LES CAHIERS DE L'ESPACE ÉTHIQUE
septembre 2015
02

Big data et pratiques biomédicales

— Implications éthiques et sociétales dans la recherche, les traitements et le soin

→ ESPACE ÉTHIQUE
RÉGION ILE-DE-FRANCE

DISTALZ
LABORATOIRE D'EXCELLENCE

Big data et pratiques biomédicales
— Implications éthiques et sociétales dans la recherche, les traitements et le soin

Réflexions, concertations et propositions tirées du workshop organisé le 16 avril 2015 par l'Espace de réflexion éthique de la région Ile-de-France et le Laboratoire d'excellence DISTALZ, avec la Fédération hospitalière de France

Sous la direction de **Emmanuel Hirsch, Léo Coutellec, Paul-Loup Weil-Dubuc**

Big data et pratiques biomédicales

— Implications éthiques et sociétales dans la recherche, les traitements et le soin

— Le Labex DISTALZ

Labellisé dans le cadre du Plan «Investissements d'Avenir», le laboratoire d'excellence DISTALZ fédère 7 équipes de recherche au plus haut niveau international.
Il offre, sous une identité commune, une masse critique compétitive de chercheurs et de capacités de recherche, de visibilité internationale, avec une attractivité accrue capable de rivaliser et de collaborer avec d'autres centres d'excellence sur les maladies neurodégénératives dans le monde.

Objectifs du Projet DISTALZ :

- Explorer les hypothèses actuelles et nouvelles de la physiopathologie de la maladie d'Alzheimer, notamment les voies métaboliques de la protéine amyloïde et de la protéine Tau, enrichies des découvertes génétiques récentes
- Tirer de ces connaissances des hypothèses biologiques nouvelles transférables en clinique au travers de biomarqueurs ou de cibles thérapeutiques potentielles
- Permettre, par une approche transdisciplinaire, la mise en place des bases biologiques, médicales, sociales et éthiques d'essais cliniques recrutant des individus et des patients identifiés comme présentant un risque maximum de maladie d'Alzheimer avant leur conversion vers la maladie d'Alzheimer

Mise en œuvre du Projet DISTALZ selon 4 axes :

1. Du gène aux hypothèses physiopathologiques : DISTALZ poursuivra la caractérisation de la composante génétique de la maladie d'Alzheimer et le décodage de l'héritabilité manquante, fournissant aux autres axes des hypothèses nouvelles

2. Des hypothèses physiopathologiques aux voies biologiques : DISTALZ étudiera l'impact de ces gènes et des voies ainsi identifiés dans des modèles expérimentaux, centrés sur les mécanismes de régulation des activités protéolytiques modulant la production/dégradation de l'Aβ, sur l'implication des nouveaux fragments de l'APP, sur les fonctions classiques et nouvelles de Tau ainsi que sur la propagation de l'agrégation des protéines

3. Des voies biologiques aux cibles concrètes : DISTALZ développera des tests génétiques et biologiques, tenant compte des interactions avec d'autres maladies neurodégénératives et cérébrovasculaires dans une logique de médecine personnalisée

4. Des cibles concrètes aux essais cliniques : DISTALZ accélérera le transfert de ces découvertes en clinique en facilitant l'accès à des patients caractérisés à un stade précoce de la maladie d'Alzheimer, anticipant les conséquences psychologiques, sociales et éthiques de ce diagnostic précoce

http://distalz.univ-lille2.fr

— L'Espace de réflexion éthique de la région Ile-de-France

Créé en 1995, l'Espace éthique de l'Assistance publique – Hôpitaux de Paris est le premier Espace éthique conçu et développé au sein d'une institution (repris en 2004 comme modèle de dispositif de réflexion éthique dans le cadre de la loi relative à la bioéthique). En 2013 il a été désigné Espace de réflexion éthique de la région Ile-de-France (ERE/IDF).

- En 2010, l'Espace éthique/AP-HP s'est vu confier le développement de l'Espace national de réflexion éthique sur la maladie d'Alzheimer (EREMA) dans le cadre du Plan Alzheimer 2008-2012.
- De 2010 à 2012, l'Espace éthique/AP-HP a fait partie des Centres collaborateurs OMS pour la bioéthique.
- Depuis 2010, son équipe de recherche développe la composante 'Éthique, science, santé et société' (ES3) de l'équipe d'accueil 1610 'Étude sur les sciences et les techniques' de l'université Paris Sud, cela dans la continuité du Département de recherche en éthique Paris Sud créé en septembre 2003.
- En 2012, l'Espace éthique avec son EA a été désigné, dans le cadre des investissements d'avenir, membre du laboratoire d'excellence DISTALZ (Développement de stratégies innovantes pour une approche transdisciplinaire de la maladie d'Alzheimer).
  Il est plus spécifiquement en charge d'une recherche portant sur les interventions et les diagnostics précoces, notamment de la maladie d'Alzheimer et des maladies associées.

L'Espace éthique se définit comme un lieu d'échange, d'enseignements universitaires, de formations, de recherches, d'évaluation et de propositions portant sur l'éthique hospitalière et du soin. Il assure également une fonction de ressource documentaire.

Ses missions :

- Observation et analyse des pratiques (recherche et expertise) des situations relevant au sein des hôpitaux de considérations d'ordre éthique
- Réponses adaptées aux sollicitations des professionnels de santé et d'associations intervenant dans le domaine médico-social : concertations, sensibilisation, formations, conseils, consultations
- Formations universitaires, séminaires interdisciplinaires et réflexions thématiques
- Développement et encadrement des recherches menées par des professionnels ou des étudiants intervenant dans le champ de l'éthique hospitalière et du soin ou du social, mais également de la bioéthique
- Synthèse et analyse de publications consultables dans un centre de ressources documentaires (matériel bibliographique, électronique, web, multimédia audio et vidéo, etc.). Le centre documentaire est installé à la Faculté de médecine de l'université Paris Sud ;
- Mise en réseau des références, des réflexions et des recherches, au moyen de sites Internet qui informent également sur les activités de l'Espace éthique et sa programmation mais également sur les initiatives nationales susceptibles d'être relayées :
  www.espace-ethique.org /www.espace-ethique-alzheimer.org
- Contribution à la concertation publique à travers l'organisation d'évènements et de rencontres thématiques
- Il procède à des publications qui restituent la diversité des réflexions et des recherches pour contribuer à l'expression, à la diffusion et au renforcement d'une culture de l'éthique hospitalière et du soin (Collection Espace éthique, éditions ères)

http://www.espace-ethique.org

Directeur de publication :
Emmanuel Hirsch

Rédaction :
Pierre-Emmanuel Brugeron
Léo Coutellec
Patrice Dubosc
Emmanuel Hirsch
Sebastian Moser
Virginie Ponelle
Paul-Loup Weil-Dubuc

Espace de réflexion éthique
de la région Ile-de-France
CHU Saint-Louis
75475 Paris Cedex 10
www.espace-ethique.org

Conception :
Zoo, designers graphiques
www.z-o-o.fr

Sommaire

Introduction 07
Synthèse générale du workshop 11
Présentation des participants 18

① Le contexte des *big data* dans les recherches sur les maladies neuro-dégénératives 21
1 — *Big data* : entre réalité et illusion 22
2 — *Big data* : conditions matérielles et limites techniques 23
3 — *Big data* : enjeux méthodologiques et fiabilité des connaissances 24
4 — *Big data* : contraintes organisationnelles et infrastructure 25
5 — *Big data* : partage et diffusion des données 27
6 — Donner un sens aux données : des data aux connaissances 28
7 — Accès aux données et protection juridique 29

② Les données massives d'imagerie : origines, intérêts, conséquences 33
1 — Produire des objets cohérents dans la complexité 34
2 — La délicate combinaison des niveaux d'explication 35
3 — Emettre des hypothèses ou multiplier les découvertes à partir des données ? 35
4 — Les bases de données et la pluralité des niveaux descriptifs 37
5 — Qu'est-ce que l'authentique multidisciplinarité ? 39
6 — Sommes nous capables de nous appuyer sur le *big data* au profit d'une politique d'anticipation responsable ? 40
7 — Le droit de savoir, l'incertitude et la vérité 41

③ Surveillance, participation, finalités. Les données massives exigent-elles de repenser l'éthique de la recherche ? 43
1 — Introduction 44
2 — L'autonomisation des données 45
3 — Le *big data* et la dilution du consentement 46
4 — Les algorithmes comme gouvernement 46
5 — Vers des formes collectives de consentement ? 47
6 — *Big Data* : intentions implicites et contrôle démocratique 50
7 — Solidarité et égalité des chances : vers une transformation par le numérique ? 51

④ Les données massives en recherche clinique et l'utilisation du séquençage haut débit 55
1 — Plan cancer et *big data* 56
2 — L'INCA et la recherche dans un environnement changeant 57
3 — La génétique et la juste appréciation de la hiérarchie des facteurs de risque 59
4 — L'appréhension collective des facteurs de risques et la hiérarchisation des priorités d'action 60

Propos conclusifs 63

Introduction
— Les *datas* au cœur des pratiques de recherche, des thérapeutiques et du soin

Léo Coutellec, Paul-Loup Weil-Dubuc,
Emmanuel Hirsch

Une démarche éthique engagée

L'Espace éthique de la région Ile-de-France n'a cessé de chercher à étendre son réseau et ses champs de compétences. Il a constamment sollicité de nouvelles personnalités afin d'élargir le périmètre de sa réflexion au-delà de ses préoccupations originelles. Initialement, l'Espace éthique était en effet confronté à des enjeux plus spécifiquement liés aux réalités de l'éthique hospitalière et du soin. Sa démarche concerne au sens le plus large la démocratie sanitaire. Le Plan Alzheimer 2008-2012 a fait évoluer l'Espace éthique dans la mesure où il lui a été demandé de prendre la responsabilité de constituer un Espace national de réflexion éthique dédié, précisément, à la maladie d'Alzheimer (EREMA). L'EREMA a développé jusqu'en 2014 son approche à la fois de la recherche, des pratiques médicales et soignantes, de l'accompagnement médico-social. En novembre 2014 il a diversifié ses missions devenant l'Espace national MND (EREMAND) dans le cadre du Plan maladies neuro-dégénératives 2014-2019. Cette ouverture illustre à la fois un engagement politique au sein de la cité ainsi qu'une méthode pluridisciplinaire soucieuse du bien commun, des valeurs de la recherche, de la médecine et des pratiques soignantes au service de la personne malade et de ses proches.

Les missions d'un Espace éthique[1] concernent la recherche et les formations universitaires, la concertation et la sensibilisation de la société aux enjeux notamment de la biomédecine du point de vue de la recherche fondamentale, de ses applications et des pratiques professionnelles au plus près de la personne au domicile ou en institution. Ainsi, dans le cadre des réflexions souvent anticipatrices menées avec les équipes de recherche dans une perspective pluridisciplinaire, avec le Laboratoire d'excellence Distalz auquel l'Espace éthique de la région Ile-de-France est associé nous avons organisé deux

1 www.espace-ethique.org

précédents workshops[2], consacrés à la problématique du diagnostic précoce de la maladie d'Alzheimer et des maladies neuro-dégénératives : ils ont donné lieu à la publication du premier numéro des *Cahiers de l'Espace éthique*[3].

Il nous tient à cœur de souligner le fait que nous encourageons les participants à nos workshops à parler vrai. La libre parole est irremplaçable et, même si des questions doivent demeurer en attente de réponses fondées, ayons le courage de les exprimer, ne serait-ce que pour mettre en valeur les enjeux sur les plans de la démocratie et des libertés individuelles. Par exemple le concept de « médecine de précision » a récemment émergé, et il convient d'expliquer pourquoi et selon quels objectifs, tout en contribuant à une réflexion pluraliste qui permette de poser les véritables enjeux sans se satisfaire de considérations générales et convenues. Cela au risque de susciter des confrontations en termes d'argumentations, indispensables à la prise en compte d'enjeux parfois négligés, notamment lorsque certaines évolutions biomédicales exposent à de nouvelles formes de vulnérabilités humaines et sociales. Il est un courage, voire une audace, dans toute démarche éthique engagée.

Big data : un phénomène à la fois culturel, technologique et scientifique

Le workshop dont nous présentons les réflexions dans ce *Cahier de l'Espace éthique* est consacré au *big data*. Depuis quelques années, nous constatons en effet l'émergence de ce phénomène qui se traduit par une « avalanche de données », une collecte systématique et massive de données et une croissance rapide des technologies de traitement. Il est possible de définir la démarche *big data* selon une dimension quantitative (basée sur le volume et le rythme de production des données, en constante croissance) et une dimension qualitative (données hétérogènes provenant de sources multiples). La dimension qualitative, moins souvent relevée, doit faire l'objet d'une attention particulière.

Appréhender *big data* seulement comme une révolution technologique – traitement majoritaire qui lui est actuellement fait – serait une réduction. Nous proposons de le comprendre comme un phénomène à la fois culturel, technologique et scientifique. Défis et enjeux auxquels, au-delà de la communauté des chercheurs et des praticiens, notre société est confrontée ne serait-ce

2 Le premier workshop s'est tenu le 4 avril 2013 à l'hôpital Bretonneau (AP-HP) ; le second le 19 mai 2014 à la Fondation Médéric Alzheimer (Paris).

3 Coutellec L., Weil-Dubuc P.-L., Hirsch E. (dir.), « Interventions précoces, diagnostics précoces. Approches éthique et sociétale de l'anticipation de la maladie d'Alzheimer et des maladies neuro-dégénératives », Les Cahiers de l'Espace éthique, n°1, octobre 2014.

que du fait des bouleversements que provoque cette mutation scientifique dans nombres de domaines qui touchent à nos représentations mais également à nos libertés fondamentales.

Nous inscrivons ce workshop dans la dynamique initiée avec la réflexion sur le diagnostic précoce dans le contexte de la maladie d'Alzheimer. Cette approche a permis de faire émerger deux champs de questionnement. Le premier est celui de l'anticipation comme problématique transversale, transdisciplinaire, et qui donne lieu depuis 2014 à un séminaire de recherche en éthique : « Anticipation, penser et agir avec le futur »[4]. Le second porte sur la place et le statut des données dans la recherche biomédicale, les essais cliniques et l'accompagnement concret des patients. Et, plus précisément, sur la façon dont ces domaines sont bouleversés par ce que l'on a coutume d'appeler *big data*. Un phénomène qui correspond à la disponibilité récente de données massives, dynamiques et hétérogènes.

Plusieurs enjeux de fond on émergé des travaux préparatoires à ce workshop :

1. Enjeux éthiques

Nous pouvons identifier plusieurs enjeux d'ordre éthique/juridique :
- banalisation de l'enregistrement de données biométriques ;
- problèmes éthiques de la collecte non sélective et des découvertes fortuites ;
- traçage, surveillance, contrôle par les données ;
- droit à l'oubli et protection des données ;
- quantification de la personne par ses données biométriques ;
- automatisation des relations par l'intermédiaire des données.

2. Enjeux sociaux

Nous pouvons identifier plusieurs enjeux d'ordre social/culturel :
- une mise en données du monde qui donne le primat à la quantification ;
- une algorithmisation des existences et des institutions (ex. : logiciel de décision en médecine) et « gouvernementalité algorithmique » (gouverner par les chiffres) ;
- un usage abusif du « data mining » (profilage et traçage par les données) ;
- un futur réduit à une prédiction algorithmique ;
- *Big data* → *Open Data* ?
- *Big data* → bio-marqueurs → médecine personnalisée ?
- Retombées réelles pour les populations ?

4 http://www.espace-ethique.org/seminaire14

3. Enjeux scientifiques

Nous pouvons identifier plusieurs enjeux d'ordre scientifique/épistémologique :
- distance entre objectifs de collecte massive et objectifs de résultat ;
- place du cadre théorique, des hypothèses et de l'imagination scientifique ;
- place du contexte et des filtres interprétatifs ;
- statut et place des données dans la recherche (*data-driven science ?*) ;
- passage difficile des données aux connaissances ;
- maintenir l'hétérogénéité des données vs standardisation des protocoles de collectes et de traitement des données.

Comment les données, les datas, font-elles évoluer les pratiques de recherche, d'accompagnement thérapeutique et les manières de dispenser le soin ? Ce questionnement s'inscrit dans la filiation de nos travaux passés, au cours desquels il a progressivement émergé.

Nous avons dégagé pour ce workshop quatre thèmes principauxqui structurent cette publication :
- les implications du *big data* dans la recherche relative aux maladies neuro-dégénératives ;
- les données massives d'imagerie médicale à travers une série d'exemples concrets, notamment en rapport avec le diagnostic ;
- l'éthique de la recherche et les implications des données massives sur les cadres de pensée, en convoquant les notions de surveillance, de participation et de finalité ;
- les données massives en recherche clinique et l'utilisation du séquençage haut débit, en lien avec le « Plan Cancer ».

La restitution des interventions et des échanges au cours de ce workshop reprend la méthodologie que nous avons retenue pour la diffusion des réflexions partagées dans le contexte de ces concertations dynamiques et créatives. À la relecture de leurs interventions, les participants ont apporté quelques précisions tout en préservant le caractère direct des échanges. Cette approche fera l'objet d'approfondissements dans le cadre des activités de recherche du Laboratoire d'excellence Distalz.

Nous remercions la Fédération hospitalière de France (FHF) associée à ce workshop qu'elle a accueilli dans ses locaux, ainsi que les personnes et les institutions impliquées dans cette initiative.

Synthèse générale du workshop — « De l'épistémologie des données à l'éthique de l'anticipation »

Léo Coutellec &
Paul-Loup Weil-Dubuc

Ce workshop interdisciplinaire visait à comprendre et à examiner les implications éthiques et épistémologiques de la production et de l'utilisation de plus en plus massives de données dans le champ des pratiques biomédicales.

Nous avons plus précisément souhaité comprendre quels phénomènes, quelles tendances de la recherche biomédicale et du soin *big data* – cette expression désormais consacrée – servait à désigner. Le *big data* est certes une réalité : dans le domaine de la génomique, la vitesse du séquençage s'est prodigieusement accélérée depuis 2005, ce qui a induit une baisse spectaculaire de son coût ; l'imagerie médicale connaît une évolution analogue dont témoignent l'amélioration des résolutions, le perfectionnement des images en 3D, les possibilités d'une multiplicité d'angles de vue.

Toutefois, nous sommes partis de l'hypothèse que le *big data* est autant une promesse qu'une réalité et avons tenté d'en cerner les implicites, les intentions et les impacts dans le champ de la biomédecine. Nos regards se sont orientés vers les maladies neuro-dégénératives et les cancers, deux types de pathologies chroniques qui ont pour point commun de nourrir de considérables efforts de recherche visant aussi bien leur anticipation, leur diagnostic que leurs traitements possibles.

À la lecture du workshop, quatre grands enjeux d'ordres épistémologique et éthique, quatre mutations induites par le *big data*, nous semblent se dégager.

1. Le processus de production du savoir à l'épreuve du phénomène *big data*

« Doit-on souhaiter des modèles intégratifs ou travailler sur cette matière si particulière qu'est l'hétérogénéité ? Il semble effectivement que l'hétérogénéité fasse partie intégrante de l'interdisciplinarité. » — Anne-Françoise Schmid

Comme on le sait, la production massive de données, dans le champ biomédical comme ailleurs, porte l'ambition de mettre au jour de nouvelles vérités scientifiques. Cet espoir est permis par la puissance de techniques informatiques capables de collecter et de traiter des données issues de disciplines et de terrains d'expérience embrassant, pourrait-on penser, la totalité des possibles. De ces données émergeraient des corrélations, voire des causalités jusqu'alors insoupçonnées.

Cette démarche scientifique entend rompre avec la méthodologie traditionnelle de la recherche biomédicale, « hypothético-déductive », au sein de laquelle l'hypothèse du chercheur constitue la première étape, infirmée ou non par l'expérimentation. Ce serait ici la mise en présence de données diverses qui serait le premier moment de la recherche : « Avec le *big data*, on passe du schéma rationaliste classique de l'hypothèse aux données. En un sens, on peut parler de science de la découverte par comparaison à une science de l'hypothèse » (Arnaud Cachia). On parlera également, pour désigner ce type de recherche qui se passerait d'hypothèses aprioristes, de recherche « agnostique ». Les hypothèses ne disparaissent pas mais changent de statut. Elles ne sont plus la source de l'investigation, elles sont des ingrédients d'ajustement théorique à la suite de l'exploration des données.

Plus généralement, nous pourrions avancer que la démarche *big data* vise la vérité par le biais d'un affranchissement à l'égard de tous les éléments contextuels de la recherche : aussi bien les éléments de contexte subjectifs (l'hypothèse du chercheur déterminée par sa subjectivité et sa culture) que les éléments de contexte objectifs (les théories et les terrains de l'expérimentation). Or, au plan épistémologique, cette prétention à s'affranchir des contextes, notamment ces cadres théoriques, cet « agnosticisme » revendiqué, se heurte à trois difficultés majeures.

La première réside en ce qu'une réalité « mise en données » pourrait être un reflet appauvri, voire déformé de la réalité. Le travail sur les données, depuis leur recueil jusqu'à leur agrégation et leur interprétation, suppose un effort de décontextualisation, par lequel on abstrait l'objet étudié – la séquence génétique, l'image, etc. – d'autres données qui lui sont intimement liées ou d'autres savoirs conçus sans données parce qu'orientés vers la biographie et le vécu des personnes. Autant de savoirs qu'une démarche scientifique pilotée par les données ignorera pour la seule raison que les systèmes de collecte et de calcul sont incapables de les « digérer ».

Cette première difficulté en implique une deuxième : le *big data* tend à déplacer, à brouiller, voire à effacer les frontières entre disciplines. Dans le champ biomédical, la promesse d'un croisement de données par nature hétérogènes donne crédit au projet d'une approche « intégrative » des maladies

tout autant qu'elle en révèle les difficultés[1]. Les données se présentent volontiers comme un nouveau langage, un nouveau mode de communication entre disciplines autrefois hermétiques les unes aux autres. Une question se pose pourtant : doit-on viser une compréhension unifiée des maladies ou doit-on admettre une hétérogénéité irréductible ? Hétérogénéité qui laisse inexorablement des espaces de non-cohérence entre savoirs, qui implique des ruptures de chaîne causales, qui appelle à rompre avec la vision totalisante de la science positiviste.

La visée d'une recherche sans hypothèse, enfin, pourrait être illusoire. Les données ne font pas sens par elle-même ; le savoir fait nécessairement intervenir le regard du chercheur interrogeant la base de données à partir des connaissances dont il dispose déjà : « Nous agrégeons des données d'après ce que l'on sait déjà, c'est-à-dire d'après la somme de connaissances déjà disponibles ; le risque est donc grand de conforter ce qui est déjà connu, jusqu'à la tautologie » (J.-C. Lambert).

Plus fondamentalement, la génération massive et spontanée de données pourrait laisser penser que la vérité sur les mécanismes biologiques de l'humain se situent là, dans des « lacs de données » (P.-O. Gisbert) infiniment exploitables, à portée d'algorithmes. Sans doute faut-il se garder de ces illusions épistémologiques nourrissant des espérances parfois trompeuses et entretenues par des intérêts économiques.

Le statut de la donnée est à interroger lorsqu'elle celle-ci se présente comme *data*. « Insistons sur le fait que les données sont le résultat de longs process et pipeline d'analyses. En imagerie, on ne voit pas le cerveau mais des « p value », c'est-à-dire des constructions. La donnée est elle-même le résultat d'analyses massives en quelque sorte. L'objet est un résultat et non une évidence » (A. Cachia). De la *data* à la *ficta*, c'est une forme de « phénoméno-technique » de la donnée qu'il nous faut construire collectivement afin de tracer et de prendre du recul sur l'histoire de vie de ces données devenues si précieuses pour les recherches. C'est à ce prix que l'espoir de recherche suscité par les *big data* ne sera pas sacrifié sur l'autel de l'économie de la promesse.

2. L'organisation de la recherche à l'épreuve du phénomène *big data*

« Ne négligeons pas la réalité au nom du possible, d'un horizon ou d'un futur. »
— Emmanuel Hirsch

À l'ère des données massives, le paysage de la recherche biomédicale se transforme. Les projets de recherche, par l'ampleur des moyens techniques qu'ils mobilisent, impliquent naturellement davantage d'acteurs, de dis-

ciplines, de financements. Les acteurs de la recherche sont appelés à former des réseaux dépassant les compétitions de laboratoires. Cette collaboration nécessaire pourrait notamment s'incarner dans des espaces de stockage partagés – *clouds* – entre laboratoires, éventuellement nationaux et publics, sans toutefois ignorer les limites écologiques de ce genre d'entreprise, comme l'a souligné Nicolas Lechopier : « [Les limites] ne sont pas d'ordre logique ou commerciales, elles sont d'ordre énergétique car stocker consomme beaucoup d'énergie. Concrètement, les unités de stockage abritées dans des entrepôts doivent être sans cesse refroidies. Nous touchons là l'une des vraies limites de la révolution industrielle du numérique. À ce sujet, ne perdons pas de vue que toute révolution industrielle est synonyme de choc écologique. Rien ne peut s'étendre indéfiniment car nos sociétés dépendent de contraintes énergétiques ».

Ceci étant dit, le *big data* rend possible un élargissement de la communauté de recherche aux citoyens qui, malades ou non, pourraient accepter de mettre leurs données à la disposition de l'intérêt public. C'est l'enjeu particulier de l'*open data*. L'idéal démocratique susceptible d'animer cette ouverture de l'accès aux données doit être mis en regard des menaces qui l'accompagne : intrusion dans l'intimité et discrimination. D'autant que les croisements des données permettent aujourd'hui la réidentification des personnes, l'anonymat devenant en un sens « obsolète » (C.-A. Cuenod). La notion même de « donnée sensible » devient obsolète : « le *big data* participe d'un mouvement qui rend absurde le projet de vouloir définir la sensibilité des données en les classant selon la nature du domaine qu'elles représentent. La combinaison d'un certain nombre de données individuellement non sensibles peut aboutir à une information très « parlante », par laquelle un tiers par exemple serait capable d'acquérir une emprise sur l'individu en question » (N. Lechopier).

Du reste, le projet d'un gouvernement démocratique des données ne peut être mené à bien que si les orientations mêmes des recherches biomédicales sont ouvertes au débat démocratique. « Dans la logique *data driven*, les recherches dérivent de bases de données plutôt que d'une question directrice préalablement posée. La réflexion sur l'orientation des recherches devient aussi capitale que difficile. » (N. Lechopier). À cet égard, le *big data* fournissant les conditions de projets de recherche d'une ampleur et d'un coût inédits, une vigilance démocratique s'impose. L'espérance de traitements nouveaux ne peut être érigé en intérêt supérieur aux dépens d'autres finalités tout comme la fuite en avant à l'innovation technologique. « Nous nous trouvons dans un cas de figure où la construction d'outils a pris le pas sur toute autre logique qui lui aurait été préexistante. C'est le régime techno-instrumental d'une science passée à l'échelle industrielle. » (N. Lechopier). L'espérance portée à juste titre par la recherche biomédicale et ses nouveaux dispositifs techniques ne sau-

rait discréditer d'autres formes d'espérance – celles qui concernent, par exemple, le soin de la personne[1].

L'utilisation de plus en plus fréquente d'outils de diagnostic génétique comme le séquençage très haut débit nous appelle également à interroger les frontières classiques entre recherche, clinique et soin. Ceci est tout fait prégnant dans le cas du cancer : « En cancérologie, on ne peut faire de différence entre recherche et soin. Nous ne cessons donc d'accumuler des données » (H. Nabi, N. Hoog Labouret). Recherche et soin ne forment plus deux étapes distinctes et autonomes mais constituent désormais deux faces contemporaines d'un même processus visant indissociablement la recherche à plus ou moins long terme et le soin de la personne. C'est de ce nouvel entremêlement entre recherche et soin que les acteurs doivent prendre conscience. Et ceci nous amène à décrire un troisième enjeu qui tient précisément aux différents déplacements de frontières.

3. Implications éthiques sur le consentement et la démocratie

La collecte massive de données et les outils techniques d'analyse qui l'accompagnent interrogent la notion même de consentement. Principe fondamental de l'éthique des pratiques biomédicales, le consentement implique, depuis la loi du 4 mars 2002, qu' « aucun acte médical ni aucun traitement ne peut être pratiqué sans le consentement libre et éclairé de la personne et ce consentement peut être retiré à tout moment ». Cela s'applique donc directement à la possibilité de collecter des données de santé sur la personne et pour le cas des tests. Y. Hirsch nous rappelle que des « règles applicables aux traitements des données y compris celles issues du *big data*. Par exemple, ceux qui collectent ces données et qui les traitent doivent, notamment lorsqu'il s'agit de données à caractère personnel, informer les individus concernés et obtenir leur consentement avant toute mise en œuvre du traitement ».

Nous identifions toutefois deux contextes où ces principes et règles semblent difficilement applicables, celui du séquençage haut débit du génome utilisé comme test génétique – « À l'ère du séquençage haut débit, nul doute que des anomalies vont être systématiquement dépistées » (H. Nabi et N. Hoog Labouret) et celui des objets connectés – « Avec les objets connectés, ne verrons-nous pas de nouvelles sources de données émerger, qui n'auront pas un statut comparable aux données scientifiques, mais qui n'en seront pas moins disponibles et utilisables ? » (Y. Hirsch).

En conséquence, nous sommes invités à penser de nouvelles formes de consentement et à étendre nos réflexions vers de nouveaux horizons conceptuels liant éthique et démocratie, « à envisager des formes collectives, politiques, de consentement. La vision patrimoniale des données, assimilables

à un capital ou une richesse, est bien trop étriquée pour rendre compte de ce qui se joue sous nos yeux» (N. Lechopier). Car, avec cette question du consentement, c'est un enjeu démocratique de premier ordre qui se dessine, celui de garder une forme de maîtrise collective sur l'usage de nos données et d'éviter des formes de gouvernementalité algorithmique ou de nouvelles formes de surveillance (que l'on connaît déjà avec le profilage par les données). Ceci nous conduit directement au quatrième enjeu identifié à l'occasion de ce colloque, un enjeu tout autant éthique que politique, celui des finalités.

4. Interroger les finalités de cette collecte massive et systématique de données

«Dorénavant, il faut interroger le rapport entre les sciences et des valeurs, ainsi que la manière dont sont opérés des choix de recherche. La notion de responsabilité est incontournable.» — Nicolas Lechopier

Une réflexion éthique sur les *big data* doit prendre en compte le contexte, les conséquences, les valeurs mais aussi s'interroger sur les finalités d'un tel phénomène. Entre réalité et illusion se logent les intentions implicites, les finalités assumées et celles plus difficiles à identifier. «Je peux témoigner du fait que des chercheurs en informatique (notamment) capturent des données à très grande échelle, bien souvent sans développer le moindre questionnement éthique relatif à la finalité de la captation de ces données.» (Nicolas Lechopier). Pourtant, afin de penser un tel phénomène dans toute sa complexité, il convient de se prémunir contre toute forme de déterminisme qui pourrait nous enjoindre de n'en faire qu'un élément de la longue histoire du progrès technologique. Pour penser les *big data,* «il importe de se déprendre de la force du présent, de la mode et de l'évidence» (N. Lechopier). Car le *big data* peut être investi pour des finalités multiples, parfois contradictoires entre elles.

Si les progrès sont réels (puissance, rapidité, hétérogénéité), nous devons aussi voir et interroger les limites et les problèmes des *big data*. Nous avons évoqué les enjeux épistémologiques et éthiques, mais qu'en est-il des enjeux économiques? Les infrastructures technologiques autour des *big data* appellent de lourds investissements et autant de moyens pour les maintenir et les faire évoluer. Dans une période de restriction budgétaire, des choix s'opèrent sur les orientations de la recherche et, inévitablement, des voies d'avenir entrent en concurrence. Comment éviter que l'engouement autour des *big data* ne se transforme en prophétie auto-réalisatrice qui rendrait indispensables des approches biomédicales impliquant de lourdes et coûteuses infrastructures technologiques? Comment bénéficier des espoirs suscités par ces nouvelles approches par les données tout en cultivant une

forme de responsabilité collective à leur propos ? Comment faire en sorte que la génération massive de données se solde par un véritable gain des citoyens en autonomie et qu'elle ne soit pas un facteur supplémentaire de discrimination et de déterminisme socio-économique ?

Nous le constatons, de l'épistémologie des données à l'éthique de l'anticipation, nous avons à préciser et à cerner de façon plus fondamentale les différents enjeux des *big data* dans les pratiques biomédicales. Nous espérons que ce workshop en sera une contribution utile.

Présentation des participants

Sanaa Ait Daoud
Digital & Ethics ; Laboratoire Montpellier Recherche en Management.

Philippe Amouyel
Professeur d'épidémiologie et de santé publique, directeur de l'UMR 744 Inserm, université Lille 2 – Institut Pasteur de Lille et du Laboratoire d'excellence DISTALZ.

Céline Bellenguez
Biostatisticienne, unité INSERM-U1167, Institut Pasteur de Lille.

Pierre-Emmanuel Brugeron
Chargé de mission et chef de projets à l'Espace éthique de la région Ile-de-France.

Arnaud Cachia
Professeur de neurosciences, Université Paris Descartes, CNRS LaPsyDE, Sorbonne, membre fondateur du Programme interdisciplinaire 'Imageries du Vivant' de Sorbonne Paris Cité.

René Caillet
Responsable du pôle Organisation sanitaire et médico-sociale de la Fédération Hospitalière de France.

Vincent Chouraki
Médecin de santé publique, post-doctorant à l'unité INSERM-U1167, Institut Pasteur de Lille.

David Claverau
Chargé de l'e-santé à l'Agence Régionale de Santé (ARS) – Ile-de-France.

Leo Coutellec
Chercheur en philosophie des sciences, Laboratoire d'Excellence DISTALZ, Espace éthique de la région Ile-de-France, Université Paris-Sud.

Charles-André Cuenod
Professeur de radiologie à l'Université Paris Descartes et praticien à l'HEGP, coordinateur du Programme interdisciplinaire 'Imageries du vivant' (IDV).

Danielle Geldwerth
Biophysicienne, laboratoire CSPBAT-UMR CNRS 7244, rattaché à l'Université Paris 13 et au programme interdisciplinaire 'Imageries du Vivant' de Sorbonne-Paris-Cité.

Paul-Olivier Gibert
Fondateur et directeur de Digital & Ethics.

Benjamin Grenier-Boley
Bio-informaticien, INSERM-U1167, Institut Pasteur de Lille.

Jean Hache
Physicien, actuellement en thèse de philosophie à l'Université Paris 1 sur la médecine personnalisée.

Hermann Nabi
Epidémiologiste, responsable du département Recherche en SHS, épidémiologie et santé publique de l'INCA.

Emmanuel Hirsch
Professeur d'éthique médicale à l'Université Paris-Sud, directeur de l'Espace éthique Ile-de-France et de l'Espace national de réflexion éthique maladies neuro-dégénératives.

Yaël Hirsch
Avocate en Télécommunications, Média et Nouvelles Technologie au cabinet Simmons & Simmons.

Natalie Hoog Labouret
Pédiatre, Institut du Cancer, pôle Recherche et Innovation.

Pauline Lachapelle
Médiatrice, Service Sciences et Société de la communauté universitaire Lyon/Saint-Etienne.

Jean-Charles Lambert
Biologiste, Directeur de recherche à l'unité INSERM-U1167, Institut Pasteur de Lille.

Nicolas Lechopier
Philosophe des sciences, Maître de conférences, Université Claude Bernard de Lyon.

Muriel Mambrini-Doudet
Biologiste, Directrice de recherche INRA Jouy-en-Josas, ex-présidente du Centre INRA de Jouy-en-Josas, Chercheure invitée à MINES Paristech.

Max Mollon
Designer, Doctorant à l'ENSAD, chercheur associé à l'Espace éthique de la région Ile-de-France .

Gislain Philip
Interne en médecine du travail, Espace éthique de la région Ile-de-France.

Simon Saint-Georges
Chargé de mission, Pôle Santé Publique de l'INSERM-AVIESAN.

Anne-Françoise Schmid
Maître de conférences en épistémologie, Chercheure invitée à MINES Paristech.

Henri-Corto Stoekle
Doctorant en éthique médicale au Laboratoire d'Ethique Médicale et Médecine Légale (EA4569) à l'Université Paris-Descartes.

Paul-Loup Weil-Dubuc
Chercheur en philosophie politique et morale, Laboratoire d'Excellence DISTALZ, Espace de réflexion éthique de la région Ile-de-France, Université Paris-Sud.

Édition et relecture

Alexandre Descamps
Interne en santé publique, Espace de réflexion éthique de la région Ile-de-France

Patrice Dubosc
Ressources documentaires, Espace éthique de réflexion éthique de la région Ile-de-France

Clément Landron
Stagiaire, Espace de réflexion éthique de la région Ile-de-France

1

Le contexte des *big data* dans les recherches sur les maladies neuro-dégénératives

1 — *Big data* : entre réalité et illusion

Philippe Amouyel

Nous avons à cerner la réalité du phénomène des *big data*. Quelle est sa juste portée ? Dans la mesure où le concept a été très largement diffusé dans le grand public, trop de propos relèvent du fantasme. Les limites à mettre en évidence ne sont pas tant techniques que conceptuelles. Dans le contexte des maladies dégénératives comme dans d'autres domaines, l'exécution de techniques de haut débit combinées ou non aux nanotechnologies peut générer en quelques heures des volumes de données colossaux. Pour obtenir des informations sur un gène, il fallait attendre 48 heures voici quelques années de cela. Aujourd'hui, tout le génome peut être investigué en 6 heures. Depuis 2005, la vitesse de calcul s'est prodigieusement accélérée.

Notre équipe de recherche est à l'origine d'une publication scientifique dans *Nature*[1], qui correspond à une méta-analyse des génomes de plus de 74 000 individus. Le travail a permis d'identifier 11 nouveaux *loci*[2] de susceptibilité à la maladie d'Alzheimer. Il va de soi que, dans un univers informationnel de près de 11 millions de variables, le volume d'informations à traiter pour parvenir à un tel résultat est considérable. L'utilisateur non averti est fréquemment surpris par les durées requises au téléchargement des données de travail. L'information relative à 500 000 SNPs pour 10 000 sujets représente 5 gigas. Il faut du temps pour ouvrir un fichier Excel et il convient naturellement de tenir compte de la durée d'analyse des données au cours d'un travail scientifique.

Dans le champ des affections neuro-dégénératives, le G7 (aujourd'hui G8) Dementia Research a examiné l'exploitation des *big data* dans le but de développer un traitement et de tester de nouveaux modèles de prise en charge. Pour parvenir à l'objectif visé, il fallait des données « larges » (en population) et « profondes » (sur les plans clinique et biologique). Naturellement, les questions de gouvernance et de financement d'un projet international sont incontournables. La plupart des budgets de recherche gouvernementaux ne prennent pas en charge la problématique et il convient de définir une stratégie de financement. À cet effet, il a été fait référence aux principes des Bermudes, formalisés en 1996[3]. Rappelons le contexte de concurrence à cette époque entre le projet de séquençage public du génome humain et le projet privé de Craig Venter[4]. Lorsque notre unité de recherche a mené à bien ses premiers travaux en 2008, mettant en ligne les données produites, il est apparu manifeste que les équipes anglaises et américaines financées par des fonds publics avaient un impératif de publication effective au bout de deux ans. Nous étions donc loin d'une publication immédiate. Ainsi doit-on toujours mesurer la distance entre les principes proclamés et la réalité.

Actuellement, notre équipe travaille sur le séquençage des exons du génome humain, soit 34 millions de paires de bases ou environ 1,2 % du génome. Ces dernières années, nous avons assisté à une baisse sensible du coût des opérations de séquençage.

À la demande du G8, l'OCDE a soutenu la publication d'un document : « *Big data for advancing dementia research* »[5]. Celui-ci fait état de 4 types de défis dans l'exploitation des *big data* :

- technologiques ;
- relatifs au recueil des consentements ;
- en lien avec les *process* et l'organisation (organisation d'un écosystème et de son financement) ;
- d'ordre humain (rassemblement des compétences et aptitude à les combiner dans un projet).

« Pour obtenir des informations sur un gène, il fallait attendre 48 heures voici quelques années de cela. Aujourd'hui, tout le génome peut être investigué en 6 heures. Depuis 2005, la vitesse de calcul s'est prodigieusement accélérée. »

1 http://www.nature.com/ng/journal/v45/n12/abs/ng.2802.html

2 Emplacement correspondant à un gène sur le chromosome

3 Publication automatique de toute séquence de génome assemblée supérieure à 1 kb au mieux avant 24 heures, publication immédiate des séquences annotées terminées et objectif de rendre toute séquence entière librement disponible dans le domaine public, dans des perspectives de recherche et développement avec le but ultime de maximiser les bénéfices apportés à la collectivité.

4 À ce propos, voir : John Sulston. « Le génome humain sauvé de la spéculation », *Le Monde diplomatique*, décembre 2002.

5 http://www.oecd-ilibrary.org/science-and-technology/big-data-for-advancing-dementia-research_5js4sbddf7jk-en?crawler=true

2 — *Big data* : conditions matérielles et limites techniques Benjamin Grenier-Boley

Sur un plan technique, les premières étapes d'un travail de recherche consistent à transférer, stocker, sécuriser et analyser des données. L'interprétation clinique et les applications de la « médecine personnalisée » ne constituent que l'étape finale d'un processus complexe et lent.

Dans nos travaux, nous sommes par exemple amenés à considérer un échantillon de 10 000 individus pour 600 000 variables (SNPs). Pour ces variants, les données brutes représentent 50 Go (giga-octets) et les données traitées après filtrage de l'information sont de l'ordre de 2 Go.

Si nous voulons séquencer les exons complets de 1 000 individus, on doit composer avec 30 To (tera-octets) de données brutes pour des données effectives traitées de 10 Go. Quant au génome complet de 1 000 individus, son étude nécessite un espace de travail de 500 To, pour un volume de données effectivement traitées de 200 Go.

Sur le plan du transfert des données, il faudrait un an et demi à un réseau de débit ordinaire (1 Mo/s) pour acheminer 1 000 génomes complets. Il ne faut que 56 jours avec un réseau universitaire, par exemple celui de l'Institut Pasteur à Lille dont le débit est de 10 Mo/s.

Les données ne sont pas définitives dans la mesure où l'on décline une série de programmes afin d'en contrôler la qualité, de les nettoyer en quelque sorte. La démarche consiste à sélectionner des mutations à analyser par des outils statistiques et, partant, d'en extraire une forme de connaissance. Il va de soi que l'exécution de l'ensemble des programmes requis réclame des machines très performantes. En effet, une même machine doit renfermer 100 Go de matière, là où les PC domestiques n'ont qu'une capacité de 4-8 Go. Or, pour passer des étapes clés, il faut effectivement disposer de 100 Go incompressibles.

Le traitement des données est également très exigeant. Si on devait traiter les génomes complets de 1 000 individus, sur un processeur, il faudrait 240 années afin de terminer le processus. Bien évidemment, le calcul s'appuie sur des clusters comprenant de nombreux processeurs pour conduire des tâches en parallèle.

Au terme de l'exécution de l'ensemble des processus de traitement, les données générées doivent être impérativement sécurisées. Concrètement, il s'agit de les répliquer pour les affecter à deux unités distinctes physiquement. Ce faisant, on en double le volume. La réplication des données, afin de ne pas les perdre, est très onéreuse. Les robots de sauvegarde et le recours à des bandes magnétiques coûtent cher, d'autant plus qu'à l'image des disques durs les bandes magnétiques ont une durée de vie limitée. On doit recopier les données de bande à bande continuellement, ce qui est également fastidieux. Le *cloud* a pu être évoqué comme la solution ultime aux problèmes de stockage. Or, sur les volumes de données que nous manipulons, son coût est prohibitif. Le recours à des *clouds* privés (Amazon, Google, etc.) poserait en outre des problèmes juridiques.

Une infrastructure de stockage matériel des données est garantie 7 ans. Autrement dit, tous les 7 ans son renouvellement doit être financé. Il va de soi qu'il n'existe aucune sécurité absolue. En cas d'incendie au lieu de stockage, il subsiste toujours un risque de tout perdre.

« Le traitement des données est également très exigeant. Si on devait traiter les génomes complets de 1000 individus, sur un processeur, il faudrait 240 années afin de terminer le processus. »

En définitive, le *big data* est une affaire d'équipe rassemblant les compétences nécessaires. Soit l'équipe possède toutes les compétences en interne, soit elle est dans la nécessité de nouer des partenariats. Les besoins considérables en capacité de calcul conduisent à solliciter par exemple les grands centres (IDRIS du CNRS, TGCC du CEA), les universités, l'Institut Français de Bioinformatique ou encore les centres de séquençage. Si de nombreux acteurs proposent des solutions de calcul, rares sont ceux qui offrent des possibilités de stockage pérennes.

3 — *Big data* : enjeux méthodologiques et fiabilité des connaissances

Céline Bellenguez

L'information que nous manipulons est considérable en taille, mais aussi en complexité. Ne perdons pas de vue que les biologistes s'intéressent à des phénomènes complexes. Ainsi, étant relative par exemple à l'ADN, à l'ARN ou aux protéines, l'information à traiter est hétérogène. L'origine génétique des individus constitue un autre facteur d'hétérogénéité. Ajoutons que la technique n'est pas homogène, les procédés de séquençage ne sont nullement uniformes. De ce fait, lorsque des consortiums internationaux travaillent sur de larges échantillons, les données sont générées par plusieurs laboratoires opérant différemment. Finalement, quand bien même de l'information nouvelle serait produite de nature à répondre à de nouvelles questions scientifiques, encore faudrait-il l'exploiter statistiquement avec des méthodes valides. Ainsi, les méthodes statistiques doivent-elles évoluer.

La pratique statistique est donc impactée par les caractéristiques des *big data*. La statisticienne que je suis doit détenir des compétences en informatique afin d'exploiter au mieux les capacités du cluster de calcul. Méthodologiquement, nous sommes face à la limite de l'intégration de données de natures différentes. Il n'existe pas aujourd'hui d'approche intégrative globale. Souvent, le statisticien ne travaille pas sur l'intégralité des données mais sur une information résumée. Or, en diminuant le « bruit de fond » des données, on perd immanquablement des éléments d'information. Sur un plan éthique, un des enjeux essentiels correspond à la protection de la confidentialité. Cette fois, le fait de travailler sur une information résumée et non sur une information exhaustive y contribue.

Le statisticien est constamment aux prises avec les « faux positifs » et les « faux négatifs ». Sur ce plan, l'hétérogénéité non détectée de la matière rassemblée est redoutable. Disposer d'échantillons de grande taille permet en effet de déceler des effets de faible ampleur, y compris ceux résultant de faibles biais, nous exposant ainsi au risque de générer des faux positifs. Pour conférer un caractère probant à des résultats, on demande donc qu'ils soient reproduits dans un échantillon indépendant. Il est cependant souvent malaisé de trouver un autre échantillon de grande taille, indépendant, pour répliquer ses conclusions. De nouvelles approches statistiques sont régulièrement proposées, mais nous n'avons encore que peu de recul sur leur utilisation. De même, de nouveaux outils et *softwares* sont régulièrement mis à la disposition du statisticien, mais ils ne sont pas toujours matures. En d'autres termes, nous n'avons qu'une connaissance très imparfaite des biais susceptibles d'entacher nos travaux.

« De nouvelles approches statistiques sont régulièrement proposées, mais nous n'avons encore que peu de recul sur leur utilisation. De même, de nouveaux outils et *softwares* sont régulièrement mis à la disposition du statisticien, mais ils ne sont pas toujours matures. En d'autres termes, nous n'avons qu'une connaissance très imparfaite des biais susceptibles d'entacher nos travaux. »

4 — *Big data* : contraintes organisationnelles et infrastructure

Vincent Chouraki Mon intérêt personnel pour les statistiques et l'informatique m'a permis de travailler sur des fichiers qu'un ordinateur individuel, en raison de leur taille, ne peut manipuler sans se bloquer. Personnellement, j'ai rapidement opté pour une station de travail Linux et j'ai souvenir que la première analyse que j'ai dû exécuter sur cette plateforme a réclamé 15 jours. Chaque collaborateur a des centres d'intérêts variés, mais on doit s'appuyer sur des compétences cardinales bien particulières. Ainsi, dans le monde du *big data*, il faut savoir déléguer car il est impossible de maîtriser l'ensemble des domaines mobilisés dans un projet commun. Un statisticien n'est pas biologiste et vice-versa. Chacun a besoin de l'autre dans l'exécution de processus selon une logique volontiers décrite comme celle du *pipeline*. Toute l'information n'est pas stockée, parfois elle doit être traitée en direct. En règle générale, l'information passe surtout à travers une succession de filtres. Même sous un format résumé, elle se trouve par exemple sous la forme d'un fichier de 2 Go comportant plusieurs millions de lignes que l'on s'efforce de mettre en rapport avec des processus physiopathologiques.

Aujourd'hui, on ne cesse d'annoter le génome humain. De plus en plus de *loci* sont associés aux maladies dégénératives. Ce faisant, la cartographie des associations entre pathologie et génome se complexifie et l'accumulation d'informations résumées reconduit au *big data* suivant une sorte de circularité.

Sans doute la période 2010-2020 pourra-t-elle être essentiellement une étape d'acquisition de connaissances, tant l'émergence de nouvelles pratiques médicales prendra du temps. Actuellement, les équipes transmettent aux biologistes des données dans l'espoir de mieux prédire des risques. Notons que ces données peuvent être relatives à la génétique ou à d'autres marqueurs que des séquences de génome.

Dans une perspective de médecine génomique, la croissance spectaculaire des données disponibles soulève la question des modalités de la prise de décision médicale. Il va de soi que les médecins auront besoin d'outils capables d'extraire l'information pertinente de la profusion qui ne manquera pas de régner à l'avenir. Les cliniciens ont besoin d'éléments compréhensibles, exprimés sur une page et, de surcroît, disponibles dans un délai acceptable. De plus, des données en apparence pertinentes émergeront, mais sans rapport avec le contexte clinique justifiant les consultations. Que faudra-t-il faire de ces données ?

Dans l'état actuel des choses, force est de constater que les outils de prédiction du risque de développer une maladie d'Alzheimer d'après des déterminants génétiques ne fonctionnent pas très bien.

Enfin, les stratégies relatives à l'information disponible ne sont pas toutes clarifiées. Doit-on stocker indéfiniment l'information génétique brute ou bien en extraire des éléments jugés pertinents pour l'effacer ? Comment faut-il former les étudiants ? De quelles compétences devront-ils disposer ? De quels outils auront-ils besoin ? Toutes ces questions demeurent ouvertes.

« Dans une perspective de médecine génomique, la croissance spectaculaire des données disponibles soulève la question des modalités de la prise de décision médicale. »

Paul-Olivier Gibert Disposez-vous d'une authentique infrastructure spécifique permettant de stocker des données massivement distribuées et d'effectuer des calculs à partir de la base stockée ?

Jean-Charles Lambert Nous sommes capables de gérer correctement la quantité d'informations que nous manipulons grâce à des solutions d'externalisation. Une solution interne au laboratoire ne serait pas concevable pour une double raison de gestion financière et de gestion des compétences.

Disons que nous sommes une petite structure réactive, qui s'appuie sur un groupe restreint de personnes. Nous sommes nécessairement contraints d'aller chercher des ressources à l'extérieur. Pour le moment, nous disposons d'équipes adaptées à des besoins qui sont évolutifs. Il est donc concevable que nous grandissions au-delà des limites de la solution que nous avons développée. Plus nous élargirons les possibilités de développement, plus nous serons satisfaits. Notre solution nous convient aujourd'hui, mais nous ne saurions prétendre qu'elle nous conviendra encore dans 5 ans.

Danielle Geldwerth Comment résumer la nature de votre infrastructure ?

Jean-Charles Lambert Nous nous appuyons sur le Centres de ressources informatiques de Lille (CRI) où nous avons la chance de pouvoir compter sur des personnes parfaitement qualifiées. Grâce à cette ressource externe, nous avons accès à une infrastructure de stockage et à des clusters de calcul en phase avec nos besoins. Même lorsqu'à partir de 2011 il nous a fallu conduire des analyses sur génome entier, nous avons pu disposer des capacités de calcul nécessaires. Parallèlement à la génomique, elles sont mises à disposition des physiciens ou des spécialistes de la modélisation moléculaire. Toutefois, nous sommes confrontés à des contraintes croissantes en termes de stockage, les volumes entrant en jeu ne cessant d'augmenter.

Vincent Chouraki Nous confions désormais les fonctions d'administration système, de calcul et de stockage à des partenaires extérieurs. Ainsi, notre collaborateur potentiellement compétent se consacre intégralement à la bio-informatique.

Philippe Amouyel L'externalisation n'est pas non plus une solution parfaite. Nous avons déploré, par exemple, le fait de ne plus avoir accès à nos données pendant 3 mois en raison d'un problème relatif à la sécurité d'un centre de calcul. À ce moment-là, nous avons souhaité disposer de notre propre cluster. Toutes les solutions ne sont pas tenables. Amazon et Google nous ont formalisé des propositions, mais elles ne sont pas acceptables du fait de la nécessaire confidentialité des données.

Arnaud Cachia Nous avons insisté sur le fait que de multiples compétences étaient techniquement incontournables. Comment ces dernières sont-elles susceptibles d'être gérées sur le plan institutionnel ? N'est-on pas, avec le *big data*, passé d'une recherche organisée en laboratoires autonomes à une recherche collaborative, plus distribuée, en réseau ?

Philippe Amouyel Sur le plan de l'organisation de la recherche, les biologistes ont beaucoup à apprendre des physiciens. Si notre unité de recherche disposait d'une solution d'externalisation parfaite, elle ne manquerait pas d'y recourir. Malheureusement, nous n'avons pas en vue de réponse adaptée à nos besoins, notamment sur le plan du stockage, qui ne soit hors de prix. A un moment ou à un autre, l'accroissement du volume des donnés implique un changement d'échelle. En France, une réflexion est en cours au sujet d'un *cloud* souverain. Nous ne demandons qu'à disposer d'une solution de stockage valide pour ne plus consacrer de temps à l'infrastructure mais plutôt à ce que nous savons le mieux faire.

Jean-Charles Lambert Interrogeons ce que l'on entend par « donnée ». On postule qu'une donnée n'est pas automatiquement agrégeable à une autre, et à juste titre. En effet, trois paramètres déterminants sont incontournables :

— les processus de traitement de l'information ;

– l'harmonisation des données ;

« La biologie est une science constamment aux prises avec l'hétérogénéité. Deux ou trois laboratoires travaillant en parallèle sur un même questionnement produisent fréquemment des résultats non superposables. »

– la qualité du travail à fournir pour disposer de données utilisables et agrégeables.

La biologie est une science constamment aux prises avec l'hétérogénéité. Deux ou trois laboratoires travaillant en parallèle sur un même questionnement produisent fréquemment des résultats non superposables. Prenons un exemple dans le contexte de la maladie d'Alzheimer : le dosage de deux marqueurs dans le liquide céphalo-rachidien des patients. En France, une vingtaine de centres effectuent le dosage et nous nous sommes aperçus que les données générées n'étaient pas superposables. Rappelons qu'il existe des dizaines de milliers de biomarqueurs dans les sciences biomédicales et que leur nombre est en expansion. Les logiciels employés en bioinformatique ne génèrent pas toujours la même information. En protéomique, en transcriptomique, le biologiste est confronté en permanence aux variabilités interindividuelle et intra-individuelle. On ne peut qu'agréger ce qui est normalisé et la normalisation constitue un enjeu de première importance car nous touchons là à la possibilité même d'agréger.

5 — *Big data* : partage et diffusion des données

Hermann Nabi Méthodologiquement, il faut reproduire des résultats donnés sur une population suffisamment importante en taille pour qu'ils soient acceptables pour la communauté scientifique. Cependant, la variabilité constitue un obstacle de taille car, de fait, les populations sont distinctes. On ne saurait par exemple agréger sans précautions des données en provenance de la population française avec d'autres données issues de la population finlandaise.

Pauline Lachapelle Le *big data* induit des changements dans nos cultures universitaires. Quel est le périmètre actuel de la collaboration scientifique et qu'en est-il du partage de données entre les pays du nord et du sud ?

Philippe Amouyel Il a été précédemment fait mention de principes éthiques généraux qui doivent gouverner la recherche : ceux dits des Bermudes adoptés par le G8. À titre personnel, j'ai milité pour l'ouverture au public non seulement des données produites par les universités, mais encore de celles produites par les laboratoires pharmaceutiques privés. À ce sujet, je me suis toujours heurté à une fin de non-recevoir. On peut parler d'asymétrie dans la mesure où le monde public doit mettre ses données en ligne, alors que la sphère privée ne partage même pas ses données froides dénuées d'enjeu commercial.

Jean-Charles Lambert Notre laboratoire participe à une étude africaine pilotée à Brazzaville, dont le but est d'évaluer la pertinence d'un facteur génétique dans un processus physiopathologique. Il est possible de transférer des compétences dans un pays du sud, mais c'est là une tâche très complexe. Concrètement, il faut disposer d'une personne compétente et capable, culturellement, de remplir la fonction d'interface. Notons que la problématique de la diffusion de la connaissance est délicate, dans les pays du nord comme dans ceux du sud. Notre équipe sait à quel point il est délicat de communiquer sur une éventuelle prédisposition génétique à la maladie ou sur des facteurs de risques. Schématiquement, la population a trop tendance à assimiler le facteur de risque au risque lui-même. La diffusion de la culture scientifique n'est pas chose simple. Ensuite, il existe des enjeux de rapport entre le Nord et le Sud. Nous n'avons pas formalisé de réflexion à ce sujet.

Philippe Amouyel Nous avons mené à bien une étude en partenariat avec une équipe algérienne. À vrai dire, le pilotage a été opéré depuis la France, des personnes relais étant sur place. Beaucoup de temps a été nécessaire en vue de former les collaborateurs algériens. On peut estimer le temps de transfert de compétences à au moins 10 ans avant que l'équipe formée ne devienne autonome.

Philippe Amouyel La notion de donnée utile est d'une importance capitale. Aujourd'hui, l'internaute est face à une information dont souvent il ne sait que faire et, désormais, il est question de son propre génome. La société 23andMe est typique de la logique qui dévoile le génome des individus sans aucune médiation. En d'autres termes, nul n'intervient entre la production et la réception de l'information, ce qui n'est pas sans soulever une difficulté éthique majeure. On peut par exemple annoncer à quelqu'un qu'il a un gène de prédisposition au glaucome sans que la personne exprime pour autant la maladie.

Henri-Corto Stoekle Générer, traiter et stocker des données renvoient à des compétences bien distinctes. C'est en effet la stratégie de Google qui émancipe ce marché au travers de ses filiales comme 23andMe. Son objectif est en effet de générer des données valorisables à travers des offres de service pouvant séduire un consommateur avide de nouvelles technologies. Toutefois, quel est l'état des certifications ou de la labellisation de ces données générées? Google est dans une phase majeure de production, il est important dans ce cadre qu'il se porte garant de la qualité et du respect des données. C'est dans cette réflexion qu'un réel partage de données serait possible.

Philippe Amouyel Google a soutenu à l'origine le lancement de 23andMe. De façon intéressante, Sergueï Brin a relevé qu'il était porteur de la mutation associée à la maladie de Parkinson dont sa mère a été frappée. Il s'est donc publiquement engagé contre cette maladie et a soutenu des études. Quand Sergueï Brin s'est ensuite séparé de sa compagne, à l'origine de 23andMe, il a fondé Calico. À n'en pas douter, les propositions de dévoiler les génomes individuels qui ont fleuri sur Internet relèvent de logiques commerciales et bien des discours relèvent du fantasme.

Il n'existe pas de voie conduisant tout droit de la génétique à la maladie d'Alzheimer ou au cancer. L'ADN des tumeurs recèle sa propre complexité. Le positionnement d'acteurs tels que 23andMe est très éloigné de la clinique qui demeure le seul point d'aboutissement de la génétique. **À quoi rime le fait d'adresser à domicile des tests BRCA1 aux consommateurs? Que faire concrètement de l'information selon laquelle on est porteur d'une grave mutation?**

6 — Donner un sens aux données : des *data* aux connaissances

Jean-Charles Lambert Prenons l'exemple de quelqu'un qui se verrait révéler la présence d'un gène de prédisposition au glaucome via un test à puce affymetrix. Si la présence avérée de ce gène, signifiant un risque trois fois supérieur à la normale, s'inscrit dans une histoire familiale marquée par la pathologie, alors la personne sera encline à la dépister très tôt. Concrètement, elle demandera à son ophtalmologiste de mesurer sa pression intraoculaire à un âge qui surprendra ce dernier. D'ailleurs, comment un ophtalmologiste comprendra-t-il une demande d'examen hors de tout contexte clinique et uniquement basé sur l'invocation d'un facteur de risque génétique ? Manifestement, il faudra le convaincre de la pertinence d'une investigation clinique *a priori* en l'absence de fondement logique. **L'information génétique doit par conséquent être maitrisée non seulement par son destinataire – la personne ayant demandé à ce que des séquences de son génome soient révélées – mais aussi, ensuite, par les cliniciens. Il existe donc deux niveaux de lecture.**

Nicolas Lechopier Je m'interroge sur la réduction de notre discussion aux seules données génétiques. Les prédictions génétiques souffrent de nombreuses incertitudes. Certes, nous devons progresser en matière d'éducation à l'usage de l'information génétique, mais les données génétiques doivent être associées à d'autres données pour être parlantes. Ce ne sont pas des données détenant, en tant que tel, un savoir privilègié. Ne sommes-nous pas en train de chercher à leur conférer un tel privilège en affirmant qu'elles sont susceptibles de modifier des pratiques cliniques ? Ne risquons-nous pas d'estomper ainsi l'incertitude ?

Philippe Amouyel Nous avons conscience de ces questions. Une chose consiste à faire mention du *big data* à propos de la génétique, une autre d'évoquer ce phénomène au sujet de l'hypercholestérolémie. En aucune façon nous devons inférer un diagnostic d'une information non validée. Nous parlons aujourd'hui de grands volumes de données qui, naturellement, ont une portée qu'il conviendrait effectivement de clarifier.

Vincent Chouraki La prédiction de risques par la génétique n'a fait ses preuves que dans de rares cas. Force est de constater que, n'étant reproductibles nulle part ailleurs, de nombreux résultats de recherche ne sont valables que pour la population étudiée. Rappelons qu'en 2013, la FDA a interdit à 23andMe d'utiliser les données génétiques en provenance de ses clients pour effectuer de la médecine prédictive. Cette société a donc été cantonnée au périmètre de la généalogie jusqu'en début d'année 2015, date où la FDA a autorisé 23andMe à commercialiser un test génétique dans le cadre d'une maladie génétique rare.

Philippe Amouyel De fait, c'est bien encore l'hétérogénéité des populations (européenne, américaine, asiatique…) qui prévaut en matière de génétique.

Jean-Charles Lambert Dans le cas de l'hypercholestérolémie, on utilise un seuil quelque peu arbitraire pour distinguer le normal et le pathologique s'agissant d'une variable continue. Ce n'est absolument pas le cas de la génétique.

Muriel Mambrini-Doudet Le monde académique n'a-t-il pas la responsabilité de promouvoir un autre rapport aux données ? Le débat pointe la nécessité de faire la part des choses. D'une part, entre ce qui est significatif et ce qui ne l'est pas et, d'autre part, entre ce qui ne vaut que pour une population et sur ce qui est de portée générale. Les statisticiens ont bien insisté sur le fait que des conclusions ne sont pas reproductibles d'un groupe à l'autre. **Manifestement, nous avons besoin d'échanges entre les différents domaines de connaissance. Où sont les spécialisations aujourd'hui ? Qui est responsable de quoi, notamment dans le cadre de la relation médecin/patient ?** Manifestement, le monde académique doit intervenir afin de déterminer quels messages méritent d'être diffusés et défendus. Certes, nous sommes dans un monde de compétences et de spécialisations, encore faut-il s'appuyer sur un langage et sur des références partagées, selon une certaine fluidité.

« Certes, nous sommes dans un monde de compétences et de spécialisations, encore faut-il s'appuyer sur un langage et sur des références partagées, selon une certaine fluidité. »

Jean-Charles Lambert Au-delà de la reproductibilité ou non des conclusions d'une investigation d'une population à l'autre, nous avons à appréhender la qualité des publications scientifiques et la démocratisation de l'accès aux données. Surtout, le fait que les données soient accessibles ne signifie nullement qu'elles soient analysées. Une étape sera franchie dans la démocratisation de l'accès aux données avec celle de leur analyse. Concrètement, n'importe qui pourra conduire des analyses en exécutant des outils. Reconnaissons qu'aujourd'hui, des publications scientifiques fleurissent se bornant à de nouvelles investigations sur des données existantes, sans rien apporter de nouveau. Dans le bruit de fond actuel de publications, le scientifique est mis au défi de mettre en lumière ce qui est authentiquement digne d'intérêt. Dans cette perspective, il incombe à la recherche publique de mettre en place des formes de filtres pour que le public ne se trouve pas noyé dans la surabondance d'informations. Par exemple, la génétique est concernée par des dizaines de publications de « méta-analyses » mensuelles, souvent d'origine chinoise. Or, on constate qu'elles sont sans intérêt et risquent d'occulter des travaux qui méritent l'attention.

Vincent Chouraki Nous touchons là une des caractéristiques du *big data*. Il suffit d'exécuter des outils pour les faire parler presque à l'infini.

Philippe Amouyel Effectivement, nous ne saurions nier la responsabilité du monde académique pour faire la part des choses. La problématique de l'homogénéité dépasse le champ du *big data.* Aujourd'hui, les objets connectés sont à la mode. Les *start-ups* se multiplient. Or, une dizaine d'objets censés mesurer la même chose sont en réalité à l'origine de différences considérables. Tout est affaire d'usage. On a pu imaginer le pire à partir d'informations génétiques dévoilées par la société 23andMe. Distinguons bien le savoir brut du savoir-faire.

Dans le monde scientifique, les listes d'auteurs s'allongent, mais les laboratoires tendent à publier isolément. Prenons le champ de la maladie d'Alzheimer. Entre 1994 et 2009, 587 gènes ont été proposés comme déterminants potentiels de la maladie. Or, aucun n'a été vérifié. Pour apporter les publications collectives dont la communauté a besoin, il est nécessaire de diffuser un certain état d'esprit. Que signifie d'être l'un des contributeurs d'une étude engageant la contribution d'un millier d'auteurs ? La question mérite d'être posée et on doit s'honorer d'avoir apporté sa pierre à un édifice qui fait sens. En effet, le *big data* est porteur de changements dans la manière dont la communauté de la recherche biomédicale fonctionne. Contribuer à un immense effort collectif porteur de sens est bien plus que de vouloir publier isolément, dans son laboratoire, des éléments parcellaires.

« Contribuer à un immense effort collectif porteur de sens est bien plus que de vouloir publier isolément, dans son laboratoire, des éléments parcellaires. »

7 — Accès aux données et protection juridique

Yaël Hirsch Qu'en est-il de l'encadrement légal de l'utilisation des données relevant du phénomène du *big data ?* Il convient sur ce point d'être clair, les données traitées dans le cadre d'un projet de *big data* ne sont pas laissées sans protection juridique au libre jeu des acteurs. Il existe en France et en Europe des règles applicables aux traitements des données y compris celles issues du *big data*. Par exemple, ceux qui collectent ces données et qui les traitent doivent, notamment lorsqu'il s'agit de données à caractère personnel, informer les individus concernés et obtenir leur consentement avant toute mise en œuvre du traitement. En revanche, de par sa taille, le *big data* pose la question de la mise en œuvre de règles adaptées aux activités envisagées.

Sur ce point, tant sur le plan national (activité de la CNIL) que sur le plan européen (activité du G29 et projet de règlement européen relatif à protection des personnes physiques à l'égard du traitement des données à caractère personnel), les entités compétentes cherchent à créer un équilibre entre la protection de la vie privée des individus et la volonté de favoriser l'essor de nouvelles activités liées au développement du *big data*.

Les objets connectés s'inscrivent-ils dans un vide juridique ? Les objets connectés sont devenus une des sources du *big data* dans le domaine de la santé. On ne compte plus le nombre d'applications et autres montres permettant aux utilisateurs non professionnels de mesurer un ensemble de données les concernant (nombre de pas, rythme cardiaque,

etc.). Leur intérêt pour la médecine et la recherche est évident, encore faut-il pouvoir s'assurer de la fiabilité des données captées. Il existe une volonté européenne, déclinée en France, consistant à instaurer un cercle de confiance entre les objets connectés et leur éventuel usage médical. Au-delà de l'application purement médicale qui pourra être faite de ces données et qui en principe devrait être soumise à la règlementation déjà existante en matière de santé, la question se pose, le plus souvent pour les concepteurs d'applications mobile ou autres objets, de la délimitation entre les données dites « de bien-être » et les données de santé. La tâche est d'autant plus difficile qu'aujourd'hui le Code de la santé ne fournit aucune définition légale de la notion de donnée de santé. **À mon sens, au-delà du caractère récréatif des objets connectés, dès lors qu'on en attend un avis ou une consultation médicale conduite par un professionnel de la santé, on devrait tomber dans le champ régulé de l'acte médical.** Telle serait la ou une des limite(s) pertinente(s) sur le plan juridique : le but recherché par le concepteur ou l'utilisateur du produit.

Emmanuel Hirsch Dès lors qu'une pratique est considérée innovante et intervient dans un contexte dépourvu de repérages efficients, la tentation est grande de considérer que les encadrements seront produits à l'expérience ou à l'épreuve des évolutions. De telle sorte qu'on engage des dispositifs dans un contexte de précipitation et de compétition bien souvent rétif au temps d'une pondération éthique, pour ne pas dire d'une réflexion nécessaire. Il est du reste assez décevant de constater le peu d'engagement des instances à vocation éthique dans ces domaines qui justifieraient de leur part une implication plus forte que les quelques recommandations convenues dont on sait qu'elles ne résisteront pas à la montée en puissance d'enjeux notamment d'ordre financiers qui risquent d'imposer leurs logiques et leurs règles. **Lorsque l'on observe le caractère encore approximatif des approches effectives de l'usage des données informationnelles que nous évoquons, mais tout autant les développements qu'elles permettent d'envisager, il me semble qu'il y aurait urgence à mettre en œuvre un travail d'anticipation, d'observation et de suivi de leur implémentation d'un point de vue sciences humaines et sociales. Car il y va de toute évidence d'aspects qui concernent nos droits et nos libertés fondamentales.**

« Les objets connectés sont devenus une des sources du *big data* dans le domaine de la santé. »

Paul-Olivier Gibert La directive de 1995[1] est antérieure à l'émergence du *big data*. Le cadre conceptuel réglementaire européen est en phase avec la réalité des années 2008-2009. Peut-être ne pose-t-il pas tous les bons points de contrôle. Quant à la loi française sur la santé, des enjeux politiques ont prévalu sans doute sur des enjeux éthiques spécifiquement liés à l'*open data* et, *a fortiori*, au *big data*.

Philippe Amouyel La « donnée » existe par elle-même. Lorsque l'on évoque l'*open data*, on fait référence à la problématique de l'accès public aux données qui est transversale à de nombreux sujets. On l'a dit, les données liées à la santé sont sensibles. N'est-il pas depuis longtemps courant de recueillir des valeurs de pression artérielle et de les consigner ?
Aujourd'hui, nous sommes aux prises avec des données d'un volume inédit et leur disponibilité pose problème. Sans doute la problématique de l'*open data* est-elle plus décisive que celle du *big data* à proprement parler.

Paul-Olivier Gibert La « donnée » telle que définie par la loi de 1978, modifiée en 2004, ne correspond pas à la *data* contemporaine. Le cadre conceptuel, à l'époque, était déterminé par la carte perforée. Autant dire que c'est le traitement qui primait sur la donnée. Nous sommes loin des bases de données qui apparaissent présentement comme des gisements indéfiniment ré-exploitables. Sans doute la réglementation n'a-t-elle pas pris la mesure du phénomène. **De fait, il existe des gisements ou des « lacs » de données.** Par exemple, on peut analyser l'activité de millions de compte Twitter. De même, on peut conduire des analyses sur des données publiques et naturellement sur des génomes. **Dès lors, ou bien ces gisements sont ouverts à tous, ou bien ils sont d'accès restreint.**

Emmanuel Hirsch On perçoit de multiples ambigüités dans l'appréhension des données relatives à la santé. Pour le moment, les autorités publiques ne semblent pas avoir fait preuve de suffisamment de volontarisme dans la loi de modernisation du système de santé, compte tenu des enjeux liés à ces données. Prenons le cas de celles qui proviennent des 1,3 milliards de feuilles de soin de l'Assurance Maladie. Pour le moment, aucune position convaincante, autre que l'annonce de dispositifs de contrôle dont nul n'ignore les faiblesses, ne nous a permis d'être assurés que l'appariement du contenu de différents fichiers informatiques relevant de données tant sanitaires que médico-sociales ne seraient pas de nature à produire des informations sensibles qui, ainsi cumulées et accessibles y compris de manière sélectives, s'avèreraient préjudiciables à l'intérêt direct de la personne. Il nous faudrait nous satisfaire de résolutions publiques qui se veulent rassurantes en termes de mesure de protection des libertés individuelles alors que l'on sait d'expérience les données informatiques sujettes à toutes formes d'intrusions et de détournements. **Quant à la compétence des décideurs politiques dans des domaines technologiques particulière-**

ment sophistiqués, elle est susceptible d'être pour le moins déconsidérée par ceux qui détiennent l'expertise scientifique et la capacité d'intervention.

Philippe Amouyel On se doit de vivre avec son temps et de clarifier les droits d'accès aux donnés.

Charles-André Cuenod La discussion éthique a évoqué l'*open data*, les principes des Bermudes sur le plan de l'accès aux données et de leur publication. Toutefois, comment gérer leur anonymisation ? Nous savons qu'il suffit de 5 informations pour identifier une personne à coup sûr. Le croisement de données, même apparemment anonymes, peut fort bien aboutir à désigner quelqu'un.

Philippe Amouyel Avec seulement 1000 SNPs, on peut déterminer l'identité de quelqu'un, où il est né et si le père déclaré à l'état civil est le père biologique.

Charles-André Cuenod La notion d'anonymat serait donc en un sens obsolète. Il y a 3 ans, une communication scientifique est parue, qui s'est basée sur des informations d'état civil publiées par les mormons aux Etats-Unis. En testant un millier de séquences ADN disponibles sur le web d'une manière ou d'une autre, il est apparu que les individus pouvaient être retrouvés dans 10 % des cas par croisement de données.

Paul-Olivier Gibert L'information n'est-elle pas entachée d'une marge d'erreur ? Comment peut-on effectuer un rapprochement systématique entre une séquence de génome et l'état civil d'un individu ? N'est-ce pas ce que la protection « informatique et libertés » est censée prévenir ?

Philippe Amouyel On ne peut défendre que des rapprochements puissent être systématiquement effectués. Ce qui est problématique à l'heure actuelle, c'est d'y parvenir au moins dans quelques cas. Nous voyons qu'une évolution est à l'œuvre. Le généticien sait que les séquences Y sont très caractéristiques des individus. Il n'a nullement besoin de très longues séquences pour procéder à une identification. Avec des données collectées par les Mormons à Salt Lake City, il est tout à fait envisageable qu'un membre d'une fratrie soit contacté pour l'informer d'une mutation décelée chez son frère ou sa sœur.

Jean-Charles Lambert Avec les avancées actuelles en génomique, il n'est pas absurde de penser qu'un jour le génome d'un individu sera disponible sur sa « carte Vitale ». Certes, on peut penser

Philippe Amouyel qu'une information demeurera d'accès restreint aux seuls médecins, mais d'innombrables corrélations pourront être faites pour lever des voiles de doute. Nous ne sommes pas loin du scénario du film « Bienvenue à Gattaca ».

Charles-André Cuenod Manifestement, une tendance est à l'œuvre qui tend à rendre tout un chacun nu sur la toile. Il ne semble guère possible de lutter contre elle.

Philippe Amouyel Elle est déjà bien enclenchée.

Charles-André Cuenod Avec l'*open data*, on peut tout à fait imaginer voir des individus chercher des données génétiques sur un conjoint potentiel avant de se marier.

1 La directive 95/46/CE constitue le texte de référence, au niveau européen, en matière de protection des données à caractère personnel. Elle met en place un cadre réglementaire visant à établir un équilibre entre un niveau élevé de protection de la vie privée des personnes et la libre circulation des données à caractère personnel au sein de l'Union européenne (UE).

« Comment gérer l'anonymisation des données ? Nous savons qu'il suffit de 5 informations pour identifier une personne à coup sûr. Le croisement de données, même apparemment anonymes, peut fort bien aboutir à désigner quelqu'un. La notion d'anonymat serait donc en un sens obsolète. »

②

Les données massives d’imagerie : origines, intérêts, conséquences

1 — Produire des objets cohérents dans la complexité

Arnaud Cachia

Que signifie l'émergence des données massives dans l'univers de l'imagerie cérébrale et, plus généralement, dans la discipline de la neuro-imagerie ? Dans ma pratique quotidienne, je suis affilié à deux laboratoires utilisant l'imagerie cérébrale (Imagerie par Résonnance Magnétique, IRM) dans lesquels nous examinons des IRM de sujets sains ou des IRM de patients souffrant de troubles psychiatriques. **La question des données massives en neuro-imagerie est déjà très présente dans le domaine de la pathologie ; toutefois, cette question commence également à émerger chez le sujet sain, en particulier dans le domaine de la neuro-éducation.**

Comme dans d'autres domaines des sciences biomédicales, nous avons assisté à une croissance exponentielle de la quantité de données générées. Certes, on peut invoquer l'augmentation de la résolution des clichés. Comme pour les images des appareils photos numériques, nous avons assisté à un accroissement spectaculaire des quantités de voxels (pixel en trois dimensions) par image. Toutefois, un nouveau type de données est apparue pour caractériser des fibres de substance blanche avec l'IRM de diffusion qui nécessite d'explorer un objet selon plusieurs directions. En d'autres termes, afin de préciser l'orientation des fibres, il est nécessaire d'échantillonner l'espace selon de nombreuses directions. Au début des années 1990, l'orientation des fibres était estimée avec des séquences d'IRM de diffusion utilisant 3 ou 4 directions. Aujourd'hui, nous pouvons utiliser plusieurs centaines de directions pour explorer les détails des fibres de matière blanche. Nous disposons donc, pour chaque sujet, d'images gigognes où, en chaque voxel[1],de plusieurs centaines d'informations. Dans la mesure où la résolution angulaire augmente avec le nombre de directions étudiées, nous sommes bien en présence du *big data*.

Les données supplémentaires servent-elles à quelque chose ? Leur utilité a-t-elle été démontrée ? Tout d'abord, plus on augmente le nombre d'éléments corroborant une hypothèse, plus on renforce sa fiabilité. Prenons le cas des investigations d'imagerie fonctionnelle qui s'intéressent au fonctionnement du cerveau. Classiquement, on procède aux investigations sur 15 à 20 sujets. Des statistiques paramétriques sont donc conduites avec une population somme toute très limitée en nombre. Comme il a été souligné précédemment, le risque de « faux positifs » et de « faux négatifs » est redoutable sur les échantillons limités. En 2013, une contribution scientifique dans *Nature* a suscité de longs débats car elle insistait sur les limites des connaissances en neurosciences — et en particulier en neuro-imagerie —, les expérimentations étant conduites sur des groupes trop restreints pour être probantes. De la même manière, dans d'autres domaines, publier des données relatives seulement à une quinzaine d'animaux de laboratoire attire les critiques des statisticiens. La communauté des scientifiques travaillant dans les domaines de l'imagerie est donc invitée à mutualiser les données de recherche pour augmenter la taille des échantillons.

Observons qu'il existe différentes échelles d'investigation et nos objets sont caractérisés dans une perspective « multi-échelle ». Prenons l'exemple de la maladie d'Alzheimer. On conduit des investigations sur le génome, le protéome, sur des aspects cérébraux, cognitifs, jusqu'à la caractérisation clinique du syndrome. Les niveaux d'analyse se démultiplient. C'est dans cette perspective que le *big data* s'inscrit, afin de faciliter une description multi-échelle avec une analyse multimodale des phénomènes. Actuellement, de nouvelles méthodologies se développent en imagerie multimodale, afin de parvenir à des niveaux de résolution spatiale et temporelle bien particuliers. **Naturellement, une voie de recherche consiste à coupler des données d'imagerie avec des données génétiques. Le *big data* s'inscrit donc dans un contexte où l'accent est mis sur des approches intégratives des phénomènes.** Toutefois, si l'on combine sans cesse les données, on arrive à un schéma de *big data* au carré ou même au cube, car chaque niveau d'analyse s'appuie sur des données massives. Ceci n'est pas sans poser un triple problème de stockage, d'analyse et de définition de modèles théorique interprétatifs.

Pour illustrer notre propos, nous pouvons retenir le cas de la schizophrénie. Bien des travaux de recherche ont en quelque sorte déconstruit l'objet qu'est cette maladie, en le décrivant à différents niveaux. Ainsi, des études s'intéressent aux facteurs de risque génétiques, à la dimension cognitive de la pathologie, aux mécanismes développementaux, à la symptomatologie, etc. Il importe donc de réunir des niveaux bien distincts ou de les combiner dans des représentations cohérentes.

1 Pixel en trois dimensions.

« Avec le *big data*, on passe du schéma rationaliste classique de l'hypothèse aux données. En un sens, on peut parler de science de la découverte par comparaison à une science de l'hypothèse. »

2 — La délicate combinaison des niveaux d'explication

Arnaud Cachia Des éléments produits par imagerie sont parfois assimilés à des biomarqueurs. Ainsi s'intéresse-t-on au diagnostic précoce de la maladie d'Alzheimer au moyen de tels éléments. Typiquement, on cherchera des plaques amyloïdes sur le cerveau des personnes chez qui on suspecte un processus dégénératif. Toutefois, dans ce cas, il est bien difficile de transposer des conclusions valables dans des analyses de groupe au niveau individuel.

Plus généralement, une grande quantité de données est requise pour identifier en imagerie les paramètres pertinents et spécifiques, utilisables au niveau individuel. Dans le même ordre d'idées, il faut des masses de données pour répliquer les biomarqueurs dont on essaie de vérifier, sur plusieurs bases de données, la validité dans un contexte biologique donné. En outre, la compréhension statistique des phénomènes, sur des populations nombreuses, ne signifie pas que l'on soit en mesure de prédire l'évolution des processus individuels. On se doit d'insister sur le fait que nous ne disposons pas de biomarqueur diagnostique déterministe. C'est pourquoi on se contente d'évoquer un « risque de développement de la maladie ». L'exemple de la schizophrénie est, à ce titre, évocateur. On a objectivé des étapes précoces de son développement, désignées comme « phases prodromales » chez les sujets à risque. Idéalement, on tente d'identifier parmi ces sujets à risque ceux qui sont particulièrement à même de développer un trouble psychotique, même quand les symptômes ne sont pas assez intenses ou continus. Chez les sujets à risque, il existe un groupe d'individus qui ne va pas développer *in fine* de trouble psychotique. Toutefois, on sait qu'une proportion d'entre eux va l'exprimer. Sur la base de cette certitude partielle ne concernant qu'une population et non chaque individu directement, on peut tenter d'utiliser l'imagerie comme outil d'investigation paraclinique, dans le but de compléter les informations cliniques. Si, dans une population, le risque d'exprimer la schizophrénie est de 10 à 20 %, ne peut-on pas déterminer un sous-groupe dans lequel ce ratio s'élèverait à 80 %? L'imagerie est complémentaire de la génétique. Méthodologiquement, on essaie de superposer les grilles de lecture pour une meilleure compréhension des phénomènes.

3 — Emettre des hypothèses ou multiplier les découvertes à partir des données?

Arnaud Cachia On peut distinguer des méthodologies « *hypothesis-driven* » et « *data-driven* ». Tant que l'on raisonne sur des échantillons restreints, on doit formuler une hypothèse et expérimenter, en vue de sa confirmation ou de son infirmation, au moyen de tests statistiques. Avec le *big data*, on passe du schéma rationaliste classique de l'hypothèse aux données. En un sens, on peut parler de science de la découverte par comparaison à une science de l'hypothèse. Sur un plan nosologique, avec le *big data*, nous espérons identifier des mécanismes physiopathologiques que nous ne connaissons pas encore en mettant en évidence des sous-groupes dont nous n'avions pas précédemment conscience. Ce sont les données qui génèrent d'elles-mêmes des hypothèses. Le procédé n'est toutefois pas dénué de limites. Dans le champ de l'imagerie, la pratique du *big data* correspond en réalité à l'agrégation multicentrique de données acquises par l'exécution de différents protocoles. Par conséquent, on doit veiller à ce que les « lacs de données » obtenus ne le soient pas au moyen de protocoles par trop dissemblables. Souvent, le processus d'acquisition est découplé de l'analyse. Mieux vaut donc préciser les questions auxquelles on souhaite répondre avant de se lancer dans l'acquisition de données. Trop souvent, on produit sans réfléchir préalablement aux questions à se poser. Or, la façon dont on s'interroge conditionne sensiblement, en amont, la manière dont on va acquérir les données. Précédemment, l'enjeu de l'harmonisation a été souligné car on ne saurait agréger des éléments de formats disparates. Le problème de l'harmonisation invite à évoquer celui de la collaboration. Les chercheurs en imagerie doivent apprendre à travailler ensemble au bénéfice de la somme de données générées dans les différents centres.

Ne perdons pas le recul théorique nécessaire à la compréhension de ce qui réside dans le matériau brut des données. La connaissance ne découle pas de ces dernières automatiquement. Plus que jamais, il faut des modèles, des

hypothèses pour comprendre. Aujourd'hui, il est fréquemment question de modèles intégratifs. Le thème est fascinant sur le plan intellectuel, mais très complexe à mettre en oeuvre. Pourquoi est-il si difficile d'intégrer ? On invoque des démarches interdisciplinaires ; multidisciplinaires, transdisciplinaires. On doit associer des intervenants qui n'ont pas l'habitude de travailler ensemble. Or, chacun doit s'efforcer de comprendre comment travaille l'autre. On n'est jamais expert que de son domaine.
La démarche de Distalz correspond à l'application d'un programme interdisciplinaire. Les défis à relever ne sauraient être sous-estimés, tant sur le plan humain que sur le plan méthodologique. Quand bien même on amasse des données, on a besoin de modèles. Aucune donnée ne véhicule *per ipse* la façon dont elle est interprétable. Les spécialistes de l'imagerie ont besoin de grandes bases de données et le partage n'est pas sans mettre en jeu des contraintes techniques majeures, tout comme naturellement le stockage et l'analyse.
Enfin, tous semble appeler de leurs vœux une plus grande mutualisation des données. N'est-ce pas là un vœu pieux ? Quel mécanisme récompensera les chercheurs qui partagent leurs données et, de ce fait, fera changer les pratiques ? Qui est prêt à partager la matière brute de ses investigations, préalablement à toute analyse ? Les difficultés sont évidentes sur le plan de l'*authorship* et de l'évaluation académique. Un défi collectif consiste à valoriser tous ceux qui travailleront à une vaste entreprise dans le cadre d'un consortium. Quelle solution avons-nous à notre disposition, au-delà des listes d'auteurs de plusieurs pages ? Il faudrait une profonde évolution culturelle, mais la reconnaissance de la contribution de chacun dans un travail collectif a toujours été problématique.

Jean-Charles Lambert Sans doute allons-nous assister à un changement de paradigme sur le plan de l'*authorship*. Il n'est pas évident à accepter. Les réticences sont évidentes, mais on ne saurait aller à l'encontre d'une tendance de fond. Les physiciens ont l'habitude des listes d'auteurs très longues, agencées par ordre alphabétique. On voit mal comment on pourra faire l'économie de pratiques nouvelles découlant du besoin d'agréger des données massives. Certes, nous manquons de recul sur le plan méthodologique et conceptuel. Surtout nous agrégeons des données d'après ce que l'on sait déjà, c'est-à-dire d'après la somme de connaissances déjà disponibles. Le risque est donc grand de conforter ce qui est déjà connu, jusqu'à la tautologie. Reconnaissons que nous interrogeons les bases de données sur des fonctions bien connues des protéines et des gènes. Or dans les deux cas il y a pléiotropie. Nous sommes loin de maîtriser cet aspect du vivant et défions-nous du biais consistant à procéder par agrégation à partir de connaissances déjà acquises. Le danger à conjurer est celui d'une stérilisation de la recherche.

Arnaud Cachia Nous allons sans doute vers la multiplication des *data papers*. Ces articles ont vocation à présenter les bases de données et leur évolution. Naturellement, chaque contributeur verra son nom mentionné, ce qui sera une manière de le valoriser.

Hermann Nabi Il n'est pas toujours si simple d'opposer la formulation d'hypothèses à la méthode *data-driven*. Dans le domaine de la recherche contre le cancer par exemple, des consortiums se sont mis en place à partir de directions de recherche sous-jacente. On n'accumule que très rarement des données sans finalité ou direction sous-jacente. Le *big data* n'est sans doute pas un changement de paradigme complet en recherche biomédicale.

Leo Coutellec **Les exposés ont mis en évidence la question du traitement de l'hétérogénéité. Nous avons mis en relief la différence entre la démarche hypothético-déductive et une logique empirico-inductive, où la connaissance jaillirait spontanément des données.** Or, des pathologies ont été citées à propos desquelles des filtres sont appliqués lorsque l'on passe d'une échelle à l'autre, d'une discipline à l'autre. Chaque filtre reflète pourtant une théorie ou un choix.
Avec les données massives et diverses, la question centrale n'est pas tant celle du volume que celle de l'hétérogénéité. Comment les choix sont-ils opérés pour conférer un ordre au divers ? Comment les biologistes retiennent-ils une annotation plutôt qu'une autre ? Sur un plan épistémologique, le « data » est-il bien un donné ? Après tout, il faut bien en faire quelque chose, et peut être à un stade très précoce des expérimentations. Enfin, l'hétérogénéité de la masse de données n'est pas sans soulever des questions éthiques.

Vincent Chouraki L'approche de l'hétérogène renvoie à un certain historique. Considérons les rapports entre généticiens et l'étude de la maladie d'Alzheimer. Durant une quinzaine d'années, les équipes ont produit des résultats non reproductibles sur des cohortes restreintes de patients. Ils ont résolu de se fédérer pour disposer d'une source d'informations d'une autre échelle. Ce faisant, un immense effort d'homogénéisation a été initié. La seule solution aux problèmes communs n'est autre que la constitution de consortiums.

« Quand bien même on amasse des données, on a besoin de modèles. Aucune donnée ne véhicule *per ipse* la façon dont elle est interprétable. »

Jean-Charles Lambert On tendait à commettre les mêmes erreurs. Effectivement, pour cesser de générer une information non pertinente, il est souhaitable de travailler en consortiums. Relevons que les acteurs qui recherchent des biomarqueurs en protéomique répètent les maladresses de méthode qui ont été commises en génétique. Pourtant, ils ont été avertis du risque de diluer les efforts en une multitude d'entreprises qui ne mènent à rien sur le plan de la connaissance. Très humainement, les chercheurs pensent tous qu'ils feront mieux en suivant leurs propres inclinaisons et en travaillant comme ils ont l'habitude de le faire, mais ils commettent systématiquement les mêmes fautes.

Vincent Chouraki Existe-t-il une méthode entièrement « *data driven* » ? C'est loin d'être évident. Les hypothèses ne sont certes pas toujours explicites, mais elles sont bien là dans la structuration des données et la manière dont elles sont analysées. Ainsi, pouvons-nous conduire des investigations à l'échelle du génome entier, mais sur la base d'hypothèses sur les déterminants génétiques de maladies complexes. Le programme de recherche consiste à étudier un nombre minimal de marqueurs. L'hypothèse est implicite. Si nous ne mettons rien de significatif en évidence sur le plan statistique, il est possible que cette hypothèse sous-jacente soit tout simplement fausse.

« Nous agrégeons des données d'après ce que l'on sait déjà, c'est-à-dire d'après la somme de connaissances déjà disponibles. Le risque est donc grand de conforter ce qui est déjà connu, jusqu'à la tautologie. Reconnaissons que nous interrogeons les bases de données sur des fonctions bien connues des protéines et des gènes. »

4 — Les bases de données et la pluralité des niveaux descriptifs

Anne-Françoise Schmid Que devient le savoir clinique dans cette évolution ?

Arnaud Cachia La clinique est aux prises avec des évaluations plus fines des situations. Par conséquent, les collaborateurs doivent être formés à un raffinement accru des tableaux cliniques. Observons qu'une évaluation clinique fine prend du temps, est coûteuse, comme le fait de multiplier les IRM… Dans le champ des pathologies neuro-dégénératives, nous commençons à avoir affaire à des bases de données d'imagerie très riches, alors que la caractérisation clinique et cognitive des situations est beaucoup plus sommaire. La technique d'accumulation des données est une chose, la caractérisation phénotypique des situations en est une autre.

L'imagerie est particulièrement intéressante sur le plan nosologique. En psychiatrie, on a depuis longtemps l'habitude de caractériser les syndromes au niveau phénotypiques. Originellement, les pathologies étaient décrites au niveau de leurs symptômes cliniques et non au niveau physiopathologique. De ce fait, il est concevable d'identifier des sous-groupes plus homogènes précisément sur la base de critères physiopathologiques. Logiquement, on doit disposer de thérapeutiques plus ciblées si l'on isole des sous-groupes spécifiques sur le plan étiologique. Ainsi, l'intérêt de l'imagerie apparaît évident, pour aller au-delà de la seule description clinique de syndromes.

Anne-Françoise Schmid Ne voit-on pas, dans bien des cas, la caractérisation des maladies basculer de la psychiatrie vers la neurologie avec les *big data* ?

Arnaud Cachia De manière schématique, on considère souvent que la différence entre une pathologie dite psychiatrique et une maladie neurologique réside dans la connaissance de son substrat cérébral, des mécanismes physiopathologiques impliqués.

Anne-Françoise Schmid La définition des syndromes tend à glisser d'un niveau de description à un autre.

« Dans le champ des pathologies neuro-dégénératives, nous commençons à avoir affaire à des bases de données d'imagerie très riches, alors que la caractérisation clinique et cognitive des situations est beaucoup plus sommaire. La technique d'accumulation des données est une chose, la caractérisation phénotypique des situations en est une autre. »

Arnaud Cachia De fait, les conséquences d'une atteinte cérébrale peuvent se décrire à différents niveaux. Tandis que la psychiatrie s'intéresse à l'ensemble du cerveau, la neurologie se focalise sur une région ou un mécanisme physiologique bien particulier.

Anne-Françoise Schmid Les deux savoirs entrant en jeu sont extrêmement différents.

Arnaud Cachia Notons que le contexte culturel joue en vue de déterminer les périmètres des disciplines. En France, la démence relève de la neurologie, tandis qu'elle relève de la psychiatrie en Allemagne.

Anne-Françoise Schmid Les problèmes de détermination de périmètres et de choix du niveau de description fondamental n'ont-ils pas été exacerbés ?

Arnaud Cachia Nous restons en présence de disciplines distinctes qui ne s'attachent pas aux mêmes atteintes et aux mêmes anomalies.

Anne-Françoise Schmid Il a été question de modèles intégratifs. Pourtant, les objets de la neurologie et ceux de la psychiatrie clinique semblent extrêmement difficiles à intégrer.

Arnaud Cachia Ne nous leurrons pas, la question de l'existence de modèles intégratifs est des plus ardues. Nous n'en avons pas encore qui soient opérants.

Jean-Charles Lambert On parle de syndrome pour la maladie d'Alzheimer, mettant en jeu différentes entités cliniques. Par ailleurs, en psychiatrie, on parle de bipolarité et, dans ce cas, la définition même du patient est très délicate. C'est d'autant plus vrai lorsque l'on veut définir des sous-entités psychiatriques. On peut rapprocher la maladie d'Alzheimer de la bipolarité sur le strict plan épistémologique de la difficulté de classification. Qu'apporterait le *big data* en pareil contexte si ce n'est ajouter de l'hétérogénéité à celle qui préexiste ?
A contrario, on évoquera le cas de figure des démences fronto-temporales qui ont la particularité d'être des syndromes où la génétique, associée à l'anatomopathologie et la biologie, a permis de définir clairement des sous-classes de pathologies. Nous ne sommes plus ici dans la configuration du syndrome multiforme, mais d'entités cliniques distinctes par rapport à des profils définis. Même si tout n'est pas clair dans de nombreuses circonstances, en principe une direction est définie en vue d'interpréter les données préalablement à leur collecte.

5 — Qu'est-ce que l'authentique multidisciplinarité ?

Muriel Mambrini-Doudet L'exposé interroge l'évolution des relations entre les disciplines. Le *big data* a beaucoup à voir avec la façon dont on organise la pensée et les questions que l'on se pose. Constate-t-on une sorte de réagencement, de redistribution de la pensée en imagerie ? Les disciplines se recombinent-elles pour aller de pair avec une nouvelle manière de voir stimulante ?

Arnaud Cachia Dans nos disciplines, nous devons composer avec la force de l'image. En effet, l'image attire l'attention. Elle a un grand pouvoir et exerce une profonde attraction sur les médias et le grand public. L'image aide-t-elle les chercheurs à mieux cerner leurs objets ? On doit apprendre à se méfier des images trop séduisantes. Le visuel ne fait pas tout. Rappelons que, tout comme la génétique, l'imagerie est interdisciplinaire. Nous avons besoin d'une multitude de compétences pour travailler efficacement. L'image n'est que l'aboutissement d'un processus d'exploration, de traitement, d'intégration et d'interprétation. On doit avoir à l'esprit ce caractère authentiquement interdisciplinaire de l'imagerie.

Muriel Mambrini-Doudet Dans ma perspective, l'interdisciplinarité peut être renforcée. Dans l'élaboration traditionnelle de la connaissance, chacun a l'habitude de considérer l'objet du point de vue de sa discipline, tout en échangeant avec les autres. Avec le *big data* et l'imagerie, on perçoit les véhicules de changement qui sont à l'œuvre. Des zones d'ombre semblent révélées dans ce qui perturbait auparavant les relations entre les disciplines de la connaissance.

Arnaud Cachia Le travail intégratif sera mené à bien en communiquant mieux. Il existe des obstacles techniques et épistémologiques. Ne le nions pas. Mais ne nions pas non plus la force des obstacles institutionnels. Il est notoire que la recherche est organisée par discipline et, partant, divisée en territoires. Certes, tous souhaitent davantage d'interdisciplinarité aux interfaces. Toujours est-il que dès qu'une discipline doit céder des crédits au bénéfice d'une autre ou d'un projet transversal, on sait bien quels mécanismes de défense entrent en jeu. Ne sous-estimons donc pas les blocages épistémologiques, mais gardons bien en tête la prégnance des blocages institutionnels.

Anne-Françoise Schmid Les commissions compétentes du CNRS ont bien analysé la question.

Jean-Charles Lambert L'imagerie est toutefois très intéressante pour illustrer les évolutions à l'œuvre sur le plan méthodologique.

Anne-Françoise Schmid Il existe un décalage entre le discours sur l'interdisciplinarité et sa pratique. Souvent, on affirme qu'il faut trouver un langage commun aux disciplines pour désigner la même chose. Or, le niveau pertinent est le travail concret de l'objet. Doit-on souhaiter des modèles intégratifs ou travailler sur cette matière si particulière qu'est l'hétérogénéité ? À bien écouter les exposés, il semble effectivement que l'hétérogénéité fasse partie intégrante de l'interdisciplinarité. Ne centrons pas le débat sur des difficultés institutionnelles bien connues depuis des décennies.

Arnaud Cachia Elles sont pourtant une réalité évidente.

Anne-Françoise Schmid On ne cesse de reproduire les mêmes schémas.

Arnaud Cachia Nous avons tous plus ou moins vécu ces blocages. Examinons l'exemple du programme « Imageries du vivant ». Près de deux années ont été nécessaires pour comprendre ce que chacun faisait. Souvent, tous ne mettent pas les mêmes mots derrière un même concept pour l'expliciter.

Jean-Charles Lambert J'ai eu la chance de participer à un projet pluridisciplinaire, dans le cadre d'un programme européen. En réalité, la pluridisciplinarité n'est autre qu'une démarche dans laquelle on accepte de ne pas tout maitriser, ce qui conduit à faire confiance à l'autre. Il faut accepter de se faire mutuellement confiance. Là est la clé. En tant que biochimiste, j'aurai pu perdre énormément de temps à vouloir devenir informaticien. Certes, j'ai consacré du temps afin de progresser en bioinformatique, mais très vite il est apparu patent que je ne serais jamais bioinformaticien de métier. Par conséquent, j'ai dû faire confiance à un collaborateur. Il ne sert à rien de vouloir communiquer au même niveau scientifique avec ses partenaires en permanence. On avance bien trop lentement.

« Le *big data* a beaucoup à voir avec la façon dont on organise la pensée et les questions que l'on se pose. »

Arnaud Cachia En effet, la confiance est primordiale. Chacun doit toutefois comprendre le questionnement de l'autre.

Anne-Françoise Schmid Le comprend-on ou bien croit-on le comprendre ?

Hermann Nabi On ne saurait amener tout le monde au même niveau de technicité dans l'ensemble des disciplines. Cependant, chacun doit nécessairement réfléchir au *big data* et à ses implications, qu'il soit généticien, statisticien, spécialiste de l'imagerie, etc. Au-delà de ce que chacun sait, ce sont les enjeux sociaux et éthiques de ses recherches qui impliquent de recourir à un langage commun.

Jean-Charles Lambert Intéressons-nous également à la temporalité dans laquelle s'inscrivent les collaborations. On ne peut pas expliciter les tenants et aboutissants d'un projet de recherche en quelques réunions. Dans un contexte d'industrialisation, on doit laisser le temps aux équipes d'apprendre à travailler ensemble. Désormais, nous disposons d'outils capables de générer exponentiellement des données. Tous les goulots d'étranglement limitant cette production vont progressivement sauter. Intéressons-nous au savoir-faire individuel et aux collaborations humaines sur le long terme, pour ne pas nous limiter à une société de l'industrialisation de la connaissance.

«Le niveau pertinent est le travail concret de l'objet. Doit-on souhaiter des modèles intégratifs ou travailler sur cette matière si particulière qu'est l'hétérogénéité ?»

6 — Sommes nous capables de nous appuyer sur le *big data* au profit d'une politique d'anticipation responsable ?

Emmanuel Hirsch L'enjeu de la lisibilité est central. Tant de choses se défont aujourd'hui pour se refaire le lendemain. Ayons le courage d'affirmer des considérations et des exigences d'ordre politique. Au nom de quelles valeurs et pour viser quelles fins se mobilise-t-on ? Tient-on suffisamment compte de la personne malade, de ses proches et des principes qui conditionnent les possibilités du vivre ensemble ? Les professionnels du soin sont aux prises au quotidien avec des éléments d'évidence, avec des responsabilités immédiates que les nouvelles perspectives évoquées en terme de cumul des connaissances semblent davantage complexifier que soutenir du point de vue d'enjeux concrets. Il importerait donc de penser les phénomènes et les évolutions liés au cumul de données à explorer et à exploiter au service d'une fin à strictement encadrer, en termes de montée en puissance de nos responsabilités. S'il est question de performances, en termes de finalités, de quelle manière accompagner les mutations qu'elles provoquent et ne pas déroger, au nom de finalités séduisantes et encore approximatives, aux principes mêmes de la vie démocratique ?

Paul-Loup Weil-Dubuc Je souhaiterais revenir à une dimension de votre exposé sur l'identification des sujets à risques. En l'absence de syndrome visible, l'imagerie peut suggérer des diagnostics et conduire les individus à s'approprier une maladie qu'ils n'expriment pas encore. Que sait-on, au plan psychologique, de cette identification précoce à la maladie ? N'induit-elle pas l'expression prématurée des symptômes ou l'accélération du développement de la maladie ?

Arnaud Cachia Le sujet à risque n'est pas encore un patient. Effectivement, en psychiatrie, poser un diagnostic trop vite sur un sujet à risque peut modifier la trajectoire clinique du sujet. Le fait de détenir l'information est porteur de stress. Or, il est notoire que le stress est un facteur de développement des troubles psy-

chotiques. Il n'est pas surprenant de voir les cliniciens opter dans les phases précoces de la maladie pour un diagnostic non tranché laissant la place à une part d'incertitude.
Sur le plan de la recherche, on ne sait pas bien quelle attitude adopter en présence de sujets à risques. Doit-on intervenir précocement ou pas ? En tout état de cause, la question pertinente est : a-t-on une intervention à proposer ou non ? En psychiatrie, le recours très précoce aux antipsychotiques n'a pas donné les effets attendus. La gestion du stress en psychothérapie précoce est en revanche envisageable. Enfin, il existe des cas de figure comme dans la maladie de Huntington où on ne peut rien faire.

Muriel Mambrini-Doudet Dans l'exposé, la difficulté du passage d'un savoir relatif à une population à un savoir portant sur un individu a été pointée, de même que le décalage entre l'acquisition et l'analyse des données. Qu'est-ce que la précision aujourd'hui, dans un univers caractérisé, d'une part, par la multiplication des données et, d'autre part, par la persistance de l'incertain ?

Arnaud Cachia La première question incontournable consiste à se demander si les données sont précises. Ceci revient à interroger leur qualité. Ensuite, la précision interroge l'attitude qu'il faut avoir par rapport au risque et, notamment, par rapport au patient à risque. Le médecin a affaire à des probabilités. Quant à l'individu, il attend qu'on lui dise s'il est malade ou non. Comment imaginer un clinicien dire à un patient qu'il a une probabilité d'exprimer la schizophrénie de 12 % dans les années à venir ?

« Le médecin a affaire à des probabilités. Quant à l'individu, il attend qu'on lui dise s'il est malade ou non. Comment imaginer un clinicien dire à un patient qu'il a une probabilité d'exprimer la schizophrénie de 12 % dans les années à venir ? »

7 — Le droit de savoir, l'incertitude et la vérité

Emmanuel Hirsch Les représentations de la maladie ont un rapport évident avec l'image, voire l'estime de soi. Il importe d'être attentif à ce que clichés d'imagerie révèlent et à ce qu'ils d'objectivent de la maladie. Cette médiation du virtuel pour mettre en évidence ce qui ne se voit pas mais se perçoit de différentes façons, justifie une démarche clinique attentive à la complexité de la transmission d'un savoir, de son explication et de son incorporation par la personne malades. Les évolutions technologiques, on le constate, justifient des modalités d'accompagnement, des commentaires et des attentions qui parfois font défaut. À cet égard, les sciences humaines et sociales peuvent être sollicitées pout étayer des élaborations nécessaires.

Arnaud Cachia Sur ce plan, les retours des patients souffrant de troubles psychiatriques contemplant des images de leur cerveau objectivant des anomalies sont très positifs. Le fait de leur montrer une déviance structurelle ou fonctionnelle par rapport à la normale est très bien perçu pour deux raisons :

– manifestement, il existe un dérèglement objectif, qui n'est donc pas le seul fruit de la subjectivité du malade — ayant de surcroît des difficultés d'introspection ;
– la prise de distance qu'offre le cliché est précieuse car il aide à externaliser la pathologie.
Très simplement, on dira que l'on est en mesure, au moyen de clichés, de montrer « une vraie maladie » à une personne en difficulté car sa capacité d'introspection est altérée. À ma grande surprise, j'ai constaté que le fait d'illustrer la pathologie est très bien perçu par les malades, leurs entourages et les associations. De fait, les personnes concernées au premier chef et leurs proches ont quelque chose de tangible sur lequel s'appuyer.

Emmanuel Hirsch Sur le plan juridique, le *big data* n'interroge-t-il pas le problème de la responsabilité ? On tend à démultiplier les mécanismes d'assistance. Ne dira-t-on pas un jour : « on savait, des données étaient disponibles et on n'a rien fait » ? Le fait de disposer de données, d'un tableau de la réalité cohérent, oblige-t-il à intervenir ? Ce questionnement est en rapport

avec ce que l'on a coutume de désigner comme le droit à l'information du malade.

Arnaud Cachia Dans de très nombreux cas (sauf tumeur par exemple), l'imagerie cérébrale ne met en évidence qu'un risque. On peut parler de déviances par rapport à une normale. Souvent, l'imagerie a donc un statut très incertain par rapport à celui d'une séquence d'ADN, donnée une fois pour toutes. Le clinicien a affaire à des anomalies plus ou moins manifestes et plus souvent silencieuses.
Dans l'état actuel des choses, une question se pose immédiatement. Le *big data* est de plus en plus onéreux du point de vue du coût de stockage des données. Qui va le financer et au nom de quels principes ?

Jean-Charles Lambert Il existe une compétition des différents domaines de recherche entre eux pour l'accès aux sources de financement. Plus on souhaite tendre à une connaissance exhaustive via le *big data*, plus les coûts augmentent. Immanquablement, une limite est atteinte à partir de laquelle il n'est plus rationnel d'investir davantage. L'information supplémentaire générée devient de moins en moins essentielle. Bon nombre d'arbitrages consisteront à déterminer s'il est judicieux, ou non, de dépenser plus pour accroître le périmètre de données à investiguer.

Pauline Lachapelle Lorsqu'il est question de diffusion de la connaissance, on doit partir de l'espace d'échange dont on se sert et non de la connaissance elle-même. Ce sont d'espaces de démocratie dont nous avons besoin, ou d'instanciation de la démocratie. Certes, prendre part à des discussions techniques peut réclamer une certaine formation, mais le modèle des conférences de consensus de citoyens est très bon et nous devrions y penser davantage.
Les enjeux sociétaux autour de la connaissance sont complexes. Nous songeons à la dimension sociétale de la recherche et de l'innovation. Toutefois, les comités d'investissement sont orientés par des intérêts économiques et industriels. Les modes de gouvernance classiques prévalent. Les espaces démocratiques sont plus que jamais à promouvoir et on peut se demander légitimement qui est le mieux à même d'en défendre le principe. Quantité de débats sont complexes (à titre d'exemple, on pourrait citer celui relatif aux nanotechnologies).

Jean-Charles Lambert L'univers de la connaissance scientifique et des publications scientifiques a lui-même basculé dans le *big data*. Cette multiplicité ne sert nullement la science. On peut fort bien mobiliser le *big data* pour faire valoir des hypothèses fausses. Des études erronées des années 80 ont été brandies par les détracteurs de la trithérapie pour la disqualifier. En l'occurrence, ils s'étaient servis de publications obscures s'appuyant sur des données parcellaires sans aucun recul scientifique, contestant le lien entre sida et infection par le VIH.
Nous aurions tout à fait besoin d'espaces régulateurs de l'information scientifique, dans un cadre démocratique, afin de dépasser les expertises.

Emmanuel Hirsch Il paraît important, à travers nos échanges, d'identifier des champs qui justifient des approfondissements du point de vue notamment des conséquences politiques d'évolutions dont on cerne encore mal les impacts à tant de niveaux également autres que ceux de la biomédecine dont nous traitons. Ou nous sommes partie prenante sans réserve d'une dynamique qui nous mène vers des champs inédits de connaissances et de possibles qui nous satisfont en tant que tel, et dont nous renonçons à évaluer la recevabilité en ce qui concerne leurs impacts péjoratifs, notamment s'agissant des principes de la vie démocratique. Ou nous estimons indispensable de produire un appareil critique et une veille susceptibles d'éclairer et d'accompagner, tant que cela est encore possible, les processus décisionnels afin de préserver des exigences à réaffirmer. Dans cette seconde hypothèse, il conviendrait de ne pas s'en remettre qu'aux seules instances ayant mission d'intervenir d'un point de vue formel ou du fait de leurs compétences dans les domaines juridiques, voire éthiques. Car à ce jour leurs positions semblent assez peu explicites et engagées au regard d'enjeux et de risques qui se précisent chaque jour davantage. Il semblerait justifié d'être inventif de dispositifs innovants intervenant au plus près des chercheurs impliqués dans ces domaines, et de nature à créer les médiations nécessaires avec ceux qui dans la société doivent être à la fois informés et associés à des choix arbitrés selon des règles démocratiques, notamment de justice et de transparence.

«Ce sont d'espaces de démocratie dont nous avons besoin, ou d'instanciation de la démocratie. Certes, prendre part à des discussions techniques peut réclamer une certaine formation, mais le modèle des conférences de consensus de citoyens est très bon et nous devrions y penser davantage.»

③

Surveillance, participation, finalités. Les données massives exigent-elles de repenser l'éthique de la recherche ?

1 — Introduction

Nicolas Lechopier

Je me suis intéressé à l'éthique de la recherche et, notamment, à l'utilisation des données personnelles dans la recherche épidémiologique. La question de la bonne utilisation des données est en fait déjà ancienne. Je l'ai appréhendée sous les angles de l'éthique et de l'épistémologie. Aujourd'hui, je participe à un groupe de travail sur l'éthique de la recherche en informatique qui réinterroge l'enjeu des données de recherche.

Je voudrais commencer par rappeler que le terme de « *big data* » autour duquel nous discutons est archi-contemporain. On le retrouve dans la presse, les revues spécialisées, la communication institutionnelle, bref dans de multiples milieux. On l'a dit, le projet de loi sur la santé faisait mention de l'open data plutôt que du *big data*. Le projet de loi sur la surveillance assigne au *big data* la tâche de prédire et détecter les comportements terroristes. Les traders y ont recours pour analyser les flux financiers. Ajoutons que pour nombre d'entreprises multinationales, le *big data* est devenu un modèle économique et une promesse de croissance. Les universités s'intéressent aussi au *big data* pour l'enseignement et la recherche. **Bref, le *big data* a cette dimension ubiquitaire, cette capacité à circuler d'un univers à l'autre. N'aurions-nous pas besoin de davantage de recul face à ces mots si contemporains ? A-t-on bien une idée claire des promesses qui se dessinent ? Ne sont-elles pas des mirages ? Sommes-nous sur un chemin bien tracé, dont nous connaîtrions l'issue, comme si le *big data* n'appelait rien d'autre qu'une voie unidirectionnelle ? Il importe de se déprendre de la force du présent, de la mode et de l'évidence.**

Sur le plan de l'éthique de la recherche, nous sommes quelque peu contraints d'aller au-delà du cadre classique de la déclaration d'Helsinki, centré sur le droit des patients et sur des enjeux de justice. **Dorénavant, il faut interroger le rapport entre les sciences et des valeurs, ainsi que la manière dont sont opérés des choix de recherche. La notion de responsabilité est incontournable.** Nous sommes à présent réunis car titulaires d'une expertise ou d'une fonction académique. Comment concevons-nous nos responsabilités individuelle et collective ?

Précédemment, il a été fait mention du problème du stockage des données, comme contrainte au déploiement du *big data*. Les capacités de stockage ne sont en effet pas indéfiniment extensibles. D'ailleurs, quelles sont les limites de *Google ?* Elles ne sont pas d'ordre logique ou commerciale, elles sont d'ordre énergétique car stocker consomme beaucoup d'énergie. Concrètement, les unités de stockage abritées dans des entrepôts doivent être sans cesse refroidies. Nous touchons là l'une des vraies limites de la révolution industrielle du numérique. À ce sujet, ne perdons pas de vue que toute révolution industrielle est synonyme de choc écologique. Rien ne peut s'étendre indéfiniment car nos sociétés dépendent de contraintes énergétiques.

Je voudrais également faire un lien avec l'ouvrage de Jean-Baptiste Fressoz, *l'Apocalypse joyeuse,* une histoire du risque industriel à travers laquelle il montre que le progrès procède par enfouissements successifs de savoirs. Ces derniers (les savoirs des artisans, ou encore tous ces « savoirs assujettis » dont parle Michel Foucault) sont, par temps de révolution industrielle, activement oubliés. Je pose donc la question : avec l'industrialisation de la recherche, prend-on conscience de ce qu'implique une « révolution industrielle » ? À ne voir que ce que l'on gagne (puissance, rapidité, hétérogénéité), nous occultons ce que l'on peut perdre (les savoir-faire artisanaux, la recherche artisanale, à échelle humaine).

Maintenant, considérons les recherches en épidémiologie. Les données personnelles manipulées par les épidémiologistes sont naturellement susceptibles d'intéresser des acteurs sociaux ayant des fonctions et buts différents (la police, les assureurs, etc.). Dans les années 80 et 90, la recherche épidémiologique se développe et la problématique du secret médical se trouve reposée à nouveau frais du fait du développement d'une collecte de données qui étaient jusqu'alors sanctuarisées ou — en théorie du moins — préservées. Cela dit, à l'époque, on put aboutir à un consensus sur la ou les finalités épidémiologiques à travers lesquelles ces données pouvaient être légitimement récupérées. Qu'en est-il aujourd'hui ? Je peux témoigner du fait que des chercheurs en informatique (notamment) capturent des données à très grande échelle, bien souvent sans développer le moindre questionnement éthique relatif à la finalité de la captation de ces données. Cela témoigne des mutations actuelles dans les modes de production des connaissances. Les recherches autour du *big data* sont faites avec des mathématiciens qui sont aussi technologues, qui évoluent dans un certain éclatement disciplinaire et qui construisent des outils. Comme l'a montré le sociologue des sciences Terry Schinn, dans le régime de production scientifique « instrumental », les objets technoscientifiques ont pour propriété de pouvoir concerner en même temps recherche fondamentale et applications industrielles. Ces objets (dont *big data* est un bon exemple) peuvent donc être investis en même temps pour des finalités multiples, voire contradictoires. Ce régime instrumental de la science rend difficile de canaliser ou d'apprivoiser ces instruments dans le cadre d'un projet collectif conscient et lucide.

« Rien ne peut s'étendre indéfiniment car nos sociétés dépendent de contraintes énergétiques. »

Je voudrais maintenant mentionner trois concepts posant problème : celui des données (2), celui du consentement ou de la participation (3), et celui du gouvernement par la surveillance (4).

2 — L'autonomisation des données

Nicolas Lechopier

Les chercheurs en informatique s'interrogent sur le sens de la notion de donnée à caractère personnel. L'engagement de la CNIL à partir de ce concept fait-il encore sens ? L'ensemble du dispositif Informatique et Libertés repose sur la donnée personnelle, mise en tension aujourd'hui. L'arrêté du 22 décembre 1981 sur l'enrichissement du vocabulaire informatique a défini la donnée comme la *« représentation d'une information sous une forme conventionnelle destinée à faciliter son traitement »*. Cette définition est-elle encore valide ? Les données personnelles font référence à un individu, à un être humain. En d'autres termes, elles sont rattachables à une personne, qui a un nom inscrit dans un registre d'État civil. Or, nous assistons à une reconfiguration du public et du privé qui vient questionner cette définition de la donnée. Traditionnellement, la sphère publique renvoyait à l'agora, à la rue et aux médias, tandis que la sphère privée était celle du domicile. Ainsi, le droit avait-il l'habitude de distinguer le public et le privé. Désormais, la distinction est brouillée. Les chercheurs en informatique voient des données circuler. Elles sont accessibles sans être pour autant publiques au sens juridique. Doit-on inventer d'autres concepts, au-delà de la distinction public/privé, pour rendre compte de la dissémination actuelle de données qui ne sont plus seulement individuelles ? Chacun sait que celles détenues par les réseaux sociaux sont relationnelles au sens où elles mettent en jeu les amis, les contacts, les relations d'un individu. Dans la sphère médicale, les données relatives au génome de quelqu'un impliquent ses affiliés génétiques (ascendants, descendants,…). Autant dire qu'elles ne sont plus individuelles, mais comportent une dimension collective.

La légitimité de la loi « Informatique et Libertés » réside dans sa mission de protection des individus vis-à-vis des pouvoirs (notamment de l'État) qui pourraient exploiter des données sensibles comme la vie sexuelle, les opinions politiques, l'appartenance syndicale, etc. Or, aujourd'hui, le *big data* participe d'un mouvement qui rend absurde le projet de vouloir définir la sensibilité des données en les classant selon la nature du domaine qu'elles représentent. La combinaison d'un certain nombre de données individuellement non sensibles peut aboutir à une information très « parlante », par laquelle un tiers par exemple serait capable d'acquérir une emprise sur l'individu en question. La combinaison de données non sensibles aboutit à générer des données sensibles, par exemple sur les plans juridique, politique ou économique. Nous voyons donc que la notion de donnée personnelle au sens où nous l'entendons habituellement est actuellement sous forte tension.

« Ainsi, le droit avait-il l'habitude de distinguer le public et le privé. Désormais, la distinction est brouillée. Les chercheurs en informatique voient des données circuler. Elles sont accessibles sans être pour autant publiques au sens juridique. Doit-on inventer d'autres concepts, au-delà de la distinction public/privé, pour rendre compte de la dissémination actuelle de données qui ne sont plus seulement individuelles ? »

3 — Le *big data* et la dilution du consentement

Nicolas Lechopier Nous avons évoqué l'enjeu du consentement lorsque les personnes prennent part à un programme de recherche dans le but d'accroître les connaissances de la communauté scientifique sur un problème. Quelle est la part d'autonomie du sujet dans l'univers du *big data*? Les chercheurs en informatique reconnaissent ne pas être en mesure d'entamer une discussion et de recueillir le consentement, à chaque opération de traitement, avec la foule d'individus concernés. Cette impossibilité pratique remet en cause le modèle classique du respect du choix de chacun à contrôler les données qui se rapportent à lui-même. Faut-il donc chercher à abaisser l'exigence, à rechercher un accord moins formalisé ou moins individuel? On pourrait convoquer ici la notion de non-opposition, qui existe dans la loi «Informatique et Libertés». Elle consiste à informer plus ou moins précisément les personnes concernées, par le biais de publications ou sur Internet par exemple, dans le but de leur dire qu'elles ont la possibilité de s'opposer à un usage de leurs données.

Mais, au-delà des modalités pratiques et plus fondamentalement, que signifie cette participation dans le contexte qui nous intéresse? Fait-on référence à un contrôle des individus sur leurs données et sur ce qu'ils communiquent sur eux-mêmes? Le but est-il en dernière instance de permettre aux intéressés de participer à l'interprétation des connaissances qu'ils ont contribué à produire? Les personnes dont on utilise les données dans un protocole de recherche peuvent-elles *ipso facto* bénéficier de ses résultats? Doit-on rester dans un modèle de participation «individuelle»? Dans le cas des biobanques, les personnes participantes pourraient s'affilier sélectivement à des associations pour aider à orienter l'usage de leurs données, en préférant que celles-ci servent dans le domaine des maladies infectieuses ou neuro-dégénératives plutôt que celui du cancer (par exemple). Mais cette affiliation sélective rencontre de nouveau l'écueil des finalités. Dans la logique *data driven*, les recherches dérivent de bases de données plutôt que d'une question directrice préalablement posée. La réflexion sur l'orientation des recherches devient aussi capitale que difficile.

«Quelle est la part d'autonomie du sujet dans l'univers du *big data*? Les chercheurs en informatique reconnaissent ne pas être en mesure d'entamer une discussion et de recueillir le consentement, à chaque opération de traitement, avec la foule d'individus concernés. Cette impossibilité pratique remet en cause le modèle classique du respect du choix de chacun à contrôler les données qui se rapportent à lui-même.»

4 — Les algorithmes comme gouvernement

Nicolas Lechopier Lorsque nous désignons le *big data*, nous ne faisons pas référence à un champ disciplinaire homogène. Nous nous trouvons dans un cas de figure où la construction d'outils a pris le pas sur toute autre logique qui lui aurait été préexistante. C'est le régime techno-instrumental d'une science passée à l'échelle industrielle. Or, comme chacun le sait, les outils peuvent être utilisés de façons diverses. Un même algo-

rithme peut être utilisé pour rendre service à un usager ou à des fins de contrôle et de surveillance par les pouvoirs.
Mais quelle société se dessine à travers la généralisation du *big data* ? Si surveiller, c'est faire faire quelque chose à autrui, surveiller c'est donc gouverner. Quel gouvernement pouvons-nous donc voir se préciser à travers le *big data* ?
Plusieurs formes de gouvernements se profilent. Tout d'abord, évoquons le gouvernement de soi-même. Chacun découvre son profil de risque et se forge son idée de ce à quoi il est exposé. Il se pense dans telle ou telle catégorie de risque et trouve (dans les cas de figure positifs) des ressources pour agir sur lui-même ou sur son environnement.
Le gouvernement, c'est aussi celui qu'exercent les puissances, qu'elles soient garantes de l'ordre public (l'État) ou cherchant des opportunités d'affaires. Ainsi, la police utilise les *big data* en prétendant prévenir les attentats. Le moteur de recherche Google a recours à ces mêmes algorithmes pour prédire des comportements d'achat. Or, ces puissances sont accrues par l'outil, en même temps qu'elles échappent au contrôle des citoyens, ce qui pose un sérieux problème démocratique.
Il faut enfin parler, avec Antoinette Rouvroy, du gouvernement algorithmique, c'est-à-dire de ce gouvernement qui n'est ni de soi sur soi-même, ni d'autrui sur soi-même. Disons qu'il s'agit d'un gouvernement « machinique » et aveugle, dans lequel le droit est réduit au fait. Pour comprendre ce point, considérons des entreprises comme Linkedin qui proposent des solutions de recrutement automatisées aux recruteurs, avec les ressources d'un réseau professionnel. Le but de cette entreprise est de prédire ce que sera le bon recrutement dans tel ou tel cas, et de s'adresser aux professionnels du secteur en leur proposant les profils les plus adaptés à des besoins déterminés. Ce service à des entreprises est certes digne d'intérêt, mais il pose un problème de taille. En effet, il consiste à recommander des actions futures en vertu des actions passées « réussies ». Ainsi, comme les recrutements passés et actuels sont *de facto* biaisés (phénomène de discrimination à l'embauche sur des bases raciales, de genre ou autre), tous les recrutements qui leur succéderont le seront également, sans que nous ne nous rendions compte de ce phénomène de reproduction. **Le *big data*, en tant qu'outil de prédiction, est donc aussi un outil de reproduction de l'existant et du déjà-donné, ou du déjà numérisé. Une société injuste le restera si on la laisse être gouvernée par des algorithmes qui partent de l'existant.** Nous avons donc une « responsabilité épistémique » de prendre la mesure de l'intérêt et des dangers de ces différentes formes de gouvernement, notamment d'une fétichisation des connaissances émergeant à partir des données prises comme un substitut du réel lui-même.

« Mais quelle société se dessine à travers la généralisation du *big data* ? Si surveiller, c'est faire faire quelque chose à autrui, surveiller c'est donc gouverner. Quel gouvernement pouvons-nous donc voir se préciser à travers le *big data* ? »

5 — Vers des formes collectives de consentement ?

Leo Coutellec Nous avons donc à nous interroger sur le consentement et sur les finalités en matière de recherche.

Hermann Nabi La discussion sur le consentement et sur l'usage de l'existant peut être illustrée avec l'exploitation scientifique des bases de données médico-administratives de l'assurance maladie. Quel type de consentement pourrait donner des millions de personnes directement concernées par l'exploitation de leurs données ? Manifestement, la conscience des risques potentiels existe, mais on escompte des bénéfices en termes de santé publique. Au nom d'une plus grande efficacité collective, on accepte un risque et on vit avec.

Yaël Hirsch Des éléments essentiels méritent d'être rappelés, même s'ils ne sont pas aisément applicables dans le contexte du *big data*. La loi

n°78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés (dite « loi Informatique et Libertés ») précise que sauf exceptions, l'individu doit avoir consenti de manière éclairé au traitement par un tiers de ses données à caractère personnel[1]. En d'autres termes, avant toute collecte, enregistrement, conservation, communication, destruction, etc., le responsable du traitement doit avoir donné à l'individu des informations sur le traitement et sur la personne qui est responsable de ce traitement avant de recueillir son consentement. Cela étant dit, le fait de souscrire aux conditions générales d'utilisation d'une plateforme Internet d'une dizaine de pages de façon quasi-automatique est difficilement assimilable à un véritable consentement éclairé. Toujours est-il que sur le plan des principes fondamentaux, le recueil du consentement des individus a bel et bien un sens et des conséquences juridiques.

La loi « Informatique et Libertés » prévoit des dispositions particulières pour les traitements de données à caractère personnel ayant pour fin la recherche dans le domaine de la santé. Pour ce qui est du consentement, l'individu concerné a uniquement le droit de s'opposer à ce que ses données à caractère personnel fassent l'objet de la levée du secret professionnel. Dans ce cas, aucun consentement express n'est nécessaire pour débuter le traitement des données à caractère personnel. Par exception, dans le cas où la recherche nécessite le recueil de prélèvement biologique identifié, un consentement éclairé et express est requis préalablement à la mise en œuvre du traitement. Il est donc possible d'en conclure qu'une position équilibrée a été mise en place pour préserver le consentement des individus sans entraver la recherche.

Il convient enfin de souligner que le fait qu'une donnée soit publique n'enlève rien à son caractère personnel. La distinction public/privé n'est nullement déterminante pour la qualification de donnée à caractère personnel. En d'autres termes, la protection de la loi s'étend aux données publiées, donc publiques. La disponibilité d'une donnée ne change rien au fait qu'en raison de son caractère personnel elle est protégée par la loi.

Paul-Olivier Gibert À propos du contexte de la loi de 1978, il faut préciser que l'on appréhendait l'information à travers quelque chose de matériel : la carte perforée. La terminologie de 1981 qui a été précédemment mentionnée n'est pas fausse, mais elle occulte un élément désormais capital. Alors qu'originellement la donnée était accessoire par rapport à son traitement, c'est l'inverse qui, aujourd'hui, est vrai. Le statut de la donnée a changé. Les individus génèrent des données en temps réel, ne serait-ce qu'en se déplaçant. Une dynamique de la coproduction de données entre individus et plateformes n'a émergé que récemment. Elle ne renvoie pas à la production classique par l'intermédiaire du travail salarié.

Dans ce contexte, c'est le caractère « personnel » de la donnée qui importe, lequel n'est nullement lié à la distinction public/privé. Le critère déterminant est de pouvoir remonter à la personne physique. C'est dans cette perspective que l'on a conçu le « droit à l'oubli ». Ainsi, une personne a fait valoir ce droit, en demandant qu'un article de presse espagnole faisant état d'une faillite frauduleuse passée soit enlevé. Observons que la requête est paradoxale car cette faillite était de notoriété publique.

Dans le domaine médical, la notion de consentement revêt une importance bien particulière. Il va de soi qu'un patient doit consentir au traitement qu'on lui prescrit. Cela dit, considérons la liste des effets secondaires répertoriés des médicaments. Qui lit la liste intégrale de ces effets lorsqu'il ouvre une boîte de comprimés ? Qui est conscient des effets indésirables de *l'ibuprofène*, par exemple ? Ceux-ci ne sont pourtant pas à négliger car parmi ces derniers on recense des accidents cardiaques. Pourquoi n'en tient-on pas compte ? La réponse à cette question réside dans la confiance que l'on accorde au médecin qui prescrit un traitement. **Aujourd'hui, curieusement, les Français font davantage confiance à *Facebook*, *Twitter* ou *Linkedin* qu'à l'administration. On pourrait du moins l'affirmer compte tenu de l'échec du dossier médical personnalisé.**

« Le *big data* s'inscrit également dans un contexte de mondialisation. Prêtons la plus grande attention à la pluralité culturelle. Notre perception de l'égalité et de la laïcité n'est pas transposable sur le plan international. »

Nicolas Lechopier **Il convient de souligner que le modèle du consentement individuel n'est tout simplement pas opérant. Nous avons affaire à une logique de consentement par délégation.** Par analogie, on songera au rôle d'un parlement. Les citoyens ne consentent pas directement aux changements législatifs, ils le font par l'intermédiaire de leurs représentants, les députés et les sénateurs. En l'occurrence, c'est aux parlementaires de décider des modalités de l'usage des données que manipulent les chercheurs. Précédemment, nous avons justement discuté de la construction de formes démocratiques adaptées à l'évolution des technologies. Comment éviter que le citoyen ne soit dépossédé de sa capacité de décision ? Avec les processus industriels, on perd en effet en démocratie. Si la démarche consiste à recueillir des données préalablement à toute autre opération, alors on préempte en quelque sorte l'avenir. De fait, on collecte des éléments sur la base desquels on devra se

prononcer *a posteriori*. En ce sens, on requiert de la population qu'elle entérine des états de fait.

Vincent Chouraki Considérons le mode de fonctionnement de *Twitter*. Il est bien évident que l'ensemble des propos publiés via *Twitter* relève de données publiques. Ceci figure du reste dans les conditions générales d'utilisation de la plateforme. Des réseaux sociaux fonctionnent de façon plus compartimentée, à l'instar de *Facebook* qui permet d'instaurer différents filtres et de distinguer des catégories de proches. Il demeure que la présence sur un réseau social demeure largement une expérience sans consentement. Le fait d'être sur *Gmail* implique de donner son consentement implicite à l'analyse de ses courriers électroniques.

Jean-Charles Lambert Le *big data* s'inscrit également dans un contexte de mondialisation. Prêtons la plus grande attention à la pluralité culturelle. Notre perception de l'égalité et de la laïcité n'est pas transposable sur le plan international. L'analyse des données et la perception de leurs modes de traitement dépendent de catégorisations culturelles. Par exemple, peut-être que la culture asiatique qui insiste beaucoup sur le collectif est moins interrogée par les tentions entre *big data*, consentement et responsabilité individuelle. On ne saurait oublier qu'il existe des prismes culturels en contexte mondialisé et que leur superposition donne parfois lieu à une forme de violence.

Leo Coutellec En substance, la conception classique du consentement n'est plus adaptée au contexte contemporain. Le cas du séquençage haut débit comme tests cliniques est typique de ce à quoi nous avons désormais affaire. Les individus consentent-ils à chaque test individuellement ou bien à une série d'examens ? Certes, nous avons de grandes difficultés à anticiper les implications futures des évolutions actuelles, à l'horizon de 5 ou 10 années. Si l'on ne peut pas accepter de ne plus tenir compte du consentement individuel, il s'avère toutefois nécessaire de le reconsidérer. Quelle serait l'alternative à la doctrine du consentement ? Lorsque l'on accepte que ses données soient utilisées, on agit d'après une sensibilité de ce que seront les conséquences de leur traitement ou de leur dévoilement. On peut évoquer une forme de sensibilité ou encore une imagination ou une prescience, mais aussi une forme d'éducation politique ou scientifique nécessaire pour (sur)vivre à l'époque des *big data*.
Lorsque nous considérons la surveillance, nous devons être attentifs au contexte dans lequel les données sont manipulées. Ceci repose la question de la finalité de la collecte et du traitement de ces données massives. Sur la base d'un questionnement sur le contexte, les valeurs et les finalités, nous pouvons jeter les bases d'un consentement renouvelé.

Pauline Lachapelle L'éducation du grand public – pas seulement du jeune public – au numérique et aux medias est essentielle. Sans doute assistons-nous à un double changement de logique et de langage. Il faut être en capacité de comprendre les caractéristiques de la donnée aujourd'hui, sans quoi le risque est grand de se trouver dépendants de ceux qui en ont la faculté. À l'évidence, nous sommes les témoins du développement de nouvelles formes de production des savoirs, dans une sorte d'économie circulaire animée par des grandes entreprises, des *start-ups* et les interactions entre individus. Sommes-nous capables de bien penser, partout, le phénomène ? Attention au vide conceptuel et au vide d'éducation quant au fonctionnement et aux enjeux du *big data.* Ne faisons pas usage de modes de pensée habituels, mais périmés, alors qu'une vaste diversité d'expériences n'est pas appréhendée.

«L'éducation du grand public – pas seulement du jeune public – au numérique et aux medias est essentielle. Sans doute assistons-nous à un double changement de logique et de langage. Il faut être en capacité de comprendre les caractéristiques de la donnée aujourd'hui, sans quoi le risque est grand de se trouver dépendants de ceux qui en ont la faculté.»

1 Les données à caractère personnel sont définies par ladite loi «Informatique et Libertés» comme «toutes information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres» (article 2 de la loi «Informatique et Libertés»).

6 — *Big Data* : intentions implicites et contrôle démocratique

Vincent Chouraki La forme de donnée que les individus produisent est une quasi-monnaie. Un adage devenu célèbre sur Internet dit en substance que si un service en ligne est gratuit, alors vous n'en êtes pas l'utilisateur mais le produit. Plus généralement se pose la question du rapport entre le service rendu par un service et l'utilisation «cachée» que fait ce service des données générées par l'utilisateur. Si on accepte d'être géolocalisé, on peut par exemple espérer une meilleure prédiction des embouteillages, mais également savoir si quelqu'un a participé à une manifestation. Si l'on est présent sur les réseaux sociaux, on conserve une forme de lien immédiat avec des proches éloignés géographiquement, mais on confie également à une entreprise privée des informations personnelles dont on ne sait pas toujours quelle utilisation elle pourra en faire ni à qui elle pourra les transmettre. Ainsi, existe-il toujours une contrepartie. Si un service est gratuit pour l'utilisateur, il est rémunéré d'une manière indirecte. Il est donc de la plus haute importance que les usagers soient conscients des contreparties qui entrent en jeu.
La problématique de la surveillance est incontournable. En France, la loi relative au renseignement a été défendue au motif qu'on n'allait pas surveiller davantage. Il s'agissait de légaliser des pratiques déjà bien vivaces. En d'autres termes, il fallait légaliser l'usage que les services secrets font des nouvelles technologies. Tel est du moins l'argument qui a été mobilisé. Lorsque les Etats-Unis ont adopté le *Patriot Act* à la suite des attentats du 11 septembre, le débat a été vif et continue de l'être, d'autant plus depuis les révélations d'Edward Snowden sur les dérives de la surveillance de masse. En France, force est de constater que le sujet n'a pas déchaîné les passions. Peut-on y voir une forme de consentement?

Nicolas Lechopier On le voit, les formes traditionnelles du consentement ne s'appliquent plus. Si l'on examine l'histoire de l'éthique de la recherche, on s'aperçoit que jusqu'à la fin du 19ème siècle, il n'existe aucune trace, nulle part, de demande de consentement de la personne. La recherche par la médecine du consentement du patient n'est que le fruit d'une évolution récente. Cette tendance a été consacrée en France par la loi du 20 décembre 1988 relative à la protection des personnes qui se prêtent à une recherche biomédicales, et par la loi du 4 mars 2002 relative aux droits des malades et à la qualité du système de santé. Durant des décennies on postulait qu'il n'y avait pas lieu de requérir l'avis des premiers intéressés car tout ce que la médecine accomplissait était par définition dans leur intérêt. Questionner le malade sur son assentiment au traitement ne pouvait qu'aboutir au pire, à entendre un «non» gênant. La médecine n'envisageait pas l'hétérogénéité des opinions face à ses interventions et une relativisation de sa légitimité.
Actuellement, nous sommes conduits à envisager des formes collectives, politiques, de consentement. La vision patrimoniale des données, assimilables à un capital ou une richesse, est bien trop étriquée pour rendre compte de ce qui se joue sous nos yeux.

Paul-Olivier Gibert Nous n'avons pas parlé de patrimonialisation du vivant à proprement parler.

Nicolas Lechopier L'idée de donnée-monnaie est populaire.

Paul-Olivier Gibert La donnée n'est pas un équivalent universel à proprement parler, comme l'est la monnaie. En revanche, intéressons-nous à des modèles émergents de la nouvelle économie. La donnée personnelle est sans doute conceptuellement moins importante que le *bitcoin*, qui correspond à un mécanisme d'échange de confiance sans connaissance individuelle de la personne avec laquelle on est en contact. Internet accélère considérablement les changements économiques. Il n'est pas évident que dans 5 à 10 ans *Google* soit toujours aussi dominant avec son modèle de captation des données d'interaction des individus avec la toile. Les navigateurs Internet peuvent bloquer la publicité. Les modèles fondés sur les annonces vont donc devoir muer. Ne soyons pas catégoriques sur l'état des lieux en présence sur Internet. Abstenons-nous d'affirmer de *Google* ce que l'on affirmait de *Microsoft* il y a 15 ans et *d'IBM* il y a 30 ans. Il se peut que le premier moteur de recherche mondial soit contourné par d'autres acteurs.

Nicolas Lechopier Les modèles économiques sur Internet ne sont pas au cœur de ce qui nous intéresse ici aujourd'hui.

Paul-Olivier Gibert Ils sont incontournables.

Nicolas Lechopier La vision patrimoniale des données est en lien avec une représentation politique et économique d'Internet qu'il nous faut questionner. La doctrine classique du consentement est proche en un sens du modèle libéral dans lequel un individu contractualise avec les autres. Or, actuellement, il est patent que le *big data* met en jeu des défis collectifs. Nous devons trouver des gouvernements collectifs qui ne se résument pas à l'acceptation des états de fait déterminés par l'évolution technologique. Dans cette perspective, notre approche du consentement est effectivement à renouveler.
En philosophie, nous sommes coutumiers de la distinction entre le fait et le droit. Nous débattons de grands enjeux

« Nous l'avons dit, les données ne sont pas neutres. Manifestement, le *big data* est révélateur de bon nombre de phénomènes et c'est un révélateur précieux. »

politiques autour de la liberté, de l'égalité, des droits fondamentaux, de la mondialisation. L'informatique ne permet pas de déterminer des idéaux ou des idées directrices. Elle décrit le réel, le divise, l'analyse et produit des relations. Elle ne nous dispense pas de choisir des institutions et des objectifs de vie commune. Le rôle du droit est de permettre d'assigner des normes à la société. Lorsque nous disons : « les hommes naissent libres et égaux en droit », nous invoquons justement le droit par-delà le fait. Le rôle régulateur du droit est tout simplement irremplaçable.

Quelle nouveauté est-elle susceptible de faire déjouer le cours d'un algorithme ? Peut-on vraiment parler, à ce sujet, de nouveauté ? La question de savoir si quelque chose d'émergent n'est pas déjà contenu dans le donné préexistant est après tout philosophique. Nous nous sommes interrogés sur le jeu social particulier qu'est *Linkedin* et sur la théorie des interactions sociales implicites dont il serait porteur. En pareil cas, n'est-ce pas le jeu social qui reproduit les mêmes schémas organisateurs qui biaise le cours des choses, plutôt que les données elles mêmes ? Nous l'avons dit, les données ne sont pas neutres. Manifestement, le *big data* est révélateur de bon nombre de phénomènes et c'est un révélateur précieux.

Muriel Mambrini-Doudet En effet il faut à cet égard distinguer donnée et *data*. Le *big data* est révélateur de notre lien aux données. Pourquoi postuler d'emblée que les données sont neutres ?

Nous sommes interpelés car nous recherchons des théories sous-jacentes. Comment les grands acteurs d'aujourd'hui se saisissent-ils du *big data* ? Il y a lieu de s'étonner sur leurs stratégies et de ne pas les recevoir comme évidentes.

A quelle théorie sous-jacente avons-nous affaire ? Celle-ci est mondiale. Il est nécessaire de clarifier l'intentionnalité qui est à l'œuvre. Où est l'intention ? Le *big data* opère comme révélateur de la manière dont nous appréhendons les données et imaginons les liens sociaux, les liens entre les structures et les infrastructures. Avec une posture de questionnement des stratégies et des intentions, nous quittons la prégnance du schéma techniciste ou technologiste car on se doit de prendre du recul.

7 — Solidarité et égalité des chances : vers une transformation par le numérique ?

Paul-Loup Weil-Dubuc Il a été souligné que les réseaux sociaux, et notamment professionnels, peuvent reproduire les inégalités et les injustices dans la société. La logique du *big data* pourrait être essentiellement conservatrice, au sens où elle favoriserait la préservation et le développement du capital biologique et culturel existant. Ce que met directement en jeu le *big data*, c'est précisément le concept d'égalité des chances ou d'égalité des opportunités, tel que l'a développé notamment le philosophe John Rawls. Par le *big data*, les opportunités qui s'offrent à un individu sont prédéterminées par des algorithmes de décision. Les données prétendent se substituer au mérite ou à l'effort individuel comme critère de distinction dans les trajectoires individuelles.

Certains font valoir, de façon plus optimiste, que le *big data* procède d'une logique de solidarité alternative au modèle de l'égalité des chances, en ce qu'il permet un partage des données supposément bénéfique à tous et à chacun. C'est notamment le discours de certaines compagnies d'assurance : le pouvoir prédictif des données rendrait possible un accompagnement médical personnalisé, plus proche des besoins des personnes. La « médecine personnalisée », aujourd'hui plus communément désignée en France comme « médecine de précision » ou « médecine stratifiée » en est une illustration parmi d'autres.

Ces deux discours nous semblent indiquer sur quelle ligne de crête nous devons nous tenir : comment promouvoir un usage des données qui soit compatible avec l'égalité des chances ?

Jean-Charles Lambert Au plan scientifique, les algorithmes peuvent aboutir à des tautologies. Notons que, par définition, le raisonnement par analogie s'épuise de lui-même. Un système fonctionnant en vase clos

s'auto-asphyxie. Ainsi, si les personnes recrutées ne sont que des clones de leurs prédécesseurs, il s'ensuit un appauvrissement spectaculaire du capital humain. À un moment ou à un autre interviendra une cassure. Cela étant dit, il n'est pas contestable que le *big data* favorise effectivement le raisonnement purement analogique.

Arnaud Cachia Nous avons longuement discuté des implications de certains algorithmes, comme ceux de *Linkedin*, qui seraient sensiblement reproducteurs des inégalités. Si le *big data* reste un outil, encore faut-il expliciter les modèles qui agissent derrière lui. Nous faisons signe vers un modèle théorique discret, dissimulé, insidieux. Pourtant, il existe bel et bien. Il est donc nécessaire, par-delà les données, de faire un effort afin de le déconstruire et de l'expliciter.

Henri-Corto Stoekle L'individu est-il systématiquement à opposer au bien commun ? Depuis plus d'une décennie, on promeut des thérapies ciblées contre le cancer. Des essais sont mis en place, auxquels les malades participent, donnant leur consentement. Force est de constater que ces personnes participent à un protocole de recherche pour un avenir meilleur. En ce sens, l'individu est bien au service du collectif. Nul n'est obligé d'accepter de prendre part à un essai clinique. Toujours est-il qu'un bénéfice est à la clé pour le malade inclus dans le protocole et pour autrui. Attention au discours sur l'autonomie de la personne qui pourrait être entendu comme un droit à l'égoïsme.

Natalie Hoog Labouret De quoi parle-t-on lorsque l'on parle de « consentement éclairé » ? Le Plan cancer met en avant la médecine de précision afin de mieux comprendre les mécanismes de réponse aux traitements et trouver de nouvelles options thérapeutiques. Cela peut être pris comme une promesse de résultats par les patients, sachant par ailleurs que les thérapies classiques ne sont pas obsolètes. Ce faisant, a-t-on bien éclairé les malades ?

Jean-Charles Lambert Nous parlons de consentement au sujet de personnes en échec thérapeutique. N'y a-t-il pas là un biais majeur ? On peut douter qu'une authentique alternative leur soit offerte.

Nicolas Lechopier Sur un plan éthique, quel modèle de solidarité en matière de soins préventifs et de soins curatifs allons-nous défendre ? Il existe comme un contrat implicite entre les patients et le système de soins public liant la gratuité des soins avec la possibilité d'expérimenter lorsqu'il le faut afin de faire progresser la médecine. Les mêmes méthodes sont naturellement transposables dans le privé, dans la mesure où on peut très bien, parfois, y déceler le même contrat implicite.
Nous évoquons des malades hospitalisés ou, en tout cas, astreints à un traitement lourd. Autant dire que ce sont des personnes en situation qui raisonnent différemment de celles qui considèrent leur cas de manière abstraite. Même si la société reconnaît le droit de dire non, qui va refuser de partager ses données médicales en situation de grande maladie ? C'est une question de principe et un acte de reconnaissance de certaines valeurs cardinales.
Le *big data* est-il en train d'affecter profondément le modèle d'égalité des chances tel que nous l'avons conçu ? Il dévoile en effet un public de malades potentiels, divisé en catégories de profils à risques. Comment maintenir l'assurance universelle face à pareille logique ? Telle est l'une des questions majeures qui est soulevée. Les thérapies standardisées ont un coût somme toute mesuré par rapport aux thérapies ciblées, plus onéreuses. Nous sommes donc confrontés au défi d'atteindre des objectifs collectifs de réduction des inégalités (définies par les plans de santé publique) tout en ne cessant d'accumuler des informations et de faire des choix de recherche qui creusent ces inégalités.

Hermann Nabi Il existe plusieurs stratégies de maîtrise des coûts. Par exemple, le NHS britannique accepte de financer un traitement à partir du moment où un gain minimal bien établi d'espérance de vie est attesté par l'utilisation des QALYS (quality-adjusted life years).En France, nous n'avons pas encore explicitement accepté ce principe.

Natalie Hoog Labouret Jusqu'à présent, nous avons toujours refusé ce critère de prise en charge en France. Pourquoi fixer arbitrairement un seuil d'amélioration de service médical rendu, par exemple à trois mois de vie supplémentaire ? La notion de progrès incrémental est aussi prise en compte. Par ailleurs, notons que les thérapies ciblées s'adressent à une population restreinte, ce qui est un argument économique. La question : « la médecine personnalisée relève-t-elle du marketing ou de la science ? » peut être formulée car il existe, notamment du côté des développeurs, un souhait de reconnaissance financier.
N'oublions pas cependant que du côté clinique les tests moléculaires préalables doivent éviter de prescrire inutilement une thérapie ciblée à un patient.

Hermann Nabi Reconnaissons que les thérapies ciblées ne concernent que relativement peu de patients. La grande majorité d'entre eux bénéficient encore des thérapies standards. Il y a sans doute un équilibre à trouver dans les arbitrages sur des les logiques thérapeutiques innovantes.

Paul Loup Weil-Dubuc Le rationnement des soins auquel il a été fait référence n'est pas forcément contraire au principe mutualiste, selon lequel tous doivent disposer d'un égal accès aux soins. Plus problématique est l'accès des soins conditionné à la possession ou à l'expression de gènes spécifiques. Nous allons être de plus en plus confrontés à cette forme de régulation de l'accès.

Vincent Chouraki C'est l'ensemble des comportements de souscription d'une assurance qui est remis en cause si les risques dont chacun est porteur sont objectivés.

Pauline Lachapelle L'égalité des chances met en jeu des rapports de pouvoir entre experts et profanes. Il ne s'agit pas de les négliger. Les *big data* questionnent nos protocoles de soin, mais qu'en est-il dans les modalités de production des diagnostics ? Nous restons dans le schéma classique où un médecin qualifié expertise les patients. On attend de ces derniers qu'ils consentent aux examens et aux traitements, mais ils sont aussi les témoins de leur propre vécu et détiennent une expertise de la maladie qui ne saurait pourtant être officiellement reconnue. Le savoir profane est intéressant à l'ère où le numérique a bouleversé les manières de rechercher et de produire du soin. Comment le *big data* va-t-il affecter cette transformation sociale ? Comment les expertises savantes et profanes vont-elle interagir ? Va-t-on vers davantage de co-construction des savoirs, au-delà du consentement à l'intervention experte de la médecine ?

« Le *big data* est-il en train d'affecter profondément le modèle d'égalité des chances tel que nous l'avons conçu ? Il dévoile en effet un public de malades potentiels, divisé en catégories de profils à risques. Comment maintenir l'assurance universelle face à pareille logique ? Telle est l'une des questions majeures qui est soulevée. »

Paul-Olivier Gibert Le consentement est sans doute à refonder avec davantage de co-activité. La notion n'est pas dépassée, elle est à redéfinir.

Leo Coutellec Nous pourrions dire que le consentement éclairé classique se mue dans une forme de consentement impliqué. Le *big data* est une technologie. C'est aussi une posture scientifique et une forme de méta-politique. Nous ne sommes pas en l'absence de théorie, dans un grand vide. Il existe des théories pour que la « data » se transforme en quelque chose d'effectif. Le grand récit auquel nous sommes confrontés est celui de la fin des grands récits, le recul du politique classique s'opérant au fur et à mesure que l'influence de la sagesse des foules grandit. Ce vaste mouvement de montée en puissance de la sagesse des foules et de l'hétérogénéité tend à bousculer les cadres classiques, jusqu'à en rendre certains obsolètes. Bien des enjeux éthiques et politiques sont à expliciter en pareil contexte.

Paul Olivier Gibert Le *big data* est effectivement l'un des aspects de « l'ère de la multitude ».

4

Les données massives en recherche clinique et l'utilisation du séquençage haut débit

1 — Plan cancer et *big data*

Natalie Hoog Labouret

L'institut national du cancer a été créé à l'occasion du premier Plan cancer. Le 3ème plan[1 2], qui a démarré en 2014[3], décline une feuille de route. Rappelons-en simplement trois objectifs majeurs :

— guérir plus de personnes malades ;
— préserver la continuité et la qualité de vie ;
— investir dans la prévention et la recherche.

Le Plan cancer a été conçu avec la participation non seulement des experts, mais encore des citoyens. Dans le cadre du plan, le *big data* est envisagé à différents niveaux, par exemple au sujet de la médecine non plus dite « personnalisée », mais « de précision » (ou moléculaire). Au cours des réunions de travail, il a été question d'anticipation à de multiples reprises. Insistons sur le fait que les objectifs du Plan ont été déterminés non pas par les seuls experts mais par les instances gouvernementales, lesquelles ont notamment sollicité les associations de patients.

Aux trois ambitions du plan s'ajoute la volonté d'optimiser le pilotage et les organisations de la lutte contre les cancers pour une meilleure efficience, en y associant pleinement les personnes malades et les usagers du système de santé.

Le plan est en rapport avec le *big data* à travers une série d'objectifs opérationnels. Le premier d'entre eux correspond à l'accélération de l'émergence de l'innovation au bénéfice des patients. Cette accélération passe notamment par des évolutions conceptuelles induites par l'émergence des thérapies ciblées. Ainsi, on a envisagé la stratification des essais sur de petites populations ou le développement d'essais incluant des malades atteints de tumeurs touchant différents organes, mais partageant les mêmes mécanismes physiopathologiques (essais cliniques transpathologies, programme AcSé). À ce propos, le programme AcSé (Accès sécurisé à des thérapies ciblées innovantes) est emblématique de la démarche qui consiste à étendre le spectre de pathologies auxquelles s'adressent des thérapies ciblées qui jouent sur un mécanisme moléculaire bien déterminé. Les nouvelles molécules arrivant sur le marché sont disponibles hors AMM (Autorisation de mise sur le marché) ou en amont par le biais d'ATU (Autorisation temporaire d'utilisation). Ces deux cadres ne permettent pas une prescription sécurisée et le recueil de données. Concrètement, si un patient se trouve dépourvu de ressource thérapeutique et si sa tumeur présente l'anomalie moléculaire ciblée par un nouveau médicament, alors il peut être fondé de lui prescrire ce médicament qui actionne cette anomalie. Si les firmes ne développent pas d'essais cliniques accessibles au patient, le programme AcSé permet, via des essais cliniques académiques, d'encadrer cette prescription dans un cadre sécurisé et de colliger les effets (positifs comme négatifs) dans un cadre sécurisé. Insistons sur le fait que, s'agissant de malades sans ressource thérapeutique validée, le problème du consentement ne se pose pas dans ses termes habituels. À l'ère du séquençage haut débit, nul doute que des anomalies vont être systématiquement dépistées. Actuellement, deux essais sont réalisés dans le cadre du programme AcSé. Les enfants pour lesquels la demande de recherche et de développement de médicaments est forte sont concernés par ce programme.

La médecine personnalisée (aujourd'hui « médecine de précision ») et son développement correspondent à l'objectif 6 du Plan. De façon très ambitieuse, il était prévu de mettre en œuvre, dès 2014, des essais cliniques incluant l'analyse de l'exome tumoral sur 3000 patients atteints de cancers du sein, du côlon, du poumon et sarcomes pour démontrer la faisabilité à grande échelle de ces approches et leur utilité dans la prise en charge des patients.

Les techniques de séquençage qui vont être mobilisées pour identifier les anomalies en rapport avec des processus tumoraux ne concernent pas seulement les personnes directement intéressées, mais également leurs familles.

L'action 6.4 du Plan (couvrant la période 2014-2019) envisage le séquençage à haut débit de l'ensemble des

« Si, dans l'esprit de nombreux patients, il faut bénéficier du "programme à la pointe", pour les cancers réfractaires, la nécessité de mieux comprendre les mécanismes afin de trouver des traitements efficaces passe par cette recherche moléculaire. En cancérologie, on ne peut faire de différence entre recherche et soin. Nous ne cessons donc d'accumuler des données. »

cancers d'ici l'année 2019. **Là encore, l'ambition est majeure et en lien direct avec le *big data*. En effet, nous souhaitons disposer de l'analyse de 10 000 tumeurs en 2015 et 60 000 en 2018.**
Si l'impression peut être donnée que l'accent a été mis de façon immodérée sur les thérapies ciblées, les patients et leurs familles ont finalement contribué très significativement à l'écriture du Plan cancer. En ce qui concerne la cancérologie pédiatrique par exemple, la chimiothérapie classique a accompli des progrès considérables, passant d'un taux de succès de 20 à 80 %, des années 1970 à aujourd'hui. La raison de ce progrès spectaculaire réside dans une meilleure association des molécules classiques. Si, dans l'esprit de nombreux patients, il faut bénéficier du « programme à la pointe », pour les cancers réfractaires, la nécessité de mieux comprendre les mécanismes afin de trouver des traitements efficaces passe par cette recherche moléculaire. En cancérologie, on ne peut faire de différence entre recherche et soin. Nous ne cessons donc d'accumuler des données.

1 http://www.social-sante.gouv.fr/IMG/pdf/2014-02-03_Plan_cancer.pdf
2 http://www.social-sante.gouv.fr/IMG/pdf/2014-02-03_Plan_cancer.pdf
3 http://www.social-sante.gouv.fr/IMG/pdf/2014-02-03_Plan_cancer.pdf

2 — L'INCA et la recherche dans un environnement changeant

Hermann Nabi Le Plan n'est pas complètement figé. Il évolue et, d'ailleurs, nous ne parlons plus de « médecine personnalisée », mais de « médecine de précision ». Le changement de sémantique n'est pas neutre. Il existe naturellement des défis techno-scientifiques à relever. L'action 6.6 du Plan est par exemple décrite de la façon suivante : « *développer de nouveaux modèles expérimentaux pour valider les données de génomique, développer de nouveaux marqueurs dérivés de la protéomique, tester le criblage de nouveaux médicaments et valoriser ces programmes*. » Bien entendu, il faut disposer d'équipes de recherche capables d'analyser les données et de les interpréter. Comme il l'a été précédemment souligné, on doit s'appuyer sur un dispositif de contrôle de qualité dans le but de standardiser les données. L'enjeu de l'homogénéité est essentiel, tout comme celui de l'identité du dépositaire des données. Les thèmes scientifiques et l'analyse statistique des bases de données sont régulièrement débattus. Les implications éthiques et réglementaires le sont beaucoup moins.
En cancérologie, le continuum entre les soins et la recherche est évident. Cela dit, quelle est la façon appropriée de s'adresser au patient ? Est-il un malade, un sujet de recherche ? Les attentes et les espoirs entrant en jeu ne sont pas identiques, suivant que l'on relève du périmètre du soin ou de celui de l'investigation scientifique. Ajoutons que la recherche ne trouve pas toujours ce qu'elle cherche et que des découvertes fortuites peuvent survenir. Faut-il nécessairement informer l'individu concerné au premier chef ? Aux Etats-Unis, on débat de l'opportunité d'informer les assurances des patients concernés. Ajoutons que la doctrine en matière de confidentialité et de sécurité des données est loin d'être clarifiée. À l'évidence, les résultats des recherches ne peuvent être pertinents que placés dans un contexte international. Quel cadre réglementaire mobiliser ? En France, le transfert de données vers l'étranger est autorisé, ce qui, naturellement, n'est pas le cas partout dans le monde.
La dimension sociale et organisationnelle des problèmes qui nous préoccupent n'est nullement subsidiaire. Dans un même pays, il existe des centres spécialisés à la pointe de l'évolution technologique et d'autres qui paraissent être en retard. Est-on pour autant en droit de parler de rupture de l'égalité dans l'accès aux soins ? L'accès des patients à un centre d'excellence est-il déterminant ? Quel est le degré d'appropriation des nouvelles technologies par les professionnels de santé ? Doit-on prévoir des formations spécifiques ? Comment intégrer les données aux pratiques cliniques ? Quels personnels seront impliqués dans la gestion des données et quand interviendront-ils ? La série des questions ouvertes est longue. L'INCA est au centre de transformations qui ont justifié des groupes de travail dédiés à la dimension humaine (éthique, réglementaire, économique…) de l'innovation dans les thérapeutiques ciblées. Naturellement, nous sommes également confrontés aux défis techniques relatifs au stockage, à l'analyse, au partage et à l'interprétation des données. Il n'est finalement pas surprenant de voir que le *big data* est quasiment le thème du meeting annuel de l'ASCO, à Chicago, en 2015[1].

Leo Coutellec Sur la problématique des données fortuites, il a été question de réunir un groupe de réflexion éthique. Qu'en est-il aujourd'hui ?

Hermann Nabi Pour l'instant, nous n'avons pas de réponse à cette question. Les découvertes fortuites ne sont pas sans soulever un problème juridique. Quels comportements adopteront les individus confrontés à des données fortuites révélées ? On évoque dorénavant « l'effet Angelina Jolie ». La question des com-

portements induits est centrale et, pour le moment, nous n'en sommes qu'au stade de la réflexion et avons désigné un groupe de travail. Des chercheurs interrogent ce champ des données fortuites. Quel encadrement et quelles recommandations proposer aux professionnels de santé ? Pour le moment, nous nous interrogeons.

« En cancérologie, le continuum entre les soins et la recherche est évident. Cela dit, quelle est la façon appropriée de s'adresser au patient ? Est-il un malade, un sujet de recherche ? »

Natalie Hoog Labouret Le Plan peut évoluer au fil du temps, en tenant compte de ce qui n'était pas prévu initialement.

Hermann Nabi La mission de l'INCA n'est pas de produire de la recherche, mais de la soutenir. Par conséquent, nous essayons d'identifier les problématiques sur lesquelles on a davantage besoin de connaissances. L'INCA doit aussi favoriser la dissémination de l'innovation, sans que soient générées des inégalités entre les territoires. **Souvent, la difficulté réside dans les dimensions éthique, réglementaire, sociale, organisationnelle et économique des problématiques.** Au cours des années, bien des changements ont affecté l'approche des thérapies (médecine personnalisée, *big data*…).

Muriel Mambrini-Doudet Beaucoup de parties prenantes ont été consultées au moment de l'élaboration du Plan, mais sans doute des difficultés d'ordre éthique ont-elles émergé après coup. Qui a été associé à la réflexion initiale ? Comment le futur a-t-il été imaginé ? Quelles capacités d'expérimentation ont été aménagées ? À titre personnel, j'ai été confrontée à la complexité dans mon champ d'investigation : l'agriculture, qui est au confluent d'une multitude d'enjeux. La complexité était telle que nous avons dû tenter des expériences locales, ciblées. Sans expérimentation, nous ne pouvons pas relever des défis complexes.

Natalie Hoog Labouret La démarche de questionnement a été initiée à partir d'un rapport de Jean-Paul Vernant portant sur les précédents Plans cancer.

Hermann Nabi En effet, la consultation de Jean-Paul Vernant a été le prétexte à l'analyse de plus de 3 000 contributions. Pour diverses raisons, les chercheurs en sciences humaines et sociales ont été moins impliqués et portés sur ces thématiques. **Les enjeux techniques et scientifiques du *big data* ont été anticipés (modèles, possibilités de stockages, etc.). En revanche, les dimensions éthique, réglementaire et sociale n'ont été que peu explorées dans les contributions, sans doute parce que la problématique à l'époque ne se posait pas encore en ces termes.** L'écriture du Plan a été dictée par un souci de cohérence, mais on n'est jamais suffisamment exhaustif.

Leo Coutellec Il existe une forme de contradiction à évoquer la médecine de précision ou la thérapie ciblée, alors qu'avec le *big data* on verse dans l'exhaustivité de la collecte de données. Cette collecte est du reste quelque peu aveugle et automatique, loin de la logique sélective du ciblage.

Natalie Hoog Labouret La médecine de précision aspire à étudier la dimension moléculaire des phénomènes. Les tumeurs sont très hétérogènes. De plus, on doit tenir compte de la variabilité interindividuelle, c'est-à-dire de l'ensemble des déterminants qui font la spécificité des individus (biologiques, environnementaux, sociaux). Le programme AcSé va au-delà des paramètres strictement moléculaires. C'est de l'ensemble des informations relatives à un individu dont on tient compte.

1 http://am.asco.org/

3 — La génétique et la juste appréciation de la hiérarchie des facteurs de risque

Hermann Nabi La sémantique va encore évoluer. On parle de « précision » dans une logique où l'on s'efforce de saisir un individu dans toute sa spécificité. C'est une question de finalité poursuivie. Les praticiens analysent plusieurs gènes, les interactions entre ces gènes souvent multiples et la relation gènes/environnement existe.

Danielle Geldwerth On peut parler d'éco-complexité.

Hermann Nabi Nous n'en sommes pas encore tout à fait là.

Leo Coutellec Le coût du séquençage à haut débit est il toujours aussi élevé ?

Natalie Hoog Labouret Le séquençage est de moins en moins onéreux. Surtout, le séquençage à haut débit opère sur de très nombreux paramètres, là où dans le passé on n'en explorait que quelques-uns. A l'époque où seulement quelques gènes étaient explorés à la fois, le séquençage était plus onéreux, mais plus facile également. Aujourd'hui, nous sommes immergés dans la complexité.

Hermann Nabi Les plateformes de séquençage disposent de budgets dédiés.

Vincent Chouraki Un partage de données sur un plan international est-il prévu ?

Hermann Nabi La France fait partir d'un consortium de plusieurs pays. La question du partage est abordée dans un chapitre dédié du Plan.

Pauline Lachapelle Explore-t-on les causes du cancer et le rôle de divers facteurs dans sa genèse ?

Hermann Nabi Le Plan comporte une dimension préventive importante, même si on peut avoir l'impression que l'accent est mis sur les thérapies curatives. L'approche préventive soulève de nombreuses questions, notamment sur l'identification de sujets « à risques » ou « potentiellement malades ».

Nicolas Lechopier **Le rapport entre le soin et la prévention est intéressant. N'accordons-nous pas trop d'importance à la génétique ? La décision d'Angelina Jolie, très médiatisée, à laquelle il a été précédemment fait référence était-elle motivée uniquement par des informations génétiques ?** Il est capital de bien positionner ce que l'on est en droit d'attendre de la génétique, notamment afin de ne pas surestimer son pouvoir prédictif. Celui-ci doit être perçu à sa juste valeur. Par exemple, la génétique ne permet pas de prédire de manière fiable la taille des individus. Les seuls bons prédicteurs ne sont pas des séquences de gènes, mais la taille des deux parents de l'enfant. Il serait dangereux de diffuser un déterminisme génétique qui n'a pas lieu d'être. Il convient donc plutôt d'insister sur le fait qu'être exposé aux polluants explique la genèse de certains cancers. Clarifions les rapports entre génétique et prévention. L'enjeu n'est pas seulement de recruter les bons patients et à temps, mais de déployer une véritable politique de prévention primaire. L'articulation de la prévention, de l'environnement, du comportemental et de la génétique est complexe.

« L'articulation de la prévention, de l'environnement, du comportemental et de la génétique est complexe. »

Hermann Nabi Le Plan ne fait pas l'impasse sur la prévention et envisage des investissements conséquents dans cette direction.

Nicolas Lechopier Est-il question du *big data* au sujet de la prévention ?

Hermann Nabi Il n'est pas désigné aussi explicitement que dans les chapitres du Plan relatifs au séquençage haut débit, mais il entre en jeu. Prenons l'exemple d'un facteur de risque modifiable de première importance : le tabac. On pourrait encore évoquer l'alcool ou l'exercice physique. Le Plan cancer n'a exclu aucune dimension du combat contre la maladie et englobe les préventions primaire, secondaire, tertiaire, les parcours de soins ainsi que l'après-cancer. Le périmètre retenu ne correspond donc pas aux seuls malades. Il comprend ceux qui ne sont pas encore entrés dans la maladie. Certes, pour le moment, les sciences humaines et sociales n'ont pas été suffisamment mobilisées et on doit faire davantage pour soutenir leurs recherches dans le domaine du cancer.

Vincent Chouraki Il existe quelques variants génétiques dont une mutation est associée à un risque extrêmement élevé de déclenchement de la maladie (BRCA1 pour le cancer du sein). Dans le cas médiatisé d'Angelina Jolie, rappelons que la génétique n'a pas été déterminante à elle seule. La mère de la célébrité est décédée d'un cancer ovarien à 56 ans et sa grand-mère est décédée à l'âge de 45 ans. C'est donc bien un contexte familial dramatique de décès successifs qui a dicté le choix d'une personne.

Dans l'état actuel des choses, les maladies chroniques ne sont pas explicables uniquement par des traits génétiques. Le cas de figure du gène BRCA1 n'est donc pas courant. Dans le contexte de la maladie d'Alzheimer, un variant a été découvert (l'apolipoprotéine E), tellement déterminant qu'il n'est pas besoin de segmenter davantage les populations. Il n'y a cependant guère d'intérêt à révéler la présence d'un variant défavorable car aucun traitement n'est disponible pour la maladie d'Alzheimer. On peut toutefois inclure des personnes à risques de développer des maladies neuro-dégénératives mais indemnes dans de longs essais thérapeutiques. Les lésions cérébrales commencent 15 années avant l'expression de la maladie, donc il existe une raison pertinente à évaluer l'impact de facteurs sur la durée.

Hermann Nabi Une chose est d'avoir un risque supérieur à la moyenne de développer une maladie, une autre correspond à la manière dont on présente le « surrisque ». La communication sur les facteurs de risques est capitale. Surtout, compte tenu de l'importance de paramètres environnementaux, il est de plus en plus accepté que le cancer ne constitue pas qu'une maladie des gènes. La réalité est autrement plus complexe que ce que révèle le seuil de dévoilement génétique.

Vincent Chouraki Dans de rares situations, ce sont les marqueurs génétiques qui sont déterminants.

Henri-Corto Stoeklé Il existe des déterminants environnementaux et des déterminants génétiques. Par exemple, à raison, on pointe le tabac comme un facteur de risque de déclenchement de cancers. Toutefois, des personnes fument et demeurent indemnes. L'explication est qu'il existe un large spectre de déterminants génétiques favorables ou non au développement d'une tumeur dans un tissu ou un organe d'un individu donné dans un environnement donné.

Vincent Chouraki Dans le cas de la maladie d'Alzheimer, les antécédents familiaux constituent le second facteur de risque en importance, le premier correspondant à l'âge. La composante génétique existe, mais vraisemblablement en interaction avec de nombreux facteurs environnementaux.

4 — L'appréhension collective des facteurs de risques et la hiérarchisation des priorités d'action

Leo Coutellec Nous avons créé des dispositifs de collecte des données. Avec les objets connectés, de nouvelles données vont être recueillies au domicile même des personnes. Bien des choix seront déterminés par nos nouvelles sources d'information, c'est pourquoi nous avons à mener une réflexion exigeante sur le tri de ce qui sera pris en considération, dans une pléthore d'éléments.

On peut estimer que le *big data* permettra d'élaborer de vastes bases de données sur les facteurs environnementaux. Il existe toutefois des dimensions de l'existence dans lesquelles on peut collecter plus aisément des données que dans d'autres. De surcroît, nous l'avons dit, certaines données ne sauraient être légitimement agrégées. Le choix des paramètres légitimes dans la prise de décision, par exemple thérapeutique, ne sera pas toujours évident.

Hermann Nabi Il arrive que l'on collecte des données en épidémiologie pour se rendre compte, *a posteriori*, que l'on a omis des variables essentielles. Le choix des paramètres dépend de la science du moment. Des projets relevant du *big data* sont soutenus par l'INCA, mais parvenir à une vision globale des problèmes demeure un défi.

Arnaud Cachia Il reste des situations dans lesquelles récolter certaines données est très malaisé. Parfois, un modèle théorique ou des *a priori* conduisent à s'intéresser sélectivement à des paramètres plutôt qu'à d'autres. Un important projet européen, *Imagen*, s'intéresse aux causes de l'addiction des jeunes[1]. Environ 2 000 adolescents ont été suivis longitudinalement pendant deux ans, dans l'espoir de trouver quelles sont les principales causes d'addiction. La méthode a pris en compte l'histoire du sujet, la génétique et des examens d'imagerie. Il est apparu que l'histoire du sujet est tellement prépondérante qu'il faudrait recruter davantage de sujets pour avoir une meilleure connaissance de paramètres qui n'expliquent qu'un faible pourcentage du phénomène qu'est l'addiction.

On peut alors questionner la pertinence de s'intéresser aux mécanismes cérébraux si la pathologie dépend en substance presque entièrement de l'histoire du sujet.

Hermann Nabi L'épidémiologie classique s'intéresse aux paramètres environnementaux. L'épidémiologie moléculaire s'attache aux paramètres moléculaires. Elles sont à combiner. Le débat s'est intéressé à l'interdisciplinarité et c'est vers elle qu'il faut tendre pour espérer avoir une vision globale des choses. Il existe plusieurs types d'expertises à convoquer pour être authentiquement généraliste.

Leo Coutellec Suffit-il de bien pondérer les différents facteurs de risque entrant en compte pour disposer d'une vision globale et d'une maîtrise de son objet ? En recherchant un plan global, ne sommes-nous pas en quête d'une illusion ? Considérons des pathologies dans lesquelles la génétique et l'imagerie constituent des moyens de compréhension. Pourtant, même associées, elles ne parviennent pas à rendre compte totalement de l'objet. Nous sommes aux prises avec une difficulté épistémologique ou anthropologique. Si nous renonçons à la compréhension complète de l'objet, alors nous évitons certaines erreurs.

Paul-Loup Weil Dubuc Le but poursuivi est-il véritablement de trouver une cause ou de faire une synthèse de l'objet ? Dans un contexte clinique, nous sommes plutôt à la recherche de traitements qui fonctionnent dans la majorité des cas. L'enjeu ne réside pas tant dans la relation cause/effet que dans la quête d'interventions qui sont statistiquement pertinentes.

Hermann Nabi D'un point de vue de santé publique, nous avons besoin d'établir des priorités. Sur quoi devons-nous mettre l'accent ? Quelles finalités méritent les plus importants efforts collectifs ? Nous avons à répondre à ces interrogations.

Anne-Françoise Schmid Longtemps, nous avons considéré sur le plan épistémologique les objets complexes avec l'idée que des perspectives disciplinaires différentes pouvaient converger. Cette vision n'est plus très pertinente.

Pauline Lachapelle Les études relatives aux facteurs de risques n'ont-elles pas des angles morts ? Leur focale n'est-elle pas restreinte en raison d'impératifs économiques ? On pourrait évoquer ici les pesticides et divers polluants. Les angles morts sont rarement innocents. Le plus souvent, ils sont liés à des choix économiques.

Hermann Nabi Il va de soi qu'il faut faire preuve de vigilance. De nombreuses études autour du diabète sont par exemple financées par l'industrie agro-alimentaire. Il faut donc tenir compte des influences.

Muriel Mambrini -Doudet Le Plan cancer, décidé au niveau politique, envisage l'expérimentation. C'est très heureux car l'expérimentation ciblée est finalement très économique. On ne saurait prendre des décisions hâtives à grande échelle. Il y a moyen de tester des hypothèses à moindre frais, dans un périmètre bien ciblé. Dans cette perspective, le *big data* est riche d'opportunités.

Paul-Loup Weil Dubuc Le savoir ou le dévoilement de données peut faire brutalement irruption dans la vie d'un individu, soit par la découverte inattendue de prédispositions, soit par information de la parentèle[2]. Il y a lieu de s'interroger sur la rupture que peut constituer cette irruption, singulièrement dans le cas des maladies neurologiques dégénératives, qui sont encore irréversibles, mais aussi dans le cas des cancers. **Le fait de savoir qu'on est atteint d'une maladie ouvre paradoxalement de nouveaux champs d'incertitude, pour les patients comme pour les entourages ; il induit sans doute une transformation dans le rapport à soi et aux autres ; dans le regard porté par l'autre sur soi.**

Hermann Nabi A ce sujet, il existe un projet de recherche porté par une équipe de Marseille et financé par l'INCA afin de déterminer l'incidence de la révélation de la présence du gène BRCA1 sur la santé mentale et les choix de vie des femmes.

Leo Coutellec La réception d'un savoir et la manière dont on se l'approprie sont essentielles. Sur ce point, le phénomène du *big data* complique les choses. Comment bien s'approprier les enseignements découlant d'une agrégation massive de données ? La qualité de ce que l'on communique dépend de la pertinence des modèles statistiques et de leur bonne utilisation, d'autant que les statistiques ne sont presque jamais univoques.

Hermann Nabi La médecine est constamment aux prises avec l'incertitude ou avec la définition de seuils arbitraires. Certes, les seuils peuvent évoluer. Toujours est-il que l'on a affaire à des intervalles de confiance confortables. Malgré tout, un intervalle de confiance de 95 % ne fait pas certitude.

Leo Coutellec Le problème est aussi celui de la mesure incertaine. Le choix d'un modèle statistique a des implications méthodologiques. La taille de l'échantillon est capitale et, en fonction des contextes, une donnée n'est plus vraiment une donnée. Pour reprendre une expression de Canguilhem, une « phénoméno-technique » de la donnée est à construire.

Hermann Nabi Considérons par exemple la manière dont les enquêtes qualitatives sont restituées. De fait, les restitutions changent du tout au tout en fonction des chercheurs. Chacun est influencé par ses inclinaisons propres et par son enracinement professionnel. Que l'on parle d'une enquête qualitative sur 50 sujets ou de *big data*, l'incertitude est toujours présente. On peut parler de *big data* horizontal ou de *big data* vertical selon qu'on dis-

pose de grandes quantités d'informations sur une vaste population ou sur quelques individus.

Arnaud Cachia Insistons sur le fait que les données sont le résultat de longs process et pipeline d'analyses. En imagerie, on ne voit pas le cerveau mais des « p value », c'est-à-dire des constructions. La donnée est elle-même le résultat d'analyses massives en quelque sorte. L'objet est un résultat et non une évidence.

« La médecine est constamment aux prises avec l'incertitude ou avec la définition de seuils arbitraires. Certes, les seuils peuvent évoluer. Toujours est-il que l'on a affaire à des intervalles de confiance confortables. Malgré tout, un intervalle de confiance de 95 % ne fait pas certitude. »

Muriel Mambrini-Doudet Quand on parle de risque ou de « p value », on interroge le problème de la précision. Mais la précision ne relève-t-elle pas d'une bonne orientation du regard de l'investigation plutôt que de l'application de techniques statistiques ?

Hermann Nabi Le fait qu'une association entre une exposition et un résultat de santé soit statistiquement significatif ne veut pas dire qu'il le soit aussi cliniquement. On peut très bien recourir à des chiffres fiables selon la méthodologie statistique, mais cliniquement sans valeur. La précision n'implique pas la pertinence clinique, loin s'en faut.

Yaël Hirsch Nous avons débattu de la qualité des données et de la nécessaire confiance à instaurer. Intéressons-nous à l'origine des données. Certes, elles peuvent provenir d'instruments et de technologies dédiés aux activités de santé. Par exemple, les données génétiques à la disposition des chercheurs, dont il était précédemment question, qu'ils produisent au moyen d'équipements conçus pour la recherche médicale. **Avec les objets connectés, ne verrons-nous pas de nouvelles sources de données émerger, qui n'auront pas un statut comparable aux données scientifiques, mais qui n'en seront pas moins disponibles et utilisables ?** On songera aux données de géolocalisation, de trafic, à celles produites par des capteurs de montres, etc. À quelles conditions jugera-t-on qu'une source est pertinente dans un contexte de santé publique ou de recherche biomédicale ?

Hermann Nabi Aux Etats-Unis, on expérimente actuellement un dispositif nommé « *cancer link* », dont le but est de collecter et rendre disponible des données structurées et non structurées récupérées un peu partout. Elles peuvent être en provenance du secteur privé, de l'éducation, etc. Une démarche consiste à croiser les sources de données afin de conférer une intelligibilité nouvelle à des phénomènes. Prenons l'exemple des données de l'Assurance Maladie. Elles sont agencées dans une optique comptable, mais on pourrait très bien chercher à les transformer en données exploitables. Dans le même ordre d'idées, les données économiques autour du PMSI hospitalier pourraient être mobilisées toujours dans le but de déterminer des corrélations. En d'autres termes, des données non médicales pourraient voir un sens médical leur être assigné. Le même problème ne manquera pas de se poser avec les objets connectés (*Pass Navigo* par exemple). Un collègue de l'INSERM tente d'étudier les déplacements des habitants de l'Ile-de-France pour les connecter à des risques ou à des pratiques d'activités physiques. Le *big data* permet de multiplier les associations à l'infini.

1 http://www.imagen-europe.com/

2 Décret n° 2013-527 du 20 juin 2013 relatif aux conditions de mise en œuvre de l'information de la parentèle dans le cadre d'un examen des caractéristiques génétiques à finalité médicale.

Propos conclusifs

Du *big data* au *big ficta* : un itinéraire scientifique, technique et social à partager — Léo Coutellec

De la possibilité d'un questionnement éthique partagé

Parmi toutes les initiatives et publications récentes qui traitent du thème des *big data*, et elles sont nombreuses, nous ne trouvons pas de réponse à cette question aussi fondamentale que difficile : sous quelles conditions un questionnement éthique partagé est-il possible dans le contexte des *big data ?* Car c'est bien d'abord sur la possibilité d'une éthique que doivent se concentrer nos efforts, la possibilité de questionner les valeurs, les finalités, le contexte et les conséquences d'un tel phénomène. La démarche entreprise avec ce workshop nous montre que la mise en œuvre d'une approche transversale, interdisciplinaire et participative est une première condition importante pour une expression éthique véritable. Néanmoins, mettre en présence des points de vue et partager des expertises ne peut suffire. Nous devons aussi assumer une posture, un certain engagement intellectuel pour le traitement d'une telle question.

Notre engagement a été celui de traiter conjointement les enjeux épistémologiques, éthiques et politiques du phénomène *big data*, dans la considération profonde de leur enchevêtrement. C'est une forme de *philosophie politique des sciences et des techniques des données* que nous avons expérimentée. L'hypothèse initiale a été de ne pas réduire ce phénomène, comme cela est trop souvent fait, à sa dimension technologique. Dans la continuité de Boyd & Crawford, nous avons considéré *big data* comme un enjeu à la fois culturel, scientifique et technique[1] afin d'interroger ses hypothèses, ses implicites et aussi ses illusions. Pour ces deux auteurs, ce que l'on appelle de façon assez pauvre *big data* repose sur l'interaction (i) de technologies pour maximiser la puissance de calcul et la précision algorithmique permettant de recueillir, analyser, relier et comparer de grands ensembles de données

1 Boyd & Crawford. Critical Questions for Big Data, *Information, Communication & Society*, 15 :5, pp. 662-679, 2012.

hétérogènes, (ii) d'une forme d'analyse scientifique qui travaille sur de grands ensembles de données hétérogènes pour identifier les tendances afin de faire des diagnostics économiques, sociaux, techniques et juridiques et (iii) d'une forme de mythe qui se caractérise par la croyance largement répandue que de grands ensembles de données offrent une forme supérieure de l'intelligence et de la connaissance qui peut générer des idées qui étaient auparavant impossibles, avec une prétention à la vérité, à l'objectivité et à l'exactitude. Sur ce troisième aspect, le phénomène *big data* s'inscrit pleinement dans la lignée du concept de « sagesse des foules »[3] ou « sagesse collective »[2] selon lequel la robustesse épistémique (d'une décision, d'un prédiction, …) est d'autant plus forte qu'elle est le résultat d'un processus collectif et d'une grande diversité cognitive.

Ainsi se dessine une deuxième condition pour la possibilité d'un questionnement éthique partagé : prendre le temps de définir l'objet de nos recherches, ne pas le réduire trop vite à l'une de ses caractéristiques et s'engager dans l'étude de ses déterminations tant épistémologiques (la façon dont les savoirs se constituent à son égard) que politiques (les valeurs et finalités que l'on projette sur lui). Finalement, à propos d'un tel phénomène, ce dont nous avons besoin c'est de penser l'éthique comme philosophie politique des sciences et construire une forme de « phénoménotechnique de la donnée ».

L'éthique des big data, une philosophie politique des sciences et techniques à l'ère des données

Bachelard nous invitait à étudier les phénomènes dans la précision de leur technicité et dans l'enchevêtrement qu'ils impliquent entre théorie et pratique, il en appelait à une « phénoménotechnique »[4]. C'est à ce type de défis que nous invitent aujourd'hui les *big data* où il nous faut, c'est une hypothèse, construire une *phénoménotechnique des données*. Parce que les données sont des créations matérielles, elles sont les productions d'une science plurielle invitant de multiples disciplines, théories et pratiques. La phénoménotechnique nous évite des illusions à propos des données. Elles ne sont plus des entités naturelles, passives, désincarnées. Les données ne nous sont pas données, elles ne sont pas des *data* ou, tout au moins, elles ne le sont plus à partir du moment où la science et la technique qui l'accompagnent y appliquent leur marque.

2 James Surowieski, *La sagesse des foules*, Paris, Jean-Claude Lattès, 2008.

3 Jon Elster, Hélène Landemore (dir.), *Collective Wisdom. Principles and Mechanisms,* Cambridge, Cambridge University Press, 2012.

4 Gaston Bachelard, *L'activité rationaliste de la physique contemporaine*, Paris, PUF (1965), pp. 9-10, 1951.

Considérer la donnée comme nous étant donnée, en cela qu'elle serait *data*, reste toutefois une possibilité qui relève d'un choix tout autant épistémologique que politique. Du point de vue de nos rapports aux savoirs (dimension épistémologique), c'est le choix de la neutralité des dispositifs techniques de collecte et d'analyse de ces données. C'est, pour nous, l'impossibilité de comprendre qu'à chaque étape de l'itinéraire technoscientifique de la donnée s'immiscent des valeurs et des choix qui guident, filtrent, classent, trient les données pour leur donner un sens pratique et une opérabilité scientifique. Considérer l'immanence de la donnée depuis les dispositifs techniques de collecte, c'est rester aveugle à l'interaction étroite entre technique et politique qui détermine ce qui est désormais devenu l'architecture numérique de nos existences[5].

Le choix de considérer la donnée comme *data*, plutôt que comme *ficta*[6], c'est en effet le choix politique de considérer les démarches *big data* comme agnostiques. Pour nous, c'est l'impossibilité de comprendre en quoi ces mutations scientifiques et techniques sont aussi des mutations culturelles et politiques. Nous l'avons vu avec ce workshop, le phénomène *big data* n'est que le nom d'un nouveau contexte de développement scientifique et technique où les enjeux socio-économiques prennent une place de plus en plus grande et où la donnée devient tout autant un objet de convoitise économique qu'une nouvelle espérance scientifique. Il y aurait donc matière à développer une posture critique et constructive à propos des *big data*, où l'éthique se confondrait alors à une philosophie politique des sciences et des techniques des données.

Pourtant, selon certains auteurs, le *big data* « n'est plus une hypothèse que l'on peut infléchir, mais une certitude à laquelle nous devons nous préparer »[7]. Le choix, s'il en reste un, ne peut alors se cantonner qu'aux « modèles d'émergences de ces techniques ». Inutile ainsi d'interroger les présupposés, les implicites, les valeurs, les finalités des nouvelles technologies des données, le seul espace réflexif que nous ayons est celui de l'anticipation de cette « révolution », car « cela va vite, de plus en plus vite ». Selon nous, une telle posture marque l'effacement de l'éthique au profit de la gestion (du risque, des bénéfices et des conséquences). Le phénomène *big data* n'est ni une hypothèse, ni une certitude, c'est un symptôme. C'est le symptôme d'une double difficulté : (i) d'abord, une difficulté chronique à accueillir au sein du

5 A ce propos, voir : Cédric Biagini. *L'emprise numérique : Comment internet et les nouvelles technologies ont colonisé nos vies*, L'échappée, 2007.

6 « Ficta », du latin *fictum* : fait ou façonné, autrement dit ce qui émerge d'une intention ; là où « data », du latin *datum*, est ce qui est donné, un présent immanent de l'existence.

7 Gilles Babinet, *Big data, penser l'homme et le monde autrement*, Le passeur, 2015, p.19.

paysage scientifique une nouvelle réalité autrement que dans le registre de la solution miracle ou celui de l'approche par « révolution scientifique », comme si la science ne pouvait pas se penser comme plurielle autrement que dans le registre du chaos[8] ; (ii) ensuite, une difficulté à mener une véritable réflexion éthique qui ne se réduise pas à une gestion des risques ou à une étude des conséquences mais qui puisse aussi mettre en lumière les valeurs et les finalités de ces démarches de recherche et de développement guidées par les données.

Notre workshop aura essayé, précisément sur ces difficultés, d'apporter quelques éclairages qu'il nous faudra approfondir collectivement par la suite. Parce que ce que l'on appelle le phénomène *big data* est aujourd'hui un enjeu partagé par toutes les disciplines scientifiques, par les praticiens comme par les chercheurs. C'est aussi une réalité quotidienne, là où l'avalanche de données se manifeste par des changements de pratique, des saturations technologiques ou de nouveaux dispositifs de surveillance, de traçage et de ciblage marketing. Au-delà d'un mot-valise difficile à circonscrire, d'une nouvelle mode ou promesse, nous avons vu qu'il y a un réel intérêt à appréhender ce phénomène d'un point de vue véritablement interdisciplinaire, en croisant les enjeux épistémologiques, éthiques et politiques. Au-delà de l'économie de la promesse, mais aussi au-delà d'une critique fermée qui condamnerait sans condition ce phénomène socio-techno-scientifique, une éthique de l'espérance et de la responsabilité reste à construire.

8 À ce propos, voir : Léo Coutellec. *La science au pluriel. Essai d'épistémologie pour des sciences impliquées*, collection Sciences en Questions, Editions QUAE, novembre 2015.

— Anticipation(s) : penser et agir avec le futur
Fondements et pratiques d'une éthique de l'anticipation
Séminaire de recherche interdisciplinaire d'éthique
Deuxième année / 2015-2016

En collaboration avec la Revue française d'éthique appliquée (RFEA)

Coordination scientifique du séminaire : Léo Coutellec, Sebastian Moser, Paul-Loup Weil-Dubuc & Emmanuel Hirsch [Pôle recherche – Espace éthique Île-de-France / Laboratoire d'excellence DISTALZ / EA 1610 « Études sur les sciences et les techniques », université Paris Sud]

L'Espace éthique Île-de-France – dans le cadre de ses missions au sein du Laboratoire d'excellence DISTALZ – est engagé dans une réflexion de fond, en lien étroit avec la pratique de recherche et de soin, sur la question des diagnostics précoces, notamment dans le cas de la maladie d'Alzheimer. Les questions relatives à la nature de ces savoirs d'anticipation et à leur impact éthique sur la personne et la société se posent, dans ce domaine comme dans d'autres, de manière pressante : cela nous engage à des approfondissements dans une perspective large et interdisciplinaire.

La première année du séminaire Anticipation(s) (2014-2015) – qui a rassemblé 16 intervenants, philosophe, sociologue, psychologue, juriste, médecin et praticien du soin – a fait émerger la nécessité de construire une éthique de l'anticipation là où la pluralité des conceptions de l'anticipation et les dispositifs techniques qui l'accompagnent nous mettent devant le défis majeur d'imaginer un futur souhaitable, vivable et commun.

C'est à cette éthique de l'anticipation que la deuxième année du séminaire Anticipation(s) devra sera consacrer. Une éthique de l'anticipation qui ne soit pas une éthique appliquée aux questions du futur mais u**ne éthique des temps présents qui permettent de mettre en partage des tensions essentielles sur notre conception du futur, sur notre conception des savoirs d'anticipation, sur le choix des valeurs qui doivent guider nos actions anticipatives.**

Une éthique de l'anticipation ne peut être ni une éthique sectorielle (appliquée à un champ particulier) ni une éthique purement théorique, elle devrait se construire au croissement des savoirs, des disciplines, des pratiques et des enjeux. Une éthique de l'anticipation ne peut faire l'économie d'une confrontation entre l'espérance – qui nous invite à l'utopie et à la construction de futurs souhaitables – et la responsabilité – qui nous enjoint à limiter nos champs d'action pour préserver ce qui est encore vivant et vivable dans ce monde.

Une approche de l'éthique de l'anticipation se dessine : dans une contexte où de fortes mutations anthropologiques, sociales et techniques accompagnent les processus d'anticipation, où le futur se dérobe devant toutes les incertitudes du présent, le rôle d'une éthique de l'anticipation serait d'identifier les conséquences, de clarifier le contexte, de mettre en lumière les valeurs et les finalités des démarches d'anticipation. Le défi est de taille puisqu'il s'agit de construire un espace de réflexion collective sur notre capacité à penser un avenir véritablement commun. Le séminaire 2015-2016 en sera une contribution que nous espérons significative.

Information et contact : leo.coutellec@u-psud.fr

Par extension du séminaire, nous proposerons cette année des **ateliers de création éthique** organisés en collaboration avec un designer dont le champ de recherche se construit autour du design spéculatif (démarche de "Design Fiction").

En complémentarité avec le séminaire, l'idée est d'ouvrir un espace de réflexion plus participatif et créatif, avec une méthodologie surprenante et innovante.
Pour cette première année, les ateliers mettront en discussion les enjeux éthiques propres aux maladies du plan MND (Maladies Neurologiques Dégénératives). Chaque atelier débutera par la présentation de scénarios et objets spéculatifs, illustrant une scène de vie de patient, aidant, soignant ou autre acteur impliqué par le sujet MND. L'objectif est de prendre appui sur ces images fortes, concrètes et f(r)ictionnelles pour stimuler la projection, l'empathie, l'implication et l'échange d'opinions parfois divergentes.
Deux raisons justifient le recours à cette démarche de "Design Fiction". Premièrement, dans la diversité des maladies regroupées dans le plan MND, nous avons à être créatif pour imaginer des communs qui puissent permettre de penser ensemble ces maladies sans effacer leurs spécificités. Le design fiction, comme outil d'anticipation, peut nous aider à ouvrir des possibles et à interroger certaines voies d'avenir. Deuxièmement, ces ateliers de création éthique, avec une méthodologie de design fiction, peuvent nous permettre de renouveler nos approches de l'éthique.
L'idée est ici d'interroger de façon très fine et profonde les différentes trajectoires possibles face à un même problème. Car faire du Design Fiction ce n'est pas créer des objets qui résolvent les problèmes de notre quotidien, mais des objets qui révèlent et explorent ceux de demain. L'enjeu est de co-construire un autre regard sur des questions parfois dépréciées, omniprésentes ou inédites.

Les objectifs sont les suivants :
- Créer une dynamique collective de co-construction autour du plan MND ;
- Permettre la participation de divers publics ;
- Susciter l'intérêt pour la démarche de réflexion éthique ;
- Optimiser la compréhension des enjeux ;
- Intégrer les résultats de ces ateliers aux réflexions en cours ;
- Permettre une autre approche de manipulation de concepts philosophique ;
- Défricher la notion de MND ;
- Confirmer la posture innovante de l'EREMAND.

Les ateliers dureront 2 heures et s'organiseront en trois tiers temps de durée similaire, comprenant : la présentation des projets de Design Fiction ; le débat ; la production d'une synthèse/cartographie des concepts évoqués.
Trois ateliers seront proposés en 2015-2016, chacun sur un thème différent et complémentaire :
- Moi malade (les apparences) : au yeux de soi et des autres, mon identité est faussée, recouverte d'une couche d'illusions (les symptômes invisibles ou incompris contrastent avec les symptômes manifestes qui constituent la façade publique). Ma différence est synonyme d'exclusion autant que de rareté.
- L'autre (la dépendance) : dans la relation aidant-patient, l'empathie est une notion centrale, permettant de se comprendre mutuellement ; de même pour la confiance en soi et en l'autre, elle permet au patient de "s'abandonner".
- Le temps (l'inéluctable) : la dégénérescence peut être perçue comme une chute inévitable, appelant à l'urgence de vivre, l'anticipation et la planification. D'autre part, le caractère évolutif de la maladie est soumis au hasard et à l'(in)fortune. L'annonce de ces diagnostics met la personne aux prises avec deux sentiments contradictoires : le déterminisme d'une maladie incurable et la libération d'avoir identifié la cause de ses maux ; elle met en résonnance l'impuissance des traitements et l'espoir de la recherche.

Programmation des 3 ateliers (16H00-18H00, le jour du séminaire)
• Octobre 2015, thème : Moi malade (les apparences)
• Novembre 2015, thème : L'autre (la dépendance)
• Décembre 2015, thème : Le temps (l'inéluctable)

— Revue française d'éthique appliquée

La *Revue française d'éthique appliquée* est une publication universitaire francophone à comité de lecture. Sa vocation est de contribuer à la valorisation et la diffusion de la réflexion et de la recherche en éthique appliquée.
Espace public d'analyse, d'approfondissements et d'échanges ouvert à la diversité des domaines de l'éthique appliquée et des approches disciplinaires[1], la RFEA souhaite également témoigner d'engagements concrets soucieux du bien commun.
La RFEA procède d'une démarche éthique «en acte» attentive à la fois aux expériences de terrain, aux innovations dans les pratiques et aux études académiques. À ces fins la RFEA entend couvrir transversalement quatre grands champs de l'éthique appliquée : l'éthique de la santé et du soin, l'éthique économique et sociale, l'éthique environnementale et animale et l'éthique des sciences et technologies.

La RFEA est une initiative du Département de recherche de l'Université Paris Sud et de l'Espace de réflexion éthique/Ile-de-France (ERER/IDF). La RFEA est libre d'accès et fonctionne sur le modèle de l'*open access* ; sa publication est semestrielle.

La RFEA a publié son premier numéro en ligne en février 2015.
Le dossier thématique est consacré aux «ambivalences contemporaines de la décision». Au-delà de la nécessité de repenser aujourd'hui la décision et ses processus, dans le double contexte de la démocratie et des transformations techniques contemporaines, ce thème de la décision nous semble emblématique d'un questionnement concernant tous les champs de réalité : au-delà de la santé, la décision politique, l'environnement, l'éducation, etc.

Pour y accéder : http://www.espace-ethique.org/revue

Directeur de la publication Emmanuel Hirsch
Rédacteur en chef Paul-Loup Weil-Dubuc
Rédacteurs en chef adjoints Pierre-Emmanuel Brugeron, Léo Coutellec
Comité éditorial Edgar Durand, Alexia Jolivet, Virginie Ponelle, Anne-Françoise Schmid, Thibaud Zuppinger
Conseil scientifique Elie Azria, Bernard Baertschi, Jean-Michel Besnier, Jean-Pierre Cléro, Hervé Chneiweiss, Sophie Crozier, Éric Fiat, Fabrice Gzil, Lyne Létourneau, Marc Lévèque, Muriel Mambrini, Emilio Mordini, Marie-Hélène Parizeau, Corine Pelluchon, Avner Pérez, Marie-Geneviève Pinsart

1 Notamment : sciences politiques, philosophie, sociologie, droit, économie, anthropologie, communication, histoire, etc.

Big data et pratiques biomédicales
— Implications éthiques et sociétales dans la recherche, les traitements et le soin

Depuis quelques années, l'émergence du phénomène *big data* se traduit par une « avalanche de données », une collecte systématique et massive de données et une croissance rapide des technologies de traitement. Il est possible de définir la démarche *big data* selon une dimension quantitative (basée sur le volume et le rythme de production des données, en constante croissance) et une dimension qualitative (données hétérogènes provenant de sources multiples). La dimension qualitative, moins souvent relevée, doit faire l'objet d'une attention particulière.

Appréhender *big data* seulement comme une révolution technologique – traitement majoritaire qui lui est actuellement fait – serait une réduction. Nous proposons de le comprendre comme un phénomène à la fois culturel, technologique et scientifique. Défis et enjeux auxquels, au-delà de la communauté des chercheurs et des praticiens, notre société est confrontée ne serait-ce que du fait des bouleversements que provoque cette mutation scientifique dans nombre de domaines qui touchent à nos représentations mais également à nos libertés fondamentales.

Ce numéro 2 des *Cahiers de l'Espace éthique* reprend les analyses proposées dans le cadre du workshop consacré le 16 avril 2015 aux « enjeux éthiques, scientifiques et sociaux du *big data* ». Il poursuit la réflexion développée par l'Espace dé réflexion éthique de la région Ile-de-France dans le cadre du laboratoire d'excellence DISTALZ, à propos de la maladie d'Alzheimer et plus largement des maladies neuro-dégénératives.

Sous la direction de **Emmanuel Hirsch, Léo Coutellec, Paul-Loup Weil-Dubuc.**

ESPACE NATIONAL DE RÉFLEXION ÉTHIQUE ET MALADIES NEURO-DÉGÉNÉRATIVES

ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS

ars Agence Régionale de Santé Île-de-France

UNIVERSITÉ PARIS SUD → DÉPARTEMENT DE RECHERCHE EN ÉTHIQUE

Espace de réflexion éthique Ile-de-France
CHU Saint-Louis – 75475 Paris cedex 10
www.espace-ethique.org