

Thématique : Enjeux imposés par l'évolution des pratiques scientifiques (données ouvertes, sciences participatives, publications de résultats négatifs)

**Organized Session – The genetic data values  
SPS 2021**

Titre : **Les valeurs des données génétiques**

Thématique générale (Résumé court)

En situant les sciences dans leur contexte social, politique et économique, la philosophie des sciences contemporaine dégage l'effectivité des valeurs épistémiques et contextuelles opérant dans le pilotage, les organisations et les pratiques scientifiques ordinaires. La génomique participe au projet de modernité thérapeutique d'État caractérisée par une montée en puissance de la médecine des preuves et de l'intelligence artificielle, une bioéconomie et une éthique procédurale, où les flux de données – *big data* – jouent un rôle structurant. Cette session analyse les processus complémentaires de valorisation et d'illusion objective des *données* génétiques : bien loin de constituer des données neutres et objectives au sens traditionnel des données scientifiques, les données génétiques résultent d'opérations souvent invisibles, qui leur confèrent leur *valeur* d'information et de prédiction. Données par les individus avec leur consentement éclairé, elles ne gagnent leur valeur informative et prédictive que par le travail des organisations (hôpitaux, instituts, biobanques, *biotechs*, consortium), et leur ancrage dans un contexte épidémiologique et clinique précis. Cette session analyse le travail de performance des données génétiques à partir de trois angles :

- Le travail de classification des données (primaires, incidentes, secondaires) et les conditions pour qu'elles acquièrent une signification de santé (préventive ou thérapeutique).
- Le travail (notamment clinique et statistique) de construction de la donnée pour qu'elle puisse valoir comme critère de décision.
- La valeur de connaissance des données génétiques.

- Argumentaire

Depuis les années 1970, la génétique a profondément renouvelé la manière de comprendre et pratiquer la médecine. Elle participe à une transformation du soin et de la prévention, mais aussi à une reconfiguration de l'organisation de la médecine (autour de Plateformes telles qu'Auragen ou Sequoia dans le cadre de Plan France génomique, impliquant des

collaborations entre instituts, hôpitaux, biotechs, consortiums, biobanques) justifiée au nom d'une modernité thérapeutique d'État caractérisée par une montée en puissance de la médecine des preuves et de l'intelligence artificielle, une bioéconomie et une éthique procédurale. Cette transformation ne résulte pas seulement d'un déploiement de connaissances biologiques nouvelles, mais aussi de la mise en œuvre de méthodes, structures et projets de recherche, où les flux de données – *big data* jouent un rôle structurant. Largement portée par la rhétorique de la promesse technoscientifique, la montée en puissance de la génomique tend néanmoins à rendre invisible le travail de valorisation des données génétiques pour mettre en lumière seulement leur fonction informative et prédictive en tant que données scientifiques et produire ainsi une illusion d'objectivité de l'information portée par la donnée. Ainsi l'OCDE définit les données scientifiques ou *research data* comme « des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider des résultats de recherche. » (Principes et lignes directrices pour l'accès aux données de la recherche financée sur fonds public. OCDE 2007 <http://www.oecd.org/fr/science/inno/38500823.pdf>) En étudiant le cas des données génétiques, cette session analyse et critique précisément le statut « factuel » et la fonction de « source » pour la recherche scientifique, en montrant comment elles sont construites.

En effet, la génétique humaine fabrique les données en conjuguant deux types d'opérations. D'une part, elle obtient ces données grâce au recueil de matériel biologique auprès des individus avec leur consentement éclairé, le séquençage à haut débit du génome, le stockage dans des banques de données, le traitement bio-informatique par analyses statistiques. D'autre part, elle cible et recherche ces données dans le cadre d'un travail épidémiologique et clinique contextualisé qui conduit à classifier les résultats selon trois types (données primaires, secondaires, accidentelles), qui seul permet de conférer à la donnée génétique sa valeur d'information et de critère de décision. Cette double opération s'effectue à travers une organisation complexe d'acteurs (patients, médecins, chercheurs, associations, instituts, hôpitaux, biotechs, consortiums, biobanques, Etat, Europe) qui alternent des régimes de gratuité et des logiques de capitalisation, en créant de l'anonymat, en garantissant la continuité et la massification des dons, en stockant, agrégeant, analysant, distribuant et recevant. Or l'invisibilisation des processus de construction des données génétiques crée un décalage entre les représentations citoyennes, la réalité du travail scientifique et les enjeux scientifiques embarqués par les données génétiques. Cette session propose de rendre visible ces opérations de valorisation qu'on nommera de façon générique « phénotéchnique » ou « performance » à travers trois angles d'analyse.

Speaker N°1 Catherine Bourgain INSERM CERMES 3 Villejuif

### **Beyond technology: the labor of building the value of genetic prediction**

The value of genetic data largely relies on the quality of the predictions that it enables. Frequent discourses on genetic insist on its capacity to produce solid and reliable predictions. Yet, whether it concerns disease, geographical origins or the identification of a criminal suspect, genetic predictions are associated with strong uncertainties, notably reflected by the importance of statistics in the production of genetic analysis results.

We argue here that the image of genomic technologies as producers of reliable predictions is not the consequence of sole technoscientific and commercial promises. It also results from a

significant amount of constant but largely invisible labor performed by the professionals involved in the making of these predictions. By successively exploring the evolution of professional practices in the fields of rare diseases and oncology, we describe a variety and evolving forms of professional labor. They range from the articulation of molecular, familial, clinical and epidemiological information, to the collective production of clinical decision associated with particular predictions, and rely on the central role of professional network to produce standards and databases. This labor conditions the production of genetic predictions for which the uncertainties are deemed manageable enough to make trustworthy clinical decisions. The process through which this professional labor and the role of other forms of knowledge, particularly clinical knowledge, are made invisible to external and non expert circles, is turning the reliability of predictions into a technological quality of the sole genetic tests. Henceforth, genetic tests can be taken out of their context of use, promoted and sold on markets, equipped with qualities that are not intrinsic to them, but co-produced by medical and scientific groups in the actual use of the tests.

Keywords: prediction, uncertainty, clinics

Speaker N°2 Marie Darrason – HCL – IRPHIL – Lyon

### **Next Generation Sequencing (NGS) techniques and the “information illusion”**

The development of next generation sequencing techniques (NGS) fostered the rise of Big data in genomics : it is increasingly easier and cheaper to get much more abundant and diverse genomic data. Indeed, not only is it possible to run a panel of specific genes, but also to run whole exome (WES) and whole genome sequencing (WGS).

These changes in the amount of available genomic data foster the “information illusion”: the idea that more genomic data amount to more information about the genomics of health and disease, upon which we may base our actions. The main point of my talk is, following Floridi, that information equals data plus significance and that one cannot have significance outside of context : a specific question asked to answer a specific problem for a specific individual. Therefore, more WES and WGS may amount to more uncertainties rather than more information.

The first part of our talk will be dedicated to describe what kind of data is retrieved from NGS techniques : primary data (in relation with the initial purpose of the test), but also additionnal data (data that are not related to the initial purpose of the test) including both incidental data (data that were unsolicited) and secondary data (data that are not related to the initial purpose but were actively sought). I will then review what kind of ethical issues are raised by these additionnal data and how these issues are linked to the “information illusion”.

The second part of this talk will be to describe the mechanisms of the “information illusion” : the illusion of reliability (data always depend on which technique and which resolution is used to produce it), the illusion of completeness (no type of sequencing technique may yield “the whole data”), the illusion of simplicity (interpreting genomic data requires to take into account complex interactions) and the illusion of agnosticism (interpreting genomic data without a clinical phenotypic context is very difficult).

Finally, I will sketch what kind of problem this “information illusion” raises about the concept of “free and informed consent” of patients whose genomic data are sequenced to answer a medical question, then harvested and collected in order to answer other medical questions or research purposes.

Keywords: "big data", "NGS techniques", "information"

Speaker N°3 Sarah Carvalho UFC – Logiques de l’agir - Besançon

### **Du don aux informations clinique et scientifique : phénoménotechnique et performance des données génétiques en valeurs**

Que donne le patient quand il donne directement un matériau biologique à partir duquel seront extraites et traitées « ses » données génétiques ? Pourquoi donne-t-il ? Le don à l’origine de la donnée génétique diffère des autres dons (organe, sang), dans la mesure où la valeur du don ne réside pas dans le matériau donné (pour qu’il soit perfusé, greffé, reproduit), mais dans l’information immatérielle qui en est extraite sous forme de données génétiques. La notion de données génétiques implique ainsi une première ambiguïté en ce qu’elle superpose, d’un côté, ce qui est donné directement et indirectement par le patient, et, de l’autre côté, son statut de donnée scientifique, au sens d’un enregistrement factuel, qui est agrégé, comparé, traité avec un ensemble d’autres données scientifiques dans le contexte de la recherche. Dans le droit européen, la notion de portabilité de la donnée vise précisément à tenir compte de cette ambiguïté en ce qu’elle autorise théoriquement de pouvoir retirer le don à tout moment, mais sa mise en œuvre pratique s’avère en fait limitée. Une deuxième ambiguïté concerne le donneur, en ce que l’information traitée ne le concerne pas seulement elle ou lui personnellement, mais implique un ensemble de personnes défini comme une population génomique, dont le consentement n’est pas requis. Cette deuxième ambiguïté justifie l’introduction dans le droit français d’une obligation compensatoire d’informer les apparentés adultes concernés par les risques ou les réflexions en cours sur le consentement communautaire. Enfin, une troisième ambiguïté concerne la dialectique entre don et valeurs de la donnée génétique : le patient vise essentiellement une valeur clinique, le généticien vise, outre la valeur clinique, une valeur de connaissance. La recherche a besoin d’accroître sans cesse et rapidement la masse des données pour affiner ses analyses et résultats qui produisent des connaissances nouvelles sur les maladies, les populations génomiques, les effets déclencheurs d’expression génétique. A travers ces différentes opérations, la génomique transforme le don en données génétiques et produit une valeur de connaissance actionnable : Bachelard (*Le rationalisme appliqué*, 1949) nomme ce processus opératoire de la science contemporaine capable de créer des valeurs de connaissance, une phénoménotechnique ; Ph. Steiner (*Donner...une histoire de l’altruisme*, 2016) nomme performance cette création de valeurs qui suppose des dispositifs complexes de transformation du don. Ces deux concepts permettent d’identifier le régime d’actionnabilité qui préside la production de connaissances génomiques.

Mots clés : don, données génétiques, valeur, phénoménotechnique, performance