réel sont aussi très intéressantes dans le cadre de l'analyse des stratégies comportementales. En général, ce type de données est très volatile, les individus pouvant répondre de façon assez différente à deux heures d'intervalle à des questionnaires administrés. Le Big Data permet de prendre des photos instantanées répétées de ces comportements à faible coût, et d'en dégager des tendances.

La variété, une fiabilité des données

Un des enjeux du Big Data est de garantir le croisement et l'analyse de toutes les données de différentes natures et provenance (données extérieures, données internes, structurées ou non). Le croisement de ces données, s'il peut entraîner une reproductibilité des erreurs, est un outil de précision et de fiabilité. Les corrélations émanant des analyses de ces bases de données (pouvant comporter 1 million de variables) peuvent permettre de mettre au jour des liens entre des phénomènes, et ainsi de faire émerger de façon très inductive des hypothèses de recherche *a priori* inconnues. En milieu organisationnel, l'analyse rapide, en temps réel, des données permet notamment de soutenir les managers en développant des modèles d'aide à la décision à la fois beaucoup plus englobants et beaucoup plus précis. *"Le big data est aussi une machine à produire de la micro décision, des millions de micro-décisions"* (Bénavent, 2014).

En complément des grands traits dressés autour de la notion du Big Data, il est intéressant de conclure cette partie sur les évolutions du marché engendrées par son essor.

Le bouleversement des forces traditionnelles de la concurrence

Dans un article de la *Havard Business Review*, les experts en stratégie (Porter & Heppelmann, 2014) décrivent la manière dont le Big Data vient modifier les règles traditionnelles de la concurrence. En dessinant de nouvelles alliances, on entre dans une dynamique de marché où le produit est associé à un service. Pour être précis, ce n'est pas à un remplacement du produit, mais plutôt une interconnexion et/ou à des recombinaisons de produits à laquelle le Big Data renvoie. L'entreprise UBER, par exemple, ne supprime pas les chauffeurs mais connecte les utilisateurs à ceux-ci. Cette interconnexion permet de fournir des synthèses et des analyses de données afin de concevoir des solutions prédictives et en temps réel pour optimiser les opérations complexes du client. Des entreprises élargissent ainsi leur champ de compétences.

Dans ce contexte, chaque organisation craint que les autres n'utilisent les données, ou ne définissent des alliances avec des organisations capables de le faire, comme un avantage concurrentiel (Roski et al., 2014). De nouvelles dynamiques de marché apparaissent en réaction. Ces nouvelles stratégies d'alliance contribuent à la montée en puissance d'un modèle collaboratif qui va progressivement marginaliser un capitalisme traditionnel : la «*valeur d'échange*» sur le marché va peu à peu être détrônée par la «*valeur partageable*» dans des écosystèmes collaboratifs qui ne cessent de grossir¹.

Une telle évolution du marché ne laisse pas pourtant d'inquiéter certains. Pour le chercheur et essayiste Evgeny Morozov (2014/2013), «*ces objets personnalisés à l'extrême vont certainement achever de détruire les quelques restes de solidarité et de respect mutuel laissés par des années de dogme néolibéral [...] On vit dans une période très confuse, caractérisée par ce que j'aime appeler «l'asymétrie épistémique» : les citoyens sont entièrement transparents, quantifiables, tandis que les institutions sont plus opaques que jamais [...] Prenez n'importe quel gadget de la Silicon Valley et vous verrez l'ensemble des présomptions qu'il contient sur le citoyen, le régime politique et plus largement sur la vie publique. Sans changement radical, Google fournira dans dix ans des services bancaires, des systèmes de santé et d'éducation.*» Les propos de Morozov cristallisent les tensions citoyennes face à ces "monstres" de technologies qui envahissent, bon gré mal gré, notre quotidien².

II. Des pistes de développement pour la Mutualité française

Dans le contexte qui vient d'être décrit, au moins dans ses grands principes, la Mutualité française peut avoir un enjeu à développer une stratégie dans le domaine du Big data. Le rapport précédent a souligné les impacts potentiels de la santé numérique dans le domaine de l'offre de soins et des métiers de l'assurance. Le traitement des données de santé à grande échelle peut représenter de multiples

¹ Interview d'Evgeny Morosov, Libération, mardi 20 avril 2015. Disponible en se référant au lien : http://ecrans.liberation.fr/ecrans/2015/04/20/les-technologies-sont-des-concentres-d-ideologies_1254606

² *Ibid*

enjeux comme nous allons le voir. Deux autres arguments peuvent d'emblée venir aussi justifier un tel investissement.

D'une part, la production des données de santé est amenée à être profondément modifiée. Le volume de données devrait être multiplié par 50 en 8 ans passant de 500 petaoctets de données (l'équivalent de 10 milliards de dossiers médicaux) à 25000 petaoctets (soit 500 milliards de dossiers médicaux). A titre d'exemple, le système intégré américain Kaiser Permanente qui possède un peu moins de 10 millions de membres, peut gérer jusqu'à 44 petaoctets de données. Dans ces conditions, de nouvelles bases de données vont exister. Des données remontées par les patients eux-mêmes, donnant une information de « *première main* », sont susceptibles de voir le jour. Le débat autour de l'accès aux données existantes risque aussi de perdre de son intensité, un cas emblématique étant celui de l'accès aux données de la base SNIIRAM de l'assurance maladie obligatoire.

D'autre part, la crainte du *statu quo* se répand largement face à l'évolution du marché décrite précédemment. Comme le souligne Michael Porter (2008), de nouvelles formes de concurrence et d'alliances vont se dessiner, présageant de larges restructurations, l'ensemble incitant à ne pas rester en marge du mouvement.

Différents objectifs peuvent être poursuivis dans la mise en œuvre d'une politique du Big Data au sein de la Mutualité française. Ils correspondent aux exploitations possibles dans le domaine de la santé.

Les exploitations possibles du Big Data dans le domaine de la santé

Soutien à l'intégration de l'offre de soin

Le Big Data présente une valeur économique et clinique dans le développement des offres de soin intégrées. Le rapport précédent avait souligné le rôle qu'il peut tenir dans les échanges entre professionnels et entre professionnels et patients via de systèmes de télémedecine et d'outils connectés. L'ensemble contribue à orienter et accompagner le patient durant son parcours (comme le réalise déjà le groupe américain Kaiser Permanente). Il faut ajouter à ce tableau l'apport du Big Data dans l'utilisation de systèmes de décision médicale qui joue sur l'automatisation de l'analyse des imageries et de la littérature médicale (comme par exemple, le logiciel « *Uptodate* »). De même, le Big Data introduit une valeur nouvelle par sa capacité à automatiser les processus de production (Peres, 2015). Précisément, le Big Data offre aux organisations la possibilité de traquer et d'étiqueter l'origine et l'utilisation des données dans tous les processus de production de service. Les applications sont potentiellement nombreuses, allant de la gestion de services techniques comme la stérilisation, à la gestion des rendez-vous, en passant par la traçabilité en matière de matériovigilance. Ceci est particulièrement adapté au cas des maladies chroniques, comme le rappelle Philippe Lemoine dans son rapport sur l'économie numérique rendu au gouvernement en novembre 2014.

Service de prévention personnalisée, Evaluation Comportementale et Prédiction

Le Big Data permet de détecter des modèles comportementaux ou de profils patients plus imperceptibles à petite échelle. Par exemple, en utilisant un logiciel de prédiction

en temps réel, il est possible de déceler, via les réseaux sociaux ou téléphones mobiles, des tendances suicidaires chez les vétérans de guerre (Roski et al., 2014). Le Big Data joue également sur la capacité à traiter le génome humain développant une nouvelle médecine dite prédictive. De ce fait, le risque encouru de développer une maladie ou de la voir s'aggraver devient de plus en plus connu, entraînant la possibilité de développer des services de prévention personnalisée.

Mais outre ces services, le Big Data peut avoir une capacité à évaluer le risque et le comportement des assurés dans la gestion de sa maladie ou en matière de prévention. Les intérêts sont, dans ce domaine, multiples. Ils peuvent se situer dans l'analyse des consommations de soins et des profils de patients dans l'idée de rompre avec cette image d'un payeur aveugle. Ils peuvent aussi améliorer la connaissance du risque individuel avec toutes les retombées au niveau de la couverture assurantielle. La disparition de l'aléa moral autour du risque modifie les comportements entre l'assureur et l'assuré. L'assureur peut être tenté de sélectionner les profils à faible risque. Mais l'assuré peut aussi sélectionner une assurance en fonction de la connaissance de son risque : faible coût si le risque est faible, large couverture en cas de risque élevé. La personnalisation du contrat d'assurance en fonction du risque renvoie à de nouveaux comportements dans lesquels des équilibres sont à trouver. Le respect de la valeur de solidarité dans l'accompagnement d'un assuré tout au long de son parcours de vie (incluant la phase préventive et curative) peut représenter un argument.

Le patient autonome : Une nouvelle figure à accompagner

Le patient est de plus en plus acteur de sa santé *via* les outils connectés et les forums communautaires. Le Big Data peut permettre d'aider le patient dans sa démarche d'autonomie, notamment par sa capacité à proposer des services d'analyse comparée avec d'autres patients atteints de maladies similaires, ou confrontés aux mêmes questions de santé. Les plateformes « *opensource* » sont ainsi souvent proposées afin d'aider les utilisateurs volontaires dans leur démarche (Hué, 2015), et font l'objet de développements spécifiques en santé. Mais on peut suspecter des usages des données ainsi remontées par les opérateurs de ces plateformes autres que le service de partage de données proposé. Nous avons souligné dans le rapport précédent le cas de *23andme* où les données remontées par les patients sur les effets de toxicité de certains médicaments sont ensuite revendues à des laboratoires pharmaceutiques. Cet exemple témoigne de l'apparition d'un nouveau modèle où le service procuré est offert, ce qui le rend très attractif pour l'utilisateur, mais n'est pourtant pas tout à fait gratuit dans la mesure où l'utilisateur délivre des données sur son comportement, ses habitudes et son état de santé. La relation entre le patient et l'opérateur peut y apparaître déséquilibrée en ce sens qu'elle est transparente dans un sens (le client livre beaucoup de lui-même *via* ses courriels et requêtes Internet) et très opaque dans l'autre (comment accéder aux usages que l'opérateur connaît de lui ?). L'accompagnement du patient dans sa démarche d'autonomie engage donc une certaine transparence, voire de confiance, quant à l'exploitation faite des données.

Des alliances stratégiques

La Mutualité peut trouver avec les industriels impliqués dans le marché du Big Data un objectif commun d'alliance stratégique. Elle peut jouer sur le plan de la prévention comme de l'offre de soins. Le rôle de la Mutualité n'y est probablement pas de

développer de nouveaux outils mais plutôt de considérer des services, par exemple en proposant des solutions de centralisation des données produites par les différents outils connectés sur des plateformes interopérables et modulaires. Cette vision pro-active dans la recherche d'alliance est portée par un objectif de développement partagé par de nombreux acteurs du système de santé. " *Mieux vaut s'intéresser à la façon dont nous exercerons la médecine demain plutôt que vouloir reproduire à tout prix un modèle ancien, au prétexte qu'il nous est familier.*" explique le Dr Jacques Lucas, vice-président du Conseil National de l'Ordre des Médecins (Berrier, 2015). Mais si le constat d'aller de l'avant est somme toute assez partagé, la recherche d'alliances d'un nouveau type avec les entreprises qui portent l'innovation dans ce domaine reste pour sa part moins courante, au moins à l'heure actuelle.

Formulation pour une stratégie de développement

Sur la base des exploitations possibles du Big data qui viennent d'être formulées, la Mutualité française semble avoir un intérêt à développer ses propres bases de données. Ces bases de données pourraient, grâce notamment à la participation de patients, devenir des sources précieuses d'information dans les cinq ans à venir. En s'engageant dans cette voie, la Mutualité ne ferait probablement que précéder un mouvement que l'on perçoit. L'arrivée du Big Data dans le domaine de la santé semble en effet devoir imposer de toute manière aux organisations du secteur une restructuration de leurs architectures IT. Ces dernières doivent à présent s'adapter à de nouveaux protocoles, de nouveaux délais, de nouveaux facteurs de risque, et de nouveaux mandats pour la gestion des données, en particulier dans le domaine de la vie privée et de la sécurité (Roski et al., 2014). Comment alors envisager une stratégie de développement ?

Une hypothèse souvent avancée pour accéder à un rôle de leadership dans le domaine du Big Data consiste à développer un modèle gratuit d'*Open Data*. C'est le mode de fonctionnement qu'ont adopté en leur temps Google ou Twitter. Le schéma consiste à obtenir par ce biais une position de croissance dans la remontée de données des assurés ou des patients dans des laps de temps relativement court. D'une manière générale, cette stratégie de développement est généralement déficitaire dans les premiers temps, la valorisation boursière aidant à assumer l'investissement initial. Cette stratégie concerne évidemment les sociétés capitalistiques, mais elle peut néanmoins également intéresser un acteur comme la Mutualité française. Le contenu de cette plateforme d'*Open Data* pourrait concerner des services correspondant aux trois axes de l'intégration de l'offre de soins, de l'accompagnement d'un patient autonome, et du développement de service de prévention personnalisée dans le cadre de l'activité d'assurance. Ce développement pourrait faire l'objet d'une phase d'expérimentation. Une deuxième étape viserait à assembler par la suite ces données d'un nouveau type ainsi obtenues avec des données existantes, notamment au niveau des systèmes d'information des opérateurs de la Mutualité française. Enfin, une troisième étape s'attacherait à développer de nouveaux modes de traitement des données.

La définition de la gouvernance d'une telle démarche, entre Mutualistes, avec d'autres acteurs, et avec les assurés, constitue d'emblée une étape cruciale. Certains

exemples, comme le Dossier Médical Personnel, témoignent de la difficulté à organiser le traitement de données partagées entre différents protagonistes.

D'une manière plus générale, le choix d'entrée dans une stratégie de développement du Big Data constitue pour toute organisation un triple enjeu: technologique, économique, et organisationnel. Technologique, car à ce jour la majorité des systèmes d'information des organisations (et le domaine de la santé ne fait pas exception) est incapable de traiter les flux de données massifs. Economique, car il s'agit de définir la valeur ajoutée de ces données et de leur utilisation au sein de l'organisation, et ce en conservant la confiance du client (Karoui, Devauchelle & Duzert, 2014). Et, enfin, organisationnel, car l'entreprise doit réfléchir à la nature des données à traiter et aux conditions de ces nouvelles formes de collecte de données. Dans ce cadre général, trois points contribuant à la stratégie de développement nous semblent émerger, notamment suite aux entretiens réalisés : (i) sur le plan technologique, les besoins en infrastructures de technologie de l'information nécessaires, et en matière de ressources humaines d'un nouveau genre à allouer à leur développement ; (ii) sur le plan économique, la méthode pour établir l'investissement le plus rentable ; (iii) sur le plan organisationnel, la confiance accordée par les utilisateurs, en l'occurrence les patients et les assurés, aux producteurs de données et à leur usage.

Aspects Technologiques : Les infrastructures technologiques de l'information et les ressources humaines dans le traitement des données

Face à la variété des données, les architectures de technologies de l'information doivent garantir l'interopérabilité des sources et formats de données pour en extraire un corpus exploitable. Les investissements à faire de ce point de vue peuvent s'envisager de deux manières : soit par l'apport de prestataires extérieurs, soit par une internalisation de l'ensemble de la démarche (stockage de données partagées impliquant notamment un lien avec les systèmes d'information existants, traitement à partir d'algorithmes, diffusion de l'information). Si le pilotage semble être logiquement du ressort de la Mutualité, il est probable que les compétences requises à certaines étapes nécessitent d'externaliser une partie du développement. Enfin, il est intéressant de remarquer que cette infrastructure reste aussi très dépendante d'un maillage territorial national en matière d'équipement des flux de données (passage à la fibre). Comme nous l'a rapporté Pierre-Jean Benghozi³, membre du collège de l'ARCEP (Agence de Régulation des Communications Electroniques et Postales), toute initiative d'intégration de données entre plusieurs structures reste très dépendante du niveau d'équipement dans la zone territoriale concernée. Des organisations situées dans des zones mal équipées peuvent freiner des développements régionaux ou nationaux.

Le Big Data réclame également une expertise en analyse des données à grande échelle que les entreprises possèdent généralement très rarement en interne. Ces experts sont appelés *datascientists* et possèdent des compétences variées. Ainsi, ils doivent posséder des connaissances importantes en mathématiques et informatique mais aussi en stratégie pour pouvoir relever les indicateurs pertinents. Ce type de profil très polyvalent est cependant difficile à trouver (Karoui, Devauchelle &

³ Interview réalisée par notre équipe de recherche le 26 Mai 2015.

Dudezert, 2014). Il est donc plus pertinent, de l'avis d'Alexandre Templier, co-fondateur de Quinten, société spécialisée dans la valorisation de données, de *“constituer ou faire appel à des équipes d'experts réunissant des compétences ingénieriques, de calcul ou informatiques et stratégiques ou managériales, avec si possible deux des trois compétences par personne à chaque fois”*⁴.

Aspects économiques

L'investissement dans le développement d'une ou plusieurs plateformes « *opensource* » qui rassembleraient des données de patients et d'autres existantes peut être important. Cette importance dépend de l'ambition donnée. Les dépenses sont liées à l'infrastructure technologique et aux ressources humaines nécessaires telles que décrites précédemment. Comme nous l'avons évoqué, l'investissement dans le Big Data n'est pas synonyme de rentabilité immédiate. Il existe un laps de temps qui peut s'avérer plus ou moins long, lié à la proposition de services offerts gratuits, ou à faible coût.

D'une manière générale, le retour d'investissement peut jouer sur plusieurs leviers. Certaines entreprises jouent sur les revenus de publicité liés au recours sur des sites Internet. De même, certaines entreprises (comme mentionné précédemment) jouent sur une revente des bases de données à des fins commerciales (c'est le cas de *23andme*). Comme l'a souligné Laurent Carreda, directeur d'Amerys, société spécialisée dans la sécurité des données, ces leviers ont en commun le risque de nuire à la confiance accordée par les utilisateurs. Certains retours d'investissement peuvent s'observer également par les bénéfices obtenues dans l'automatisation et la gestion en temps réel de différents processus de production. L'automatisation dans la traçabilité, l'étiquetage, et d'autres modalités, peut offrir un gain financier comme décrit dans d'autres secteurs (Peres, 2015). La gestion en temps réel de certaines maintenances peuvent également y contribuer. Enfin, des recettes générées par de nouveaux services, notamment dans le cadre de la prévention personnalisée, peuvent s'envisager. Mais sur ce dernier aspect, la définition du coût du service peut dans certains cas être influencé par la volonté de collecter de nombreuses données rapidement, et donc de limiter les frais d'accès.

Aspects organisationnels: La confiance numérique

La confiance accordée à toutes les données issues du Big Data est cruciale. Comme le disent Mc Afee & Brynjolfsson (2012) *“une violation grossière du sentiment que le public a de ses droits pourrait tromper la confiance minimale qu'on accorde aux entreprises qui collectent ces données”*. Laurent Carreda, directeur de la société Almerys, va plus loin en soulignant que cette confiance constitue un, voire le principal, enjeu du développement du Big Data, notamment dans le domaine de la santé⁵. Il propose en réponse la notion de « *confiance numérique* » qui revêt trois aspects.

Cette notion concerne d'abord la sécurité dans le stockage des données. Différents mécanismes peuvent octroyer aux organisations un contrôle supplémentaire sur la capacité de partager des informations sans compromettre la vie privée des patients ou libérer de données exclusives (Roski et al., 2014). Il y a évidemment les

⁴ Interview réalisée par notre équipe de recherche le 12 Mai 2015.

⁵ Interview réalisée par notre équipe de recherche le 22 Mai 2015.

autorisations d'accès, mais aussi des formes de stockage qui sanctuarisent dans des lieux inviolables les données tels que les « *data center bunkerisés* ».

Cette notion touche également la capacité offerte au patient de rester propriétaire dans le suivi de ces données. Comme dans de nombreux autres domaines, le patient devient un « *prosomateur* », à la fois consommateur et producteur⁶. A cet égard, Laurent Carreda souligne l'importance de la démarche contractuelle d'agrément avec l'utilisateur. Si celui-ci est le patient, il est fondamental de garantir un système de signature électronique combiné à un face-à-face avec l'opérateur avant toute transaction de données.

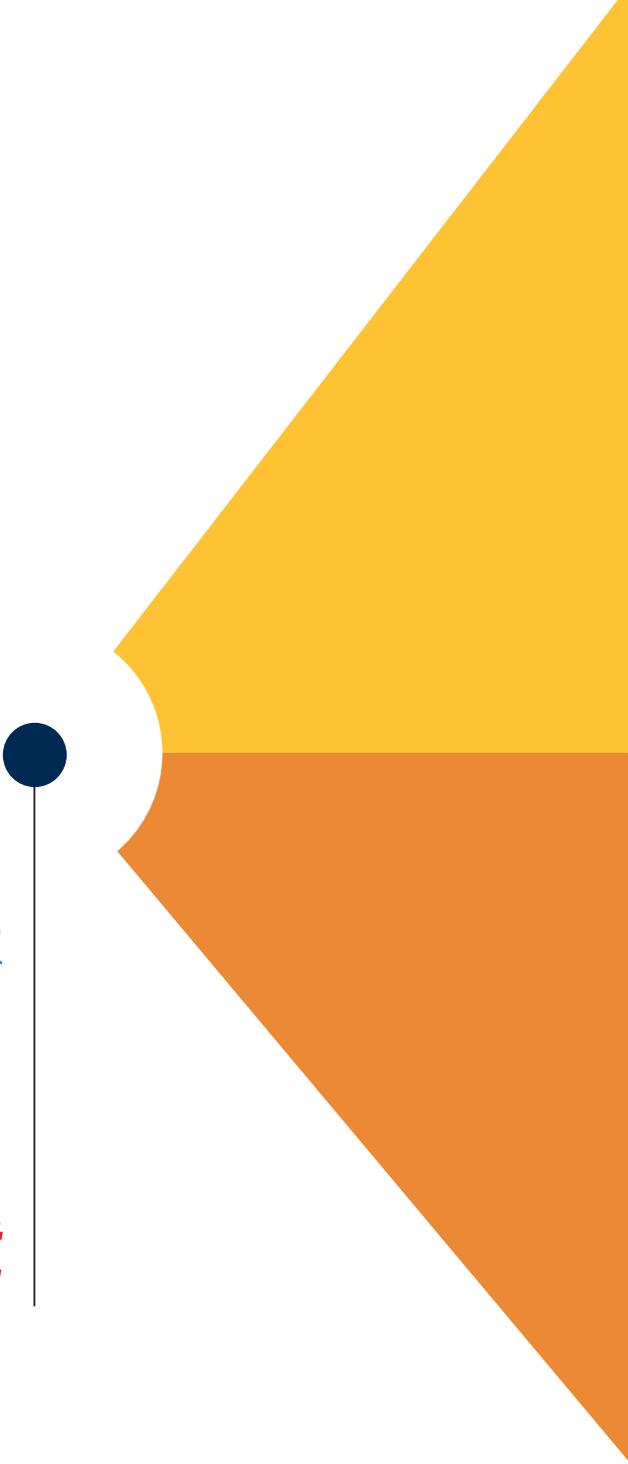
La confiance numérique touche enfin à la transparence sur les usages des traitements de données. Nous l'avons déjà souigné avec l'exploitation à des fins commerciales de certaines plateformes.

Sur la base de ces trois attributs, sécurité du stockage, maintien de la propriété par l'utilisateur, et transparence dans les usages, l'idée d'un tiers de confiance qui accompagne le développement de la stratégie du Big Data semble essentielle. Ainsi définie, la confiance numérique pourrait être une déclinaison future des valeurs mutualistes.

⁶ «Le modèle collaboratif va marginaliser le capitalisme» interview de Jeremy Rifkin, Libération, mardi 19 Octobre 2014, disponible au lien : http://www.liberation.fr/economie/2014/10/19/le-modele-collaboratif-va-marginaliser-le-capitalisme_1125202

Références bibliographiques

- Bénavent, C. (2014), "Big data: no best way", *le Libellio d'AEGIS*, hiver, 10 (4): 5-14.
- Berrier, S. (2015), "Les débats de l'Ordre. Santé connectée : enjeux et perspectives", *Médecins*, Avril, (39) : 4-6.
- Hué, B. (2015), "Research Kit, la plateforme d'Apple pour révolutionner la recherche médicale", site web RTL, lien source : <http://www.rtl.fr/culture/web-high-tech/research-kit-la-plateforme-d-apple-pour-revolutionner-la-recherche-medicale-7776969772>
- Karoui, M., Devauchelle, G., Duzert, A. (2014), « Big Data : Mise en perspective et enjeux pour les entreprises », N° Spécial "Big Data", *Revue Ingénierie des Systèmes d'Information*, 19 (3), 73-92.
- Laney, D. (2001), "3D data management: Controlling data volume, velocity and variety", *META Group Research Note*, 6.
- Mayer-Schönberger, V., Cukier, K. (2013), *Big data: A revolution that will transform how we live, work, and think*, Houghton Mifflin Harcourt.
- McAfee, A., Brynjolfsson, E. (2012), "Big data: the management revolution", *Harvard business review*, (90): 60-6.
- Peres, E. (2015), "Les données du numériques : un enjeu d'éducation et de citoyenneté", *journal officiel de la république française*, Mandature 2010-2015 – Séance du 13 janvier.
- Porter, M. (2008), « The Five Competitive Forces That Shape Strategy », *Harvard Business Review*, janvier 2008, 78-93.
- Porter, M. E., Heppelmann, J. E. (2014), "How Smart, Connected Products Are Transforming Competition", *Harvard Business Review*, 92 : 11-64.
- Roski, J., Bo-Linn, G. W., Andrews, T. A. (2014) "Creating Value In Health Care Through Big Data: Opportunities And Policy Implications", *Health Affairs*, 33(7) : 1115-1122. <http://doi.org/10.1377/hlthaff.2014.0147>



**FONDATION
DE L'AVENIR**
Accélérateur de progrès médical

