

Introduction

Avant-propos

D'une manière générale, la technologie permet à l'Homme, en lui procurant des outils toujours plus performants, de construire un univers socio-économique propre et d'innover par rapport à cette construction. Les nouvelles technologies favorisent la circulation d'une information multiforme : celle-ci nécessite moins de place et devient plus facile à déplacer, à partager, voir même à vendre. Des codes morse en passant par des fils télégraphiques, nous sommes désormais propulsés dans l'ère du numérique où jpeg et mp3 se côtoient dans les réseaux d'Internet où l'information est devenue plus que jamais immatérielle.

Conformément à la loi de Moore (1985), la technologie permet désormais de traiter des millions d'informations en quelques instants. Chaque jour, nous semons sur la toile des dizaines de petites pierres numériques qui dessinent et consolident un peu plus l'édifice des Big Data¹. Comme une « pyramide du savoir » surgissant un beau jour dans le désert égyptien, voici un nouvel univers en train d'éclorre sous nos yeux. Et contrairement aux monuments des pharaons, ce secteur croît de manière exponentielle et sans limites.

En effet, avec la croissance d'Internet, de l'usage des réseaux sociaux, de la téléphonie mobile, du Cloud Computing², des objets connectés et communicants,

1. Le terme *Big Data* est un nom anglais (traduit par « grosses données » ou « méga données » en français et qui ne nécessite pas d'être employé au féminin ni au masculin. Le mot *data* étant le pluriel latin de *datum*, nous avons décidé de garder ce nom au pluriel dans cet ouvrage.

2. Le Cloud Computing, ou « informatique des nuages » en français, désigne un modèle de télécommunication, né au début des années 2000, en perpétuel mouvement. Ce modèle propose l'accès libre à des services depuis n'importe quel dispositif numérique connecté à Internet. De l'entreposage de fichiers à la gestion de contenu en passant par l'exécution d'applications en ligne. Le terme de *cloud* ou « nuage » vient de la représentation métaphorique de ces services.

l'automatisation des échanges de données, l'information est aujourd'hui plus abondante que jamais et sa croissance est chaque jour plus rapide. L'informatique des Big Data se renouvelle tous les jours pour traiter une énorme quantité de données, souvent non structurées, en des temps record. Dès lors, l'entreprise Ericsson prédit qu'il y aura 50 milliards d'objets connectés dans le monde d'ici à 2020. Le monde crée ainsi en deux jours autant de données que toute l'humanité en a créées pendant deux mille ans. Chaque jour, nous générons 2,5 trillions d'octets de données. A tel point que 90 % des *data* dans le monde ont été créées durant les deux dernières années. L'accroissement des données brutes produites par les particuliers, les entreprises, les institutions publiques et les acteurs scientifiques offre de nouvelles perspectives de monétisation, d'analyses et de traitement. Les Big Data induisent une transformation majeure dans l'usage du numérique par les entreprises de tous les domaines économiques. Ces derniers ont des répercussions considérables en termes de développement, de recherche ou d'amélioration des services, de leur gestion, et de création d'emploi (World Economic, 2012).

Ceci aboutit donc à une société dans laquelle les nouvelles technologies de l'information et de la communication (NTIC) jouent un rôle majeur et central notamment dans le domaine de la santé.

Du fait de la valorisation intrinsèque de l'information médicale touchant à l'intimité de la personne, notre ouvrage dresse une analyse éthique des Big Data principalement dans l'écosystème de la santé. Toutefois, on peut noter que toutes nos réflexions, méthodes et outils qui découlent de ce livre peuvent être adaptés et extrapolés à d'autres secteurs connexes impliquant la production et la diffusion d'autres types de données personnelles, tels que : le commerce, les transports, la finance, la distribution, le *manufacturing*, les services, les *utilities*, les télécoms, le secteur public, et l'éducation.

Aujourd'hui, la médecine moderne est devenue presque inconcevable sans l'utilisation des données personnelles numérisées. L'émergence de la e-santé, la télémédecine, la m-health, les NBIC (nanotechnologies, biotechnologies, informatique, et sciences cognitives), et des Big Data modifient la prestation de santé, la relation médecin-patient, et la compréhension scientifique du corps humain et des maladies. L'exploitation des données personnelles est un sujet sensible, du fait que ces dernières touchent directement à l'intimité de chaque individu. Les situations dans lesquelles se posent de difficiles problèmes de choix stratégiques en matière de gestion de la donnée personnelle sont chaque jour plus nombreuses.

Dans ce contexte, l'interaction de la société avec les NTIC représente un système instable voire précaire. Ainsi, les enjeux associés aux Big Data sont importants, tant

sur le plan économique que pour garantir un espace numérique protecteur de la vie privée et des libertés fondamentales.

Par ailleurs, la révolution de ces immenses volumes de données brutes et hétérogènes va de pair avec le développement d'une nouvelle science des données. La valorisation des Big Data passe par la mise en place d'analyses sophistiquées, véritable « passage à l'échelle » dans la conception et l'usage des modèles d'analyse et la mise en œuvre des algorithmes. Désormais, les logiciels doivent avoir la capacité de détecter l'information intéressante pour obtenir un traitement optimum des données : c'est le *datamining*. Cette approche utilise une méthode inductive et non plus déductive, en cherchant à établir des corrélations entre plusieurs informations sans hypothèses prédéfinies. Notons que cette technique n'a qu'une valeur descriptive sur le plan scientifique, car elle identifie une liaison entre deux variables, mais ne l'explique pas.

Enfin, les techniques d'« analytics avancés »³ s'appuient sur ces larges volumes de données pour chercher des « signaux faibles »⁴ au sein d'une arborescence de catégories identifiées.

A partir de ces outils, les structures peuvent désormais détecter et optimiser, tracer et cibler, voire prédire et prévoir des informations précises (Institut de l'Entreprise, 2014). En effet, l'afflux et le croisement de *data* en temps réel permettent une compréhension plus fine de l'environnement. La prise de décision est améliorée et les actions ou services peuvent être pilotés plus efficacement. De plus, la granularité et le large spectre de sources des données étudiées autorisent la découverte et le suivi à un niveau très fin.

Dans ce contexte, le numérique bouleverse l'écosystème de la santé. Qu'il s'agisse de prévention, de diagnostic ou de soin, il ne se passe pas un jour sans que l'annonce d'une innovation ne vienne apporter sa part de transformations à la médecine. Ce qui sonne comme l'ultime provocation de l'ère du numérique pour les professionnels de la santé, résonne comme la promesse d'un diagnostic plus rapide et précis, voire d'un traitement totalement personnalisé pour les patients. Mais l'intérêt des Big Data ne s'arrête pas à la santé individuelle. En termes de santé

3. Le terme « analytics avancés » comprends les techniques et les méthodes suivantes : statistiques non-paramétriques, réduction de dimension, règles d'association, analyse de données réticulaires (*network analysis*), classification non supervisée (*cluster analysis*), algorithmes génétiques, etc.

4. Le signal n'est pas faible par la nature de la source d'information (formelle ou non) mais par le rattachement entre cette source et une entité en mesure de prendre une décision après avoir mis en relation le signal et un scénario. Un signal faible est difficilement interprétable, informel, improbable mais généralement annonciateur d'un évènement à venir.

publique, l'exploitation de ces données massives peut contribuer au développement de nouveaux produits de santé, détection des signaux faibles lors d'épidémies ou d'effets indésirables graves, etc. Cependant, une telle révolution comporte également des risques éthiques autour des données de santé à caractère personnel comme l'intégrité, la fiabilité, la sécurité, le respect de la vie privée et des libertés individuelles (Goodman, 2014), la réputation, la régulation, etc.

A partir de ce constat, l'ossature de notre livre se découpe selon le triptyque suivant :

- il définit tout d'abord les applications des Big Data en santé. Il montre en quoi, les Big Data constituent un phénomène nouveau, un changement de paradigme médical et à quelles évolutions sociales et techniques il est associé. Il détaille ensuite les usages et les possibilités offertes par les analyses de masses de données et des utilisations concrètes qui en découlent. L'ouvrage s'attache à signaler les principaux risques et enjeux relatifs à ces usages. L'analyse des Big Data peut engendrer des inquiétudes du fait du croisement d'un grand nombre de données. Ainsi, se pose un certain nombre de questions des conditions nécessaires, notamment, au respect de la vie privée (Schadt, 2012) et à la sécurité des données ;

- puis, ce livre présente la valeur éthique de la datasphère médicale *via* la description d'une modélisation d'analyse éthique des Big Data. Ce chapitre décrit, d'une part, les principes d'une hiérarchisation sélective des données de santé et de l'*Ethical Data Mining* et d'autre part, les outils d'une évaluation éthique des Big Data en santé ;

- enfin, ce manuscrit présente les recommandations et les étapes indispensables pour une gestion et une gouvernance réussies des données personnelles de santé, et détermine quelles sont les conditions indispensables au développement de l'étude des Big Data.

Les caractéristiques des Big Data

Le terme de Big Data désigne une nouvelle discipline qui se situe au croisement de plusieurs secteurs tels que les technologies, les statistiques, les bases de données et les métiers (marketing, finance, santé, ressources humaines, etc.). Concrètement, cette approche est une réponse à l'explosion des données non structurées voire parfois multi-structurées observée dans l'univers numérique (Internet, RFID, mobile). Cette activité permet de capter les données numériques, de les traiter à très grande vitesse et de les rendre ainsi exploitables pour les organisations, les entreprises et les institutions publiques, quelle que soit la nature de cette donnée. Les Big Data sont avant tout un dispositif technologique dont l'objectif est de transformer une donnée

brute (par exemple, de localisation, de navigation, des métadonnées, issue des réseaux sociaux ou de l'administration) en connaissance directement exploitable, jusqu'au fondement de la méthode scientifique, par une structure (Guillaud, 2011). Cela recouvre une grande diversité de traitements de données mis en œuvre à l'aide de techniques d'apprentissage automatique, d'analyse prédictive ou préemptive, de fusion ou de recherche des données. Ces NTIC ont toutes pour objectif de donner du sens à la *data*.

Les outils à l'œuvre lors de ce processus constituent la réelle innovation de ces dernières années. Ces Big Data ont été rendus possibles grâce à une puissance technologique (comme le Cloud Computing et les outils comme Hadoop, MapReduce ou Cassandra) qui a rendu possible des applications et des services qui jusque-là n'étaient que théoriques. Cela est principalement associé à deux enjeux : le volume des données et leur complexité (Bensabat *et al.*, 2014).

A l'origine des Big Data, il y a la matière de base : la *data*. Cette donnée provenant des différents supports que sont Internet, les capteurs RFIF ou les smartphones, s'exprime selon différentes formes de plus en plus complexes *via* les vidéos, les forums de discussion, réseaux sociaux. Hadoop est rapidement devenue la référence en matière de parallélisation des Big Data. Les *Big Four* américains GAFA (Google, Apple, Facebook, et Amazon), les trois géants chinois BAT (Baidu, Alibaba, et Tencent), et les quatre grandes entreprises emblématiques de la *disruption* numérique à savoir NATU (Netflix, Airbnb, Tesla, et Uber) mettent en place des technologies de promotion et de ciblage de leurs utilisateurs. Les éditeurs se sont employés à adapter l'offre initiale Open Source Hadoop en solution commerciale, customisée en fonction des applications clientes et des indicateurs de performance. C'est aujourd'hui la principale offre qu'on retrouve sur le marché, soit au travers de solutions Cloud Computing, soit par l'intermédiaire d'*appliances* (boîtier hardware intégrant la technologie Hadoop) et la mise au point de modes de calcul à haute performance (MapReduce⁵). D'autres technologies s'ajoutent à Hadoop afin de spécialiser le traitement (tels que Pig, Cassandra, Hive, etc.).

Ces logiciels sont des systèmes de gestion de bases de données non relationnelles distribuées, employant des modalités de requête NoSQL⁶ (*Not Only SQL* en anglais), qui dépassent donc les codifications du langage SQL (*Structured Query Language*),

5. Le MapReduce est une technique qui segmente le traitement d'une opération (appelée « job » chez Hadoop) en plusieurs étapes, dont deux élémentaires, afin de faciliter la parallélisation des traitements sur les données : le « Mapping » et le « Reducing » (Corinus, 2012).

6. Le terme « NoSQL » désigne une catégorie de systèmes de gestion de base de données (SGBD) destinés à manipuler des bases de données volumineuses pour des sites de grande audience. Ces SGBD ne sont plus basés sur l'architecture classique de bases relationnelles.

en français, « langage de requête structurée ». Cette approche NoSQL est d'ailleurs une des autres principales caractéristiques de l'émergence des Big Data. Généralement, ces bases de données NoSQL sont classées selon la manière dont elles stockent les *data*. On retrouve ainsi des catégories comme les bases orientées clé-valeur, document, colonne ou encore les bases structurant les données en se fondant sur la théorie des graphes. On peut donner comme exemples des solutions telles que : SenseiDB ; Voldemort (LinkedIn) – Cassandra ; Hive ; HBase (Facebook) – Dymomo ; S3 (Amazon) – CouchDB (Ubuntu One) – MongoDB (SourceForge.net) – MapReduce ; BigTable (Google) – No Database SQL (Oracle) – Storm ; FlockDB (Twitter) – Hadoop ; S4 (Yahoo).

En définitive, ces différents outils technologiques permettent de traiter tous types de données, dans de considérables quantités, et en un temps limité. Les Big Data promettent de réconcilier, d'harmoniser, d'unifier, d'interconnecter, et de fluidifier d'importants volumes de données digitales, dans un monde où tout est désormais numérisé. Elles permettent, notamment, de connecter de nombreux flux de données par exemple de la géolocalisation, des données CRM (*Customer Relationship Management*), le suivi des interactions des utilisateurs sur le site de l'entreprise et au sein des newsletters envoyées. La problématique du temps réel est d'ailleurs de plus en plus prégnante dans les offres des éditeurs, qu'il s'agisse de traitement sur des serveurs dédiés ou de traitement *in-memory* (dans la mémoire de l'ordinateur).

De plus, il n'existe pas de mesure unique pour caractériser les Big Data, mais plusieurs. Historiquement, ce phénomène a été défini selon trois mots, les « 3V » (Volume – Variété – Vitesse), par Doug Laney (2001), analyste chez Gartner. Depuis, ce concept s'est systématiquement complété. Nous en avons comptabilisé trois supplémentaires, à savoir : Véracité, Visualisation et Valeur. Dès lors, c'est l'interaction et la combinaison de ces « 6V » spécifiques qui définissent ces « méga données » plutôt que de la taille d'une base de données spécifique. Nous allons maintenant définir ces six caractéristiques propres aux Big Data.

Volume

Notre usage accru des NTIC (smartphones, réseaux sociaux, objets connectés, etc.) nous incite à produire de plus en plus de données lors de nos activités quotidiennes aussi bien professionnelles que personnelles. Aujourd'hui, le monde produit un nombre considérable de données en tout genre. La première est la « loi de Moore » (1965), qui observait empiriquement que le « nombre de transistors par circuit de même taille doublait tous les dix-huit mois, sans augmentation de leur coût ». Cette constatation n'a jamais été démentie jusqu'à présent, et a depuis été élargie aux mémoires de stockage des données.

Selon l'Organisation des Nations unies (ONU), plus de données ont été créées en 2011 que dans toute l'histoire de l'humanité⁷. La masse de données numériques est passée de 480 milliards de gigaoctets en 2008 à 2,72 zettaoctets en 2012. Jusqu'en 2020 cette masse va continuer à progresser à une vitesse exponentielle (Bail, 2014) pour atteindre les 40 zettaoctets (Aboulican, 2013). D'après le rapport d'étude de Global Investor (Crédit Suisse) en juin 2013, le monde digital aura grandi en 2020 de 300 fois sa taille de 2005, grâce notamment à l'émergence des objets connectés. Ainsi, le volume de données produites chaque année dans le monde devrait être multiplié par 44, d'ici 2020. La quantité d'information archivée croît quatre fois plus vite que l'économie mondiale, pendant que la puissance de traitement informatique croît neuf fois plus vite. On estime que 90 % de l'ensemble des données disponibles, aujourd'hui, ont été créées ces deux dernières années (Brasseur, 2013). D'après le Dr Laurent Alexandre⁸, le volume total des données d'e-santé dans le monde double tous les 73 jours. Cette importante volumétrie de données ouvre dès lors le champ aux systèmes experts.

Variété

Les Big Data représentent une grande diversité des contenus ou formats (texte, mails, vidéos, logs, images, son, etc.) et de sources (*Machine To Machine*, smartphones, etc.) de données. On parle alors de données non structurées ou multi-structurées. En effet, les *data* du marché sont omniprésentes : elles ne proviennent plus uniquement de sources internes mais également des forums de discussion, réseaux sociaux, ou d'autres sources externes qui favorisent une information des applications en temps réel. Cette notion de sources multiples est un des fondements du concept des Big Data. Désormais, il apparaît illusoire d'établir une décision qui serait fondée sur une seule source de données qui détiendrait la vérité. La nécessité de croiser plusieurs sources pour dégager des conclusions exploitables, s'impose à nous. Notons qu'une logique d'efficacité souhaite que l'on cherche à réduire sensiblement le temps nécessaire de ces traitements afin de se rapprocher du temps réel.

Les données de santé peuvent être issues des instituts de recherche, centres épidémiologiques, laboratoires pharmaceutiques, centres d'imagerie, comptes rendus hospitaliers, compagnies d'assurance, des dossiers clients, mais également des réseaux et forums sociaux. Les types de sources de données sont nombreux. Ils peuvent

7. « Robert Kirkpatrick, Director of UN Global Pulse, on the Value of Big Data », theglobalobservatory.org, 5 novembre 2012.

8. Table ronde intitulée : « Entre secret et transparence, quel espace ? », lors du CHAM : *Convention on Health Analysis and Management*, le 26 septembre 2015.

provenir de données sociales (*via* les réseaux et autres médias), personnelles (par exemple des données de dispositifs de repérage), des capteurs transactionnels, ou administratives (Kinney *et al.*, 2009).

Naturellement, les approches axées sur les Big Data dépendent de la qualité intrinsèque des données qui les sous-tendent.

Vitesse (ou vélocité)

Les Big Data sont générées et évoluent très rapidement. Le traitement accéléré des données peut aller jusqu'au temps réel. Cela impose donc un traitement rapide presque à flux tendu afin de pouvoir en exploiter des informations précises et en dégager des conclusions pertinentes. D'après le rapport de l'Institut Montaigne⁹, le « calcul de Grötschel » établit que « la vitesse de calcul des algorithmes progresse quarante-trois fois plus vite que la puissance des microprocesseurs, les algorithmes pouvant être définis comme les séquences d'opérations et d'instructions d'un programme informatique ». La prise de décision d'un être humain concernant notamment un achat, est de l'ordre de la dizaine de minutes : cela laisse le temps aux dispositifs décisionnels de croiser des informations et de tendre vers une décision. Avec Internet, ce processus voit ses repères complètement chamboulés. Désormais, tout va plus vite ! L'ordre de grandeur de quelques minutes a laissé la place à quelques secondes. En effet, l'interaction entre l'internaute et son environnement s'oriente progressivement de la sphère de la cognition plus lente vers la sphère de la motivation voire celle de l'émotion à forte célérité. Sur le plan de la psychologie sociale, le phénomène est encore démultiplié par la vitesse de propagation de l'information au sein des communautés connectées. Désormais, en quelques centaines de secondes, des millions d'internautes peuvent être informés, modifiant potentiellement leur comportement. Acquérir de l'information ne suffit plus : il faut aussi anticiper ses impacts sociaux. L'équation liant la vitesse et la contribution est primordiale dans la vie sociale. Toute société vit à travers cet équilibre instable entre partage et compétition. C'est dans le même état d'esprit que sont développés les algorithmes qui permettent la célérité.

Véracité (qualité d'origine)

Les données issues des applications centrales d'un système d'information (SI) sont limitées en nombre mais contrôlées en termes de niveau qualité, et de cohérence.

9. Rapport de l'Institut Montaigne intitulé : « Big Data et objets connectés : faire de la France un champion de la révolution numérique », p. 1-228, avril 2015.

A contrario, des données publiques associées au comportement ou à des sentiments, peuvent être abondantes mais soumises à des prismes déformants. Dans l'utilisation qui en sera faite, il apparaît indispensable de pouvoir neutraliser ces phénomènes sans pour autant modifier la donnée source. La gestion des critères de véracité et d'origine des *data* manipulées devient fondamentale. Les Big Data présentent des incertitudes qui sont imputables au manque de cohérence, à l'ambiguïté, à la latence, et à l'incomplétude, des éléments d'information. Les processus décisionnels doivent prendre en considération ce degré variable d'incertitude. Pour cela, ces dispositifs doivent avoir la capacité de distinguer, d'évaluer, de pondérer ou de trier différentes catégories de données afin de conserver une certaine authenticité.

Visualisation

La visualisation des données est un des critères de base dans le succès du traitement des Big Data. Cette *dataviz* (ou datavisualisation) s'est développée au croisement du design et de la statistique. Elle constitue une démarche structurante et collaborative dans l'accompagnement des données produites par les objets connectés. Sa valeur ajoutée réside dans la représentation ou la personnalisation de la donnée et la diffusion de son contenu aux décideurs opérationnels et au grand public pour que les Big Data soient considérés comme utiles par ces derniers. Cette datavisualisation procède donc à la fois d'une analyse et d'une mise en forme graphique lisible, notamment *via* des tableaux de bord ou des représentations en radar. Le véritable enjeu pour ce marché des Big Data est de réaliser des outils directement en lien avec la perception du destinataire de l'information reçue. C'est en élaborant des ergonomies adaptées à leur utilisateur que les *dataviz* pourront s'implanter durablement dans le champ communicationnel et décisionnel des entreprises et des organisations (Hamel et Marguerit, 2013). Enfin, cette visualisation des données doit répondre à deux impératifs : d'une part, être suffisamment complète pour gérer des corrélations complexes dans des ensembles de données très volumineux ; d'autre part, pouvoir traduire ces corrélations en visualisations pertinentes mais assez simples pour réellement appuyer la prise de décision d'une structure.

Valeur

La cohérence, la fiabilité, la prédictibilité et la qualité des données sont devenues des critères essentiels dans le traitement des grands ensembles de données. En effet, les Big Data se définissent par une valorisation de la donnée, c'est-à-dire une transformation de celle-ci en information qui générera par la suite d'importants bénéfices *via* les usages que l'on en fait (GFII, 2012). S'il est difficile de juger *a priori* de la valeur d'une donnée brute, il semble décisif de s'attacher à intégrer des

sources de données susceptibles de générer une information dont la valeur ajoutée est avérée. Les Big Data comportent des valeurs intrinsèques (de conception) et extrinsèques (d'usage) à la donnée. Il ne faut jamais sous-estimer l'importance positive ou néfaste d'une donnée. Une source de données sans usage interne peut avoir une valeur monétisable pour un collaborateur. Et une autre source de données, *a priori* sans valeur, peut s'avérer dans le cadre d'un partenariat, être porteuse d'un signal discriminant.

Qu'est-ce que l'éthique ?

Le mot « éthique » prend son origine dans le terme grec *ethos* signifiant « les mœurs » (Cicéron), « les habitudes » (Platon et Aristote). L'éthique concerne « l'habitat » et « le caractère d'une personne ». Ainsi, la manière dont nous habitons le monde représente la manière dont nous sommes quelqu'un. L'expression « être habité » prend tout son sens et sa valeur symbolique. Vue sous cet angle, l'éthique est une réflexion sur les habitudes qu'il faut contracter pour rendre un espace habitable. L'éthique apporte donc un questionnement sur les valeurs qui sous-tendent l'action, propice à un conflit de valeurs dans un monde des idées. Elle « trouve naturellement sa source de réflexion dans l'action » (Hervé, 1997). Son objectif est donc de donner du sens aux actions. L'éthique est une disposition individuelle à agir selon les vertus dans une situation donnée afin de rechercher la bonne décision. Elle n'a de sens que dans une situation propre dans laquelle elle admet l'argumentation, la discussion et les paradoxes.

L'éthique renvoie aux conditions d'une vie bonne, pour soi et pour les autres. Elle est « le désir d'une vie accomplie, avec et pour les autres ; dans des institutions justes » (Ricoeur, 1991). Elle est l'ordre de l'interprétation et/ou de la pratique. L'acte éthique est d'abord une réponse (du latin *respondere* : répondre de, répondre à, d'où responsabilité) à une situation limite et complexe. L'éthique assume trois fonctions principales, à savoir : la détermination de ce qu'est la morale, la connaissance des raisons justifiant l'effort d'une personne pour vivre moralement, et l'application à la vie pratique des résultats obtenus dans les deux premières fonctions.

Toute personne cherche les valeurs qui l'animent, choisit les principes d'action qui doivent prévaloir, veille aux conditions de leur mise en œuvre, se rend sensible à leur réalité. L'éthique est avant tout une aventure, une boussole, la recherche d'une interprétation et d'une posture adéquates par rapport à la réalité à laquelle on appartient. Nous nous confrontons à la réalité *via* le prisme de nos ressentis, émotions, de nos objectifs, de nos modèles de pensée, et nos représentations, qui nous interpellent et nous mobilisent. L'interprétation et l'analyse comportent à la fois une part d'intellect et d'affect. C'est l'ensemble de ces deux composantes qui

donne une valeur à la réalité et s'articule avec des idées ou des processus d'idées dans lesquels il trouve un sens cohérent. C'est ce dispositif complexe de valorisation et de dévalorisation, entre rationnel et sensible, qu'il nous importe d'appréhender et de nous approprier. Cela passe nécessairement par un système de médiation jouant un rôle dans les processus de signification afin de conditionner et orienter la production du sens. Dans ces conditions, l'éthique peut être définie comme « un mode de régulation des comportements qui provient de l'individu et qui met l'accent sur des valeurs coconstruites et partagées pour donner un sens à ses décisions et à ses actions, faisant ainsi appel à son jugement personnel et à sa responsabilité » (Boisvert *et al.*, 2003).

Dans ce livre, notre schéma de réflexion prend en partie sa source et ses inspirations sur les théories classiques de l'éthique que nous sommes habitués à parcourir. C'est-à-dire au modèle grec de la vertu¹⁰, où l'éthique s'intéresse d'abord à l'individu (l'agent) qui exécute une action, ou aux théories dites relationnelles (comme l'utilitarisme¹¹, le contractualisme¹² et le déontologisme¹³) dont la préoccupation majeure est la nature et la valeur morale des actions exécutées par l'agent. Notre raisonnement se fonde plus globalement sur une éthique orientée vers les personnes qui créent ou reçoivent l'action impliquant les Big Data et en subissent ses effets.

Pour cela, nous appliquons des principes universels à la fois consensuels et régulateurs tendant vers une cohésion sociale. En éthique, le principe constitue le socle qui se représente « sous la figure d'un commandement » (Le Coz, 2007). Il est immuable, universel, intangible et sa valeur n'est pas influencée par le cours de l'histoire. C'est pourquoi toutes les sociétés tendent à cette universalité illustrant cette unicité qui nous entoure. L'universel est présent dans la multiplicité des choses et donc de l'être humain. Le terme de principe provient du latin *principia*, lui-même emprunté du grec *arché* qui peut se traduire en deux significations :

– premièrement, il désigne « ce qui vient en premier, ce qui est à la source » (Le Coz, 2007). On revient aux origines de l'architecture culturelle, aux fondements des mœurs, des règles du droit, des us et coutumes d'une société donnée ;

– deuxièmement, il signifie « ce qui fait autorité » en faisant référence au « prince » qui « vient en premier » et qui est investi de l'autorité légitime suprême.

10. Principe moral de tenter d'être vertueux et de questionnement casuistique universel.

11. Principe moral universel de maximiser les conséquences.

12. Principe moral universel qui affirme que toute société est fondée sur un contrat social.

13. Principe moral universel d'impératif catégorique (Kant). Cette théorie éthique affirme que chaque action humaine doit être jugée selon sa conformité (ou sa non-conformité) à certains devoirs. Elle est centrée sur le respect des droits et des devoirs.

Si nous consultons la littérature de bioéthique internationale, nous constatons que quatre constantes reviennent constamment selon les pays. Ainsi, les références aux principes d'« Autonomie », de « Bienfaisance », de « Non-malfaisance » et de « Justice » (Beauchamp et Childress, 2001) apparaissent inlassablement dans tous les ouvrages quel que soit son lieu d'origine, sa culture, ses croyances, sa philosophie ou sa religion (voir encadré 1).

- **L'autonomie**

Elle désigne le fait qu'une personne se donne à elle-même sa propre règle de conduite, puisque les termes grecs *autos* et *nomos* signifient respectivement « soi-même » et « loi, règle ». Ce principe a pour vocation de faire participer le patient au processus décisionnel.

- **La bienfaisance**

Elle contribue au bien-être d'autrui. Elle doit répondre à deux règles bien précises : l'action entreprise doit être bénéfique et utile, c'est-à-dire avoir un rapport coût-bénéfice positif.

- **La non-malfaisance**

Elle a pour objectif d'éviter le mal à celui dont on a la responsabilité (le patient) et de lui épargner des préjudices ou des souffrances qui n'auraient pas de sens pour lui. Sa finalité implique donc que l'on fasse du bien et que l'on s'abstienne de nuire. Ce principe apparaît dans la maxime hippocratique *primum non nocere*¹⁴, dont la conséquence est de faire du bien aux patients et de les écarter du mal et de l'injustice.

- **La justice**

Elle a pour vocation de partager entre tous les patients les ressources disponibles¹⁵. Ce principe est étroitement lié aux notions d'égalité et d'équité qui interviennent directement dans le processus d'une décision de justice. Idéalement, toute action devrait tendre vers une égalité parfaite, mais selon les circonstances et la nature des personnes, l'équité s'impose souvent afin d'établir des priorités et une certaine hiérarchie dans les actes à réaliser. Ce principe renferme une portée que l'on peut désigner comme « macro-éthique » concernant l'ensemble des patients alors que les trois autres précédents principes ont une dimension beaucoup plus individuelle et relationnelle considérée comme « micro-éthique ».

Encadré 1. *Vocation des quatre principes éthiques*

14. « D'abord ne pas nuire ».

15. Ressources en temps, en argent, ou en énergie.

Les sciences humaines et sociales sont évidemment impliquées dans les aspects éthiques : juristes, sociologues, épistémologues, philosophes, chercheurs en sciences de l'information et de la communication, en sciences cognitives, psychologues, géographes, gestionnaires et économistes, anthropologues, ethnologues, linguistes, usagers et patients dont le témoignage est précieux... Tous apportent des visions et des arguments essentiels à une réflexion éthique sur l'usage et l'expansion de ces nouveaux outils qui vont profondément modifier la société et la vie de chacun d'entre nous, notamment dans le cadre de la santé.

Dans ces conditions, nous développons l'idée d'une éthique qui se déploie dans un dialogue. Il en découle donc une approche délibérative où la vision éthique la plus appropriée émane d'une discussion entre tous les acteurs concernés par la conception, la mise en place et l'usage d'un SI destinés à la prise en charge du soin. Notre éthique véhicule une vision à la fois standardisée et algorithmique de la décision médicale, dissociée d'une approche clinique personnalisée et synthétique élaborée autour des besoins du patient.

Enfin, l'éthique peut être définie comme une réflexion sur l'action pour laquelle il faut chercher le sens. Le « comment » d'un acte se transforme en « pourquoi ». L'éthique devient la recherche d'une justification des normes que nous établissons. Ces normes ne sont pas une condition *sine qua non* de la solution éthique des dilemmes pratiques, mais plutôt le résultat du processus de prise de décision lui-même (Spranzi, 2013). Pour cela, il est essentiel d'identifier et de caractériser les conditions et la méthode qui nous permettent d'aller à la recherche des quatre principes éthiques vus auparavant pouvant donner un sens au choix d'agir d'une manière ou d'une autre.

L'éthique dans le numérique

Y a-t-il une éthique propre au numérique ? Cette question revient régulièrement et fait débat, tellement il apparaît non naturel d'associer une science humaine à une science technologique que presque tout oppose. Et pourtant, le numérique crée de toute part des injonctions contradictoires qui ont par conséquent des répercussions éthiques spécifiques aux technologies de l'information et la communication (TIC). Si les Big Data sont éthiquement neutres, leurs usages ne l'est pas. Des comportements singuliers naissent des usages de ce nouvel espace-temps que génère le numérique. Les NTIC sont un phénomène culturel voire anthropologique. Ils produisent de nouveaux comportements, de nouvelles visions du monde, et de nouvelles normes sociales.

On peut prendre l'exemple, de l'anonymisation qui pose la question de la responsabilité des personnes dont l'invisibilité peut dédouaner de certaines règles de bienséance. L'instantanéité et l'ubiquité que permet l'Internet, répercutent nos actes de paroles et de pensée de manière conséquente et irréversible. Désormais, l'éthique et la technologie ne doivent plus être mises en relation selon un dispositif à deux étapes. Les questions éthiques doivent faire partie intégrante de leur mission et ainsi construire une réflexion éthique orientée. Dès lors, on ne parle plus d'une approche interdisciplinaire mais plutôt d'une fusion aboutissant à une véritable éthique du numérique où la question des implications sociales et morales s'intègre dans les NTIC.

Dans ces conditions, il devient essentiel d'établir des attentes et des préconisations éthiques spécifiques au monde numérique et de réifier des nouveaux systèmes de valeurs d'éthique et de droit, en gardant toujours en tête cette question : est-ce que le numérique peut induire un risque de mésusages de nos comportements éthiques ?

En éthique, le terme de « valeur » est de l'ordre du devoir-être. C'est un étalon de mesure qui permet de jauger les faits. Il indique des idéaux à poursuivre. Ce mot a une connotation générale et dynamique ; il a d'abord une évocation philosophique avant d'avoir une retombée éthique. Un des fondements de l'éthique est cette impérativité à faire appel à la rationalité des acteurs. Cette idée se structure *via* une entente dans la coordination, l'échange et le partage entre les protagonistes. Chaque personne contribue à la recherche d'une intercompréhension de la situation à analyser. Cela présuppose donc un certain consensus et solidarité entre les interlocuteurs qui partagent une même finalité. Si l'éthique est déjà par nature complexe à définir, sa mise en perspective avec le numérique relevait d'un autre défi. L'éthique demande une vision, un dessein, une ambition qui se concrétise dans une orientation.

Internet et les Big Data devenant omniprésents dans nos vies quotidiennes, les préoccupations éthiques autour de la sécurité de l'information sont devenues l'une des tendances les plus chaudes dans le tourbillon de la recherche et de la pratique de la technologie de l'information (Taherdoost *et al.*, 2011). Ceci est principalement dû aux progrès technologiques qui ont permis la production, la collecte, le stockage, le traitement et la transmission de données à un rythme conséquent sans précédent de diverses sources (Hamid, 2007). De nombreuses études sur l'éthique des technologies de l'information et la communication (TIC) visent à clarifier si l'éthique des TIC est différente de l'éthique dans d'autres champs. James Moor (1985) affirme que les NTIC ont généré des problèmes éthiques « uniques » ; car ces technologies sont « logiquement malléables » et offrent des possibilités nouvelles pour des comportements éthiques. Les applications technologiques ont dévoilé de nouveaux comportements éthiques et donc favorisé la génération d'uniques et nouveaux problèmes éthiques. Les questionnements éthiques associés aux TIC et l'émergence des applications ont été appelés l'« éthique de l'information » (Mason, 1986).

Par ailleurs, toute technologie définit une relation entre les êtres humains et leur environnement, tant humain que physique. L'idée d'une dynamique technologique suivant sa propre trajectoire reste très forte. L'héritage de McLuhan selon lequel les technologies se développent et ont une répercussion sur la société demeure prégnant (Badillo et Péliissier, 2015). Les TIC ont le pouvoir d'hypnotiser la société : « toute nouvelle technologie diminue donc l'interaction des sens et de la conscience, et plus précisément dans le domaine nouveau des innovations où se produit une sorte d'identification du sujet et de l'objet » (McLuhan, 1977).

Aucune technologie ne peut être considérée comme purement instrumentale. Ceci est particulièrement pertinent lorsqu'il s'agit de grands SI automatiques, mis au point pour contribuer à la gestion et l'intégration des grandes organisations, comme les structures de santé. Dans un tel contexte, l'environnement est principalement composé d'êtres humains. En faisant évoluer les SI, les facteurs humains président simplement des facteurs techniques. Même si la satisfaction de ce dernier est obligatoire, ils ne sont jamais vraiment suffisants. Dans tout Big Data, le facteur humain et l'interaction homme-ordinateur sont fondamentaux. Toutefois, dans un contexte multi-utilisateurs simultanés, l'interaction homme-homme est la principale question à résoudre. L'évaluation des grands ensembles de données numériques, tels que ceux trouvés dans le domaine de la santé, est fondée sur le concept de *relationship inter-humaine* (Fessler et Gremy, 2001) qui sous-tendent la conception, la mise en place et l'utilisation des Big Data. Dans ces conditions, ces « méga données » apparaissent principalement comme un système social, avec ses caractéristiques psychologiques, sociologiques et éthiques. L'écart entre l'espace d'action et de représentation de l'action s'est fortement réduit. Une ambiguïté s'installe entre action réelle du numérique et représentation d'elle-même, ou entre mouvement continu et gestion discrète de ce mouvement. Dès lors, le numérique est un lieu où la liberté semble être sans limites, du fait que tout est dans la continuité du mouvement. En même temps, ce mouvement est finalement discret et donc facilement contrôlable. L'éthique du numérique doit « essayer de penser le rapport entre geste et cristallisation du geste » (Vitali Rosati, 2012). C'est à partir de cette approche que notre réflexion éthique doit commencer sa démarche et poser les principes éthiques spécifiques à l'action numérique. L'éthique des NTIC peut se découper en trois grandes thématiques :

- l'éthique des données : définissant les principes éthiques garantissant le traitement équitable de données et la protection des droits individuels, tout en utilisant des Big Data à des fins scientifiques ou commerciales ;

- l'éthique des algorithmes : traduisant l'étude des problèmes éthiques et des responsabilités des concepteurs de données scientifiques, concernant les conséquences imprévues et indésirables, ainsi que les occasions manquées sur la conception et le déploiement d'algorithmes complexes autonomes ;

– l'éthique des pratiques : représentant l'identification d'un cadre éthique approprié pour façonner un code déontologique sur la gouvernance et la gestion des données, favorisant à la fois le progrès de la science des données et la protection des droits des personnes concernées.

Avant de rentrer directement dans le vif du sujet en décrivant les enjeux et les risques éthiques fondamentaux qui alimentent notre réflexion autour des Big Data en santé, il nous a semblé indispensable d'en expliquer leurs finalités propres, leurs visées et plus généralement leurs raisons d'être, c'est-à-dire la personne. Désormais, les dirigeants d'entreprise évoquent un recentrage de leur stratégie sur les besoins des personnes (par exemple, des usagers de santé pour le domaine médical). Pour le Dr Channin, « la priorité numéro un d'un SI est le patient » (Channin, Bowers et Nagy, 2009). La révolution technologique concernant le secteur de l'information doit être menée dans l'intérêt des patients et d'une meilleure prise en charge. En d'autres termes, la seule valeur à prendre en compte, en vue de la conserver, est la personne humaine considérée dans sa dignité d'être moral. Cette dignité humaine constitue une valeur absolue que l'on donne à la personne. Ainsi, des principes éthiques, pratiques, techniques et ergonomiques doivent être imposés afin que les patients et leurs familles restent les principaux bénéficiaires de cette évolution technologique. Cela est d'autant plus vrai que toute réflexion éthique est un conflit entre des valeurs humaines. Quelles que soient nos pensées religieuses, nos cultures, nos influences politiques ou domaines d'activité, ce sont nos émotions qui révèlent nos valeurs profondes. Comme le souligne si bien Pierre Le Coz (2010) lors de la première journée d'éthique « Cancer et fertilité » de l'Institut Paoli-Calmettes, « s'il n'y a pas d'émotion, il ne peut y avoir de valeurs formalisées et donc pas d'éthique ».

En effet, chaque grand principe éthique peut être associé à une émotion particulière. On peut faire le couplage suivant :

- le respect pour : le principe d'autonomie ;
- la compassion pour : le principe de bienfaisance ;
- la crainte pour : le principe de non-malfaisance ;
- l'indignation pour : le principe de justice.

Pour Davis et Patterson (2012), quatre éléments définissent, pour les particuliers et les organisations, ce qui peut être considéré comme le cadre éthique des données numériques :

- l'identité : quelle est la relation entre notre identité *offline* et notre identité *online* ?
- la confidentialité : qui doit contrôler l'accès aux données ?

– la propriété : qui possède les données, les droits de les transférer, et quelles sont les obligations de personnes qui génèrent et utilisent ces données ? Est-ce que notre existence se constitue d'actes créatifs sur lesquels nous avons un droit d'auteur ou d'autres droits relatifs à la création ?

– la réputation : comment pouvons-nous déterminer quelle donnée est digne de confiance ? »

Dans le cadre de la technologie, l'éthique a affaire à des actes, des actions qui ont une portée causale sociale incomparable en direction de l'avenir et qui s'accompagnent d'un savoir prévisionnel qui, peu importe son caractère incomplet, déborde lui aussi tout ce qu'on a connu autrefois. Elle peut donc se définir comme étant un dispositif de réflexion sur la signification morale de l'action. Cette définition est destinée à être large et fondamentale et à intégrer plusieurs composants de l'éthique informatique (Waskul et Douglass, 1996). On compte principalement, cinq applications éthiques relatives aux TIC :

– l'éthique de l'*empowerment* : associée au patient acteur (e-patient) qui demande son autonomie et sa dignité (respect de ses droits) ;

– l'éthique de l'accès : avec le droit fondamental et la transparence (*Universal Design*) ;

– l'éthique de la dissémination : relative à une mutation évolutive de l'informatique de contrôle vers l'informatique de service (centralisation et distribution) ;

– l'éthique de la réappropriation : centrée sur les mutations comme potentiels (littérature numérique) ;

– l'éthique du collaboratif : entourant le partage d'information (sur le Web avec notamment les forums en ligne ou les réseaux sociaux).

Généralement, l'éthique du numérique se traduit par des questionnements, d'une part, sur le comportement et l'usage des individus face aux NTIC, et d'autre part, sur le comportement de plus en plus autonome des outils technologiques en tant que tels. Ces technologies sont, le plus souvent, programmées dans le but de mener à bien des actions indépendamment d'une intervention humaine (comme les algorithmes de recommandations et de décisions). Dans ce cadre, l'éthique, en tant que telle constitue un mode de régulation des comportements basé sur le respect de valeurs que l'on juge essentielles, et aurait dû empêcher cette objectivation de l'être humain et apporter un cadre à l'utilisation des données (Eynard, 2012).

Il faut y ajouter l'ordre de grandeur des actions à long terme et très souvent également leur irréversibilité. Tout cela place la responsabilité au centre de l'éthique, y compris les horizons d'espace et de temps qui correspondent à ceux des principes éthiques. Ainsi, la responsabilité d'une entreprise c'est de savoir comment bien utiliser éthiquement parlant le numérique et les données. Dans ces conditions, le directeur du système d'information (DSI) doit veiller à l'éthique des projets Big Data.

Notons que la plupart des pratiques professionnelles autour des Big Data sont encadrées par des lois qui diffèrent selon la culture et les mentalités des pays. Dans l'ensemble, celles-ci visent à nous prémunir contre des comportements inappropriés et à préserver l'ordre éthique de la société. Cependant, ces lois ne couvrent qu'une partie des situations où est en jeu l'éthique.

En corrélat, puisque dans cette société où l'information prédomine et où les entités sont faites d'information, notre éthique s'oriente également vers le courant de pensée de l'éthique télématique instaurée par Luciano Floridi, professeur à l'Université d'Oxford, sur ce que serait l'éthique dans une société de l'information et dont les caractéristiques sont plus en adéquation avec notre sujet d'analyse (Floridi, 1998). En effet, contrairement, aux modèles classiques intrinsèquement anthropocentriques, individualistes et de nature sociale, l'éthique télématique s'intéresse avant tout à l'environnement (appelé aussi « infosphère¹⁶ ») où se génère et se propage l'information, notamment les Big Data. Cette infosphère représente un espace numérique constitué par un patrimoine persistant et volatile dans un espace géographique souvent indéterminé. Elle est par essence un environnement intangible et immatériel, ce qui ne le rend pas pour autant moins réel ou moins primordial (Béranger, 2015). Cet environnement informationnel est composé par l'ensemble des processus d'information, des services et des entités, y compris les agents d'information, leurs interactions, leurs propriétés, et leurs relations mutuelles.

A cette infosphère est relié l'ensemble des logiciels et autres outils technologiques gérés par ce responsable (Carley, 2000) ainsi que l'ensemble de leurs utilisateurs légitimes. Cette infosphère se compose d'un ensemble de sujets et d'objets qui gravitent autour des dispositifs informatiques. Elle comprend également toutes les données appartenant à un individu (ou personne morale) et toutes les données qui le concernent, mais qui sont en dehors de son centre de gravité (sécurité, politique, etc.). En résumé, l'infosphère d'un établissement de santé regroupe tous les objets communicants de la structure, de toutes les données et des connexions associées au SI (Béranger, 2015).

16. Ce mot est inventé par Dan Simmons (1989) pour désigner un environnement informationnel.

Pour cela, ce philosophe italien se base sur la théorie de l'information et plus particulièrement sur le concept d'entropie de l'information introduit par Shannon¹⁷ au milieu du XX^e siècle. Pour l'auteur, l'entropie d'information mesure, par analogie à l'entropie thermodynamique, le degré de désordre d'un système, ou plus exactement la connaissance que nous en avons. En effet, si nous connaissons parfaitement une chose, nous savons y percevoir et localiser tous les détails, nous pouvons en énumérer la succession dans l'ordre ; elle nous apparaît donc ordonnée. Il existe donc une relation directe entre l'organisation d'un système et la connaissance que nous en avons. De la sorte, plus l'entropie est faible, autrement dit plus un système est ordonné, plus il est connu. Et vice-versa, plus l'entropie augmente, moins on le considère comme ordonné et moins il est connu. A partir de ce postulat, Luciano Floridi développe sa réflexion autour de « l'éthique de l'information » en prenant comme critère l'entropie globale d'information qu'il applique à son concept d'« infosphère » : l'environnement où se développe l'information. Selon lui, « un comportement éthique diminuerait l'entropie puisqu'il rendrait l'information plus signifiante, tandis qu'un accroissement de l'entropie serait néfaste à tous ». Ainsi, une action dite « juste » ou considérée comme éthique conduirait à diminuer l'entropie globale, et à accroître la connaissance que nous en avons. *A contrario*, une mauvaise information ou des données déjà connues augmenteraient l'entropie en désorganisant l'infosphère.

Ainsi, l'éthique télématique désigne le bien du mal, ce qui doit être réalisé et les devoirs de l'agent moral, à partir de quatre lois fondamentales : on ne doit pas causer d'entropie dans l'infosphère, on doit prévenir la production d'entropie dans l'infosphère, l'entropie doit être exclue de l'infosphère, et la bonne condition de l'information doit être favorisée par l'extension de l'information (quantité), son perfectionnement (qualité) et son élargissement (variété) dans l'infosphère (Floridi, 1998).

Dans ce contexte, les enjeux éthiques¹⁸ qu'engendre cet univers numérique sont mieux appréhendés et compris si nous les associons à des événements ou des faits concrets de l'environnement du réel (Floridi, 2002). Par exemple, la notion de confidentialité peut être rattachée à la présence ou non, dans l'outil de traitement des Big Data, de paramètres permettant de masquer l'identification du patient pour

17. Claude Shannon, ingénieur à la Compagnie des Téléphones Bell, détermine l'information comme grandeur observable et mesurable (1948) ; celle-ci devient la poutre maîtresse de la théorie de la communication qu'il élabore avec Weaver. Ce concept d'information a été l'objet d'une théorie, appelée « théorie de l'information ». C'était une théorie mathématique appliquée aux techniques de la télécommunication. Cette théorie mathématique, née de préoccupations techniques de la télécommunication, reste à ce jour la base du concept dit scientifique d'information.

18. Tels que la confidentialité, le secret médical, la protection des données médicales, le respect de la vie privée, l'accessibilité à l'information médicale, le partage des responsabilités, le respect et le maintien de l'autonomie du patient, etc.

laquelle les informations médicales sont reliées. Concernant l'accessibilité à l'information médicale, celle-ci peut être associée par l'existence au sein du SI d'une plateforme dite de partage et d'échange dans laquelle l'utilisateur peut avoir accès selon son statut, son profil, et son niveau d'habilitation.

En complément, d'après Fessler et Grémy (2001), le concept et le mot d'« info-éthique » peuvent être examinés selon deux niveaux :

- les relations humaines existantes dans tous les Big Data en santé ;
- les structures de santé ont une dimension supplémentaire vis-à-vis de la personne : elles visent les changements des êtres humains.

Comme nous venons de le voir, l'information devient le principal objet de l'action morale. Introduire de l'éthique dans le numérique est un acte non naturel du fait que les TIC seraient dépourvus de toute valeur sociale et humaine. Cette idée provient d'une réflexion commune qui considère que toute technologie est éthiquement neutre, car seul l'être humain peut apporter du sens à ses actions. Pourtant, nous constatons que les Big Data diffusent également des valeurs dans la mesure où ils impactent et conditionnent la manière dont leurs utilisateurs se comportent. Par conséquent, aucune donnée numérique n'est jamais neutre (Fischer, 2014). C'est pourquoi il ne faut pas réduire l'éthique du numérique à l'expression de valeurs extrinsèques de bons usages de la technologie, mais également aux valeurs intrinsèques à celle-ci. Enfin, avec l'avènement du numérique et des « données massives », c'est toute une éthique qui est à inventer parce que les NTIC dessinent un nouveau paradigme relationnel et sociologique (Doueïhi, 2013). Nous n'avons pas la prétention d'inventer une nouvelle éthique mais plutôt de repenser et réinventer l'éthique existante afin de la faire évoluer vers ce que nous appelons l'« éthique algorithmique » appliquée exclusivement au numérique. Cette nouvelle approche a pour finalité d'intégrer des valeurs et principes éthiques sur la conception, la mise en œuvre, à l'usage des Big Data notamment dans le domaine de la médecine.

Les questionnements autour des Big Data en santé

S'interroger *via* une approche éthique sur la collecte, le stockage, l'utilisation et la mise à disposition des données de santé à caractère personnel, n'a rien de contraignant et peut même ressembler rapidement à de la valeur ajoutée, du sens, voire même faire exploser le retour sur investissement. Chaque technologie performe des pratiques, produit des valeurs, des comportements et par conséquent immisce de nouvelles normes sociales. Les Big Data obligent à de nouvelles considérations sur nos valeurs et le comportement de nos actions, du fait qu'ils donnent à un plus grand nombre de personnes, plus de moyens pour communiquer et interagir entre eux.

Dans ce cas, nous pouvons nous demander si ces « données massives » posent des problèmes éthiques propres. La vie privée, par exemple, a-t-elle la même valeur sur Internet que dans la vie quotidienne ?

Dans l'exercice quotidien de leurs activités, les personnes en relation avec les *data* en santé sont soumises à des interrogations multiples, successives, complexes où se confrontent en permanence le respect des valeurs humaines considérées comme universelles et les limites contraignantes des décisions concrètes à prendre. De nombreuses questions relatives au volume conséquent de données, à leur collecte, leur utilisation, leur stockage notamment par des outils informatiques vont en effet se poser dans des secteurs très variés.

Dans ces conditions, cet ouvrage vise à refléter des aspects éthico-technologiques sur la conception, la gestion, la maîtrise et l'usage des Big Data en santé. Nous proposons d'articuler nos réflexions autour d'une série de questions.

– Quels sont les changements que le Big Data et l'analyse de données apporteront aux soins de santé ? Comment les « données massives » impactent-elles la pratique médicale ? Quels changements peut-on attendre pour les services aux patients et de la santé ?

– Où trouver des données sur la santé et la façon de l'utiliser pour améliorer les soins ?

– Doit-on laisser guider les jugements et les décisions médicales uniquement par le traitement des Big Data en santé ?

– Doit-on informer les personnes des risques de maladies qu'elles courent du fait de leur patrimoine génétique ?

– Toutes ces données accumulées ont une valeur, comment et où pouvons-nous trouver cette valeur pour améliorer les soins ?

– Comment combiner la nécessité du partage et la vie privée, lors de stockage des données dans un Cloud Computing ?

– Comment conjuguer intérêt collectif et protection des personnes ?

– Faut-il adapter l'organisation du système de soins à ces évolutions technologiques ?

– Comment chaque pays sécurise, d'un point de vue juridique, ses données de santé ?

– Dans quelle mesure les pratiques professionnelles autour du numérique en santé doivent-elles répondre à un critère d'objectivité, de neutralité et/ou de rationalité ?

– Jusqu'où est-il possible de valoriser l'approche et l'engagement éthique dans l'usage des Big Data en santé ? Quelle est la valeur de l'éthique associée aux données médicales ?

– Comment transmettre une éthique aux professionnels de l'information et de la communication et/ou aux professionnels de la santé ? Comment en arrive-t-on à une prise de conscience en matière d'éthique ? A l'accroissement d'une exigence ?

– Comment penser l'éthique en NTIC en santé ? L'éthique est-elle inhérente à l'émergence des « méga données » ? Est-elle un dispositif de signification spécifique ? Comment circulent les valeurs éthiques ? Comment articuler la transformation numérique de la société à des principes éthiques ? Comment concilier les Big Data et l'éthique au sein d'une entreprise ?

– Existe-t-il un référentiel reconnu sur les bonnes pratiques éthiques de ces énormes volumes de données ?

– Comment classifier les données que nous générons ?

– Les analyses de l'ADN et du génome humain à grande échelle aideront-elles à soigner les maladies ? Ou bien cela aboutira-t-il à une nouvelle vague d'inégalités et d'injustices médicales ?

– L'étude des Big Data rendra-t-elle l'accès des usagers de santé à l'information plus efficace et effective ?

– Comment faire pour donner des limites à de l'information, des données qui peuvent être parfaitement insignifiantes prises individuellement ?

– Comment faire en sorte que la nouvelle technologie de traitement du Big Data utilise les données et les technologies existantes ?

– Comment permettre aux nouvelles formes d'analytique et aux applications d'exploiter à la fois les nouvelles et les anciennes données ?

– Ces nouveaux acteurs vont-ils s'interfacer avec les forces déjà en présence sur les segments du suivi et du traitement tels que les laboratoires pharmaceutiques et les équipementiers médicaux ? S'agira-t-il plutôt de concurrence ou de partenariats ? Quelles responsabilités auront ces acteurs dans l'utilisation de ces données ?

– La responsabilisation semble une réponse un peu courte aux croisements et aux traitements sans limites ! Que signifie réguler les acteurs, quand demain, tout à chacun sera capable de lancer ses propres fouilles et extractions de données en ligne ? Quand les systèmes seront capables de récupérer les données de n'importe qui en ligne pour en construire un profil ?

– Aura-t-on toujours un modèle financier solidaire pour payer une médecine 4.0¹⁹ basée sur une individualisation des risques ?

– Qui régulera les sociétés qui vont fouiller le Web pour affiner nos profils de recrutement ? Quelles instances doivent les gérer, en autoriser l'accès, assurer une efficacité et une sécurité optimales tout en garantissant le caractère démocratique et transparent de cette gestion des Big Data ?

– Comment arriver à créer de la richesse par le Big Data, tout en créant une économie de la confiance qui est indispensable pour que les nouveaux usages puissent se développer ?

Autant de questions qui méritent d'être examinées et débattues pour que ces enjeux déterminants autour des Big Data dans le domaine médical fassent l'objet de choix éclairés. C'est pourquoi il est indispensable de favoriser une approche technico-éthique de ces questions afin de nourrir une réflexion riche, ouverte et féconde.

Les objectifs et les apports du livre

Le développement du numérique et son omniprésence dans notre société moderne créent un besoin grandissant de poser des repères éthiques. Dans ce contexte, on peut s'interroger sur la spécificité des Big Data à créer de nouvelles problématiques éthiques ou encore à renforcer certains dilemmes moraux classiques. Le présent ouvrage souhaite fournir des outils à la réflexion éthique sur l'élaboration, la mise en place et l'usage des « méga données » en santé.

L'objectif de ce livre est de donner aux acteurs concernés par ces Big Data, les premières clés de lecture qui permettront aux lecteurs d'acquérir une approche éthique des NTIC. De définir un cadre moral et humain qui gère l'intérêt général et les droits individuels des données de santé à caractère personnel, et d'esquisser un nouvel espace de confidentialité de nos données, en se concentrant plus sur la responsabilisation de l'usage des données, que sur le consentement individuel, lu et éclairé, lors de la collecte de ces dernières (Fischer, 2014).

Nous souhaitons apporter un certain équilibre et une harmonie entre les intentions humaines et la finalité des outils technologiques associés aux Big Data. L'enjeu est

19. Le concept d'Industrie 4.0 a été énoncé pour la première fois lors du salon de la technologie industrielle à Hanovre, en 2011. Il correspond à une nouvelle manière d'organiser les moyens de production à partir de bases technologiques telles que l'Internet des objets, les technologies relatives aux Big Data, et les systèmes cyber-physiques. L'objectif est de mettre en place des entreprises dites « intelligentes » (*smart factories*) capables d'une plus grande adaptabilité dans la production et d'un usage plus efficace des ressources dont elles disposent.

de renforcer le sens de nos actions afin de permettre au lecteur de prendre conscience, voire de valider, les premières orientations d'une intégration saine et maîtrisée des « données massives » à l'écosystème de la médecine. En résumé, les attentes concernant notre ouvrage sont multiples, en permettant :

- d'identifier et caractériser les enjeux associés à ces gigantesques volumes de données ;

- d'obtenir un éclairage précis sur le cycle de vie des Big Data afin comprendre les attentes éthico-techniques des acteurs directement ou indirectement associés à ces données ;

- de comprendre les modèles d'organisation, les acteurs, les méthodes et les démarches liées aux données ;

- d'élaborer une modélisation et une évaluation éthique d'analyse des Big Data destinées à conduire des actions concrètes et sélectives de bonnes pratiques et de réduction des risques ;

- d'introduire une méthodologie de guidage technico-éthique qui permet l'alignement de manière cohérente des actions technologiques avec les valeurs morales *via* la mise en relation des ontologies des risques et des objectifs éthiques associés à l'exploitation des Big Data ;

- d'aider les organisations à élaborer un cadre pour des discussions et des réflexions éthiques explicites afin de trouver un juste équilibre entre promesse de l'innovation technologique utile et risque de préjudice dans le traitement de la donnée ;

- de lister des règles de conduite normatives pour un traitement éthique des Big Data ;

- de construire une « éthique algorithmique » sur les recommandations et les évolutions à apporter concernant les données de santé numérique à caractère personnel ;

- de sensibiliser l'ensemble des acteurs et de promouvoir une culture orientée vers la sensibilisation, l'implication, l'appropriation et la responsabilisation autour des « données massives » ;

- d'aider les entreprises à développer une aptitude à effectuer un questionnement et une analyse éthique explicite dans ce nouveau contexte des Big Data.

L'intérêt final est donc d'aider le lecteur à avoir une connaissance précise sur les enjeux éthiques qu'un tel sujet suscite. Ceci étant à notre sens, la condition essentielle d'une approche humaine qui irait au-delà des considérations financières et matérielles d'aujourd'hui.

En mettant en lumière une trame de fond basée sur une réflexion innovante éthico-technique, notre ouvrage souhaite apporter les fondements d'un changement sensible des mentalités ainsi qu'une transformation environnementale à « visage humain » des Big Data et de leurs utilisations dans le domaine médical. L'objectif est de trouver une certaine cohérence et un certain sens dans ce paysage en perpétuelle évolution technologique afin d'apporter la meilleure prise en charge possible pour l'utilisateur de santé (Béranger, 2015).