Submission to the Nuffield Council on Bioethics' Consultation on "Biological and health data--The collection, linking, use and exploitation of biological and health data: ethical issues"

Shawneequa Callier[1], JD, MA, Harald Schmidt[2], PhD, MA

January 10, 2013

Dear Prof Richards, dear colleagues in the Council's Secretariat:

Thank you for the opportunity to contribute a consultation response to the current Working Party on Biological and Health Data. We are enclosing here comments on the discreet issue of the taxonomy used for genetic data, focusing in particular, on so-called 'anonymous' samples. While we provide more detailed comment below, our basic points can be summarized as follows:

- High level guidance documents suggest that withdrawal of consent and samples and the provision of feedback are impossible in the case of anonymous samples. In view of recent developments in science and consumer-driven genomics such statements are misleading. They only muddle complex ethical questions about possible entitlements to control over samples.
- We therefore propose that terms such as 'anonymised', 'anonymous' or 'non-identifiable' be removed entirely from documents describing research samples—especially from those aimed at the public. This is necessary as a matter of conceptual clarity and because failure to do so may jeopardise public trust in the governance of large scale databases.
- As there is wide variation in the taxonomy for tissue samples and no uniform national or international standards, we also propose that a numeral-based universal coding system be implemented that focuses on specifying incremental levels of identifiability, rather than use terms that imply that the reidentification of research samples and associated actions are categorically impossible.
- Given the data-sharing and data-mining practices of today, standards for deidentification may be inadequate to protect individuals and participants in population research.

For full disclosure, we should also add that that this contribution has another report of the Council at its origin, and involved several individuals associated with the Council in different roles in the past: Harald Schmidt first came to think about the issue while working on the Secretariat's side on the Council's 2003 Report on Pharmacogentics. A short section along similar lines found its way into the report (see paragraph 3.33 and environs). The first draft of the paper that forms the basis of this submission was then produced with Caroline Rogers, one of the Council's Research Officers. The paper was later continued with Shawneequa Callier, a former intern at the Council. Julia Trusler and Varsha Jagadesham, both Research Officers, also provided critical comment on different stages of the manuscript.

[1] Assistant Professor, Department of Clinical Research and Leadership, School of Medicine and Health Sciences, The George Washington University, USA.
[2] Assistant Professor, Department of Medical Ethics and Health Policy, Research Associate, Center for Health Incentives and Behavioral Economics, Perelman School of Medicine, University of Pennsylvania, USA.

We hope that our submission is of use to the Working Party's deliberations, and would be happy to clarify any points that should require further discussion.

Sincerely,

Shawneequa Callier, Harald Schmidt

What are the new privacy issues?
The storage, re-use, and linking of biological data with health information, along with the magnitude of data being stored, are intensifying existing challenges related to privacy for individuals and groups.

The summary below describes our work highlighting threats to the concept of "anonymity" of samples used in research. The possibility of reidentifying sample contributors to research because of global data-sharing practices is the subject of much debate and a source of public consternation about genomic research.(1-8)

Note: the text below includes excerpts from a paper published as: Schmidt, H., Callier, S. 2012. How anonymous is 'anonymous'? Some suggestions towards a coherent universal coding system for genetic samples. *Journal of Medical Ethics*. 38.5 (2012): 304-309.

1. **Re-identification of alleged anonymous samples**

While many research participants' names and identifying characteristics are routinely removed from genomic samples and data used in research, reidentification of participant samples and data is possible using publicly available databases, thus raising the question of whether genetic information can ever truly be considered deidentified.(3, 9)

The use of so-called 'anonymous' tissue samples and data have historically justified the sharing, storage, and re-use of patient and research participant materials with limited consent. For example, using 'anonymous' or 'anonymised' biological samples is widely perceived as an appropriate way of reconciling conflicts between the control, privacy and confidentiality interests of those from whom the samples originated and the public (or commercial) interest in carrying out research.(10) Although scholars debate the gravity of the risks raised by reidentification, with some challenging the sufficiency of deidentification and anonymization for protecting privacy and autonomy interests (3, 11) and others maintaining that re-identification is difficult and rare (12)—scientific discoveries made in recent years show that data can indeed be reidentified if several steps are taken by third parties.(1, 4) Claims that feedback or return of results is impossible due to the "anonymous" status of samples is therefore disingenuous because of reidentification capabilities.

As we explain in "How anonymous is 'anonymous'?" significant ambiguity and semantic confusion exists internationally because of the diferent terms used to describe de-identification. The best way of avoiding this situation would be to remove terms such as 'deidentified,' 'anonymised' and 'anonymous' from policy documents at all levels, especially those aimed at patients or research participants. The same goes for 'non-identifiable' samples or samples where links to individuals are 'irreversibly deidentified.' Instead, it would seem possible to adopt a numerical instead of a verbal classification system to signify different degrees of identifiability.

2. **The adequacy of standards for anonymization/de-identification**

As global data-sharing continues, the following will be critical (1) international privacy laws (2) data-mining and data-selling, and (3) inconsistencies in data-sharing policies globally.

a. Privacy Rules

Regulatory consent policies typically only govern identifiable samples. For example, in the USA and the UK, consent is usually required for the use of research samples that can be linked to the person from whom they came. Where biological samples undergo a process of deidentification or anonymisation, however, individual consent is typically not required by law. Under policies promulgated by the Office of Human Research Protections in the US, research involving samples and data that "cannot be linked to specific individuals by the investigator(s) either directly or through coding systems" are not considered to be human subjects research –and therefore do not require consent to be used and shared in different research protocols.(13) Similarly, under HIPAA, the US privacy and security law that governs heath care providers, insurers, and health care clearinghouses, "protected health information" that is no longer deemed 'individually identifiable health information' may be used in research without authorization. Equally, under the UK's Human Tissue Act 2004, tissue from the living may be stored and/or used without consent for health related research when such projects have been ethically approved and 'the tissue is anonymised such that the researcher is not in possession of information identifying the person from whose body the material has come and is not likely to come into possession of it.'[3]

In the United States alone, however, numerous laws govern the use of samples and data inconsistently—and coverage and protection will depend on the original use and procurement of samples and medical information. For example, data containing identifiable beneficiary or physician information are protected by the Privacy Act of 1974 (which binds only Federal agencies and applies only to records in the possession and control of Federal agencies),[4] the Health Insurance Portability and Accountability Act of 1996 (HIPAA), state law, and other federal laws, such as the Common Rule and the FDA Protection of Human Subjects Regulations. Medicare providers, for instance, must comply with the Privacy Rule and the Privacy Act of 1974.

A number of scholars have debated whether these privacy rules adequately secure patient and research participant data, (3, 11, 14, 15) with many of the earlier concerns raised largely within the context of hypothetical scenarios. Research published recently, however, details how investigators reidentified research participants using genealogy information publicly available on the internet and comparing said information with deidentified genomic data.(1, 2, 12, 16) While genetic information does not itself identify an individual in the absence of other identifying information, it is clear that a person's genetic code could be construed as a unique identifier in that it could be used to match a sequence in one biospecimen bank with another sample from a databank that does include identifiers (17-19).

b. Data-selling and data-mining

Data-mining and data-sharing practices aggravate the privacy challenges raised by the inadequacies of current privacy laws. The Privacy Rule, for instance, does not apply to everyone in the health-care industry. For example, employers often house large amounts of health information, but generally employers are not covered health entities under HIPAA. Other parties

[3] http://www.hta.gov.uk/_db/_documents/2006-07-04_Approved_by_Parliament_-_Code_of_Practice_1_-_Consent.pdf

[4] http://www.hhs.gov/foia/privacy/

such as advertising companies, life insurance companies, data-mining companies, financial institutions, and social networking websites operate outside of the jurisdiction of most health privacy laws, including HIPAA. Further, many of these entities'' privacy policies could be changed at any time.

Meanwhile, deidentified data is sometimes sold to insurance companies and data mining companies. Pharmacies, for instance, sell prescription drug data to data mining companies, which in turn sell this data to insurers and drug manufacturers. Drug manufacturers are not covered entities, but insurers may use patients' identifiable health information for operational purposes. While it is clear that insurers may not sell protected health information or use protected health information in ways that contravene the Privacy Rule, they may sometimes buy it and access PHI in ways that insured individuals do not anticipate. Importantly, we have limited information about the extent of data-mining and the impact of data-selling among marketing professionals on individuals' privacy.

3. Inconsistent data-sharing policies used in global research

Despite all the great efforts to engage and partner with communities, it is unclear how well the populations who participate in genomic research are protected under current legal and policy regimes. The enhanced capability of investigators globally to combine large and complex datasets in an environment when genomic data taken from different populations are subject to local beliefs and regulatory systems is one major hurdle.

Because of the social and scientific value of gathering and analyzing large data sets containing genomic information—prevailing international laws favor a collaborative research environment over an absolute individual right to control one's genetic samples and data. Due to these principles, investigators often have free access to pre-publication sequence data that was rapidly released into public nucleotide sequence databases (20)--many of which contain information from various global populations. Data from genome-wide association studies (GWAS)— studies that measure hundreds of thousands of genetic mutations to identify common genetic traits that influence diseases—for instance, are stored in such databases. These same types of studies were singled out as possibly creating reidentification risks for research participants, causing national funders to limit access to genomic data repositories using controlled-access databases that require a data access committee to grant access.(4, 21)

Proposed changes to current rules require that individuals be informed explicitly that their data will be deposited into open or controlled-access databases,(22) but such processes do not protect individuals who belong to a group that is characterized as being genetically predispositioned to alcoholism or crime, for instance, and who could be stigmatized whether or not they participate in research.(23, 24). Nor do these policies guarantee that third parties, especially those located in other countries will follow the rules and policies of the original procurers of samples and data. In the United States, for instance, personal health information belonging to a person who has been deceased for 50 years may be used in research without an authorized representatives' consent.(25) Such a policy could raise controversy among members of specific groups in indigenous communities in the US or abroad.

Generally, the legal system is not set up to handle international or indigenous communities' claims to property rights and privacy rights. Currently, multi-institutional agreements govern investigators, and require them to follow data-release policies that are sensitive to the needs of local participants in research. But as we have seen in the case of Henrietta Lacks, various gaps in US and international laws provide opportunity for the unethical sharing of genomic data. Henrietta Lacks and her family are now famously known among the general public due to Rebecca Skloot's 2010 best-selling book, "The Immortal Life of Henrietta Lacks," which details the emotional struggles of the Lacks family in grappling with the unethical handling of Ms. Lacks' identity, privacy, and self-determination over her biological materials. Members of the public and research community have similarly pointed to disrespect for Lacks' private life and autonomy, and the unjust enrichment of investigators who have benefited from their research on the HeLa cells while her family remained poor.(27, 28) Although the HeLa story has caused a stir among members of various socioeconomic communities throughout the United States—leading to a public apology by Johns Hopkins University to the Lacks family—Skloot's publication did not stop researchers in Germany from publishing her full genome earlier this year in 2013.(29)

As a result of much public criticism, the investigators removed the data from the public domain and other papers were put on hold while the National Institutes of Health (NIH) initiated a negotiation with the Lacks family.(30) After a series of talks, NIH established the HeLa Genome Data Access Working Group of the Advisory Committee to the Director, which is a special committee charged with reviewing the release of data from the HeLa Genome with input from two serving members of the Lacks family.(30) These laudable efforts on the part of leadership in the biomedical research community demonstrate a swift and thoughtful response to the Lacks family's disgruntlement with the investigators' failure to consult with them about publicizing information that could reveal heritable traits and information about Ms. Lacks and her descendants. They also represent a response from the research community to the public's outcry over the privacy violations experienced by the Lacks family.

Considering the international nature of the recent HeLa controversy and the privacy implications for the Lacks family, challenges related to global data sharing and the potential for family and group harm resulting from data sharing are clearly important. As other countries build their own genomic research laboratories and data collection processes, more guidance is needed on how to protect research participants globally.(26)

References

1. Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y. Identifying personal genomes by surname inference. Science. 2013 Jan 18;339(6117):321-4.
2. Rodriguez LL, Brooks LD, Greenberg JH, Green ED. Research ethics. the complexities of genomic identifiability. Science. 2013 Jan 18;339(6117):275-6.
3. Schmidt H, Callier S. How anonymous is 'anonymous'? some suggestions towards a coherent universal coding system for genetic samples. J Med Ethics. 2012;38(5):304-309.
4. Couzin J. Whole-genome data not anonymous, challenging assumptions. Science. 2008;321(5894):1278.

5. Wjst M. Caught you: Threats to confidentiality due to the public release of large-scale genetic data sets. BMC Med Ethics. 2010 Dec 29;11:21.
6. Elger BS, Caplan AL. Consent and anonymization in research involving biobanks: Differing terms and norms present serious barriers to an international framework. EMBO Rep. 2006 Jul;7(7):661-6.
7. Oliver JM, McGuire AL. Exploring the ELSI universe: Critical issues in the evolution of human genomic research. Genome Med. 2011;3(38).
8. Callier SL, Schmidt H. Managing patient expectations about deidentification. Am J Bioeth. 2010 Sep;10(9):21-3.
9. Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y. Identifying personal genomes by surname inference. Science. 2013;339:321-4.
10. Lowrance WW, Collins FS. Ethics. identifiability in genomic research. Science. 2007;317:600-2.
11. Rothstein MA. Deidentification and its discontents: Response to the open peer commentaries. Am J Bioeth. 2010 Sep;10(9):W1-2.
12. Knoppers B, Dove E, Litton J, Nietfeld JJ. Questioning the limits of genomic privacy. The American Journal of Human Genetics. 2012 9/7;91(3):577-8.
13. Abel E, Browner CH. Selective compliance with biomedical authority and the uses of experiential knowledge. In: Lock M, Kaufert P, editors. Pragmatic Women and Body Politics. p. 310-26.
14. Rothstein MA. Is deidentification sufficient to protect health privacy in research? Am J Bioeth. 2010 Sep;10(9):3-11.
15. Hoffman S. Electronic health records and research: Privacy versus scientific priorities. Am J Bioeth. 2010 Sep;10(9):19-20.
16. Gutmann A, Wagner JW. Found your DNA on the web: Reconciling privacy and progress. Hastings Cent Rep. 2013 May-Jun;43(3):15-8.
17. Malin B, Sweeney L. Determining the identifiability of DNA database entries. Proc AMIA Symp. 2000:537-41.
18. Malin B, Sweeney L. Re-identification of DNA through an automated linkage process. Proc AMIA Symp. 2001:423-7.
19. Malin BA. An evaluation of the current state of genomic data privacy protection technology and a roadmap for the future. J Am Med Inform Assoc. 2005;12:28-34.
20. Erwin C, Hersch S. Monitoring reportable events and unanticipated problems: The PHAROS and PREDICT studies of huntington disease. IRB. 2007;29(3):11-6.
21. Homer N, Szelinger S, Redman M, Duggan D, Tembe W, Muehling J, et al. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. PLoS Genet. 2008 Aug 29;4(8):e1000167.
22. DHHS. Draft NIH genomic data sharing policy request for public comments. Federal register. 2013 September 20;78(183):57860-5.
23. Hausman DM. Group risks, risks to groups, and group engagement in genetics research. Kennedy Institute of Ethics journal. 2007;17(4).
24. Hausman D. Protecting groups from genetic research. Bioethics. 2008 Mar;22(3):157-65.
25. DHHS. US department of health and human services policy for the protection of human research subjects code of federal regulations, 45 CFR 160.103. . 2013.

26. The Department of Health and Human Services and the Food and Drug Administration. Human subject research protections: Enhancing protections for research subjects and reducing burden, delay, and ambiguity for investigators. Federal register. 2011;76(143):44512-3.
27. Javitt G. Why not take all of me? reflections on the immortal life of henrietta lacks and the status of participants in research using human specimens Minnesota Journal of Law, Science & Technology. 2010;2(11):713.
28. Skloot R. The immortal life of henrietta lacks. New York: Crown; 2010.
29. Skloot R. The immortal life of henrietta lacks, the sequel. [Internet]. 2013 March 23 [cited July 19, 2013]. Available from: http://www.nytimes.com/2013/03/24/opinion/sunday/the-immortal-life-of-henrietta-lacks-the-sequel.html?pagewanted=all&_r=0.
30. Charge [Internet]. Available from: http://acd.od.nih.gov/HeLaGenomeDataAccessWG-charge.PDF.