



Big Data Sharing: A Crucial Democratic Issue for Genomic Medicine

*Benjamin Derbez**

Université de Bretagne Occidentale, Brest, France

Keywords: big data, genomics, BRCA, oncogenetics, database

INTRODUCTION

Big data are often viewed as responsible for major upheavals in many aspects of contemporary life (1) and in the health sector in particular (2). For instance, in medicine, big data are perceived as one of the major drivers of genomic medicine (3). Indeed, rapid genomic data collection on a large scale, made possible by the use of high-throughput sequencing technologies, has made the production of new medical knowledge possible. This knowledge has helped to improve disease prevention, risk prediction, individualized care, and patient involvement (4, 5). One of the conditions of such progress, however, is the need to create databases large enough to enable successful comparative analyses (6). While some initiatives seeking to share different national databases have been launched at the international level (7), the sharing of data between public institutions and private organizations remains a critical question.

Drawing on the example of databases of variants in breast and ovarian cancer predisposition BRCA 1–2 genes, we will show that genomic data is a techno-scientific democracy issue worth discussing. In this case, the recent evolution of patenting legislation has led to a shift from gene sequencing to the clinical interpretation of its results as the key activity of oncogenetics (8). Database access, which is necessary to estimate the risks associated with sequenced genetic variants, has become a critical issue, especially for private firms wishing to break into the market. In this context, the partial privatization of public databases, such as that of the French consortium that will be discussed later, is proof that there is a growing movement of public-private hybridization of these infrastructures. This shift, accentuated by the developments of high-throughput sequencing and genomic medicine, needs to be accompanied by reflection about the public health system user information contributing to the constitution of these databases.

PATENTING GENES

The controversy that shook the world of genetic cancer for years is well known. Indeed, the American company Myriad Genetics filed a patent application claiming BRCA1, BRCA2, and genetic methods of diagnosing a predisposition for breast and ovarian cancer (9, 10). Thanks to the legal ownership of these genes which had been designed as biotechnologies, the start-up from Salt Lake City sought to have a global monopoly on the hereditary breast cancer market, which was expected to experience robust growth. In the face of this offensive, institutional resistance (bringing together hospitals, ministries, associations, etc.) arose in the early 2000s in Europe and then in the United States (11). This resistance has often been interpreted as paradigmatic of the opposition between an “open science,” regulated by peers respecting the law of priority, and a “proprietary science,” regulated by the market, and respecting intellectual property (12). There was thus concern that the production of public knowledge would decline because of the legal appropriation of genes by private organizations (13).

OPEN ACCESS

Edited by:

Thomas Lefèvre,
Université Paris 13, France

Reviewed by:

Nicole C. Nelson,
University of Wisconsin-Madison,
United States

***Correspondence:**

Benjamin Derbez
benjamin.derbez@univ-brest.fr

Specialty section:

This article was submitted to
Digital Health,
a section of the journal
Frontiers in Public Health

Received: 03 July 2018

Accepted: 31 October 2018

Published: 28 November 2018

Citation:

Derbez B (2018) Big Data Sharing: A
Crucial Democratic Issue for Genomic
Medicine. *Front. Public Health* 6:334.
doi: 10.3389/fpubh.2018.00334

An analysis of the British case, however, helps to get a more balanced view of this dichotomy. Indeed, (14–16) has shown that patents are perceived as legal weapons by private organizations as well as by public scientific, medical, and social institutions. Moreover, actors from private and public groups cannot be radically distinguished insofar as each defines the other in a complex network of negotiated interrelationships. In line with the studies undertaken on the role of patents in management science between academic circles and the business world (17, 18), Parthasarathy calls attention to how the NHS and Myriad reached an agreement in the early 2000s, making it possible to connect the “moral order” of the former, based on the principle of equal access to healthcare for all citizens, to the freedom of consumers valued by the latter. Among the negotiated items, it appears clearly that the issue of the transfer of data from Myriad to the NHS was essential and intended to add onto the public BRCA mutation databases. Beyond the issue of monopoly over the gene sequence through the patenting of genes or methods, this example clearly shows that the ownership of data is of crucial importance to both groups. With high-throughput sequencing technology, it has become a major issue.

NEXT GENERATION SEQUENCING

Two major developments placed the issue of the sharing of BRCA databases at the center of the debate from the 2010s. The first, naturally, was the full or partial decline in the patents claimed by Myriad Genetics around the world (19, 20). Indeed, this decline opened up the sequencing market to new private actors (GeneDx, Invitae, Pathway Genomics, Counsyl, etc.) and allowed public laboratories to carry out their activities. The second development was the progressive introduction of high-throughput DNA sequencing technology which began in the mid-2000s. The use of these “next generation” devices reinforced laboratories’ analytical capacities. It is now possible to analyse within a few hours, and at the same time, several genes (panels) of several individuals, or even the complete genome of an individual at a much lower cost—100 dollars is regularly mentioned, compared to the 3 billion dollars spent in the framework of the Human Genome Project 20 years ago (21). All these developments have led stakeholders to focus on the issue of the classification of the genetic variants in BRCA genes.

A genetic variant from a sequenced individual can only acquire the status of “mutation,” i.e., the status of “pathogenic” variant, if it is clearly linked to a history of illness, either directly (in the individual or in their family) or indirectly (in a family affected by cancer and found to have the same variant). According to the current classification in genetics, the clinical significance of these variants may vary: they can be pathogenic, probably pathogenic, of unknown significance, benign, or probably benign. As (22) have pointed out, distinguishing between these categories is a major “interpretive dilemma” for geneticists. The classification of a variant in a given category depends on available data concerning the frequency of the link associating it with a specific disease. In the absence of data, the clinical significance of the variant is deemed unknown—a Variant of Unknown Significance (VUS)—until it is identified in other individuals with similar phenotypic characteristics. The

importance of new DNA sequencing technologies thus lies in their ability to increase genetic databases more quickly in order to reduce the at times dramatic clinical uncertainty associated with diagnosed genetic anomalies (23). The sharing of information among geneticists, thanks to databases fed on an international scale, is a central issue¹. This sharing of information, however, is now problematic.

GENOMIC DATABASES

For several years now, science and technology studies have been stressing that physical infrastructure plays a central role in the production of knowledge (24–27). In this area, the study of genetic databases serves as a model (28–31). Indeed, the first molecular biology databases were launched by different public institutions around the world in the early 1980s [(32): 75]. With the spread of the Internet and the Human Genome Project in the 1990s, they quickly developed as a form of support for new open “communication regimes” between scientists, likely to encourage the emergence of new knowledge (33). However, an analysis of the construction of this information infrastructure shows that the modes of data publishing remain a major source of tension between different actors.

This tension has been highlighted by Bruno Strasser, for instance, in his study on the development of the comprehensive GenBank sequence database (32). This historian of life sciences argues that tensions linked to the different conceptions of data ownership arose from the outset of the project. Participants engaged in a “moral economy of natural history,” i.e., in a “system of values that places emphasis on the exchange of scientific knowledge” inherited from the naturalists of the eighteenth century, considered that the sequences published in scientific journals should be freely accessible data. Other participants, advocates of a “moral economy of experimentation” which has garnered momentum among molecular biologists, view sequences as the products of scientific activity and as the property of their authors. According to Strasser, GenBank embodies a form of hybridization of these two value systems. It appears that those who conceived it succeeded in taking advantage of the “ambiguity” of the very notion of “data,” owing to the fact that what seems “literally given” is at the same time “the result of an organized action” (34): 248). In the context of the Human Genome Project, this ambiguity has manifested itself in the emergence of information control modes which involve a complex interplay of revelation and concealment (35). Nowadays, as seen previously, in addition to the tensions inherent in the moral economies of science, other tensions associated with the political economy of knowledge resulting from the growing role played by private firms in the production of knowledge emerged from the early 1990s (36, 37). Beyond the question of the patentability of living organisms, it is now the question of sharing that is in front of the debate, like the case of BRCA1 and BRCA2 genes clearly shows it.

¹For example: Human Gene Mutation Database (HGMD) or Online Mendelian Inheritance in Man (OMIM).

DATA SHARING

In the present case, i.e., the focus on BRCA1 and BRCA2 genes, there is no unique and comprehensive database of BRCA variants accessible to all professionals around the world. On the contrary, different databases developed by consortia of multinational public institutions or private organizations exist, but their access is generally limited. This is the case of the database developed by Myriad Genetics throughout the period the patents were under discussion. Although this is the largest database in the world, Myriad Genetics has exclusive access to it. This has given the company a major competitive asset in the BRCA testing market insofar as the database offers a solid basis on which to interpret results. According to genetics professionals, the main issue is not the sequencing itself. Rather, what matters most is the interpretation of the results intended to give clinical significance. This has turned out to be the most costly activity, both in terms of the recruitment of highly qualified personnel and for the development, maintenance, and access to huge databases that list the known variants of specific genes. Certain professionals estimate that there is a 1 to 10 ratio with regards to the cost of complete genome sequencing and its interpretation. In this context, ownership and the opening up of genetic variants databases emerges as a crucial issue.

From this context, the example of the future of the UMD BRCA base—*Universal Mutation Database-BRCA*—speaks volumes. Developed in the 1990s by a public consortium of French geneticists, it was considered to be one of the most important global databases until 2015. Driven by two major players in genetic testing in the United States [Quest Diagnosis and Laboratory Corporation of America (LabCorp)], the database was partially privatized in 2015. These two companies purchased the right to obtain access to data in exchange for funding the database. While the French sought to finance over the short- and medium-term an activity that had become too costly for public finances to sustain, the Americans' objective was to quickly be able to compete with Myriad Genetics by improving the quality of their analyses. The question that arises, then, is: How will this be handled over the long term? Will the French geneticists at the origin of the database still be able to access it? Will French patients still benefit from the knowledge generated thanks to the data they provided? What justifies this privatization if we consider the donations made by patients who agreed to have their data kept in this database? Similar questions had already been raised by the NHS during its negotiations with Myriad in the early 2000s, when the issue of the privatization of access to BRCA testing for British citizens arose (16). Questions revolving around access (currently and in the future) to genetic databases thus remain relevant.

REFERENCES

1. Mayer-Schönberger V, Cukier K. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Boston, NY: Eamon Dolan /Mariner / Houghton Mifflin Harcourt (2013).
2. Groves P, Kayyali B, Knott D, Van Kuiken S. *The Big Data Revolution in Health Care*. Seattle, DC: McKinsey Quarterly (2013).

CONCLUSION

At a time when the opening up of public data has become common practice in the field of administration (38), the example of the genetics of breast cancer shows that data sharing is still a major issue in research (39). The question here is the extreme overlapping of public issues and private interests. In this case, there is a need to go beyond a simple comparison between the open regimes of data publication associated with academic institutions, and the closed regimes of the privatization of knowledge developed by business communities. Hybrid forms of database ownership such as those mentioned earlier, highlight the need to pay attention to the significance given to data sharing during the initial negotiations underpinning their establishment. Once these databases are filled by voluntary citizens who provide their DNA data, data sharing becomes a crucial issue in terms of technical democracy (40). Once again, however, citizens seem to be largely absent from the debate about the ownership and use of the genomic data stored in these databases. With increased power given to major programmes seeking to collect big data in genomics, it may be time to reflect on how citizens can be informed and involved in the decisions that will be made in this area.

At the very least, it seems necessary to provide people with information about the future of their genomic data: in which databases will the data be stored? For how long? Who will be able to use them? Can they be exploited for commercial purposes by private firms? As in the field of the Internet, database contributors should be able to oppose the reuse of their "data" for the benefit of private interests. The information challenge involves the very value of consent (41).

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This opinion paper is based on a research funded by the Fonds Avenir/Masfp pour la Recherche, 2016.

ACKNOWLEDGMENTS

The author would like to thank Emmanuel Rial-Sebbag, S. de Montgolfier, Pr Dominique Stoppa-Lyonnet, Pr Eric Vilain, and Dr. Zaki El Haffaf for their help and collaboration. Thank you to Catherine Davies from UBO BTU for the translation.

3. Guttmacher AE, Collins FS. Genomic medicine – a primer. *N Engl J Med*. (2002) 347:1512–20. doi: 10.1056/NEJMra012240
4. Alyass A, Turcotte M, Meyre M. From big data analysis to personalized medicine for all: challenges and opportunities. *BMC Med Genomics* 8:33. doi: 10.1186/s12920-015-0108-y
5. Flores M, Glusman G, Brogaard K, Price ND, Hood L. P4 medicine: how systems medicine will transform the healthcare sector

- and society. *Personal Med.* (2013) 10:565–76. doi: 10.2217/PME.13.57
6. He KY, Ge D, He MM. Big data analytics for genomic medicine. *Int J Mol Sci.* 18:412. doi: 10.3390/ijms18020412
 7. Scollen S, Page A, Wilson J. From the data on many, precision medicine for “one”: the case for widespread genomic data sharing. *Biomed Hub* (2017) 2:15. doi: 10.1159/000481682
 8. Stoppa-Lyonnet D. Tests génétiques: le défi n'est plus le séquençage mais l'interprétation. *Pour Sci.* (2014) 439:12–3.
 9. Parthasarathy S. *Building Genetic Medicine. Breast Cancer, Technology, and the Comparative Politics of Health Care.* Cambridge, MA: MIT Press (2017).
 10. Sherkow JS, Greely HT. The history of patenting genetic material. *Annu Rev Genet.* (2015) 49:161–82. doi: 10.1146/annurev-genet-112414-054731
 11. Cassier M, Stoppa-Lyonnet D. L'opposition contre les brevets de myriade génétiques et leur révocation totale ou partielle en Europe: Premiers enseignements. *Med Sci.* (2005) 21:648–62. doi: 10.1051/medsci/2005216-7658
 12. Dasgupta P, David PA. Towards a new economics of science. *Res Policy* (1994) 23:487–521. doi: 10.1016/0048-7333(94)01002-1
 13. Orsi F, Coriat B. Are Strong Patents beneficial to Innovative Activities - Lessons from genetic testing for breast cancer controversies. *Indus Corp Change* (2005) 14:1205–21. doi: 10.1093/icc/dth086
 14. Parthasarathy S. The patent is political: the consequences of patenting the BRCA genes in Britain. *Community Genet.* (2005) 8:235–42. doi: 10.1159/000087961
 15. Parthasarathy S. Architectures of genetic medicine: comparing genetic testing for breast cancer in the USA and UK. *Soc Stud Sci.* (2005) 35:5–40. doi: 10.1177/0306312705047172
 16. Parthasarathy S. Reconceptualizing technology transfer: the challenge of building an international system of genetic testing for breast cancer. In: Guston DH, Sarewitz D, editors. *Shaping Science and Technology Policy: The Next Generation of Research.* Madison, WI: University of Wisconsin Press (2006).
 17. Huang KG, Murray FE. Does patent strategy shape the long-run supply of public knowledge: evidence from human genetics. *Acad Manag J.* (2009) 52:1193–221. doi: 10.5465/amj.2009.47084665
 18. Murray F. The oncomouse that roared: hybrid exchange strategies as a source of distinction at the boundary of overlapping institutions. *Am J Sociol.* (2010) 116:341–88. doi: 10.1086/653599
 19. Cassier M, Stoppa-Lyonnet D. Un juge fédéral et le gouvernement des États-Unis interviennent contre la brevetabilité des gènes. *Med Sci.* (2011) 27:662–7. doi: 10.1051/medsci/201228s204
 20. Pollack A. *Myriad Genetics Ending Patent Dispute on Breast Cancer Risk Testing, New York Times.* (2015). Available online at: <https://www.nytimes.com/2015/01/28/business/myriad-genetics-ending-patent-dispute-on-breast-cancer-risk-testing.html> (Accessed 18/09/18)
 21. Reuter. *New Illumina Tech Could Usher in \$100 Gene-Sequencing Era.* (2017). Available online at: <https://www.reuters.com/article/us-illumina-stocks/new-illumina-tech-could-usher-in-100-gene-sequencing-era-idUSKBN14U1PO> (Accessed 18/09/18)
 22. Timmermans S, Tietbohl C, Skaperdas E. Narrating uncertainty: variants of uncertain significance (VUS) in exome sequencing. (2016) *BioSocieties* 12:439–58. doi: 10.1057/s41292-016-0020-5
 23. Stivers T, Timmermans S. Negotiating the diagnostic uncertainty of genomic test results. *Soc Psychol Quart.* (2017) 79:199–221. doi: 10.1177/0190272516658770
 24. Star SL, Ruhleder K. Steps towards an ecology of infrastructure: design and access for large information spaces. *Inform Syst Res.* (1996) 7:111–34. doi: 10.1287/isre.7.1.111
 25. Star SL. The Ethnography of Infrastructure ». *Am Behav Sci.* (1999) 43:377–91. doi: 10.1177/00027649921955326
 26. Bowker G, Baker K, Millerand F, Ribes D. Towards information infrastructure studies: ways of knowing in a networked environment. In: Hunsinger J, Klastrup L, Allen M, editors. *International Handbook of Internet Research Dordrecht.* Dordrecht: Springer (2010). p. 97–117.
 27. Bowker GC, Star SL. *Sorting Things Out: Classification and Its Consequences.* Cambridge, MA: MIT Press (1999).
 28. Brown C. The changing face of scientific discourse: analysis of genomic and proteomic database usage and acceptance. *J Am Soc Inform Sci Technol.* (2003) 54:926–38. doi: 10.1002/asi.10289
 29. Bowker G. *Memory Practices in the Sciences.* Cambridge, MA: MIT Press (2005).
 30. Hine C. Databases as scientific instruments and their role in the ordering of scientific work. *Soc Stud Sci.* (2006) 36:269–98. doi: 10.1177/0306312706054047
 31. Dagiral E, Peerbaye A. Making knowledge in boundary infrastructures: inside and beyond a database for rare diseases. *Sci Technol Stud.* (2016) 29:44–61.
 32. Strasser B. The Experimenter's Museum: GenBank, natural history, and the moral economies of biomedicine. *Isis* (2011) 102:60–96. doi: 10.1086/658657
 33. Hilgartner S. Biomolecular databases: new communication regimes for biology? *Sci Commun.* (1995) 17:240–63. doi: 10.1177/1075547095017002009
 34. Desrosières A. *The Politics of Large Numbers. A History of Statistical Reasoning.* Boston, MA: Harvard University Press (2002).
 35. Hilgartner S. *Reordering Life: Knowledge and Control in the Genomics Revolution.* Cambridge, MA: MIT Press (2017).
 36. Cassier M, Gaudillière JP. Recherche, médecine et marché: la génétique du cancer du sein. *Sci Soc Santé* (2000) 18:29–49. doi: 10.3406/sosan.2000.1504
 37. Huang KG, Murray FE. Entrepreneurial experiments in science policy: analyzing the human genome project. *Res Policy* (2010) 39:567–82. doi: 10.1016/j.respol.2010.02.004
 38. Goëta S, Davies T. The daily shaping of state transparency: standards, machine-readability and the configuration of open government data policies. *Sci Technol Stud.* (2016) 29:10–30.
 39. Nature Editorial. The ups and downs of data sharing in science. pooling clinical details helps doctors to diagnose rare diseases — but more sharing is needed. *Nature* (2016) 534:435–6. doi: 10.1038/534435b
 40. Callon M, Lascoumes P, Barthe Y. *Acting in an Uncertain World. An Essay on Technical Democracy.* Boston, MA: MIT Press (2009).
 41. Ducournau P. The viewpoint of DNA donors on the consent procedure. *New Genet Soc.* (2007) 26:105–15. doi: 10.1080/14636770701218191
- Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Derbez. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.