

# Vers une gouvernance éthique ( $\Psi$ , $G$ , $\Phi$ ) de l'exploitation des Big Data en santé

---

*La numérisation des données personnelles médicales peut-elle se passer de la main de l'homme ? Une sélection des données complexes et une hiérarchisation est nécessaire. La mise à disposition des données est contradictoire avec leur confidentialité. La sécurisation semble utopique. L'éthique est une réponse dans la recherche d'un équilibre permanent. Réflexions de Jérôme Béranger, chercheur en éthique du numérique.*



*Monsieur Jérôme Béranger, chercheur (PhD) et expert en éthique du numérique (SI, m-Health & Big Data) en santé - Keosys*

A l'heure de l'explosion des volumes, de l'avènement du *Big Data* et du *Cloud Computing*, la hiérarchisation et la sélection des données apparaît comme fondamentale dans les Systèmes d'information (SI). Cela nécessite d'adapter une stratégie de gestion des bases de données de production en déplaçant les données entre diverses catégories de stockage, afin d'optimiser les performances réalisées par l'utilisateur. Dès lors, un tel processus permet de soulager l'utilisateur du besoin de gérer l'emplacement des données. Ainsi, au-delà de la valeur informative, la communication numérisée a désormais pour enjeu de restaurer la confiance des utilisateurs, tout en optimisant ses fonctionnalités. Une valeur ajoutée éthique de la communication *via* un SI consiste à prendre ces données puis la transformer en

informations en lui donnant une autre dimension pour atténuer, modifier, amplifier la perception que sa réception peut entraîner. Ainsi, l'éthique réfléchit aux fondements des systèmes de règles et sur la résolution des conflits de normes. C'est la nature même de l'éthique que d'avoir des conflits notamment dans le domaine de la santé.



Dans un univers aussi changeant que celui des Nouvelles technologies de l'information et de la communication (NTIC), l'éthique impose une continuité. Il y aurait donc une tension entre les technologies, vecteur d'évolution, et l'éthique, en quête de pérennité. Or cette tension nous apparaît féconde non seulement pour l'Intelligence artificielle (IA), afin que les techniques produites ne contrarient pas nos valeurs et qu'elles répondent aux aspirations humaines, mais aussi pour l'éthique, susceptible de se renouveler grâce à l'IA et la puissance algorithmique. Pour prendre la mesure des apports de l'IA à l'éthique et inversement, soulignons d'abord que les philosophes distinguent aujourd'hui trois secteurs de l'éthique :

- **L'éthique descriptive ou appliquée** fondée sur les finalités, les moyens, les dispositifs, les voies et les procédures : Transparence Vs Protection de la vie privée / Conservation des données / Responsabilité (dimension téléologique) ;
- **L'éthique normative** basée sur les normes, les limites, les standards, les codes et les règles de conduite / Systèmes cohérents de justifications adoptées (dimension déontologique) ;
- **L'éthique réflexive ou Méta-éthique** centrée sur les principes et les valeurs morales / Vise à préciser le statut des énoncés moraux (dimension axiologique).

La seule application des codes et des règles conduites, aussi éthiques soient-elles, ne suffit pas à élaborer des SI réellement éthiques. Il faut non seulement révéler l'intention de l'action, mais également s'interroger sur les effets et les conséquences de l'action.

*La communication des savoirs dépasse les seules données objectives en préférant mettre en avant certaines données et informations plutôt que d'autres.*

De plus, si l'information regroupe des données objectives, avec le passage à la communication intégrant une hiérarchisation et une sélection de ces données, alors nous rentrons dans le secteur des connaissances subjectives. En d'autres termes, cela traduit la métamorphose des données et des informations de nature épistémologique vers des connaissances associées à l'anthropologie et au vécu expérimentiel de la personne. Ainsi, la communication des savoirs dépasse les seules données objectives en préférant mettre en avant certaines données et informations plutôt que d'autres. Elle ne peut survenir qu'au niveau de la diffusion de connaissances subjectives et cognitives. C'est pourquoi, il semble plus exact d'employer le terme de « connaissance » plutôt que d'« information » lorsque nous parlons de communication en particulier au niveau médical.

Ainsi, la hiérarchisation des données demande en prérequis d'établir une réflexion sur la valeur intrinsèque de ces dernières afin de pouvoir les évaluer :

- Pourquoi et comment évaluer ?
- Pourquoi faire, dans quel but et avec quels objectifs ?
- Comment évaluer la valeur d'une donnée et d'une information, selon quels critères ?
- Que doit-on évaluer ? (Simonnot, 2007)

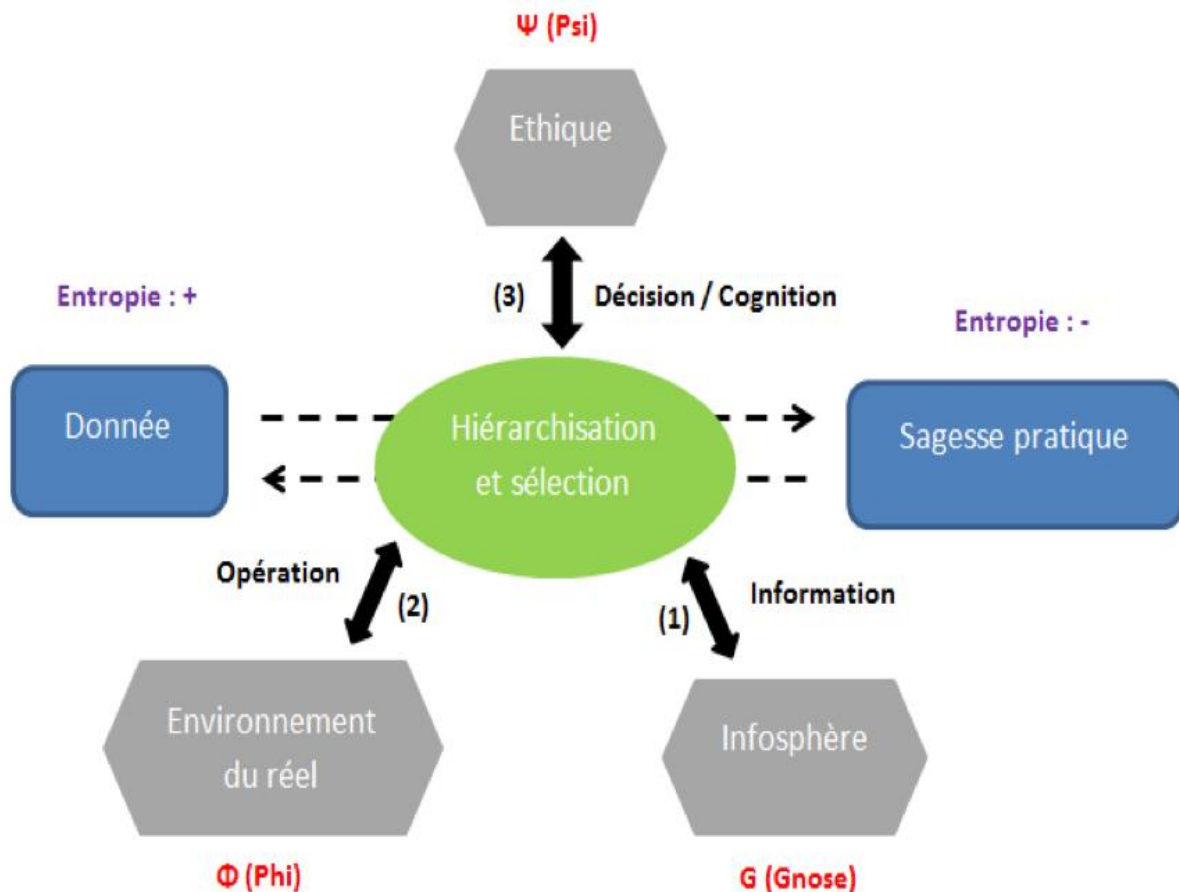
Par ailleurs, pour obtenir un équilibre pratique au sein du SI entre amélioration et surcharge des données transmises, la fonction de hiérarchisation et de sélection doit être optimisée, sur la base de deux variables :

- La fréquence de la réévaluation de l'attribution des données sur les différents niveaux ;
- Le volume de données à inclure dans l'unité minimale de stockage devant être gérée et déplacée.

Cependant, si cette réévaluation est réalisée de manière trop récurrente, la surcharge associée au déplacement des données dans un sens et dans l'autre risque d'annuler les performances obtenues grâce au déplacement des données sur les disques de stockage (SSD). L'autre difficulté concerne la complexité engendrée par un sous-système de hiérarchisation auprès du concepteur de SI. Le paramétrage de ces deux règles ainsi que d'autres associées à l'optimisation de la solution de hiérarchisation et de sélection, nécessite de dispenser au concepteur de SI (1) une formation pointue afin que le sous-système de hiérarchisation sélective propose les meilleures performances.

En outre, le processus informationnel part de la personne « utilisateur de l'information » acteur central du SI et décideur appartenant à une organisation. De ce fait, « il va devoir prendre en compte les intérêts de son organisation, intérêts définis par les dirigeants » (Lebraty, 2001). En prenant sa source à la fois dans l'environnement du réel ( $\Phi$  : Phi (2)), dans l'état donné et le monde sensible de l'infosphère (G : Gnose (3)) et dans la pensée éthique ( $\Psi$  : Psi (4)), le dispositif informationnel aboutit à la création d'une sagesse pratique (info-éthique) via la hiérarchisation et la sélection des données.

Après une phase d'observation, l'activité concrète se transforme en données, ces dernières après interprétation, évaluation, hiérarchisation et sélection deviennent des informations, qui, à leur tour, suite à un processus de cognition deviennent des connaissances nouvelles qui serviront pour l'apprentissage afin d'aboutir à une sagesse pratique. Le passage de la donnée à l'information puis à la connaissance, expliqué par une modélisation éthique systémique néo-platonicienne ( $\Psi$ ,  $G$ ,  $\Phi$ ), met en évidence qu'une hiérarchisation sélective des données contribue à faire diminuer l'entropie (degré de désordre) d'une connaissance afin que cette dernière soit utilisée de manière optimale. Le schéma ci-dessous tente d'éclairer cela :



Ainsi, la force intrinsèque de cette modélisation se situe dans son caractère à la fois interactif, multidimensionnel, actif tourné vers le sens, la connaissance et la téléologie d'un évènement. Elle se compose de trois dimensions distinctes qui s'articulent et interagissent entre elles :

- Dimension informative ( $G$ ) : Infosphère (5) ;
- Dimension environnementale ( $\Phi$ ) : Environnement du réel (6) ;
- Dimension éthique ( $\Psi$ ) : Info-éthique.

Notons que  $G$  et  $\Phi$  appartiennent à un ensemble objectif dit de « rationalité informationnelle », alors que  $\Psi$  provient d'un ensemble subjectif dit de « rationalité cognitive » (Fransman, 1994) (Roth, 2004).

Dans ces conditions, le dispositif de hiérarchisation puis de sélection des données initiales entraîne une amélioration de la valeur qualitative et entropique de la

connaissance au détriment d'une perte quantitative de la donnée et de l'information. Un système de hiérarchisation sélective automatique de données permet à une baie de stockage de migrer automatiquement des données sur la bonne classe de services en fonction des besoins de chaque utilisateur. Cette solution a donc un impact multiple sur les paramètres environnementaux qui encadrent un SI, que ce soit d'un point de vue structurel et technologique, stratégique et méthodologique, organisationnel et réglementaire, relationnel et culturel.

***Le prisme éthique permet de mieux appréhender et comprendre l'équilibre instable qui existe entre la disponibilité et la protection des données.***

Dès lors, notre approche éthique se fonde principalement sur les quatre principes éthiques de Beauchamp et Childress (2001), à savoir le principe de : bienfaisance (7), autonomie (8), non-malfaisance (9) et justice (10). Ces éléments sont conçus à la fois comme des outils dans la résolution des problèmes humains et des balises pour une intégration de la dimension éthique dans les pratiques. Dans ce contexte, l'étude de la solution de hiérarchisation sélective vue par ce prisme éthique permet de mieux appréhender et comprendre l'équilibre instable qui existe entre la disponibilité et la protection des données. Ce contre-balancement peut pencher d'un côté ou d'un autre en fonction du contexte donné.

- Une telle approche nous amène à nous poser toute une série de questions avant de mettre en application la sélection des données :
- Quels sont les objectifs, les buts, les enjeux et le sens de cette étape ?
- Que vais-je utiliser comme données ? Des données partielles ou totales ?
- Comment vais-je l'utiliser ? A quel endroit ? Auprès de quels utilisateurs ?
- Plus globalement, comment exploiter l'ensemble hétérogène de ces données médicales accumulées et stockées dans un SI ?
- Quelle sera sa pertinence par rapport à ma situation ?
- Cela ne va-t-il pas dénaturer la valeur informative initiale ?
- L'intégrité du message final sera-t-elle conservée ?
- Cela rentre-t-il dans le cadre d'un meilleur usage du SI, une communication médicale optimisée et d'une amélioration de la prise en charge des soins vis-à-vis du patient ?

***Un tel système impose des règles d'attribution et d'accès à l'information différente (et donc non égales) selon la nature de la personne.***

Ainsi, cette vision éthique démontre que la sélection des données complexes a une action positive pour le principe de :

- **Bienfaisance** : une diffusion appropriée de la connaissance envers l'utilisateur de SI constitue un bien-fondé et une légitimité d'action. La communication devient plus efficiente ;

- **Autonomie** : une information préalable (claire, précise, adaptée et compréhensible), garantir le consentement éclairé de la personne. Ce dernier agit librement en disposant d'une capacité de délibérer, de décider et d'agir. Il réalise un acte autonome : intentionnel, volontaire et indépendant avec compréhension et sans influence extérieure de contrôle ;
- **Non-malfaisance** : l'accès limité aux données selon le profil et la nature de l'utilisateur améliore donc la sécurité, la confidentialité et la protection de ces données.

Toutefois, ce dispositif sélectif des données a une répercussion négative sur le **principe de Justice**, du fait que l'information médicale transmise n'est pas la même suivant l'utilisateur de la base de données. Un tel système impose des règles d'attribution et d'accès à l'information différente (et donc non égales) selon la nature de la personne. Cela traduit d'une certaine manière une discrimination des individus entraînant une dissymétrie de connaissances, et donc qui remet en cause la transparence de l'information.



A la lumière de ces réflexions, il semble indispensable d'apporter au cœur des NTIC et notamment des *Big Data*, de la philosophie, de la sociologie et de l'éthique, qui sont en charge de produire les outils conceptuels aboutissant à la sagesse pratique notamment dans le secteur de la santé. Notre modélisation d'analyse éthique inductive passe nécessairement par cette confrontation et cet échange des genres pour lesquels nous partons de l'universel, de l'abstrait et du général pour aboutir vers la pratique, le concret et le particulier. Cette aventure épistémologique de la connaissance se réalise par l'étude et l'interprétation des données : c'est un équilibre où se contrebalancent la théorie et le terrain. Il faut donc casser cette dichotomie entre les deux. Pour cela, il faut s'interroger sur la nature intrinsèque de la donnée et de l'information.

*Il apparait difficile de trouver un équilibre entre la disponibilité des données médicales, leur confidentialité et leur protection.*

Notre modèle sert à nous rappeler que les nouvelles technologies et les processus encadrant l'information sont conçus pour être mis au service des personnes qui les utilisent et non pas l'inverse – ce qui semble souvent être le cas. C'est sur cette base, que nous cherchons à poser les jalons d'une éthique propre aux NTIC illustrée par les *Big data*. Cette modélisation systémique néo-platonicienne inductive constitue donc un outil de traduction du langage technique (Alpha (A)) vers un langage éthique (Oméga ( $\Omega$ )), et inversement. Il permet alors la conjonction entre l'« info-signal » de la technologie et de l'« info-signification » prônée par l'éthique.

Enfin, il apparaît difficile de trouver un équilibre entre la disponibilité des données médicales, leur confidentialité et leur protection. A notre sens, la technologie ne permet pas de répondre totalement à cette problématique. Nous devons également faire appel à la déontologie et aux comportements humains pour garantir la confidentialité et la protection des données personnelles. Cela passe nécessairement par une charte éthique entourant la conception, la mise en place et l'usage de la donnée personnelle de santé.

### **Notes de bas de page :**

(1) La personne en charge de la configuration et du fonctionnement du SI.

(2) Cette lettre grecque détermine le chiffre d'or mesurant 1,618. Depuis l'Antiquité, ce symbole représente la proportion harmonieuse divine. Cela constitue la base des structures dans l'architecture, l'art, l'anatomie, la musique et la littérature, c'est-à-dire les aspects concrets et pratiques de la Science et des arts.

(3) Le terme « Gnose » signifie connaissance en latin.

(4) Cette lettre grecque est souvent utilisée pour désigner la psychologie, la psychiatrie, la psychothérapie et la psychiatrie, globalement les sciences de la pensée humaine. Ce symbole signifie âme, psyché.

(5) L'infosphère amène à des actions telles que : l'actualisation (évaluation, mise à jour, enrichissement) et la valorisation (accès, diffusion, partage, exploitation et combinaison).

(6) L'environnement du réel est une dimension propice : au repérage (identification, localisation, estimation, caractérisation, cartographie et hiérarchie) et à la préservation (acquisition, formalisation et conservation).

(7) 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.

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

(9) 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* (« D'abord ne pas nuire »), dont la conséquence est de faire du bien aux patients et de les écarter du mal et de l'injustice.

(10) Elle a pour vocation de partager entre tous les patients les ressources disponibles (en temps, en argent, et en énergie). 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.

### **Bioethique.com vous conseille :**

- Beauchamp, T.L, Childress, J. (2001), *Principles of Biomedical Ethics*, Oxford University Press, 5e édition, New-York/Oxford.
- Fransman, Martin. 1994. "Information, Knowledge, Vision and Theories of the Firm", *Industrial and Corporate Change*, Volume 3, Issue 3, pp.1-45.
- Lebraty (2001), "Comprendre le concept d'information pour mieux appréhender les Technologies de l'information et de la communication", Colloque du CRIC, La communication d'entreprise - Regards croisés Sciences de Gestion Sciences de l'Information et de la Communication, Nice, 6-7 décembre 2001, p.4-6.
- Roth F., *Des mécanismes de gouvernance informationnels aux mécanismes de gouvernance cognitifs : une grille de lecture et quelques conséquences pour la communication financière des entreprises*. Congrès Gouvernance et comptabilité, HEC Montréal, Juin 2004, p.19.
- Simonnot B. (2007). *Documentaliste - Sciences de l'information*, Vol. 44, N° 3, p. 215.