

**LES ETATS GENERAUX DE LA BIOETHIQUE**  
**Espace de Réflexion Ethique de Nouvelle-Aquitaine site de Bordeaux**  
**Les Big Data**  
**Débat**

## Introduction

Les Big Data, megadonnées ou données massives, sont ces données numérisées dont le volume augmente au fur et à mesure de l'informatisation de notre monde, du développement des télécommunications, de l'addiction de beaucoup d'entre nous aux réseaux sociaux et de la croissance exponentielle du stock des objets connectés, comme des fichiers de toute nature. On évalue le nombre de données ainsi collectées en milliards de milliards d'octets et ce flux ne cesse d'augmenter. Même si leur utilité et leur pertinence sont variables, ces données représentent une richesse considérable pour ceux qui ont les moyens et le savoir-faire nécessaire à leur stockage, à leur interprétation et à leur utilisation. Des dizaines de milliers d'entreprises de toute taille s'y affairant à travers le monde mais les grands acteurs publics nationaux - notamment l'Assurance-Maladie ou les centres hospitalo-universitaires, s'efforcent aussi, et dans un état d'esprit différent, de tirer le meilleur parti des données qu'ils accumulent au fil des années dans le cadre de leur activité.

## Méthode et public

Sur le thème des Big Data, un débat a eu sur le territoire de l'ex-Aquitaine. Il a réuni environ 60 personnes et a été construit en deux temps à peu près égaux en durée :

- une introduction par des experts de différentes disciplines (médecin de santé publique, bio informaticienne, philosophe, chercheur, juriste).
- Un débat.

Liste des experts :

- Cédric Brun, maître de conférences, directeur du département philosophie, équipe SPH – science, philosophie, humanité, UFR sciences biologiques, université Bordeaux et UFR humanités, université Bordeaux Montaigne.
- Sylviane Darquy, chargée de recherche, Bordeaux Population Health Center INSERM, université de Bordeaux.
- Moufid Hajjar, praticien hospitalier, médecin de santé publique, responsable de l'unité informatique et archivistique médicales, pôle santé publique, correspondant Informatique et Liberté, CHU de Bordeaux.
- Pauline Nicolas, doctorante attaché temporaire d'enseignement et de recherche, institut Léon Duguit, université de Bordeaux.
- Macha Nikolski, directeur de recherche CNRS, responsable du Bordeaux Bioinformatics Center, université de Bordeaux.
- Frantz Thiessard, maître de conférence des universités - praticien hospitalier, médecin de santé publique, unité informatique et archivistique médicales, pôle santé publique, CHU de Bordeaux, responsable de l'équipe de recherche en informatique appliquée à la santé, Bordeaux Population Health Center INSERM, université de Bordeaux.

Le débat a été animé par Bernard Broustet, journaliste.

Le public était composé de professionnels de santé et de personnes de la société civile (grand public) s'étant librement manifestés.

Dans ce compte-rendu, seules les questions formulées lors du débat de la salle ont été rapportées et pas les interventions, commentaires ou réponses des experts invités à la tribune.

## Thèmes évoqués

### → **Le risque d'un dispositif trop protecteur des données personnelles qui entraverait la recherche.**

Un membre d'une association de patients en cancérologie et hématologie souhaite apporter le regard du patient sur ce sujet. Il craint qu'un dispositif par trop protecteur des données personnelles, n'ait un impact négatif sur la recherche. Selon lui, les laboratoires devraient pouvoir accéder sans trop de difficultés aux données pour faire de la recherche et développer des traitements. Ainsi, l'absence de transparence des données de santé lui semble pénalisante pour le patient. Il porte la même analyse sur le manque de reporting informatique qui devrait pourtant aider à mettre en évidence plus rapidement les effets secondaires des médicaments (problème de pharmacovigilance et de matériovigilance).

### → **Quelle utilisation des données, et pourquoi ?**

Une scientifique (machine learning) et diabétique se demande comment poser une limite entre les données nécessaires voire vitales et celles qui sont superflues, en particulier concernant les données collectées par les objets connectés, et comment faire la balance entre ce qui est bon pour le patient et le coût financier ou éthique. Cette intervention lance un débat sur la distinction entre la technique elle-même et l'usage que l'on en fait, avec des points de vues contradictoires (technique et usage intimement liés ou indépendants ?).

### → **Le défaut d'information du patient et sa liberté de choix.**

La présidente de l'union départementale des donneurs de sang pense que plusieurs dysfonctionnements sont manifestes lors de dons de sang : le donneur n'est pas assez informé de ce à quoi va servir son don. Il dispose en fait d'une liberté de choix qu'il n'utilise pas, faute d'information : ainsi il peut autoriser ou pas que le don serve à la recherche, mais il l'ignore.

→ **Sur les questions des outils d'aide à la décision**, un médecin pense qu'il faut se garder de laisser la possibilité aux « machines » de produire un raisonnement complexe. Il considère que les données et le raisonnement simple peuvent être laissés aux machines et que cela peut être bénéfique en libérant du temps pour les médecins (ou autres professionnels dans d'autres disciplines). Elle ne peut cependant garantir une fiabilité totale et il ne faut pas exclure des erreurs d'interprétation ou de prédiction, dont on pourrait ne pas se rendre compte. En revanche, il estime que le raisonnement complexe doit être l'apanage des professionnels de santé, qui doivent conserver toute leur autonomie. La société doit selon lui poser des limites et des gardes fous même si l'Intelligence Artificielle forte (qui raisonne) n'est pas pour demain. Un autre intervenant considère quant à lui que la machine sera plus fiable dans bon nombre de circonstances.

→ Un professeur de génétique pointe le **problème des tests génétiques disponibles sur Internet**, et s'inquiète du fait que les personnes peuvent les utiliser facilement, mais sans aucun accompagnement au moment où ceux-ci reçoivent les résultats. Le sujet de la découverte fortuite de maladie génétique est notamment évoqué. Le problème de leur coût, de la confidentialité et du stockage de ces données est également évoqué. Il indique que cette inquiétude sera exprimée lors de l'audition prévue de sa société savante par le CCNE.

→ **Les craintes d'utilisation frauduleuse ou abusive des données de santé ont également été exprimées :**

- Un membre d'une association de patients se demande si cela vaut la peine de dépenser tant dans la protection des données de santé quand en réalité, les assureurs connaissent l'état de santé de leurs clients et que cela représente déjà une perte de chance pour ces personnes.
- Le représentant d'une autre association de patients se demande si un hacker qui récupérerait des données ne pourrait pas les vendre à des assureurs, des banques etc...
- Le risque de l'élaboration d'un profilage a également été mis en avant, avec le passage des données entre plusieurs mains (problème de la multitude des données et de leur utilisation).
- Enfin, un médecin et philosophe fait part de son inquiétude quant à l'utilisation des Big Data : risque de collusion entre les grosses entreprises détenant les données personnelles et certains états dans lesquels les législations sont moins protectrices des libertés individuelles. (ex. Chine).

## Propositions

Il faudrait reconnaître le principe que les données personnelles appartiennent à la personne qu'elles concernent, sous réserve de limitations apportées par la loi (comme le droit de la propriété). Cela serait sans doute plus protecteur et cela devrait être reconnu comme un principe constitutionnel.

## Synthèse

Les discussions ont majoritairement porté sur les données personnelles et les données de santé, leur utilisation, et les dispositifs notamment juridiques à même de les protéger. En réalité, le problème de la massification des données a peu été abordé, ce qui ne semble pas surprenant car beaucoup de problèmes soulevés par les Big Data ne sont pas spécifiques à ce type de données et concernent les données sensibles en général.

Lors de ce débat, il a pu être constaté que le public présent méconnaissait les droits fondamentaux du patient, et tout le corpus législatif français concernant la protection des données de santé (valeur du consentement du patient et possibilité de le retirer à tout moment ; ou principe de minimisation de la collecte de données *i.e.* seules les données nécessaires à la réalisation du service peuvent être collectées). Cependant cela semble compréhensible car les obligations liées à la collecte et l'utilisation des données personnelles sont encore trop rarement présentées aux patients (ex. : faire un retour d'information aux personnes ayant autorisé l'utilisation de leurs données pour de la recherche ; informer si les données collectées par une société vont être utilisées par une autre...). Les individus ont des droits quant à l'utilisation de leurs données et il est important qu'ils les connaissent pour les exercer.

La médecine est passée d'une logique d'explication à une logique de prédiction. Le développement du machine learning correspond à cette évolution et la recherche de l'explication des phénomènes biologiques et physiologiques n'est plus centrale (problème de la boîte noire des algorithmes). Dans ce contexte, des fantasmes existent car l'Intelligence Artificielle est médiatisée mais les professionnels du domaine ont eu l'occasion lors de ce débat d'expliquer que beaucoup de projections ne sont en réalité pas possibles.