



Published in final edited form as:

*Sci Transl Med.* 2010 January 20; 2(15): 15cm3. doi:10.1126/scitranslmed.3000361.

## Meeting the Governance Challenges of Next-Generation Biorepository Research

Stephanie M. Fullerton<sup>1,2,\*</sup>, Nicholas R. Anderson<sup>1,3</sup>, Greg Guzauskas<sup>1</sup>, Dena Freeman<sup>1</sup>,  
and Kelly Fryer-Edwards<sup>1,2</sup>

<sup>1</sup> Institute of Translational Health Sciences, University of Washington, Seattle, WA 98195, USA

<sup>2</sup> Department of Bioethics and Humanities, University of Washington, Seattle, WA 98195, USA

<sup>3</sup> Department of Medical Education and Biomedical Informatics, University of Washington, Seattle, WA 98195, USA

### Abstract

Advances in clinical translational research have led to an explosion of interest in infrastructure development and data sharing facilitated by biorepositories of specimens and linked health information. These efforts are qualitatively different from the single-center sample collections that preceded them and pose substantial new ethics and regulatory challenges for investigators and institutions. New research governance approaches, which can address current and anticipated challenges, promote high-quality research, and provide a robust basis for ongoing research participation, are urgently required.

### INTRODUCTION

Researchers have long made use of stored biospecimens (Fig. 1), and associated phenotypic and clinical data, to study the interplay of genetic and environmental influences on disease risk. Until recently, most biorepository research was based in resources created and maintained by single investigators or research teams, typically for the retrospective analysis of specific diseases, and used samples collected via established (often clinical) relationships with participants. Such “first generation” biorepository research is now giving way to forms of population-based investigation that require access to very large numbers of research participants who are followed prospectively for a wide range of traits and diseases (1,2). These “next generation” biorepository-based approaches promise substantial scientific benefit and are being enabled by national funding initiatives, such as the Clinical Translational Science Awards, that aim to create large networks of collaborating institutions and correspondingly large national data sets (3). The specimen- and data sharing entailed by these new initiatives is unprecedented and, alongside the obvious logistical and technical issues, poses a range of important ethical and regulatory challenges for investigators and institutions (4,5). Developing a productive environment for next-generation biorepository research will involve identifying and addressing these important research governance challenges.

Here, we address four common areas of concern—privacy, institutional review, informed consent, and data stewardship—and discuss how biorepository governance will need to adapt to keep pace with evolving developments in translational science (Fig. 2).

\*Corresponding author. smflrtn@u.washington.edu.

## ANONYMIZATION WILL NO LONGER SUFFICE

Research participants often cite a fear of having their genetic or personal health information used against them should it fall into the wrong hands (6,7). In much first-generation biorepository research, researchers chose to protect the confidentiality of contributors' data by one of two methods: Anonymization (identity irreversibly severed to prevent any future re-identification) or de-identification (codification but with the retention of identifying information) of linked data and specimens (8). Although some evidence suggests that research subjects may not readily distinguish between the two approaches (9), there has been an administrative preference for anonymization as the primary means to protect privacy and limit potential misuse (10,11). In the United States this preference derives, in part, from an Office of Human Research Protections policy that allows institutions to exempt from human subjects review research that uses data "recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects" [45 CFR 46.102(f)].

However, despite the best efforts of researchers to meet the demands of anonymization, a number of features of biorepository research have combined to make it increasingly hard to achieve in practice (12). These include the generation of dense genotypic information from biospecimens (13), linkage to richly detailed clinical data, and the use of sophisticated bioinformatics tools for data mining and amalgamation (14). The recent demonstration that individual participants' identities can be determined from aggregate genotypic data (15,16) has underlined further the infeasibility of most traditional approaches to anonymization.

It is no longer clear that we can promise anonymity to participants, nor is it obvious that we should: There are compelling reasons to prefer the retention of coded identifiers to anonymization. First, anonymization interferes with the ability to recontact individuals or cohorts when clinically relevant information is obtained (2,17). Second, anonymization renders withdrawal from future research, a key tenet of voluntary research participation, impossible (18). Third, when it is uncertain whether we can honor assurances regarding anonymity, it is arguably unethical to rely on such assurances at the time of recruitment (19). From a scientific standpoint too, the analytic utility of individual samples is maximized if participant health information is kept current, and this is typically only possible when identity is maintained. [The Vanderbilt BioVU model (11), in which anonymized samples are linked to a "synthetic derivative" of the electronic medical record, represents a possible exception.]

For next-generation biorepository research, anonymization will therefore no longer suffice as a means of protecting participants' privacy, nor will it provide a satisfactory basis for forgoing research oversight, particularly when broad data sharing is anticipated. Instead, renewed attention to the control and retention of coded identifiers, combined with innovative approaches to data security and research oversight, will be required.

## BENEFITS OF INSTITUTIONAL REVIEW

On the basis of current guidance, research using coded rather than anonymized data will necessarily be subject to institutional oversight and ongoing review. Some will view this inevitability as undesirable—even potentially disastrous—for the pursuit of next-generation biorepository research. Not only will many more individual research projects require review, but a sea change could entail new (potentially burdensome) interactions with participants and the restriction or delay of particular kinds of sensitive or risky research. However, many institutions already require review of research using biorepository data, whether in fact samples are anonymized, so a move to the routine use of coded de-identified data may not be as onerous as many fear. Indeed, investigators should regard oversight as an important

opportunity to align their research activities with participants' interests in, and expectations for, biorepository research.

In first-generation biorepository research, many participants were recruited by a known investigator for the purposes of advancing a specific form of, usually disease-related, research (2). Although secondary users of biorepository data might be at a physical or temporal remove from subject recruitment, the specific intentions of participants could be readily communicated by the originating investigator(s) or inferred from the types of data contributed to the resource and/or terms of consent. Next-generation biorepository research, by contrast, relies increasingly on amalgamating diverse data sources from participants recruited to independent, frequently geographically widely dispersed biorepositories, for a broad range of potential research uses (5). Even though wide sharing maximizes the range of scientific questions that can be addressed, these new research arrangements increase the distance between investigators and participants, interfering with researcher accountability (20). Explicit research oversight can help scientists keep participants' objectives and interests firmly in view as they pursue projects long after the period of initial recruitment.

Although investigators should view ongoing oversight as ultimately beneficial to the successful pursuit of biorepository research, we concede that current regulatory and institutional review procedures are not well suited to next-generation approaches. Reform is urgently needed to support harmonization across institutions (achieved currently on a more ad hoc basis by a patchwork of cooperative understandings and Data Use Agreements) and streamline review in cases in which demonstrably similar research activities or questions are being pursued. Reform will help clarify researcher responsibilities while ensuring that the rights and interests of biorepository participants are recognized and actively promoted.

## ALTERNATIVES TO ONE-TIME CONSENT

Ongoing research review also provides an important opportunity to revisit the nature of informed consent for biorepository participation. As noted above, many forms of first-generation biorepository research were invariably directed toward the investigation of specific diseases, with consent worded to reflect the anticipated research use (21). With the expansion of biorepository-related data sharing, and a desire to use stored specimens and data to address an increasingly wide range of secondary research questions, there has been a move to broaden the terms of use so that a single consent agreement will allow any secondary investigation consistent with the stored data. Although enormously convenient for investigators, such "blanket" consent approaches have been critiqued as providing an inadequate (i.e., effectively uninformed) basis for research participation (22,23). Many individuals, including those from communities that have had bad prior experiences with biomedical research, will choose not to participate on such a basis, limiting the translational promise of next-generation approaches.

A range of alternative consent models has been proposed as a result of these issues. At one extreme have been arguments for consent processes that allow for the broadest possible scope of research use and data sharing but require enrollees to demonstrate awareness of genetic principles and the potential risks of biorepository participation (19). Others favor promoting autonomy with tiered consent, so that participants indicate specific preferences for their personal data, either with respect to the types of research use or with regard to the ways that data may be shared with third parties (10,23,24). Finally, others have called for a reconceptualization of consent as an ongoing process, in which researchers remain actively engaged in communicating with participants (7,25). In the last model, it is not clear if ongoing interactions would necessarily require re-consent (26); instead, upfront agreement of participants to an explicit form of research oversight (including, ideally, direct participant

representation) might be combined with timely notification when a review identifies potential new research risks.

Each alternative has advantages and disadvantages. Nevertheless, we believe that approaches to consent that allow for the ongoing involvement of research subjects will be key to both the recruitment and long-term retention of biorepository participants. A high priority for next-generation biorepository research will be developing new methods for recontacting participants, including creative use of electronic and other modes of communication (27).

## POSITIVE DUTY FOR STEWARDSHIP

Stewardship, or the careful and responsible management of something entrusted to one's care, is central to the pursuit of research that uses stored specimens and data. Stewardship typically implies that everyone in the research workflow has a responsibility to protect human subjects' interests and well-being to the best of his or her ability (28). In most first-generation biorepository research, the burden of stewardship fell to the originating investigator or institution and was achieved by faithfulness to the terms of informed consent and the adoption of data protections like anonymization. However, with the retention of identifying information, an expectation of ongoing oversight coordinated across independent institutions, and the need to maintain communication with participants in light of the open-ended nature of the research commitment, next-generation biorepository research entails far greater demands for stewardship and researcher accountability. These responsibilities may include taking due care with the analysis and sharing of confidential genetic and linked health information, the adoption of research goals consistent with the intentions of participants, and the avoidance of forms of dissemination (publications and similar) that promote harmful or derogatory conclusions about certain populations or groups (29).

Although recognizing these expectations is essential, formally meeting the stewardship requirements of next-generation research rests critically on the adoption of defined research governance mechanisms that are effective yet flexible enough to respond to dynamic scientific, technical, and policy developments (30). Terms of data access and use must be elaborated, and governance boards or data use committees [which preferably include participant representatives (31)] established to vet proposed research uses and oversee plans for data sharing and research dissemination. Some biorepositories, such as those run by advocacy groups like PXE International (32), which promotes research about the genetic disorder pseudoxanthoma elasticum, and the Autism Genetic Research Exchange (33), review all data use applications through the lens of how likely the proposed research is to advance the science and bring society one step closer to a cure or other treatment options. These groups require data sharing and dissemination plans to facilitate accountability and speed communication through both research and practice communities. Those undertaking larger-scale efforts will no doubt find it more challenging to define the common good and target research priorities accordingly, but the increased effort will be rewarded both by heightened trust in the research enterprise (34) and by an enhanced potential to contribute to near-term translational benefits.

## CONCLUSION

It is time to acknowledge that first-generation technical and regulatory solutions are not up to the task of addressing the ethical and scientific challenges of next-generation biorepository research. Data anonymization is no longer achievable as a matter of practice, and greater attention to research review, informed consent, and ongoing stewardship of repository samples and data is urgently required instead. The pursuit of effective

translational research rests as much on our willingness to meet our responsibilities to research participants as on our ability to advance cutting-edge methodology and analytical innovation. These priorities need not—and indeed cannot—be mutually exclusive.

## Acknowledgments

This work was supported by a grant from the University of Washington (UW) Institute of Translational Health Sciences (National Center for Research Resources, UL1 RR 025014, KL2 RR 025015, and TL1 RR 025016). Additional support was provided by the UW Center for Ecogenetics and Environmental Health (National Institute of Environmental Health Sciences, P30 ES07033), the UW Center for Genomics and Healthcare Equality (National Human Genome Research Institute, P50 HG003374), and the Northwest Institute of Genetic Medicine (Washington State Life Sciences Discovery Fund). The views presented in this Commentary are those of the authors and do not necessarily reflect the views of these funding agencies.

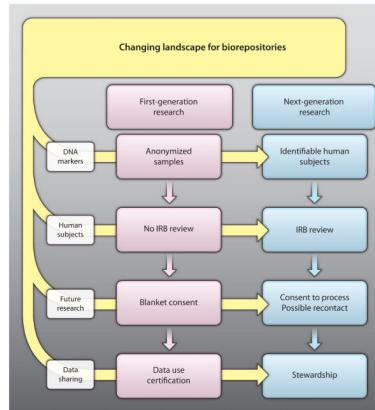
## REFERENCES AND NOTES

- Manolio TA, Bailey-Wilson JE, Collins FS. Genes, environment and the value of prospective cohort studies. *Nat Rev Genet* 2006;7:812–820. [PubMed: 16983377]
- Greely HT. The uneasy ethical and legal underpinnings of large-scale genomic biobanks. *Annu Rev Genomics Hum Genet* 2007;8:343–364. [PubMed: 17550341]
- Zerhouni EA. Translational research: Moving discovery to practice. *Clin Pharmacol Ther* 2007;81:126–128. [PubMed: 17186011]
- Gibbons SMC, Kaye J, Smart A, Heeney C, Parker M. Governing genetic databases: Challenges facing research regulation and practice. *J Law Soc* 2007;34:163–189.
- Karp DR, Carlin S, Cook-Deegan R, Ford DE, Geller G, Glass DN, Greely H, Guthridge J, Kahn J, Kaslow R, Kraft C, Macqueen K, Malin B, Scheuerman RH, Sugarman J. Ethical and practical issues associated with aggregating databases. *PLoS Med* 2008;5:e190. [PubMed: 18816162]
- Johnson JD, Case DO, Andrews JE, Allard SL. Genomics—the perfect information-seeking research problem. *J Health Commun* 2005;10:323–329. [PubMed: 16036739]
- Tutton R, Kaye J, Hoeyer K. Governing UK Biobank: The importance of ensuring public trust. *Trends Biotechnol* 2004;22:284–285. [PubMed: 15158057]
- Godard B, Schmidtke J, Cassiman JJ, Aymé S. Data storage and DNA banking for biomedical research: Informed consent, confidentiality, quality issues, ownership, return of benefits. A professional perspective. *Eur J Hum Genet* 2003;11(suppl 2):S88–S122. [PubMed: 14718939]
- Hull SC, Sharp RR, Botkin JR, Brown M, Hughes M, Sugarman J, Schwinn D, Sankar P, Bolcic-Jankovic D, Clarridge BR, Wilfond BS. Patients' views on identifiability of samples and informed consent for genetic research. *Am J Bioeth* 2008;8:62–70. [PubMed: 19003716]
- Cambon-Thomsen A. The social and ethical issues of post-genomic human biobanks. *Nat Rev Genet* 2004;5:866–873. [PubMed: 15520796]
- Roden DM, Pulley JM, Basford MA, Bernard GR, Clayton EW, Balsler JR, Masys DR. Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin Pharmacol Ther* 2008;84:362–369. [PubMed: 18500243]
- Ohm, P. Broken promises of privacy: Responding to the surprising failure of anonymization. University of Colorado Law Legal Studies Research Paper No. 09–12. Aug 13, 2009 available at SSRN: <http://ssrn.com/abstract=1450006>
- McGuire AL, Gibbs RA. Genetics. No longer de-identified. *Science* 2006;312:370–371. [PubMed: 16627725]
- Thorisson GA, Muilu J, Brookes AJ. Genotype-phenotype databases: Challenges and solutions for the post-genomic era. *Nat Rev Genet* 2009;10:9–18. [PubMed: 19065136]
- Homer N, Szelinger S, Redman M, Duggan D, Tembe W, Muehling J, Pearson JV, Stephan DA, Nelson SF, Craig DW, Visscher PM. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genet* 2008;4:e1000167. [PubMed: 18769715]
- Jacobs KB, Yeager M, Wacholder S, Craig D, Kraft P, Hunter DJ, Paschal J, Manolio TA, Tucker M, Hoover RN, Thomas GD, Chanock SJ, Chatterjee N. A new statistic and its power to infer

- membership in a genome-wide association study using genotype frequencies. *Nat Genet* 2009;41:1253–1257. [PubMed: 19801980]
17. Ravitsky V, Wilfond BS. Disclosing individual genetic results to research participants. *Am J Bioeth* 2006;6:8–17. [PubMed: 17085395]
  18. Caulfield T, McGuire AL, Cho M, Buchanan JA, Burgess MM, Danilczyk U, Diaz CM, Fryer-Edwards K, Green SK, Hodosh MA, Juengst ET, Kaye J, Kedes L, Knoppers BM, Lemmens T, Meslin EM, Murphy J, Nussbaum RL, Otlowski M, Pullman D, Ray PN, Sugarman J, Timmons M. Research ethics recommendations for whole-genome research: Consensus statement. *PLoS Biol* 2008;6:e73. [PubMed: 18366258]
  19. Lunshof JE, Chadwick R, Vorhaus DB, Church GM. From genetic privacy to open consent. *Nat Rev Genet* 2008;9:406–411. [PubMed: 18379574]
  20. Fryer-Edwards K, Fullerton SM. Relationships with test-tubes: Where's the reciprocity? *Am J Bioeth* 2006;6:36–38. author reply W10-2. [PubMed: 17085405]
  21. Chadwick R, Berg K. Solidarity and equity: New ethical frameworks for genetic databases. *Nat Rev Genet* 2001;2:318–321. [PubMed: 11283704]
  22. Caulfield T. Biobanks and blanket consent: The proper place of the public perception and public good rationales. *King's Law J* 2007;18:209–226.
  23. 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;7:661–666. [PubMed: 16819458]
  24. McGuire AL, Hamilton JA, Lunstroth R, McCullough LB, Goldman A. DNA data sharing: Research participants' perspectives. *Genet Med* 2008;10:46–53. [PubMed: 18197056]
  25. Mascalzoni D, Hicks A, Pramstaller P, Wjst M. Informed consent in the genomics era. *PLoS Med* 2008;5:e192. [PubMed: 18798689]
  26. Secko DM, Preto N, Niemeyer S, Burgess MM. Informed consent in biobank research: A deliberative approach to the debate. *Soc Sci Med* 2009;68:781–789. [PubMed: 19095337]
  27. Kohane IS, Mandl KD, Taylor PL, Holm IA, Nigrin DJ, Kunkel LM. Reestablishing the researcher-patient compact. *Science* 2007;316:836–837. [PubMed: 17495156]
  28. Jeffers BR. Human biological materials in research: Ethical issues and the role of stewardship in minimizing research risks. *ANS Adv Nurs Sci* 2001;24:32–46. [PubMed: 11763368]
  29. Goering S, Holland S, Fryer-Edwards K. Transforming genetic research practices with marginalized communities: A case for responsive justice. *Hastings Cent Rep* 2008;38:43–53. [PubMed: 18457228]
  30. Knoppers BM. Genomics and policymaking: From static models to complex systems? *Hum Genet* 2009;125:375–379. [PubMed: 19252928]
  31. Winickoff DE, Winickoff RN. The charitable trust as a model for genomic biobanks. *N Engl J Med* 2003;349:1180–1184. [PubMed: 13679534]
  32. Terry SF, Terry PF, Rauen KA, Uitto J, Bercovitch LG. Advocacy groups as research organizations: The PXE International example. *Nat Rev Genet* 2007;8:157–164. [PubMed: 17230202]
  33. Geschwind DH, Sowinski J, Lord C, Iversen P, Shestack J, Jones P, Ducat L, Spence SJ. AGRE Steering Committee. The Autism Genetic Resource Exchange: A resource for the study of autism and related neuropsychiatric conditions. *Am J Hum Genet* 2001;69:463–466. [PubMed: 11452364]
  34. Yarborough M, Fryer-Edwards K, Geller G, Sharp RR. Transforming the culture of biomedical research from compliance to trustworthiness: Insights from nonmedical sectors. *Acad Med* 2009;84:472–477. [PubMed: 19318781]



**Fig. 1. Frozen resources**  
A researcher retrieves medical samples.



**Fig. 2. Changes to accompany next-generation biorepository research**

The transition from first- to next-generation biorepository research will require innovative new approaches to privacy, institutional review, informed consent, and data stewardship. IRB, institutional review board.