

Combinaison des Big Data et de la géolocalisation dans le domaine de la santé : Etat de l'art.

Combining Big Data and geolocation in the field of health: State of the art.

Salma Lbrini

Laboratoire Géosciences, Faculté des Sciences Aïn Chock, km 8 Route d'El Jadida, B.P 5366 Casablanca, Maroc.

salmalbrini@gmail.com

Abdelhamid Fadil

Laboratoire LAGES, Ecole Hassania des Travaux Publics, km 7 Route d'El Jadida, BP 8108, Casablanca, Maroc.

fadil@ehp.ac.ma

Hassan Rhinane

Laboratoire Géosciences, Faculté des Sciences Aïn Chock, km 8 Route d'El Jadida B.P 5366 Casablanca, Maroc.

h.rhinane@gmail.com

Hassan Jarar Oulidi

Laboratoire LAGES, Ecole Hassania des Travaux Publics, km 7 Route d'El Jadida, BP 8108, Casablanca, Maroc.

joulidi@gmail.com

Résumé

Les Big Data, résultat de la révolution numérique, offre plusieurs opportunités dans le domaine de la santé. En effet, les appareils et applications connectés en permanence à l'homme et la numérisation globale des documents médicaux produisent de vastes données de santé : « Big Health Data ». Ces données font l'objet de plusieurs projets dans le monde compte tenu des opportunités offertes pour optimiser ce domaine. Le but de cette étude est d'explorer comment la combinaison de Big Data avec la géolocalisation et comment elle peut être appliquée au Maroc.

Afin d'atteindre cet objectif, Big Health Data est présenté en décrivant des projets et des expériences globales dans l'exploitation de ces données de différents pays, les opportunités de la combinaison à la géolocalisation sont soulignées en citant différents exemples concrets, et une projection de cette étude sur le Maroc est détaillée.

Abstract

The Big Data, as result of the digital revolution, offers several opportunities in the field of health. Actually, appliances and applications permanently connected to humans and the global digitization of medical documents produce a vast health data: "Big Health Data. This data is the subject of several projects in the world given the opportunities offered to optimize this area. The aim of this study is to explore how the combination of Big Health Data with geolocation can be more opportunistic and how it can be applied in Morocco.

In order to achieve this objective, Big Health Data is introduced by describing projects and global experiences in exploiting this data from different countries, opportunities of the combination with geolocation are highlighted by citing different concrete examples, and a projection of this study on Morocco is detailed.

Mots-clés

Big data, géolocalisation, santé, sig, temps réel, signes vitaux

Keywords

Big data, geolocation, Health, gis, real time, vital signs

1. Introduction

A l'ère du datamining et la surveillance médicale en temps réel, l'utilisation des données de la santé prend une dimension plus dynamique. Celle-ci est non seulement informative, non seulement curative mais aussi prédictive et préventive, et ce, à long et à court termes. Selon Mr Jonathan Epstein, de l'Institut National pour la Santé et le Développement de l'enfant, en parlant des Big Data, « Toutes les données peuvent vous dire quelque chose, même si vous pensez que cela ne sert à rien » (Coakley, Leerkes *et al.*,2013). À cet égard, le Dr Michael Rappa de l'Institut des analyses avancées de l'Université d'État de Caroline du Nord a décrit les Big Data comme « plus opportunistes que scientifiques ». (Coakley, Leerkes *et al.*,2013)

En combinant les Big Data et la géolocalisation, nous augmentons les opportunités d'exploitation de ces données.

Pour étudier ce sujet, nous présentons d'abord les Big Data en les définissant et en citant les différentes technologies, puis nous discutons des Big Health Data en décrivant des projets et des expériences mondiales dans l'exploitation de ces données. Nous montrons ensuite les opportunités de la combinaison de ces données avec la géolocalisation en présentant les définitions et les composantes de cette dernière et en citant différents exemples concrets pour conclure avec une projection de cette étude sur le cas du Maroc.

2. Background

Afin de cerner ce sujet, il est judicieux de commencer par présenter le contexte historique des Big Data, les définitions de celles-ci ainsi que les différentes technologies utilisées.

2.1 Big Data : Histoire et apparition

En 1960, l'informatisation des opérations gourmandes en temps et en paperasse des sociétés a conduit au commencement de la troisième révolution technique : la révolution numérique. La première étant la révolution de la machine et la deuxième étant la révolution mécanique de l'ère industrielle. En moins de soixante ans, l'informatisation des sociétés est devenue complète et globale, nous faisant entrer dans une ère numérique. En effet, la normalisation, en 1970, des

bases de données d'une part et la normalisation de l'architecture, à la même date, d'autre part ont permis aux applications de s'alimenter entre elles, encourageant encore plus la numérisation des données. Le développement des micro-ordinateurs et des réseaux locaux en 1980 et l'émergence de l'internet et du web ont permis l'échange de messages et de documentation entre les machines et donc l'informatisation des processus. Ce mouvement n'a pas cessé de croître depuis les décennies 2000 et 2010 avec le Web 2.0, l'élargissement des réseaux sociaux, l'apparition des terminaux mobiles et l'essor des objets connectés (Babinet, Vassoyan *et al.*, 2015).

Cette révolution a conduit à un besoin de traitement de données de différentes formes rendant le traitement classique dépassé (Hurwitz, Nugent *et al.*, 2013).

Dans la décennie 2000, les acteurs de recherche sur le Web se sont retrouvés face à un problème de « scalabilité » défini par la nécessité d'adapter la capacité de calcul aux rythmes de la demande et à la montée en charge. Ainsi, cette décennie a connu les premiers projets industriels de Big Data mais l'apparition médiatique de ce terme au grand public a eu lieu en 2011 à travers le rapport du Cabinet américain McKinsey intitulé « *Big Data : la nouvelle frontière pour l'innovation, la compétition et la productivité* ». (Huot, 2014)

2.2 Définition

Big Data est un terme composé de « big » et de « data » traduit selon l'Académie des sciences par « données massives » ou data masse (Huot, 2014). Selon cette appellation, on comprend que les Big Data sont les bases de données dont le volume dépasse la capacité de capture, de stockage, de gestion et d'analyse des logiciels typiques. Cette définition est intentionnellement subjective. En effet, avec l'avancement technologique, pour être considéré Big Data, le volume d'une base de données augmentera au fur et à mesure. Le fait de classer une base de données comme Big Data n'est donc pas fixé par un nombre précis de bits à dépasser (Manyika, Chui *et al.*, 2011).

Le concept des big data est décrit selon plusieurs niveaux. Les quatre dimensions les plus communément reconnues, représentées par 4V sont le volume, la variété, la vélocité et la véracité (Ben Salem, 2015).

Le volume représente la quantité de données stockées. Ces données ne cessent d'augmenter, elles étaient estimées à 800000 pétaoctets en l'an 2000. Ce chiffre pourrait atteindre 35 zettaoctets en 2020 (Zikopoulos, Eaton *et al.*, 2012).

La variété décrit la diversité de la nature des données d'aujourd'hui. En effet, avec l'expansion d'utilisation de capteurs, de téléphones et de réseaux sociaux, les données sont devenues complexes contenant non seulement les données traditionnelles mais également les données semi-structurées et non structurées (page Web, streaming, index de recherches, réseaux sociaux, forums, e-mail, capteurs de systèmes actifs ou passifs...) rendant l'analyse de façon traditionnelle une mission très difficile voire impossible (Zikopoulos, Eaton *et al.*, 2012).

La vitesse décrit la rapidité avec laquelle les données sont récupérées, stockées et traitées (Zikopoulos, Eaton *et al.*, 2012).

La véracité est liée à la pertinence des données (Teboul et Berthier, 2015).

A ces 4V s'ajoute une cinquième dimension adoptée par certains chercheurs qui représente le paramètre « Valeur ». L'objectif de ce 5^{ème} V est d'extraire la valeur à partir des Big Data (Monino et Sedkaoui, 2016).

2.3 Technologies Big Data

Les Big Data ne sont pas une science nouvelle, c'est le fruit d'une évolution de la gestion de bases de données de plus de 50 ans d'héritage (Hurwitz, Nugent *et al.*, 2013).

Les solutions qui ont marqué les Big Data ont été produites par des compagnies comme Yahoo, Google et Facebook. Ces sociétés étaient obligées de résoudre les problèmes de gestion liés à la quantité massive de leurs données. Ces solutions traitent l'un des problèmes fondamentaux qui consistent en la capacité de gérer les données efficacement, en optimisant le coût et le temps (Hurwitz, Nugent *et al.*, 2013). C'est ainsi que les ingénieurs de Google ont développé le modèle de programmation MapReduce pour traiter les données massives (Potisepp, 2013). En 2004, Google lance en interne le projet BigTable qui repose en partie sur l'utilisation de MapReduce. BigTable est une plateforme de haute performance pour le stockage et le traitement de vastes ensembles de données semi-structurées. L'application, qui repose sur une architecture distribuée (serveurs répartis en grappes/clusters), est conçue pour pouvoir répondre avec des temps de réponse très courts aux requêtes émanant simultanément de plusieurs milliers d'ordinateurs clients. D'un autre côté, Douglass Cutting a créé en 2004 le système Hadoop pour Yahoo. Il s'agit d'une technologie matricielle destinée à faciliter le développement de solutions optimisées pour le traitement de gros volumes de données. Les environnements Hadoop permettent d'utiliser, en couche supérieure (top level programming language), des langages de programmation simplifiés par rapport au formalisme MapReduce, dont la syntaxe se rapproche de celle des langages de développement connus (Java, SQL, ...) : Pig, Hive, Giraph, Sqoop...

(Huot, 2014). Hadoop est utilisé par presque 63% des organisations pour gérer les données massives (Khan, Yaqoob *et al.*, 2014).

On distingue trois catégories principales de plateformes existantes : les plateformes open source comme Apache Hadoop, Apache Cassandra et MongoBd, les solutions clé en main fournies principalement par Oracle, IBM, SAP et Microsoft et les plateformes cloud dont les trois grands fournisseurs sont Amazon Web Services (AWS), Microsoft Azure et Google App Engine (Verheji, 2013).

3. Big Data du domaine de la Santé (Big Health Data)

Cette section présentera les différentes sources de production des Big Health Data, ses opportunités appuyées par différents projets en cours dans le monde.

3.1 Corps humain source de Big Health Data

Les technologies qui suivent nos habitudes, notre emplacement, nos achats, nos routines, nos interactions sociales et nos comportements puisent leurs ressources dans des téléphones portables, des logiciels téléchargeables, des moniteurs et des caméras portables. Les personnes s'engagent de plus en plus à porter des capteurs spécialisés tout au long de la journée ou pendant un exercice ou encore durant le sommeil, dans le but de donner un aperçu de leurs habitudes physiques et de leur santé (Eagle et Greene, 2014). Ces applications et ces appareils sont majoritairement connectés par le biais d'internet aux serveurs du fabricant ou autres entreprises d'analyse ou de publicité et leur taux d'utilisation ne cesse d'augmenter rapidement (Adams, Purtove *et al.*, 2016).

Les paramètres vitaux du corps humain sont de plus en plus suivis et stockés en temps réel. Par exemple, le service de cardiologie de la clinique Scripps de San Diego a mis en place un système qui permet de suivre l'état de santé de ses patients même s'ils sont à la maison (Agarwal, 2016). D'autre part, les systèmes d'information de l'hôpital sont indispensables aux hôpitaux d'aujourd'hui. Ils jouent un rôle crucial en tant qu'interface d'information pour les médecins et les infirmiers. Un système d'information sur les hôpitaux comprend des dossiers de santé électroniques (DSE), un système de saisie de commandes au fournisseur, un système de comptabilité médicale, un système d'archivage et de communication d'images (PACS) et ainsi de suite. Les hôpitaux modernes ne peuvent plus se passer de leur système d'information qui génère des données volumineuses comme les valeurs d'analyse sanguine, les

électrocardiogrammes (ECG) et les rayons X (X-P) (Sawa, 2014). Cette situation a conduit à la génération de données médicales appelée « Big Health data ».

Il s'agit d'ensembles de données volumineux recueillis de manière régulière ou automatique et qui sont enregistrés et stockés électroniquement. On peut réutiliser ce concept dans le sens de données polyvalentes et y inclure la fusion et la connexion de bases de données existantes dans le but d'améliorer la santé ainsi que la performance du système de santé. Il ne s'agit pas de données recueillies pour une étude spécifique (Habl, Renner et al., 2016).

3.2 Opportunités des Big Health Data

D'après le rapport du groupe canadien « Emerging Technology Group (ETG) » spécialiste dans les technologies de l'information médicale publié en Avril 2013 et intitulé « Big Data Analytics in Health », la recherche sur l'impact des analyses poussées des Big Data sur le secteur de la santé prévoit une nouvelle vague de productivité, d'efficacité et d'innovation dans les domaines des soins personnalisés, des opérations de santé, des décisions cliniques, des politiques de santé et une meilleure prise en charge des patients. Le milieu de la recherche estime que cette utilisation peut réduire le coût des soins de santé tout en améliorant la qualité de ces soins et tout en les rendant plus préventifs et personnalisés.

En effet, les Big Health Data, mis à part leur valeur scientifique, ont une valeur économique. McKinsey Global Institute a prédit en 2011 que si les organismes de soins de santé américains utilisaient les analyses de Big Data de façon créative et efficace pour générer de l'efficacité et de la qualité, le secteur pourrait générer plus de 300 milliards de dollars annuellement. Plusieurs projets pour bénéficier du Big Health Data y sont déjà en cours, parmi lesquels on trouve les projets IBM Watson. Il s'agit d'un programme informatique d'intelligence artificielle conçu par IBM dans le but de répondre à des questions formulées en langage naturel. Il s'intègre dans un programme de développement plus vaste, le « DeepQA research project ». Le nom « Watson » fait référence à Thomas J. Watson, dirigeant d'IBM de 1914 à 1956, avant même que cette société ne s'appelle ainsi. Ce programme a plusieurs utilisations. Nous en citons ci-dessous quelques exemples.

- Watson pour le traitement génomique : Ce logiciel permet de générer des informations pour aider les médecins à identifier les mutations du cancer qui sont potentiellement traitées et ces idées pourraient alors aider le médecin à identifier des traitements ciblés.

- Watson pour découvrir les médicaments : « Watson for Drug Discovery » aide les chercheurs à identifier de nouvelles cibles de médicaments et de nouvelles indications pour les médicaments existants. La plate-forme peut aider les chercheurs à découvrir de nouvelles connexions rapidement, ce qui peut conduire à de nouvelles idées et à des percées scientifiques. Les capacités cognitives à la fine pointe de la technologie synthétisent des sources larges et diverses de contenu structuré et non structuré, public, autorisé et privé pour aider les chercheurs (Gliozzo, Ackerson *et al.*, 2017). Cette plate-forme est utilisée également par l'Institut Ontarien du Cerveau et la Clinique des troubles du mouvement du Réseau Universitaire de Santé de Toronto pour entreprendre le tout premier projet de recherche sur la maladie de Parkinson au Canada (Fortin, 2016).
- Watson pour la gestion des soins : IBM Watson Care Manager aide les entreprises à débloquer et à intégrer l'intégralité de l'information à partir des systèmes et des fournisseurs de soins multiples, d'automatiser les flux de travail de gestion des soins et d'augmenter les besoins des populations en pleine croissance (Gliozzo, Ackerson *et al.*, 2017).

En Europe, plusieurs projets « Big Health Data » sont également en cours. Ainsi, plusieurs études ont été commandées sur les Big Data dans le domaine de la santé publique, la télémédecine et les soins de santé. Parmi ces études couvrant une vingtaine de projets, dix ont été sélectionnées comme ayant le plus de valeur. En voici quelques exemples.

- Comet K-Project Dexhapp – AT est un projet autrichien qui vise à développer de nouvelles méthodes, modèles et technologies afin d'analyser le statu quo du système de santé, de prévoir les développements futurs et de comparer des scénarios en fonction d'interventions différentes dans le secteur de la santé.
- The Shared Care Platform – DK est une plate-forme informatique de soins partagée qui soutient une collaboration intersectorielle dans les soins de santé en facilitant la coordination entre le praticien général, la municipalité et l'hôpital. Le projet a été financé par le fonds national pour les maladies chroniques et développé par la Région du Danemark méridional en coopération avec IBM. L'application de la plate-forme de soins partagée initialement réservée aux patients atteints de maladies chroniques, sera élargie à la lutte contre les maladies chroniques et à l'amélioration de la qualité et de l'efficacité du traitement.

- E-Estonia – National Identity Scheme – EE : relève de la recherche sur les systèmes de santé et des services, ainsi que sur l'épidémiologie. Il est interfacé avec des données de dossiers de patients collectées de façon routinière en aidant à la durabilité du système de santé.
- PASSI – IT : est un système de surveillance des facteurs de risque comportementaux en Italie, mis en place depuis 2006. Son objectif principal est d'estimer la fréquence et l'évolution des facteurs de risque comportementaux pour la santé et la diffusion des mesures préventives au fil du temps. La diffusion d'informations pertinentes pour les professionnels de la santé publique et les communautés sont les priorités de PASSI afin de contribuer à un mode de vie plus sain.

Sur la base de ces études, l'Agence exécutive pour les consommateurs, la santé et l'alimentation relevant de l'Union Européenne (Consumers, Health and Food Executive Agency, CHAFEA) a élaboré 10 recommandations :

- Développer et mettre en œuvre une stratégie de communication pour sensibiliser à la valeur ajoutée du *Big Data* dans le domaine de la Santé et promouvoir une image positive du *Big Data* de la Santé ;
- Renforcer le capital humain par rapport au besoin croissant d'une main d'œuvre qui peut tirer profit du potentiel du *Big Data* de la Santé ;
- Développer et diversifier les sources existantes de *Big Data* de la Santé et en découvrir de nouvelles tout en assurant leur qualité et leur protection ;
- L'accès libre et le partage des *Big Data* sans compromettre le droit des patients à la protection de leur vie privée et à la confidentialité ;
- Appliquer de manière ciblée les résultats des analyses de *Big Data* en fondant sur les besoins et les intérêts des parties prenantes, y compris des patients ;
- Identifier les potentiels de l'analyse des *Big Data*, améliorer les méthodes d'analyse et promouvoir l'utilisation de méthodes novatrices ;
- Mettre en œuvre des mécanismes de gouvernance qui assurent un accès et une utilisation équitable et en toute sécurité des *Big Data* pour la recherche médicale ;
- Développer des normes techniques pour les *Big Data* de la Santé afin de simplifier leur application et améliorer leur interopérabilité ;
- Assurer un investissement réfléchi guidé par la Commission Européenne afin de garantir rentabilité et durabilité ;

- Clarifier et aligner les réglementations existantes des *Big Data* de la Santé concernant la protection de la vie privée.

Plusieurs utilisations des Big Data Santé sont en cours, mais quand ces projets sont combinés à la géolocalisation, l'aspect prend encore d'autres dimensions.

4. Big Health Data et géolocalisation

Cette section présentera plusieurs cas d'utilisation et d'illustration qui ont prouvé l'apport de la combinaison des Big Health Data et la géolocalisation, mais commençons par introduire les concepts et composantes de la géolocalisation.

4.1 La géolocalisation

Bergeron a défini la géomatique, appelée aussi SIG, comme la discipline ayant pour objet la gestion des données à référence spatiale et qui fait appel aux sciences et aux technologies reliées à leur acquisition, leur stockage, leur traitement et leur diffusion (Pornon,1998). Ainsi d'après Thériault, un SIG est un ensemble de principes, de méthodes, d'instruments et de données à référence spatiale, utilisé pour saisir, conserver, transformer, analyser, modéliser, simuler et cartographier les phénomènes et les processus distribués dans l'espace géographique.

Les données sont analysées afin de produire l'information nécessaire pour aider les décideurs (Prélaz-Droux,1995).

Les données géographiques au cœur du SIG comportent trois composantes principales :

- Une composante spatiale ;
- Une composante sémantique à travers des attributs non spatiaux ;
- Une composante métadonnées : exemple : date de collection des données (Ducommun et Lucot, 2017).

Il existe plusieurs technologies de localisation dont voici quelques exemples.

- « Enhanced Assisted » GPS : basée sur la localisation par satellites, la précision peut atteindre 5 à 20 mètres si associée à des données de correction (Hainaut, 2011).
- Cell Id : basée sur les cellules des réseaux mobiles (précision de localisation de 500 m en ville à 2,5 km en campagne et accessible sur tous les téléphones mobiles). Autour de la technologie basée sur les cellules, d'autres technologies permettent de renforcer la précision de la localisation (par exemple la triangulation) ;

Il existe plusieurs technologies de géolocalisation et le choix dépend de l'objectif d'utilisation.

Les SIG ont prouvé leur pertinence dans plusieurs domaines. Les paragraphes suivants illustrent quelques exemples relatifs au secteur médical.

4.2 Big Health Data et géolocalisation au service de la détection et prévision des épidémies

Les SIG sont un moyen efficace pour contrôler les maladies à transmission vectorielle, aérienne ou aquatique.

Ils ne peuvent pas seulement fournir une plate-forme pour la préparation du mode géo spatial mais ils peuvent être utilisés pour le développement d'un modèle prédictif à l'aide de divers facteurs environnementaux et climatiques. Les résultats de ces études consistent à lier la topographie à la maladie, aux zones de reproduction, aux analyses comparatives des maladies, etc. (Aggarwal, 2017).

La carte de Choropleth offre un moyen simple pour visualiser comment une mesure varie selon une zone géographique et montre le niveau de variabilité dans une région appliquée à la santé (Challa, Pattnaik *et al.*, 2017).

Les cartes de Choropleth ont été largement discutées et utilisées pour répondre aux besoins multiples de services de la santé publique. Ces utilisations sont de plus en plus répandues dans d'autres pays comme en Inde où ces outils cartographiques ont servi comme une base de communication compréhensible pour mettre en évidence les villes où la politique de la santé a donné ses fruits de celles où elle a échoué (Challa, Pattnaik *et al.*, 2017).

Dans un autre volet, le développement des modèles pour les habitats des insectes comme la tsétsé, responsable de la propagation des maladies, permet la description aussi bien que la prédiction et nécessite ainsi la combinaison de plusieurs données dont celles à référence spatiale (Guptill et Moore, 2013).

Grâce à la vitesse des « Big Health Data », les données peuvent mener à des interventions rapides et efficaces face à des menaces sur la santé de la population. A titre d'exemple, l'apparition du choléra en Haïti, après le tremblement de terre de 2010, a nécessité une réponse rapide et ciblée des services de la santé publique dans des conditions défavorables. L'identification des sujets et des zones infectés et le déploiement des vaccins contre le choléra auraient été aidés par l'utilisation des nouvelles technologies et la combinaison de données émanant de multiples sources et de divers types (Mooney, Westreich *et al.*, 2015).

Diverses maladies transportées par l'air, par l'eau et par les virus peuvent être prévisibles à l'avance, ce qui peut aider les décideurs à planifier des stratégies de lutte contre la propagation des maladies et particulièrement dans le cas des maladies à transmission vectorielle. La modélisation géo-spatiale des maladies peut conduire à une application robuste d'analyse pour suivre la répartition géographique et la transmission modèle de maladie (Aggarwal, 2017).

Ainsi, plusieurs projets ont-ils été mis en place pour la combinaison de la dimension spatiale avec les « Big Health Data ». On donne ci-dessous quelques exemples illustratifs.

Le paludisme étant une maladie à transmission vectorielle, la maladie inflige le plus grand nombre de morts dans la plupart des pays tropicaux. La Papouasie-Nouvelle-Guinée est l'un des pays de la région tropicale où le paludisme est la maladie la plus redoutée. Plusieurs facteurs responsables de l'attraction de ces vecteurs pathogènes ont été identifiés et étudiés à l'aide de SIG et de la télédétection. Des évaluations multicritères et analytiques ont été explorées pour déterminer le rôle de différents facteurs géographiques et non-géographiques dans les infections du paludisme (Jana, 2017).

L'outil, Google Flu Trends, est un outil Web sophistiqué pour la détection des épidémies régionales de grippe aux États-Unis. En effet, des millions de personnes dans le monde recherchent en ligne des informations sur la santé, chaque jour, ce qui rend les requêtes de recherche Web une source précieuse d'informations sur les tendances collectives de la santé. La société Internet Google a récemment lancé un outil expérimental pour la détection à court terme des épidémies de grippe en surveillant et en analysant les comportements de recherche de soins de santé sous forme de requêtes sur son moteur de recherche en ligne (Carneiro et Mylonakis, 2009).

Google trends qui se fonde sur le suivi de recherches géo localisées de ses utilisateurs dans une région précise. Il peut détecter les épidémies régionales de grippe 7-10 jours avant les systèmes conventionnels de surveillance des centres de prévention et de contrôle des maladies aux États-Unis qui se basent sur des données cliniques et de laboratoires pour publier des rapports hebdomadaires avec un décalage d'une à deux semaines (Carneiro et Mylonakis, 2009). Cette technologie unique et innovante nous rapproche un peu de la vraie surveillance des éclosions en temps réel.

Le virus du Nil occidental est une maladie transmise par les moustiques qui provoque des épidémies saisonnières aux États-Unis. Son cycle naturel se base sur la transmission des moustiques aux oiseaux et aux mammifères. L'éruption cutanée présente des motifs saisonniers

qui correspondent bien aux pics du nombre de cas de ce virus. Les augmentations du volume de recherche pour les éruptions cutanées, qui peuvent être un bon proxy pour le virus, commencent à augmenter en mai, juste avant l'augmentation des cas observés en juin de chaque année dans les données des centres de contrôle et de préventions (Carneiro et Mylonakis, 2009).

4.3 Big Health Data et géolocalisation au service de l'optimisation de la télémédecine

Grâce aux technologies d'information, la médecine à distance utilisant le « Geomapping » permet d'approcher de plus en plus les patients des services médicaux. Le projet de télémédecine de pédiatrie comportementale du développement est un exemple marquant. Cette étude qui a duré de Janvier 2009 à Décembre 2010 sur 5 sites de télémédecine de Kentucky des Etats-Unis est considérée comme la première à utiliser les SIG pour démontrer que l'accès au service pédiatrique de télésanté créerait des économies substantielles (Soares, Johnson *et al.*, 2013). Cette étude a montré l'efficacité de la télémédecine non seulement dans l'économie de temps et d'argent mais aussi dans la réduction du temps d'attente qui peut atteindre 5 semaines (Soares, Johnson *et al.*, 2013). En plus des avantages liés à la qualité de service, la télémédecine offre une base de données riche d'informations, qui peut être utilisée en temps réel pour détecter les épidémies et les attaques biologiques à l'instar du projet de surveillance des syndromes en temps réel mené en Ontario au Canada (Kieran, 2004).

4.4 Big Health Data et géolocalisation au service de la science

Des projets de collecte et d'analyse de données sont en cours afin de bénéficier à long terme des opportunités analytiques des Big Health Data. En effet, une expérience a montré comment des messages et des conseils ont contribué à la réussite de la régulation des taux de sucre dans le sang chez les patients ciblés. Pour ce test, les médecins ont donné des podomètres à 126 patients atteints de diabète de type 2 afin de suivre leurs activités physiques. En fonction des progrès réalisés par les patients dans le respect de leurs objectifs, ils recevraient des conseils de motivation via un message texte. Les chercheurs ont conclu que la messagerie texte personnalisée peut être utilisée pour améliorer les résultats chez les patients atteints de diabète de type 2 en employant des mesures optimales d'engagement des patients (Horner, Agboola *et al.*, 2016).

Parmi les projets qui se basent sur les analyses des Big Data, on trouve le projet Kavli HUMAN (KHP). Il s'agit d'agrèger les données de 2 500 ménages de la ville de New York dans les cinq quartiers (environ 10 000 individus). Le projet vise à étudier l'interaction dynamique de notre

biologie, de notre comportement et de notre environnement et lance un projet de collecte de données biologiques, comportementaux, environnementaux ainsi que les événements des membres de KHP sur 20 ans. KHP utilise une base de données d'un système d'information géographique de grande échelle. KHP vise à offrir des vues synoptiques et granulaires de la manière dont la santé et le comportement humain se développent au cours du cycle de vie et pourquoi ils évoluent différemment pour différentes personnes. Les porteurs du projet souhaitent apporter de nouvelles approches scientifiques basées sur la découverte, enracinées dans les grandes analyses de données, pour améliorer la santé et la qualité de la vie humaine, particulièrement dans les contextes urbains (Azmak, Bayer *et al.*, 2015).

Les médecins, sur la base des nouvelles découvertes, et sur la base des analyses des données cumulées seraient en mesure d'impacter la santé des patients rien qu'en leur demandant de changer certains comportements. D'après les chercheurs, des résultats financiers, opérationnels et cliniques peuvent être améliorés, chose qui permettra une meilleure performance des services de santé et la sauvegarde de vies (Parslow, 2014).

5. Big Health Data combinés à la géolocalisation : opportunités au Maroc

A l'issue de cette revue de littérature sur la combinaison des big data avec la santé et avec la géolocalisation, il s'avère que plusieurs opportunités médicales, scientifiques et économiques peuvent être explorées et analysées dans le cas du Maroc.

L'un des axes prometteurs est le suivi de l'état de santé des patients à travers la collecte des données par des capteurs que ces derniers peuvent porter.

Les informations médicales du patient pourraient être stockées et analysées de façon continue ou périodique selon les cas d'utilisation et les besoins exprimés.

Ainsi, la richesse de ces données et ces technologies s'avère opportune sur plusieurs niveaux pour le cas du Maroc. Nous les présentons dans ce qui suit.

- Au Maroc, le nombre de personnes atteintes d'Al Zheimer était estimé en 2015 à 119 000 et ce chiffre ne cesse d'augmenter (Nicaise, Palermi, 2016). La localisation en temps réel des patients contribuera à lutter contre les cas de disparitions et d'égarement de cette catégorie de patients.
- Au Maroc, les résultats de l'enquête prospective 2000 du Ministère de la Santé ont donné une prévalence globale de l'hypertension artérielle de 33,6% (Abdelkhirane, Azouzi *et al.*, 2012). Cette maladie est de 53,8% chez les personnes de plus de 40 ans et

de 72,2% chez les 65 ans et plus (Abdelkhirane, Azouzi *et al.*, 2012). D'autre part, les maladies cardiaques sont les premières causes de mortalité. En 2014, Dr Souteyrand, Représentant de l'Organisation mondiale de la santé à Rabat a estimé à 40% le nombre de décès causés par les maladies cardiovasculaires dans le Royaume (autrement dit quatre décès sur dix), contre 30% au niveau mondial. La combinaison des signes vitaux et la géolocalisation permettrait de détecter en temps réel les cas urgents et intervenir pour sauver des vies ;

- La plate-forme informatique luttera contre les pertes de documents papiers et économisera énormément de temps et rendra les services de santé plus efficaces et fournira aux médecins une vue plus pertinente et globale sur l'état du patient. Elle permettra de croiser et de synthétiser toutes les informations d'historique sur le patient (Analyses, Signes vitaux, Scanners, médicaments, traitements suivis...) afin de rendre le diagnostic et le suivi plus précis et plus efficace ;
- La disponibilité des informations numérisées et cumulées pendant des années exploitées par le Datamining contribuera à découvrir d'autres causes d'éléments aggravant des maladies suite par exemple à l'observation de répétitions de certains comportements chez les mêmes patients ;
- Le suivi personnalisé des patients, comme on a vu précédemment, influence positivement l'amélioration de la santé. Cette connexion avec le patient contribuera à donner le bon conseil au bon moment et permettra même d'anticiper certaines maladies héréditaires ;
- Différents projets d'application des « Big Health Data », à l'instar des exemples cités dans les sections précédentes, sont déjà en cours de mise en place dans plusieurs pays. Le projet KHP précité produira des données cumulées de vingt ans.

Ces données seront utilisées dans des études qui ouvriront de nouveaux horizons de découvertes et contribueront aux développements scientifiques de ces pays.

6. Conclusion

Big Data est un concept et une technique qui a montré son utilité et son apport dans divers domaines et particulièrement avec la croissance exponentielle de la quantité d'informations produites quotidiennement.

Le corps humain est source de vastes données. En effet, les systèmes d'information sont indispensables pour les hôpitaux. Ces systèmes comprennent des dossiers de santé électroniques (DSE), un système de saisie de commandes de fournisseur, un système de comptabilité médicale, un système d'archivage et de communication d'images (PACS), etc. Ce qui génère des données volumineuses comme les valeurs d'analyse sanguine, les électrocardiogrammes (ECG) ou les rayons X (X-P). D'autre part, les paramètres vitaux du corps humain sont de plus en plus suivis et stockés en temps réel à travers les capteurs. Ces données forment les « Big Health Data ».

Leur application dans le domaine de la santé ne cesse de proliférer et de produire de nouveaux exemples et de projets dans le monde développé. En effet, plusieurs projets pour bénéficier du « Big Health Data » sont déjà en cours aux Etats-Unis, parmi lesquels on trouve les projets IBM Watson. En Europe, plusieurs études ont également été lancées sur les apports des Big Data dans la santé publique, la télémédecine et les soins de santé.

L'intégration de la dimension géospatiale avec les « Big Health Data » ajoute d'autres horizons et d'autres opportunités qui ne peuvent être que bénéfiques pour l'amélioration des services de santé et la sauvegarde des vies humaines. En effet, d'une part les SIG sont un moyen efficace pour contrôler les maladies à transmission vectorielle, aérienne ou aquatique. D'autre part, grâce aux technologies d'information, la médecine à distance utilisant le « Geomapping » permet d'approcher les patients des services médicaux. Il est à souligner que des projets de collecte et d'analyse de données sont en cours afin de bénéficier à long terme des opportunités analytiques de la combinaison des Big Health Data et la géolocalisation.

Le Maroc, qui a mis en place une stratégie numérique et affiche des ambitions majeures pour l'amélioration de la performance des services de santé, peut tirer profit de ces approches s'il arrive à s'appropriier ces technologies à travers leur adaptation à son contexte et ses besoins concrets. Finalement, si on n'investit pas dans les projets « Big Health Data », le fossé numérique et technologique avec les pays développés augmentera et l'écart risquera de ne pas être facilement rattrapable.

Références

Abdelkhirane, c., Azouzi, L., Bennis, K., Bensafiddine, S., Bentalha, S., Bougteb, H., *et al.* (2012). L'hypertension artérielle de l'adulte, recommandations des bonnes pratiques médicales, affections longues durées Available at : <http://www.anam.ma/upload/document/hypertension-resume.pdf>

Adams, S., Purtova, N., Leenes, R., (2016). *Under Observation: The Interplay Between eHealth and Surveillance* (978-3-319-48342-9). Berlin: Springer

- Aggarwal, G., (2017). GIS for control of communicable diseases. Geospatial World Forum. Hyderabad India : 23-25 January. Available at <https://geospatialworldforum.org/speaker/SpeakersImages/GIS-for-control-of-communicable-diseases.pdf>
- Azmak, O., Bayer, H., Caplin, A., Chun, M., Glimcher, P., Koonin, S., et al. (2015). Using Big Data to Understand the Human Condition: The Kavli HUMAN Project. *Big Data*, Vol. 3, Num. 3, pp 173–188. Available at <http://doi.org/10.1089/big.2015.0012>
- Babinet, G., Vassoyan, R., Asséraf, A., Colligon, H., Delcroix, G., Gerard de Lescazes, F., et al. (2015), Big data et objets connectés Faire de la France un champion de la révolution numérique, Rapport Avril 2015, Paris : Institut Montaigne
- Ben Salem, A., (2015). Qualité contextuelle des données : Détection et nettoyage guidés par la sémantique des données, Thèse de doctorat en Informatique, Paris : l'Université Paris 13 Sorbonne
- Carneiro, H., Mylonakis, E., (2009), Google Trends: A Web-Based Tool for Real-Time Surveillance of Disease Outbreaks, *Clinical Infectious Diseases*, Vol.49, Num.10, pp 1557-1564. Available at <https://academic.oup.com/cid/article/49/10/1557/298019/Google-Trends-A-Web-Based-Tool-for-Real-Time>
- Challa, S., Pattnaik, S., Mathai, D., (2017). Choropleth mapping as a tool of advocacy in primary health care and public health practice. Geospatial World Forum. Hyderabad India : 23-25 January.
- Coakley, M., Leerkes, M., Barnet, J., Gabrielian, A., Noble, K., Weber, N., et al, (2013) . Unlocking the Power of Big Data at the National Institutes of Health available at <https://doi.org/10.1089/big.2013.0012>
- Ducommun, C., Lucot, E., (2017). *Les cartes et les données pédologiques Des outils au service des territoires*, Dijon : Educagri
- Eagle, N., Greene, K. (2014). *Reality Mining: Using Big Data to Engineer a Better World (978-0-262-02768-7)* Cambridge, Massachusetts: MIT Press
- Fortin, J. (2016). Les leaders canadiens en neurosciences ont recours à IBM Watson pour accélérer la découverte de nouveaux médicaments contre la maladie de Parkinson. Available at <https://www.ibm.com/news/ca/fr/2016/10/12/n006085h28737w08.html>
- Gliozzo, A., Ackerson, C., Bhattacharya, R., Goering, A., Jumba, A., Yeon KimBuilding, S., (2017). *Cognitive Applications with IBM Watson Services: Volume 1 Getting Started (073844264X)* USA: International Business Machines Corporation
- Guptill, S.C., Moore, C.G. (2013). Investigating Vector-Borne and Zoonotic Diseases with Remote Sensing and GIS. *Essentials of Medical Geology*. Chap29 Springer, Dordrecht pp 647-663
- Habl, C., Renner, A-T., Bobek, J., Laschkolnig, A. (2016). Study on Big Data in Public Health, Telemedicine and Healthcare, Rapport Final Octobre 2016. Vienne: Gesundheit Österreich GmbH
- Hainaut.C.(2011).Les services de géolocalisation :technologie et applications <https://doi.org/10.1016/j.rehab.2011.07.774>
- Horner, GN., Agboola, S., Jethwani, K., Tan-McGrory; A., Lopez, L. (2017). Designing Patient-Centered Text Messaging Interventions for Increasing Physical Activity Among Participants with Type 2 Diabetes: Qualitative Results From the Text to Move Intervention. *JMIR Mhealth Uhealth*.Vol.5, Num.4. Available at: <http://mhealth.jmir.org/2017/4/e54>

- Huot, C. (2014). Le Big Data si nous en parlions ? In Huot, C., Masanès, J., Sébillot, P., Carrive, J., André, F., Grumbach, S., et al. *Big Data, Nouvelles partitions de l'information (978_2_8041_8915_0)*. Belgique : De Boeck
- Hurwitz, J., Nugent, A., Halper, F., Kaufman, M., (2013). *Getting Started with Big Data. Big data for dummies (978-1-118-50422-2)*. New Jersey: John Wiley & Sons, Inc. 10–21.
- JANA, S., (2017) Exploring Vulnerable areas for Malaria in Morobe Province of Papua New Guinea - Remote Sensing and GIS as Potential Tools. Geospatial World Forum. Hyderabad India: 23-25 January. Abstract available at <https://geospatialworldforum.org/2017/speakers-bio.asp?id=gwf2017A392>
- Khan, N., Yaqoob, I., Abaker, I., Hashem, T., Inayat, Z., Kamaleldin Mahmoud Ali,W., et al.(2014).Big Data: Survey, Technologies, Opportunities, and Challenges. *The Scientific World Journal*. vol. 2014, Article ID 712826 Available at <http://dx.doi.org/10.1155/2014/712826>
- Kieran, M. (2004). Real-time syndrome surveillance in Ontario, Canada: the potential use of emergency departments and Telehealth. *European journal of emergency medicine: official journal of the European Society for Emergency Medicine*. Vol.11, Num.1, pp 3-11 Available at <https://insights.ovid.com/pubmed?pmid=15167186>
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., et al. (2011). *Big data: The next frontier for innovation, competition, and productivity*, June 2011, New York : McKinsey Global Institute
- Monino,J, Sedkaoui,S,(2016), *Big Data, Open Data and Data Development (978-1-84821-880-2)*, New York: John Wiley & Sons
- MooneyS, Westreich,D,El-Sayed,A et al (2015) *Epidemiology in the Era of Big Data*, *Epidemiology (Cambridge, Mass.)*,Vol. 26, Num. 3, pp 390-394 Available at <http://doi.org/10.1097/EDE.0000000000000274>
- Nicaise, S., Palermi, F. (2016) *rapport Alzheimer et méditerranée 2016 état des lieux - enjeux – perspectives*. Monaco : Association Monégasque pour la recherche sur la maladie d'Alzheimer
- Parslow,W. (2014). How big data could be used to predict a patient's future, *The Guardian* 17 June 2014 Available at <http://www.theguardian.com/healthcare-network/2014/jan/17/big-data-nhs-predict-illness>
- Prélaz-Droux, R.(1995).*Système d'information et gestion du territoire*. Lausanne : PPUR presses polytechniques.
- Pornon, H.(1998).*SYSTÈMES D'INFORMATION GÉOGRAPHIQUE, POUVOIR ET ORGANISATIONS - Géomatique et stratégies d'acteurs*. Paris : L'Harmattan.
- Potisepp, K. (2016). *Large-scale Image Processing Using MapReduce*, master of science Estonia: Tartu university Faculty of Mathematics and Computer Science
- Sawa, T (2014) *Leading Advances in the Utilization of Big Data in the Healthcare Industry*; Intel Health & Life Sciences Available at <https://www.intel.com/content/dam/www/public/us/en/documents/white-papers/big-data-healthcare-tokyo-paper.pdf>
- Soares, N., Johnson, A., Patidar, N. (2013). *Geomapping Telehealth Access to Developmental-Behavioral Pediatrics, Telemedicine and e-Health*. July 2013, 19(8): 585-590. Available at <https://doi.org/10.1089/tmj.2012.0226>
- Teboul, B., Berthier, T., (2015). Valeur et Véracité de la donnée : enjeux pour l'entreprise et défis pour le Data Scientist. Actes du colloque “ La donnée n'est pas donnée ” École Militaire – 23 mars 2015 available at <https://tel.archives-ouvertes.fr/tel-01662471/document>
- Verheji, B. (2013). *The Process of Big Data Solution Adoption, an exploratory study within the Dutch telecom and energy utility sector*, Master Management of Technology, Hollande: Faculty Policy and Management Delft University of Technology.

Zikopoulos, P., Eaton, C., Deroos, D., Deutsch, T., Lapis, G. (2012). *Big Data: From the Business Perspective. Understanding Big Data* (978-0-07-179053-6). New York: McGraw-Hill.5-7.