

PERSPECTIVE

# Controlled Access under Review: Improving the Governance of Genomic Data Access

Mahsa Shabani<sup>1\*</sup>, Stephanie O. M. Dyke<sup>2</sup>, Yann Joly<sup>2</sup>, Pascal Borry<sup>1</sup>

**1** Centre for Biomedical Ethics and Law, Department of Public Health and Primary Care, University of Leuven, Leuven, Belgium, **2** Centre of Genomics and Policy, Faculty of Medicine, McGill University, Montreal, Quebec, Canada

\* [mahsa.shabani@med.kuleuven.be](mailto:mahsa.shabani@med.kuleuven.be)



 OPEN ACCESS

**Citation:** Shabani M, Dyke SOM, Joly Y, Borry P (2015) Controlled Access under Review: Improving the Governance of Genomic Data Access. *PLoS Biol* 13(12): e1002339. doi:10.1371/journal.pbio.1002339

**Published:** December 31, 2015

**Copyright:** © 2015 Shabani et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** MS is supported by by the Interfaculty Council for Development Cooperation (IRO) funding from University of Leuven (<http://www.kuleuven.be/english/international/funding/iro>). SOMD is supported by the Canadian Institutes of Health Research (grants EP1-120608; EP2-120609) (<http://www.cihr-irsc.gc.ca/e/193.html>), the Canada Research Chair in Law and Medicine (<http://www.chairs-chaire.gc.ca/chairholders-titulaires/profile-eng.aspx?profileid=2680>), and the Public Population Project in Genomics and Society (P3G) (<http://www.p3g.org/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** We have read the journal's policy and have the following conflicts: Yann Joly is the Data Access Officer of International Cancer Genome Consortium's Data Access Compliance Office.

## Abstract

In parallel with massive genomic data production, data sharing practices have rapidly expanded over the last decade. To ensure authorized access to data, access review by data access committees (DACs) has been utilized as one potential solution. Here we discuss core elements to be integrated into the fabric of access review by both established and emerging DACs in order to foster fair, efficient, and responsible access to datasets. We particularly highlight the fact that the access review process could be adversely influenced by the potential conflicts of interest of data producers, particularly when they are directly involved in DACs management. Therefore, in structuring DACs and access procedures, possible data withholding by data producers should receive thorough attention.

## Introduction

Over the last decade, sharing the data of publicly funded genomic studies is an issue that has received widespread attention in the scientific community. Funding organizations, research institutes and journals promote data sharing to ensure the optimal (re)use of data by researchers. Nevertheless, because of inherent privacy concerns in using genomic data and the importance of respecting the wishes of data subjects, controlled-access mechanisms have been adopted to allow researchers to use such data [1,2].

Data access committees (DACs) have been set up as a major instrument in reviewing access requests. The way DACs have been designed and function, however, differs significantly from one committee to another [3]. While some DACs are constructed at an institutional or consortia level, with higher formality and regular meetings, others are formed by small groups and rely on rather ad hoc meetings [4]. The two main central genomic databases, namely, the database of Genotypes and Phenotypes (dbGaP) and the European Genome-phenome Archive (EGA), also exhibit different approaches to DACs [5]. Limited available information on the access procedures of other controlled-access genomic databases [6,7] makes evaluation of their access procedures difficult, if not impossible.

Despite the differences, the ultimate goal of access review is to grant access to qualified users for a set of approved uses. Accordingly, assessing conformity of the proposed uses of data to the original consent forms research participants signed and examining the qualifications of the

**Abbreviations:** DAC, data access committee; dbGaP, database of Genotypes and Phenotypes; EGA, European Genome-phenome Archive; IRB, Institutional Review Board; SOPs, Standard Operating Procedures.

researchers requesting access seem to constitute the main tasks of an access review. Furthermore, access review specifies the duties and responsibilities of data users by means of instruments, such as data access agreements co-signed by data users' institutions.

However, in situations in which the same researchers that have generated data also control further downstream use of the data, there is a conflict of interest, which might undermine data sharing. There are several reasons researchers might not be willing to share data they have produced, or at least, not for the time being. As Bobrow observes, "researchers who spend time, effort, and ingenuity to generate, process, and manage large research data sets expect to get appropriate credit" [8]. Data producers' concerns have also been recognized in a number of international policy statements, such as The Fort Lauderdale Agreement (2003) and the Toronto Statement on Pre-publication Data Sharing (2009) [2,9]. In this article, we suggest a number of potential solutions to address this concern: objectivity of access review, fairness of terms and conditions in access to data, and transparency and accountability. We argue that these core elements should be incorporated into the process of access review by both established and emerging DACs in order to best serve the interests of the various parties involved.

## Objectivity of Access Review

Many DACs are mainly composed of the primary investigators or researchers who produced data. This allows the primary investigators to monitor the data they have generated. Primary researchers' direct involvement is believed advantageous in order to ensure that any specific concerns expressed by research participants have been taken into account in the access review, particularly in dealing with sensitive data. Conceivably, primary researchers are trusted by research subjects in safeguarding their interests and respecting their wishes in the course of further data uses [10].

Despite the perceived advantages, objective decision-making in access review may be affected as a result of such direct involvement. Indeed, the DAC structure and composition should secure an objective access review, which is conducive to fair and responsible data access. This is in keeping with the ultimate purposes of data sharing, to optimize the use of genomic databases as community resources [9,11]. Researchers' feeling of ownership to data they generated or expectation to control downstream uses of data have been reported in previous studies and could unduly influence data access review [12]. Currently, data sharing policies and guidelines do not explicitly address this issue, so data producers are left to determine on their own how to structure their DAC.

If a DAC is independent, this concern may be mitigated. For example, dbGaP's DACs are composed of United States federal employees who conduct the access review without the involvement of the data producers. This model, however, entails a central infrastructure set up by the research institutes or funding organizations. At the moment, the prevalence of such central infrastructures is limited. It is therefore necessary that funding organizations, research institutes or relevant government bodies set up such infrastructure and steer access review. National data archives, such as the United Kingdom Data Archive or the Finnish Social Science Data Archive, or international archives, such as the Consortium of European Social Science Data Archives, could serve as potential models. Despite the similarities in general goals of data sharing in various disciplines, it is important to keep in mind that the nature of data, funding, and associated concerns and interests of the involved bodies in downstream uses may influence the structure of central infrastructures in different disciplines.

It is essential that central data access infrastructures avoid unnecessary administrative burdens on access review and potential delays in access. Indeed, reviewing a large number of access requests in a timely manner demands a significant and sustainable investment. This

underscores the importance of seeking a balanced approach that fosters both efficiency and objectivity.

A two-tiered access committee might be an alternative solution. Whereas a DAC or compliance office could review access requests on a case-by-case basis, an independent steering committee might function in an oversight and/or advisory role. Steering committees can be formed in institutions or at a consortium level and could be responsible for oversight of the access review conducted by DACs that involve data producers. The relationship between steering committees and DACs, along with the scope of their discretion, should be clarified to avoid potential redundancies in access review and establish the optimal mechanisms of oversight. Establishing clear guidelines on access review will be also crucial. Steering committees would monitor the extent to which DACs adhere to such guidelines and root out arbitrary decision-making.

A steering committee can offer consultation in controversial cases or provide broader policy-level observations and guidelines. Also, should any complaint arise from access review, the case can be referred to the steering committee for further inquiry. With this two-tiered system, users will be provided with an opportunity to appeal against the DAC decisions if they wish. This resembles the idea of “appellate” institutional review board (IRB) which was put forward by Lantos as a transparent and publicly auditable body to issue decisions in controversial ethics cases [13]. In the absence of steering committees, an independent ombudsman could take on this role.

## Fairness in Terms and Conditions of Access to Data

Researchers’ concerns about receiving credit for their data producing and sharing efforts warrant the establishment of a number of relevant policies and terms and conditions of access. For example, researchers sharing data may be reluctant to provide others with access to it before they have had time to analyze it themselves and to publish the results of their analysis. Publication embargos or moratoriums, which prohibit secondary data users from publishing on shared data for a certain amount of time, were introduced into data sharing policies to address this [2,14], though it has been noted the mechanism may lack clarity and may sometimes prevent non-competing analyses [15]. Even after publication, data withholding has been documented amongst academic geneticists, indicating that researchers sharing data may be influenced by other conflicts of interest as well [16]. While some credit for researchers sharing data can be secured via the data access procedure as discussed below, and this may help to reduce their concerns, fair data sharing requirements on the part of data producers should be established through data sharing policy and not left to DACs to oversee.

Researchers sharing data may benefit from the new collaborations this facilitates. However, collaboration, and co-authorship in particular, should follow authorship guidelines, and sharing data should not automatically entail collaborative opportunities, though it may enhance these. Incentivizing data sharing remains important, as data sharing has yet to become the norm and there are many perceived barriers and disadvantages to sharing, from a researcher’s perspective. Reward should be provided in the form of acknowledgment and citation, where possible, of shared data resources [17,18]. This may be encouraged or enforced through the data access procedure, if stipulated in Data Access Agreements, for example. There is also some evidence that sharing the data that underlie research publications may lead to increased citation of the publication itself [19]. Nevertheless, the main routes to providing credit for data sharing in academia ought to be through standard institutional and funding mechanisms [15], rather than through DACs.

Some resource projects require that users contribute the data that they have generated thanks to the resource (e.g., UK Biobank [20]). Others simply request that research using the

data be published. Obligations such as these are seen as an expression of reciprocity that could potentially be employed in other data sharing scenarios in order to demonstrate more broadly the benefits of data sharing.

## Transparency and Accountability

Only rarely do DACs provide public information about their assessment criteria or the number of data requests received (and how many were accepted or rejected). It could be argued that, just as respect for privacy and confidentiality is necessary in order to build trust and develop valuable relationships, a sufficient degree of transparency is necessary for maintaining that trust and these relationships. Transparency is also often seen as an important component of governance in health research because it is conducive to stakeholder accountability, including researchers [21]. This is particularly relevant in a field like genomic research, where soft norms and internal self-regulation still play a considerable role but need to evolve rapidly in order to keep pace with scientific and technological developments. In setting up Standard Operating Procedures (SOPs) for a DAC, transparency can also be used to reassure research participants regarding the integrity and security of the data sharing activities and to promote more harmonized processes that will facilitate international exchanges of data.

When setting up a DAC, it should be a given that the procedures and governance framework of the data access, including access and refusal criteria, number of applications, etc., are to be made available openly, and are to be presented in a way that is understandable to patients, research participants, and members of the research community [22]. In doing so, special attention should be paid to the proposed scope of the sharing and the applicable data protection framework (i.e., physical, administrative, and technological security measures). This information should also be updated at regular intervals to remain current. Furthermore, we would argue that DACs would also have much to gain by keeping a public log containing a minimum amount of key information on approved and declined data access requests, data security incidents, and the manner in which they were addressed by the DACs. This log could contain items such as the date of the incident, its general nature, the date and manner in which the incident was addressed, and any future implications for the data sharing process. For security reasons, it would be justifiable for databases to be given a reasonable delay period to make this type of information public. The significance of such public reporting and its role in fostering transparency has already been stressed by some stakeholders and addressed by some DACs [23–25].

## Conclusion

In principle, the structure of DACs should facilitate a comprehensive and fair access review. Current DACs often hinge on direct involvement of data producers, which triggers some concerns considering potential conflicts of interest. We believe such concerns could be addressed by complementary governance mechanisms adopted by the parties involved.

Funding organizations are best placed to play a key role in fulfilling fair, efficient, and responsible access review. There are, however, a number of roadblocks to the active engagement of funding organizations. There are limited examples at the moment of such extensive engagement by funding organizations across the world, due to limited infrastructure or underinvestment in this area. Funders' higher involvement in an efficient access review system should be stimulated, and current best practices seen in different fields should be considered as examples.

Funding organizations need to be assisted by other bodies, such as journals and institutions (e.g., universities, hospitals, or other research institutes) [2]. Journals are particularly well

placed to maintain oversight of publication policies embedded in access arrangements. Research institutions also have a crucial role in providing organizational support for researchers so that they can adhere to data access and sharing mandates arranged by funding organizations.

Ultimately, however, the success of data sharing policies depends on introducing streamlined and harmonized approaches to data access across research and funding organizations. Internationally recognized standards and access arrangements will invigorate interoperability among data producing entities and improve the scalability of central data archives and of fair, transparent, and accountable processes of data sharing and access.

## References

1. Church G, Heeney C, Hawkins N, de Vries J, Boddington P, Kaye J, et al. Public access to genome-wide data: five views on balancing research with privacy and protection. *PLoS Genet*. 2009; 5: e1000665. doi: [10.1371/journal.pgen.1000665](https://doi.org/10.1371/journal.pgen.1000665) PMID: [19798440](https://pubmed.ncbi.nlm.nih.gov/19798440/)
2. Birney E, Hudson TJ, Green ED, Gunter C, Eddy S, Rogers J, et al. Prepublication data sharing. *Nature*. 2009; 461: 168–170. doi: [10.1038/461168a](https://doi.org/10.1038/461168a) PMID: [19741685](https://pubmed.ncbi.nlm.nih.gov/19741685/)
3. Shabani M, Knoppers BM, Borry P. From the principles of genomic data sharing to the practices of data access committees. *EMBO Mol Med*. 2015; 7: 507–509. doi: [10.15252/emmm.201405002](https://doi.org/10.15252/emmm.201405002) PMID: [25759363](https://pubmed.ncbi.nlm.nih.gov/25759363/)
4. Lowrance WW. Access to collections of data and materials for health research. A report to the Medical Research Council and the Wellcome Trust London: MRC and Wellcome Trust. 2006: 1–36.
5. Lappalainen I, Almeida-King J, Kumanduri V, Senf A, Spalding JD, Saunders G, et al. The European Genome-phenome Archive of human data consented for biomedical research. *Nat Genet*. 2015; 47: 692–695. doi: [10.1038/ng.3312](https://doi.org/10.1038/ng.3312) PMID: [26111507](https://pubmed.ncbi.nlm.nih.gov/26111507/)
6. Yang Y, Dong X, Xie B, Ding N, Chen J, Li Y, et al. Databases and web tools for cancer genomics study. *Genomics Proteomics Bioinformatics*. 2015; 13: 46–50. doi: [10.1016/j.gpb.2015.01.005](https://doi.org/10.1016/j.gpb.2015.01.005) PMID: [25707591](https://pubmed.ncbi.nlm.nih.gov/25707591/)
7. Thorisson GA, Muilu J, Brookes AJ. Genotype–phenotype databases: challenges and solutions for the post-genomic era. *Nat Rev Genet*. 2009; 10: 9–18. doi: [10.1038/nrg2483](https://doi.org/10.1038/nrg2483) PMID: [19065136](https://pubmed.ncbi.nlm.nih.gov/19065136/)
8. Bobrow M. Balancing privacy with public benefit. *Nature*. 2013; 500: 123. doi: [10.1038/500123a](https://doi.org/10.1038/500123a) PMID: [23925206](https://pubmed.ncbi.nlm.nih.gov/23925206/)
9. Report of a meeting organized by the Wellcome Trust and held on 14–15 January 2003 at Fort Lauderdale. Sharing Data from Large-scale Biological Research Projects: A System of Tripartite Responsibility. 2003. <https://www.genome.gov/Pages/Research/WellcomeReport0303.pdf>
10. Kaye J, Heeney C, Hawkins N, de Vries J, Boddington P. Data sharing in genomics—re-shaping scientific practice. *Nat Rev Genet*. 2009; 10: 331–335. doi: [10.1038/nrg2573](https://doi.org/10.1038/nrg2573) PMID: [19308065](https://pubmed.ncbi.nlm.nih.gov/19308065/)
11. HUGO Ethics Committee. Statement on Human Genomic Databases. 2002. [http://www.hugo-international.org/img/genomic\\_2002.pdf](http://www.hugo-international.org/img/genomic_2002.pdf)
12. Fecher B, Friesike S, Hebing M. What drives academic data sharing? *PLoS ONE*. 2015; 10: e0118053. doi: [10.1371/journal.pone.0118053](https://doi.org/10.1371/journal.pone.0118053) PMID: [25714752](https://pubmed.ncbi.nlm.nih.gov/25714752/)
13. Lantos J. It is time to professionalize institutional review boards. *Arch Pediatr Adolesc Med*. 2009; 163: 1163–1164. doi: [10.1001/archpediatrics.2009.225](https://doi.org/10.1001/archpediatrics.2009.225) PMID: [19996057](https://pubmed.ncbi.nlm.nih.gov/19996057/)
14. Joly Y, Dove ES, Kennedy KL, Bobrow M, Ouellette BF, Dyke SO, et al. Open science and community norms Data retention and publication moratoria policies in genomics projects. *Med Law Int*. 2012; 12: 92–120.
15. Dyke SO, Hubbard TJ. Developing and implementing an institute-wide data sharing policy. *Genome Med*. 2011; 3: 60. doi: [10.1186/gm276](https://doi.org/10.1186/gm276) PMID: [21955348](https://pubmed.ncbi.nlm.nih.gov/21955348/)
16. Campbell EG, Clarridge BR, Gokhale M, Birenbaum L, Hilgartner S, Holtzman NA, et al. Data withholding in academic genetics: evidence from a national survey. *JAMA*. 2002; 287: 473–480. PMID: [11798369](https://pubmed.ncbi.nlm.nih.gov/11798369/)
17. Cambon-Thomsen A, Thorisson G, Mabile L, BRIF Workshop Group. The role of a Bioresource Research Impact Factor as an incentive to share human bioresources. *Nat Genet*. 2011; 43: 503–504. doi: [10.1038/ng.831](https://doi.org/10.1038/ng.831) PMID: [21614086](https://pubmed.ncbi.nlm.nih.gov/21614086/)
18. Bravo E, Calzolari A, De Castro P, Mabile L, Napolitani F, Rossi AM, et al. Developing a guideline to standardize the citation of bioresources in journal articles (CoBRA). *BMC Med*. 2015; 13: 33. doi: [10.1186/s12916-015-0266-y](https://doi.org/10.1186/s12916-015-0266-y) PMID: [25855867](https://pubmed.ncbi.nlm.nih.gov/25855867/)

19. Piwowar HA, Vision TJ. Data reuse and the open data citation advantage. *PeerJ*. 2013; 1: e175. doi: [10.7717/peerj.175](https://doi.org/10.7717/peerj.175) PMID: [24109559](https://pubmed.ncbi.nlm.nih.gov/24109559/)
20. Collins R. What makes UK Biobank special? *Lancet*. 2012; 379: 1173–1174. doi: [10.1016/S0140-6736\(12\)60404-8](https://doi.org/10.1016/S0140-6736(12)60404-8) PMID: [22463865](https://pubmed.ncbi.nlm.nih.gov/22463865/)
21. Kaye J, Whitley EA, Lund D, Morrison M, Teare H, Melham K. Dynamic consent: a patient interface for twenty-first century research networks. *Eur J Hum Genet*. 2014; 23: 141–146. doi: [10.1038/ejhg.2014.71](https://doi.org/10.1038/ejhg.2014.71) PMID: [24801761](https://pubmed.ncbi.nlm.nih.gov/24801761/)
22. Knoppers BM. Framework for responsible sharing of genomic and health-related data. *Hugo J*. 2014; 8: 3.
23. Walker L, Starks H, West KM, Fullerton SM. dbgap data access requests: A call for greater transparency. *Sci Transl Med*. 2011; 3: 113cm34. doi: [10.1126/scitranslmed.3002788](https://doi.org/10.1126/scitranslmed.3002788) PMID: [22174311](https://pubmed.ncbi.nlm.nih.gov/22174311/)
24. Paltoo DN, Rodriguez LL, Feolo M, Gillanders E, Ramos EM, Rutter JL, et al, for the National Institute of Health Genomic Data Sharing Governance Committees. Data use under the NIH GWAS Data Sharing Policy and future directions. *Nat Genet*. 2014; 46: 934–938. doi: [10.1038/ng.3062](https://doi.org/10.1038/ng.3062) PMID: [25162809](https://pubmed.ncbi.nlm.nih.gov/25162809/)
25. Ramos EM, Din-Lovinescu C, Bookman EB, McNeil LJ, Baker CC, Godynskiy G, et al. A mechanism for controlled access to GWAS data: experience of the GAIN Data Access Committee. *Am J Med Genet*. 2013; 92: 479–488.