

CORRESPONDENCE

# Towards a data sharing Code of Conduct for international genomic research

Bartha Maria Knoppers<sup>1\*</sup>, Jennifer R Harris<sup>2</sup>, Anne Marie Tassé<sup>1</sup>, Isabelle Budin-Ljøsne<sup>2</sup>, Jane Kaye<sup>3</sup>, Mylène Deschênes<sup>4</sup> and Ma'n H Zawati<sup>1</sup>

## Abstract

Data sharing is increasingly regarded as an ethical and scientific imperative that advances knowledge and thereby respects the contributions of the participants. Because of this and the ever-increasing amount of data access requests currently filed around the world, three groups have decided to develop data sharing principles specific to the context of collaborative international genomics research. These groups are: the international Public Population Project in Genomics (P<sup>3</sup>G), an international consortium of projects partaking in large-scale genetic epidemiological studies and biobanks; the European Network for Genetic and Genomic Epidemiology (ENGAGE), a research project aiming to translate data from large-scale epidemiological research initiatives into relevant clinical information; and the Centre for Health, Law and Emerging Technologies (HeLEX). We propose seven different principles and a preliminary international data sharing Code of Conduct for ongoing discussion.

## Introduction

As early as 2002, the International Ethics Committee of the Human Genome Organization (HUGO) stated that human genomic databases should be considered as global public goods [1]. In this statement, global public goods were defined as goods 'whose scope extends worldwide, are enjoyable by all with no groups excluded, and when consumed by one individual, are not depleted for others' [2]. Buttressed by the Bermuda Principles of 1996 [2] and mirrored in the Fort Lauderdale rules of 2003 [3], the common philosophy of sharing resources was reaffirmed in the 2008 International Summit on Proteomics Data

Release and Sharing Policy in Amsterdam [4] and in the Toronto International Data Release Workshop of 2009 [5].

Finally, in January 2011, 17 major health funding agencies signed a joint statement on sharing research data to promote and improve public health [6]. However, the challenge is to take these fundamental values of sharing and access and to develop guiding principles and procedures that can be used as a basis for emerging practice.

To begin, we consider data sharing as a form of data processing as defined by the EU Directive 95/46/EC on data protection [7]. In this directive, data processing refers to: 'any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as [...] retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available [...]' [7]. Data can include raw data, genotype/phenotype data and data included within governmental health administrative databases. Theoretically, personal medical records could be subsumed under this term, but we have not specifically addressed such data because their regulation is jurisdiction-specific. The code's principles, however, remain pertinent to such data. For the terms 'coded' and 'anonymized', we use the definitions provided by the 2007 International Conference on Harmonization [8].

Data sharing is regarded as essential for enabling and promoting genomic research in a way that will maximize the benefits to public health [6] and society [9]. All countries, funders and investigators are aware of the need for research ethics and governance mechanisms in research, but currently there is little policy guidance that is specific to the international sharing of genomic research data. In view of the recent calls for the development of common principles applying to data access and use [7,10], Public Population Project in Genomics (P<sup>3</sup>G) [11], European Network for Genetic and Genomic Epidemiology (ENGAGE) [12] and Centre for Health, Law and Emerging Technologies (HeLEX) [13] are working on an international data sharing Code of Conduct (Box 1). This has a dual purpose: to elucidate shared values and to provide guidance on the basic obligations

\*Correspondence: bartha.knoppers@mcgill.ca

<sup>1</sup>Department of Human Genetics, McGill University, 740 Dr Penfield Avenue, Montreal, Quebec H3A 1A4, Canada

Full list of author information is available at the end of the article

## Box 1. International Data Sharing Code of Conduct

### Preamble

This proposed international data sharing Code of Conduct seeks to promote greater access to and use of data in ways that are (as proposed by the joint statement by funders of health research [6]):

'Equitable: any approach to the sharing of data should recognize and balance the needs of researchers who generate and use data, other analysts who might want to reuse those data, and communities and funders who expect health benefits to arise from research.

Ethical: all data sharing should protect the privacy of individuals and the dignity of communities, while simultaneously respecting the imperative to improve public health through the most productive use of data.

Efficient: any approach to data sharing should improve the quality and value of research and increase its contribution to improving public health. Approaches should be proportionate and build on existing practice and reduce unnecessary duplication and competition.'

### Principles and Procedures

#### 1. Quality

Irrespective of the discipline, scientists involved in data sharing should be *bona fide* researchers.

Proof of academic or other recognized peer reviewed standing is essential.

Harmonization of data collection and archiving methods and tools ensures validation of scientific quality.

Collaboration promotes efficiency, sustainability and comparability.

#### 2. Accessibility

Facilitation of both the deposit of data and secure access to data are the foundations of data sharing.

Curators of databases should promote sharing to generate maximum value.

Harmonization of deposit, access procedures and use promotes accessibility, equity and transparency.

#### 3. Responsibility

Responsible governance should be shared between funders, generators and users of data.

Investments in databases require coordination, strategy and long-term core funding.

Mechanisms for building interoperability should be encouraged and appropriate management anticipated.

Capacity building and recognition of all the data generators contributes to best practices.

#### 4. Security

Trust and the promotion of data sharing rely on data management and security mechanisms and also on oversight of their functioning.

Mechanisms for identifying and tracking data generators and users should be international.

#### 5. Transparency

Key policies on publications, intellectual property, and industry involvement should be public.

Websites that are accessible to the general public serve to provide feedback on progress and general results.

#### 6. Accountability

Inter-agency co-operation and funding fosters streamlined and efficient monitoring and good governance.

Provisions should be made for ongoing public engagement that is tailored to the nature of the database and local cultures.

#### 7. Integrity

Mutual respect between all stakeholders is founded on personal and professional integrity.

Prevention of harms and anticipation of public concerns and scientific needs through foresight mechanisms encourage the development of common, prospective policies.

Irresponsible research practices should be reported.

Sanctions for breach of this Code or of other legal or ethical obligations must be clear.

flowing from it. Given the varied disciplinary backgrounds of researchers working in genomic research, it can no longer be presumed that all the scientists engaged in data sharing are bound by the same medical or other professional deontological frameworks or can be subject to disciplinary action for a breach. Therefore, the proposed international Code of Conduct for data sharing in

genomic research seeks to provide common guidance on the basis of two fundamental values: (i) mutual respect and trust between scientists, stakeholders and participants; and (ii) a commitment to safeguarding public trust, participation and investment. The elaboration and eventual implementation of such a code should be the object of ongoing discussion and will begin with a series

of consultative discussions at international, European and national fora.

### **Principles and procedures: background and rationale**

Although we are not attempting to prioritize or in any way create a hierarchy among various principles in the field of data sharing, they all derive from a shared belief in maximizing both scientific quality (Box 1, point 1) and public benefit through rapid release and public accessibility to data (Box 1, point 2) [14].

The assurance of quality is *sine qua non* for ethical science. Making it an explicit requirement reiterates its importance and mandates comparison, validation and replication, thereby ensuring appropriate and common standard operating procedures and the use of accredited facilities. Prospectively harmonizing procedures to facilitate interoperability and comparability is likely to promote such quality and accessibility.

There is no doubt that maximizing public benefit, investment and participation is facilitated through data sharing. Not only should access be equitable for researchers in both the public and private sectors, but ethics reviewers should have the proper training and tools to evaluate international requests. The datasets themselves may be derived from the contributions of multiple sources from different countries and projects. The current legal and ethical constraints and bottlenecks to access are obvious. Indeed, multiplicity of ethics review may well be the Achilles heel for efficient sharing.

The tripartite responsibility of the data producers, users and funders lays the foundation for data sharing (Box 1, point 3). We see data sharing, which is often a condition of funding, as part of the efficient and proper stewardship of public funds. It also binds eventual users in the recognition of a just return on public investment and participation. This responsibility is chiefly expressed both in the security mechanisms that translate the principle into the construction of information technology tools and firewalls and in the governance framework.

### **Security mechanisms**

Security mechanisms (Box 1, point 4) go well beyond the application of firewalls or de-identification techniques, such as coding or anonymization. Indeed, unique, digital identifiers (IDs) for biobanks [15,16] and for researchers [17] have been proposed not only for security purposes but to facilitate access. Such IDs would enable verification and validation of the identities and credentials of researchers by institutions and would become a mechanism for allowing, tracking and auditing access, as well as attributing contributions.

Digital identifier systems allow data tracing and prospectively limit the potential for malicious activities

involving re-identification of participants. This transparency of data flow, access and use also curtails the possibility of pre-publication scooping between producers and users (Box 1, point 5). Pre-publication data release depends on the respect by users and journals of publication moratoria that allow data producers to share data openly but provides a period of time to analyze and publish their own data before secondary users do so. Proper acknowledgement of the use of data resources also allows funders to track their 'investments.' It allows the public to see that their altruistic participation has led to fruitful scientific endeavors. Most importantly, data users agree not to use intellectual property protection in ways that would prevent or block access to, or use of, any element of the dataset or any conclusion drawn directly from it [18]. This does not prevent further research with attendant intellectual property rights in downstream discoveries provided that the best practices for licensing policies for genomic inventions are followed.

### **Governance framework**

Good governance underpins a system of data sharing that depends on trust. Approaches to governance necessarily vary between contexts and countries. Irrespective of these differences, governance should be flexible in the oversight and monitoring systems put in place. This is crucial because public trust, which is increasingly translated through broad consents, is counterbalanced by both security systems and governance. It could be asked whether in considering the longevity of large international datasets, including samples, separate governance models should be developed as distinct from local institutional mechanisms or those applicable to the oversight of clinical trials.

Good governance assures the public and funders of proper accountability and ethics review (Box 1, point 6). Although local laws and ethics review systems vary, the ethics norms and biobank policies applicable to large data repositories are beginning to emerge [19,20]. These common norms are increasingly mirrored in model material transfer and access agreements [10]. Contractual in nature, they serve to bind researchers and their institutions. Implicit in such agreements are the very principles under discussion here. By making them explicit by using such contracts, researchers, policymakers and ethics committees have tools to work with that are more transparent. For scientific integrity (Box 1, point 7) to be viable, discussion on the nature of such principles and their procedural translation in different contexts will necessarily vary. Nevertheless, mutual respect between all stakeholders and participants can be built on these fundamental principles and procedures. Integrity also entails the prevention of harms, anticipation of public concerns and scientific needs as well as the reporting of

irresponsible research practices and the creation of appropriate sanctions [21].

Most importantly, ongoing communication with the public on the 'reality' of data sharing principles and procedures is essential. Thus, lay summaries of the research proposals accessing and using data repositories should be publicly posted. Although there is no personal benefit to participants, such a public registry of research uses ultimately allows participants to withdraw if they disagree with the direction of the research. There are also other mechanisms of communication, such as bulletins and websites. Population studies recontact their participants for updates, or to take new measurements, thereby keeping ongoing consent alive and valid.

The most telling aspect of the developments described above, however, is that the underlying values presented here come from the current approaches promoted and used by the scientists and funders themselves. Concern for scientific integrity and mutual respect are then not imposed by legislative or professional fiat but rather reveal an already existing shared ethos on the proper foundations for international science in the 21st century. This augers well for the future viability of the preliminary version of our proposed international data sharing Code of Conduct in genomic research (Box 1).

## Conclusion

Addressing the issue of data sharing in the context of international genomic research requires not only a holistic approach, but also the fair balancing of the interests, rights and duties of various stakeholders involved in collaborative endeavors. We have highlighted the need for equitable, ethical and efficient access to data and proposed a Code of Conduct (Box 1) that incorporates seven principles: quality, accessibility, responsibility, security, transparency, accountability and integrity. We trust that this code will foster broader discussion involving multiple stakeholders.

## Abbreviations

ENGAGE, European Network for Genetic and Genomic Epidemiology; HeLEX, Centre for Health, Law and Emerging Technologies; HUGO, Human Genome Organization; P<sup>3</sup>G, Public Population Project in Genomics.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

BMK wrote the first draft of the manuscript; JRH, AMT, IBL, JK, MD and MHZ contributed equally to the manuscript. All authors read and approved the final manuscript.

## Acknowledgements

The authors would like to thank Michael Le Huynh for his assistance in editing this article.

## Author details

<sup>1</sup>Department of Human Genetics, McGill University, 740 Dr Penfield Avenue, Montreal, Quebec H3A 1A4, Canada. <sup>2</sup>Norwegian Institute of Public Health,

PO Box 4404, Nydalen, N-0403 Oslo, Norway. <sup>3</sup>Department of Public Health, University of Oxford, Richards Building, Old Road Campus, Headington, Oxford OX3 7LF, UK. <sup>4</sup>P<sup>3</sup>G, Suite 590, 3333 Queen-Mary Road, Montreal, Quebec H3V 1A2, Canada.

Published: 14 July 2011

## References

1. Human Genome Organisation (HUGO), Ethics Committee: **Statement on human genomic databases, December 2002.** *J Int Bioethique* 2003, **14**:207-210.
2. Human Genome Organisation (HUGO): *Principles Agreed at the First International Strategy Meeting on Human Genome Sequencing: 25-28 February 1996; Bermuda.* HUGO; 1996. [http://www.gene.ucl.ac.uk/hugo/bermuda.htm]
3. **Sharing Data from Large-scale Biological Research Projects: A System of Tripartite Responsibility. Report of a meeting organized by the Wellcome Trust and held on 14-15 January 2003 at Fort Lauderdale, USA** [http://www.sanger.ac.uk/datasharing/docs/fortlauderdalereport.pdf]
4. Rodriguez H, Snyder M, Uhlén M, Andrews P, Beavis R, Borchers C, Chalkley RJ, Cho SY, Cottingham K, Dunn M, Dylag T, Edgar R, Hare P, Heck AJ, Hirsch RF, Kennedy K, Kolar P, Kraus HJ, Mallick P, Nesvizhskii A, Ping P, Pontén F, Yang L, Yates JR, Stein SE, Hermjakob H, Kinsinger CR, Apweiler R: **Recommendations from the 2008 International Summit on Proteomics Data Release and Sharing Policy: The Amsterdam principles.** *J Proteome Res* 2009, **8**:3689-3692.
5. Birney E, Hudson TJ, Green ED, Gunter C, Eddy S, Rogers J, Harris JR, Ehrlich SD, Apweiler R, Austin CP, Berglund L, Bobrow M, Bountra C, Brookes AJ, Cambon-Thomsen A, Carter NP, Chisholm RL, Contreras JL, Cooke RM, Crosby WL, Dewar K, Durbin R, Dyke SO, Ecker JR, El Emam K, Feuk L, Gabriel SB, Gallacher J, Gelbart WM, Graneli A, et al.: **Prepublication data sharing.** *Nature* 2009, **461**:168-170.
6. **Sharing research data to improve public health: full joint statement by funders of health research** [http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Public-health-and-epidemiology/WTDV030690.htm]
7. **EU Directive 95/46/EC - The Data Protection Directive** [http://www.dataprotection.ie/viewdoc.aspx?DocID=92]
8. **International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use: Definitions for Genomic Biomarkers, Pharmacogenomics, Pharmacogenetics, Genomic Data and Sample Coding Categories E15** [http://www.ich.org/fileadmin/Public\_Web\_Site/ICH\_Products/Guidelines/Efficacy/E15/Step4/E15\_Guideline.pdf]
9. **Data storage and DNA banking for biomedical research: technical, social and ethical issues.** *Eur J Hum Genet* 2003, **11** Suppl 2:S8-S10.
10. O'Brien SJ: **Stewardship of human biospecimens, DNA, genotype, and clinical data in the GWAS era.** *Annu Rev Genomics Hum Genet* 2009, **10**:193-209.
11. **Public Population Project in Genomics** [http://www.p3g.org]
12. **European Network for Genetic and Genomic Epidemiology** [http://www.euengage.org]
13. **Centre for Health, Law and Emerging Technologies** [http://www.publichealth.ox.ac.uk/helex/]
14. Laurie G, Mallia P, Frenkel DA, Krajewska A, Moniz H, Nordal S, Pitz C, Sandor J: **Managing Access to Biobanks: how can we reconcile individual privacy and public interests in genetic research?** *Med Law Int* 2010, **10**:315-337.
15. Cambon-Thomsen A: **Assessing the impact of biobanks.** *Nat Genet* 2003, **34**:25-26.
16. Kaufmann F, Cambon-Thomsen A: **Tracing biological collections: between books and clinical trials.** *JAMA* 2008, **299**:2316-2318.
17. **GEN2PHEN Knowledge Centre** [http://www.gen2phen.org]
18. **International Cancer Genome Consortium** [http://www.icgc.org]
19. **OECD principles and guidelines for access to research data from public funding** [http://www.oecd.org/dataoecd/9/61/38500813.pdf]
20. **Guidelines on human biobanks and genetic research databases (HBGRDs)** [http://www.oecd.org/sti/biotechnology/hbgrd]
21. **Singapore Statement on Research Integrity** [http://www.singaporestatement.org/statement.html]

doi:10.1186/gm262

**Cite this article as:** Knoppers BM, et al.: **Towards a data sharing Code of Conduct for international genomic research.** *Genome Medicine* 2011, **3**:46.