

# Intra-consortium data sharing in multi-national, multi-institutional genomic studies: gaps and guidance

Jerome Amir Singh · Abdallah S. Daar

Received: 28 June 2009 / Accepted: 23 October 2009 / Published online: 14 November 2009  
© The Author(s) 2009. This article is published with open access at Springerlink.com

**Abstract** Growing investments in health research by governments and charitable organizations have fueled an increase in collaborative research projects between investigators from affluent and developing countries. Current international guidelines are silent on common intra-consortium data-sharing issues that arise in the context of such collaborations. A lack of guidance on intra-consortium data sharing threatens to undermine the success of crucial research ventures. In this work we outline some of the practical problems commonly faced by investigators working in multi-institutional, international genomic collaborations and offer recommendations on these issues. A data sharing policy should be prospectively negotiated and concluded between collaborators as early as possible. Sponsors of research, including those from developing countries, should issue detailed guidance on the above and related issues as doing so will facilitate research and catalyze scientific progress.

**Keywords** Data sharing · Genomics · Contracts, memorandum of understanding · Collaborations

## Introduction

Growing investments in health research by both governments and charitable organisations have fuelled an increase in collaborative research projects between investigators from affluent and developing countries. Data sharing—defined as “the voluntary provision of information from one individual or institution to another for purposes of legitimate scientific research” (Hogue 1991)—can be challenging in multi-institutional international genomic collaborative ventures, particularly as investigators have different expectations of their respective stakes in research outcomes, and hail from different institutional and cultural ideologies. Current international guidelines are silent on common intra-consortium data-sharing issues. This knowledge gap is fuelling uncertainty and threatens to undermine the success of crucial research endeavors. In this work we outline some of the practical intra-consortium problems commonly faced by investigators working in multi-institutional, international genomic collaborations and offer recommendations on these issues.

---

J. A. Singh (✉)

Centre for the AIDS Programme of Research in South Africa (CAPRISA), Nelson R. Mandela School of Medicine, Durban, South Africa  
e-mail: singhj9@ukzn.ac.za

J. A. Singh

Howard College School of Law, University of KwaZulu-Natal, Durban, South Africa

J. A. Singh

McLaughlin-Rotman Centre for Global Health, Dalla Lana School of Public Health, and Joint Center for Bioethics, University of Toronto, Toronto, Canada

A. S. Daar

McLaughlin-Rotman Centre for Global Health, UHN and University of Toronto, Toronto, Canada

## Drafting and sharing: involve and invest in developing world collaborators

Sponsors and principal investigators of international collaborations, such as the Centre for HIV/AIDS Vaccine Immunology (CHAVI) and Grand Challenges in Global Health (GCGH) Initiative, typically hail from affluent countries. These parties usually assume the responsibility

for drafting and driving data sharing or consortium agreements. Intellectual Property (IP) regulatory and guidance frameworks typically have a European or North American basis. Moreover, there is often no or poor intellectual property expertise available to consortium partners from developing countries. Accordingly, data sharing or consortium agreements drafted by PIs or sponsors from affluent countries are sometimes perceived by consortium partners in developing countries as being biased in favour of the interests of their consortium partners from the developed world. PIs from affluent countries should accordingly ensure that their colleagues from the developing world are prospectively involved in the drafting of consortium or data-sharing agreements. Such an approach could see data access and research output issues explicitly and meaningfully governed. Such an agreement should describe the management of intellectual property rights related to the proposed project, including plans for sharing data, information, and materials resulting from the award. The policy must also clearly govern the timing and means of data release, and any constraints on release. ([http://www.ipm.ucdavis.edu/PD/pdrfp\\_attachb.html](http://www.ipm.ucdavis.edu/PD/pdrfp_attachb.html)). We recommend that local communities have a say in the management and sharing of data relating to them.

However, involving developing world collaborators in the drafting of a data sharing agreement will be meaningless if they are unable to access or interpret data. For example, while consortium or data sharing agreements typically specify database access rules (all consortium partners usually have access to a common consortium database) and research output rules (authorship sequence in consortium publications is usually determined by the respective contributions of consortium partners), partners from developing countries often lack the resources to access the pooled data, or the experience and confidence to equally contribute to research output based thereon. This places the better-resourced and experienced collaborators from affluent countries at a distinct advantage in regard to exploiting the common database and authoring publications based thereon. As such, developing world partners are sometimes absent from consortium research outputs or relegated to junior co-authorship status in such works.

In its 2002 statement on human genomic databases the Hugo Ethics Committee declared that there is a scientific responsibility to ensure the professional competence of researchers working with data, as well as the quality and accuracy of the data. (<http://www.hugo-international.org/PDFs/Statement%20on%20Human%20Genomic%20Data%20bases%202002.pdf>). Based on this guiding principle principal investigators and partners from affluent countries have a moral responsibility to build the capacity of their developing world colleagues to enable them to competently undertake the research in question and to

contribute to consortium-wide research outputs. This may necessitate sponsors and/or principal investigators investing in the infrastructure of their developing world colleagues, for example, in compatible computer software and broadband Internet access to enable developing world partners to access consortium databases. Moreover, where needed, collaborators from developed countries should train their developing world partners to ensure they have necessary expertise to exploit the research opportunities in the consortium database.

### Who owns data derived from consortia-wide efforts?

In privately funded research initiatives (to which international agreements such as the Bermuda Accord, ([http://www.ornl.gov/sci/techresources/Human\\_Genome/research/bermuda.shtml#2](http://www.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml#2)) or similar guidelines apply) and in instances where recommended IP-related clauses in a data sharing or consortium agreement (<http://www.jisclegal.ac.uk/publications/IRPConsortia.htm>) are absent, it can be argued that data derived from consortia-wide efforts are the proprietary interests of the consortium as a whole. As such, the authorization to publish such data ought to be obtained from a specially constituted publication and data access committee comprising consortium members and/or independent experts from developed and developing countries. The decisions of such a committee should be binding.

### Who has the right to publish consortium-derived data first?

In multi-institutional collaborations, particularly those focusing on genomic variance, partners usually share their site-sourced data/specimens with all members of the consortium. These are analyzed by designated consortium partners—in many instances, the principal investigator—and ultimately form part of a common database accessible to all consortium members. Site-specific genotypic results are usually also made available to the contributing partner when these become available. Data sharing guidelines of major sponsors (NIH: [http://grants1.nih.gov/grants/policy/data\\_sharing/index.htm](http://grants1.nih.gov/grants/policy/data_sharing/index.htm); Wellcome Trust: [http://www.wellcome.ac.uk/print/WTX035045\\_print.html](http://www.wellcome.ac.uk/print/WTX035045_print.html); MRC, UK: <http://www.mrc.ac.uk/PolicyGuidance/EthicsAndGovernance/DataSharing/PolicyonDataSharingandPreservation/index.htm>; National Academies of Sciences, USA: [http://books.nap.edu/catalog.php?record\\_id=10613#toc](http://books.nap.edu/catalog.php?record_id=10613#toc)) are silent on whether parties are entitled to publish an analysis of their results using clinical and demographic data at their disposal, in combination with consortium-derived but site-specific genotypic data, prior to a

consortium-wide publication that focuses on an overall analysis of the cumulative data.

MalariaGEN investigators—who are faced with such a dilemma—recommend that a data-sharing policy begin by demarcating resources—samples, data, and infrastructure—that will be shared across the consortium and those that remain in the domain of individual investigators. (Chokshi et al. 2006) They stress that a clear distinction be made between “consortium experiments” (where data is analyzed across the whole consortium) and “investigator-initiated analyses” (where individual collaborator groups utilize the data they have collected together with any data that may have been generated on those samples in the consortium experiments. We endorse Malaria-Gen’s recommendations on how these resources should be managed. From the outset of its existence, MalariaGEN established an in-house ethics team to address data sharing and other consortium-related issues. This is a model that other consortiums may want to consider adopting.

If a consortium is not governed by a prospectively negotiated data sharing policy that governs data release, collaborating partners should weigh the costs and benefits of publishing the site-specific genotypic data before a group publication. If the publication of the site-specific data threatens the intellectual property interests of other consortium partners or the consortium as a whole, the consortium partner seeking to publish its work ahead of a consortium-wide work ought to desist from doing so. On the other hand, consortium-wide works are dependent on consensus being reached between consortium partners on key findings, are sometimes subject to authorship sequence disputes, and, as such, can take long to reach the public domain. Such instances could unnecessarily delay the reporting of important site-specific genetic data by individual partners and could be detrimental to scientific advances in that field. This could negatively impact on those who could most benefit from that information. In the event of the latter, a strong argument can be made for the site-specific data of collaborating partners to be published ahead of a consortium-wide work. In the event of a dispute arising between consortium partners on data release, the dispute should be referred to the aforementioned proposed publication and data access committee. In such instances, the onus of proving the merits of a particular data release strategy should rest with its proponents.

### **Are consortium partners entitled to share their data derived from the consortium with non-consortium members?**

Unless a consortium is bound by a data sharing agreement the sharing of consortium data with non-consortium

members by individual consortium partners could undermine consortium-wide research outputs and have intellectual property implications for those concerned. In such instances, data derived from consortium collaborations ought to be considered the proprietary interests of the consortium as a whole. CHAVI’s service contract, which is entered into by all CHAVI members and collaborators, stipulates that CHAVI collaborators and members “may at their sole discretion share their own respective other data that has not been published or otherwise publicly disclosed...” with non-CHAVI scientists subject to authorization being obtained from the non-profit entity [such as a university] that generated the proprietary material and a confidential disclosure agreement being entered into between the relevant parties (CHAVI: amended and restated research consortium agreement, paragraph 8.1.4). In the absence of being bound by a prospectively negotiated data-sharing or consortium agreement governing this issue, genetic data derived from consortium-wide endeavors ought not to be shared with non-consortium members unless authorization has been obtained from the aforementioned publication and data sharing committee. Members of this committee who have conflicts of interests in a matter at hand ought to recuse themselves.

### **Conclusion**

A data sharing policy should be prospectively negotiated and concluded between collaborators as early as possible. We urge sponsors of research, including those from developing countries, to issue detailed guidance on the above and related issues as doing so will facilitate research and catalyze scientific progress. We also invite others to contribute to this important discourse by sharing their perspectives on data sharing and research output governance.

### **Summary**

Good practice recommendations on data sharing

1. Consortium data sharing and intellectual property agreements should be negotiated and concluded among collaborators as early as possible. Research collaborators from the developing world should be prospectively involved in the drafting of such documents to ensure that data access and research output issues are explicitly and meaningfully governed.
2. A data sharing policy should describe the management of intellectual property rights related to the proposed project, including plans for sharing data, information, and materials resulting from the award. It should

- demarcate resources—samples, data, and infrastructure—that will be shared across the consortium and those that remain in the domain of individual investigators. The policy must be specific about the nature of the data to be shared, the timing and means of release, and any constraints on release.
3. Local communities should have a say in the management and sharing of site-specific data relating to them.
  4. Principal investigators and partners from affluent countries have a moral responsibility to build the capacity of their developing world colleagues to enable them to competently undertake the research in question and contribute to consortium-wide research outputs.
  5. In privately funded research initiatives which are not governed by international guidelines, and in instances where recommended IP-related clauses in a data sharing or consortium agreement are absent, data derived from consortia-wide efforts should be considered the proprietary interests of the consortium as a whole.
  6. Authorization to publish data derived from consortium-wide efforts ought to be obtained from a specially constituted publication and data access committee comprising consortium members and/or independent experts from developed and developing countries.
  7. If the publication of site-specific data threatens the intellectual property interests of other consortium partners or the consortium as a whole, the consortium partner seeking to publish its work ahead of a consortium-wide work ought to desist from doing so.
  8. Where delays in the reporting of important site-specific genetic data by individual partners could be detrimental to scientific advances in that field and to those who could most benefit from that information, site-specific data of collaborating partners should be published ahead of a consortium-wide work, subject to disputes related to such dissemination being settled by a consortium publication and data access committee.
  9. Data derived from consortium-wide endeavors ought not to be shared with non-consortium members unless authorization has been obtained from a publication and data sharing committee.
- Acknowledgments** The primary source of funding for this project is a grant from the Bill & Melinda Gates Foundation through the Grand Challenges in Global Health Initiative. Jerome Singh receives support from the Centre for the AIDS Program of Research in South Africa, which forms part of the Comprehensive International Program of Research on AIDS funded by the National Institute of Allergy and Infectious Disease at the National Institutes of Health and the US Department of Health and Human Services (grant #1 U19 AI51794). Abdallah Daar is supported by the McLaughlin Centre for Molecular Medicine, University of Toronto. The funders had no role in the decision to submit or preparation of this paper other than providing comments as described below.
- Competing Interests** The authors work on the Ethical, Social and Cultural Program for the Grand Challenges in Global Health initiative, which is funded by the Bill & Melinda Gates Foundation. An earlier version of this paper was circulated to the Grand Challenges in Global Health community, including program officers of the supporting foundations, who had the opportunity to provide comments. JAS is funded by CAPRISA, which is a CHAVI collaborating centre.
- Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.
- References**
- CHAVI: amended and restated research consortium agreement. Paragraph 8.1.4
- Chokshi DA, Parker M, Kwiatkowski DP (2006) Data sharing and intellectual property in a genomic epidemiology network: policies for large-scale research collaboration. *Bull World Health Organ* 84:382–387
- Hogue CJR (1991) Ethical issues in sharing epidemiological data. *J Clin Epidemiol* 44(Suppl 1):103S–107S