

analyser afin d'acquérir des connaissances sur eux-mêmes, et d'opérer des changements dans leur vie quotidienne.

Dans ce contexte, l'avènement du *Quantified self* rend possible la surveillance au quotidien des indicateurs de santé personnelle : santé connectée, hygiène de vie et entretien du corps, surveillance de l'évolution d'une pathologie, etc. Mais les *big data* bouleversent aussi les pratiques de recherche et la mise au point de nouveaux traitements.

En se servant des corrélations, les outils numériques ont permis à l'utilisateur de santé de réguler son poids en fonction de sa marche, de son sommeil, de son stress, et surtout de réduire un traitement médicamenteux contraignant. En produisant et en se réappropriant ses données, il s'est constitué un système informationnel personnel lui permettant de requalifier ses comportements. Sur la durée, cette expérience lui a permis de passer d'une pratique de santé « curative » à une pratique de santé « préventive ».

En outre, après la numérisation de l'ensemble des savoirs humains accumulés depuis des siècles, les services numériques, les réseaux sociaux, les *devices* mobiles, les objets et capteurs connectés produisent sans cesse de nouvelles données dont le stockage, la diffusion, la gestion et l'usage rentrent de plus en plus dans le quotidien des personnes.

Ces « métadonnées » ou *big data* représentent désormais le carburant de l'économie et la connaissance du début du XXI^e siècle. Le volume de données médicales créé chaque jour par notre civilisation explose de manière exponentielle. Lorsque nous nous connectons à Internet (réseaux sociaux ou différents sites Internet), notre navigation laisse de nombreux indices sur notre vie personnelle telle que nos besoins, nos comportements, nos pathologies, etc. Ainsi, ces « métadonnées » représentent un volume d'informations insoupçonnées que nous laissons derrière nous dès que nous nous connectons à un dispositif numérique. Ces quantités massives de données numérisées sont stockées et archivées dans des entrepôts de données (ou *data-warehouses*).

Le volume de données numériques augmente de manière exponentielle : 90 % de l'ensemble des données aujourd'hui disponibles ont été créées ces deux dernières années (Brasseur 2013). En 2013, l'humanité a stocké plus de 2 000 milliards de gigaoctets de données numériques nouvelles. On estime que le volume de données stockées dans le monde double tous les quatre ans (Blondel 2013). Les données recueillies sont souvent bruitées et imprécises et doivent être traitées pour en extraire de l'information utile. Pour cela, on utilise des algorithmes simples de traitement statistique, inductive, d'inférer des profils (patterns ou modèles de comportements). Cette information sur l'information nourrit une connaissance à très haute valeur des comportements individuels. Grâce aux *big data*, tout un chacun est en mesure de

comparer ses propres données avec celles d'autres utilisateurs, afin de se situer par rapport au reste de la population. Ces métadonnées contiennent la promesse de personnaliser nos traitements médicaux.

Dès lors, les volumes de données enregistrées, susceptibles d'être croisées et exploitées, sous couvert d'anonymisation bien sûr, pour en extraire du sens pour un professionnel de santé, sont d'ores et déjà immenses et amenés à se développer quasi exponentiellement. Elles sont issues des instituts de recherche, centres épidémiologiques, laboratoires pharmaceutiques, centres d'imagerie, comptes rendus hospitaliers, compagnies d'assurance et des dossiers clients, etc.

Depuis quelques années, les géants du numérique (Amazon, Facebook, Google, Apple) utilisent ces données pour améliorer la qualité des services proposés mais également pour les revendre à des sociétés commerciales. Ils font de ces données des biens privés qu'ils revendent. Aujourd'hui, ce ciblage marketing est réalisé uniquement dans une optique commerciale. Dans ces conditions, les entreprises et les institutions s'entourent désormais de *data scientist* capables de traiter à grande échelle le bruyant tourbillon des *big data*. L'objectif de ces spécialistes du marketing numérique est de découvrir dans ce flot d'informations le comportement des internautes et d'anticiper leurs envies afin de proposer des contenus individualisés et adaptés à leurs besoins spécifiques.

La diffusion des TIC et l'arrivée des *big data* dans les secteurs de la santé, que ce soit auprès des industries pharmaceutiques, des professionnels de santé ou des institutions de soin poseront, en 2025, « autant de défis éthiques que de défis scientifiques et industriels » (Aubert 2009). La quantité de savoir brut ne cesse de croître de manière exponentielle, mais la connaissance qui découlerait de son traitement avance nettement moins vite. L'étude de ces métadonnées au service de la santé amène des problèmes et des réflexions spécifiques d'ordre éthique et technique tels que :

- quelle valeur scientifique peut-on donner à ces données médicales traitées de manière informatique ?
- quelle sera la place du jugement du praticien dans ce nouveau cadre technologique ?
- le développement des arbres de décision personnalisés ne va-t-il pas mettre en péril le respect de la dimension humaine au sein de la relation médecin-patient ?
- l'interopérabilité des données et la sécurité des données médicales hébergées sur les différents territoires du monde ;
- les données médicales publiées dans les revues scientifiques ne reflètent pas forcément la réalité de la médecine, du fait qu'elles correspondent à des patients correspondant à des critères spécifiques de l'étude ;

- repenser les infrastructures de stockage ;
- la formation des professionnels de la santé à ces NTIC ;
- le renouvellement des outils de lecture de traduction et d'exploitation de ces bases constitue l'un des enjeux majeurs de la recherche dans les prochaines années.

Dès lors, on s'aperçoit que les dérives possibles de l'usage secondaire de nos données personnelles *via* des algorithmes prédictifs portent autant sur la notion de protection de la vie privée et de la confidentialité que sur le risque que nous soyons amenés à analyser les personnes non pas sur leur comportement réel, mais sur leur propension à avoir le comportement que les données leur donnent. On constate que cette connexion permanente génère une empreinte digitale autant par les données que nous diffusons volontairement *via* les réseaux sociaux que par notre propre comportement.

Ainsi, l'activité que nous générons en naviguant en permanence sur Internet constitue un flux d'informations qui envahit notre quotidien et perturbe les frontières entre sphères publique et privée, nous entrons dans l'univers de la « connexion permanente ». L'avènement des *big data* déstabilise notre conception de la vie privée et modifie le caractère du risque lié à la surveillance. Les *big data* facilitent tellement la ré-identification et la réutilisation des données secondaires, que toutes les stratégies qui assuraient la protection de la vie privée – à savoir l'anonymisation, le consentement préalable et l'*opt-out* – semblent inefficaces. Désormais, il est indispensable de réinventer la protection de la vie privée et de faire évoluer le cadre normatif et législatif qui y est associé. Le processus de « notification et de consentement » pour garantir la confidentialité apparaît dépassé. En effet, on ne peut imaginer Google contacter tous ses utilisateurs afin de leur demander leur permission pour utiliser leurs requêtes afin de prédire la grippe.

Par ailleurs, le *Quantified self* et les applications santé sur les smartphones ne cessent de se développer de manière exponentielle. De tels usages posent la question de l'exploitation des données confidentielles que l'on peut en faire.

De plus, les *big data* ne doivent pas nous enfermer dans un monde social prédéfini par des probabilités mettant en péril la potentialité de chaque individu. En effet, notre capacité humaine à construire notre avenir pourrait s'en trouver altérée par une technologie algorithmique qui analyserait toutes nos données personnelles (milieu social et scolaire, antécédents médicaux, cercle de relations, etc.). Le risque serait de juger les personnes uniquement par des données contextuelles et non plus par leurs actions.

Enfin, les *big data* risquent d'amplifier le phénomène de catégorisation, de classement, de discrimination, et leurs dérives au sein de la société, car les données seront de plus en plus associées à la personne.

Dès lors, avec l'émergence du *big data* alimentée par une multitude de technologies de mesure et de récupération d'informations diverses, le monde du lisible se transforme progressivement en monde du prévisible, où la maxime d'Auguste Comte : « Savoir pour prévoir, prévoir pour pouvoir », prend tout son sens et son importance.

Dès lors, ces *big data* ont le potentiel de modifier la manière dont les entreprises exploitent leurs données. Avec l'accroissement du volume, de la diversité et de la vitesse des données, les entreprises doivent adapter leurs pratiques de gestion de données au fur et à mesure qu'ils récupèrent et analysent toutes ces informations. Cette mise en profit du *big data* débute par :

- l'utilisation de logiciels qui simplifient l'intégration et la manipulation de ces métadonnées ;
- besoin grandissant d'appliquer les procédures de qualité de données ;
- ne pas appliquer des principes de gestion de projet et de gouvernance dès le début.

Dans le domaine de la santé, ces *big data* peuvent se révéler très prometteuses notamment dans les approches technologiques destinées à intégrer, traiter, analyser et mettre en perspective les grands volumes de données médicales. Cette démarche permet de mieux considérer les données. En effet, auparavant, par limitation technique, conceptuelle et humaine, la plupart des études statistiques se réalisaient sur la base d'échantillons, censés être représentatifs. Les interprétations que l'on en sortait n'étaient certes pas erronées, mais parfois approximatives. Dès lors, avec le traitement de grands volumes de données, nous observons une sensible amélioration en termes de précision d'étude par rapport à une analyse sur des échantillons de données.

Par ailleurs, nous connaissons aujourd'hui les principaux facteurs de risque de maladie existants. Nous savons les corréler à des comportements. Dans un proche avenir, nous pourrions extrapoler des données de comportements et des prédispositions *via* les réseaux sociaux, *smartphone*, e-mails, etc., et étudier la probabilité d'une personne d'avoir telle ou telle pathologie et à quel moment. On constate que ce processus est déjà en application dans le monde de la finance où des dispositifs informatiques centralisent les informations de l'ensemble des échanges boursiers de l'histoire, les croisent entre elles, et élaborent des modèles d'interprétations ciblées afin d'établir des prévisions affinées pour les traders ou les assureurs. Dans ces conditions, on peut imaginer une analyse scientifique étudiant à très grande échelle le comportement et les habitudes sur Internet des personnes atteintes de maladies chroniques comme le diabète afin de cibler et repérer les corrélations possibles entre les facteurs de risque et les habitudes de navigation. Ces métadonnées permettront à grande échelle la diffusion non commerciale de messages de santé publique afin de prévenir les risques de maladies chroniques dans le monde.