

INFORMATION ACCESS

Raw Personal Data: Providing Access

Jeanine E. Lunshof,^{1,2*} George M. Church,¹ Barbara Prainsack³

Heated debates on responsibilities in biomedical research currently focus on the end of the data and information pipeline: They revolve around issues of returning results to participants and patients (1–3, 4). Although these debates are timely, they miss a crucial point at the beginning of the pipeline: the question of whether sample donors are able to access the raw data derived directly from their stored sample. The U.S. Presidential Commission recently reviewed 32 reports from the United States and worldwide on returning of findings in diverse contexts (4); it is striking that access to raw data by participants was not addressed in any of them.

In this Policy Forum, we argue, first, that there are compelling ethical reasons for enabling donors to access the raw data derived from their material deposited in any kind of repository; second, that there is a crucial difference between providing access to data and returning findings.

A One-Way Transaction

Drawing an analogy between data banks and money banks, despite their differences in other respects, may be instructive to illustrate the crucial point of data access. In everyday life, banking is a two-way interaction. When we make a monetary deposit to a bank, we get a receipt. Furthermore, we are able to access our account and see the balance, we can track transactions, and we can terminate our relationship with the bank and have the remaining balance returned. As an account holder, we can directly access our own data.

When we make a deposit to a data bank or biorepository, in contrast, putting our data or specimen into the hands of researchers or clinicians, we lose track of it. We do not have the status of account holders. We are not even able to access the basic descriptors of our own



contribution: The raw data, directly from the deposited sample before analysis, are—with very few exceptions (5)—not made available to the depositor. Thus, a deposit to a research or clinical repository is a one-way transaction in which we give up most of our agency and control.

Data are shared and used by a large number of researchers; the individual donors are those with the least access to the data they contributed. This one-way communication is a blatant anachronism in the relationship between data donors and researchers or clinicians, and unnecessary when relatively low cost, user-friendly Web-based tools to facilitate such access are available. Today, samples and data sets are usually identified and tracked through bar codes. Handing over that bar code to any research participant or patient when the sample or the survey is taken can enable access if the system is set up for that. It is then the individual who decides whether to use the access code and look at her or his raw data.

Getting It Right from the Start

Web-based information and communication technologies render direct data access technically feasible and economically affordable in many cases. But repositories have to be designed accordingly right from the start, as later adaptation will often be expensive and difficult. Data storage formats need to be interoperable in order to build a framework

Donors should have access to raw data derived from their contributions to research or clinical repositories to increase personal choice and reciprocity.

for data sharing, a requirement that may be very challenging in a competitive setting with diverse research and commercial interests. Databases that provide public open access, as, e.g., the Personal Genome Project (5), avoid such problems, and for them, issues of donor access are moot. But also studies for which public access is not an option should pay urgent attention to the structures and conditions necessary to provide donors access to their raw data.

In actual practice, the financial and logistic challenges related to providing access might result in significant additional costs, costs that participants could decide they do not want to incur (6). But data access needs to be an option, in order to enable a personal choice.

A Fair and Reciprocal Relationship

Although we use biorepositories as an example to illustrate our argument for reciprocity and fairness, the argument is applicable to any field where researchers and participants interact. Participants' access to raw data could be a helpful mechanism to increase transparency in any study, be it genomics, other “omics,” or social psychology. Data fabrication would be much more difficult if study participants (or their proxies) had access to their raw data (7) and could verify that they were included in the study.

The possibility for research participants to access their raw data is a basic requirement for a just and reciprocal relationship, establishing at least a basic symmetry between those who donate and those who use data for their research (5, 6). Moreover, it enables individuals to act upon their considered judgments—a requirement for autonomous agency that was stated decades ago in the Belmont report (8). Providing access to the data that are derived directly from the sample, before analysis and interpretation, recognizes the donor's agency in at least three ways: freedom to decide, option of independent analysis, and informed decision about participation (see the box).

Besides these, there are many other good reasons for openness and reciprocal rela-

¹Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. ²Department of Molecular Cell Physiology, VU University Amsterdam, 1081 HV Amsterdam, Netherlands. ³Department of Social Science, Health and Medicine, King's College London, London WC2R 2LS, UK. *Corresponding author: jelunshof@genetics.med.harvard.edu

ENACTING DONORS' AGENCY: THE BIOREPOSITORY AS AN EXAMPLE

Freedom to decide: Data donors can decide whether or not to access the raw data sets (e.g., the sequencing output data) from clinical and research biobanks. As no analysis has yet been performed, the role and responsibility of the biobank is merely to facilitate data access at a technical-procedural level. Usual disclaimers concerning data accuracy will apply, as technical errors can never be ruled out entirely.

Independent analysis: Data donors can, independently of the researchers and clinicians, apply data interpretation tools or have analyses performed. High-quality genome data-interpretation tools are available in the public domain, as well as commercial service providers. In particular, after analyses have been carried out and when filtered or selected findings are being disclosed, individuals may want to use the original raw data set for more integrated and comprehensive information or an independent opinion.

Informed decision about participation: The outcomes of an independent, participant-initiated analysis of the raw data can lead to better decisions about (continued) participation in a study or the pursuance of a diagnostic process. Later, the option of tracking who the data set has been shared with, and in what context, should be available. This, too, can contribute to a more informed decision about continued participation.

tionships between researchers and study participants (5, 6). Yet, on the basis of these three arguments alone, it is clear that providing access to their raw data is essential to taking individuals seriously as partners in research not merely as sources of samples and data. Giving individuals access to raw data is less complex than reporting research findings or returning the results of clinical investigations—issues that we do not claim to solve here.

A Crucial Distinction: Data Versus Findings

For a clearer perspective, we elaborate on the crucial distinction between access to data and returning of findings. The process of providing donors with access to their raw data (i.e., before any interpretation has been performed) is, in principle, straightforward. At the stage of returning findings, however, there are additional layers of complexity: At that point, people other than the data donors have (i) made the decision on how and which data are analyzed, (ii) on the relevance of findings, and (iii) whether they are actionable or not. In addition, at the data-processing stage, there are relevant differences between research and the clinic. These features have significant ethical and legal implications for the returning of results.

In research, notwithstanding the strict demand for data-sharing and transparency, researchers need to be able to protect data they are working on while discovery is in progress. This may temporarily override the interest of others in information-sharing and openness. Yet, these concerns do not apply to individual participants accessing their raw data.

In the clinical setting, there are further degrees of complexity, as the professional role, responsibility, and liability of clinicians and other actors must be accounted for. In

returning results of diagnostic procedures to patients, a health care provider assumes a specific professional responsibility that is clearly defined by the professional associations and is subject to legal regulations concerning the practice of medicine. There are concerns regarding the risks of individuals accessing data that they may misunderstand and impulsively act upon (9). However, long-term follow-up of direct-to-consumer genetic testing has not shown any evidence of such effects, and calls have been made to widen the notion of utility in this context to include personal or social utility (10, 11). Impulsive actions as a consequence of access to raw data are even less likely, because several intermediate steps are needed before substantial action is possible.

Another important question regarding current practice is whether anticipated difficulties in returning data justify the intentional limitation of knowledge generation and repeated cycles of information-filtering (3). Reduction of potential knowledge gain by avoiding or eliminating disturbing data points that may yield “unsolicited” (3) or “incidental” findings (1) is scientifically and clinically unsound and inhibits progress in the understanding of human biology and the complexity of disease states. Filtering out genetic variants that are assumed to have limited or no clinical utility, as recommended by the European Society of Human Genetics (3), systematically precludes the possible discovery of complex genetic causation.

Besides this, in the clinical setting, the lack of access to raw data directly derived from the donor's specimen is a relic of long-outdated strong paternalism in a practice of medicine where doctors were assumed to be omniscient. Giving a clear and focused answer to a specific question

from a patient—while at the same time communicating the uncertainty and incomplete knowledge inherent in medicine—is key in clinical practice. But dealing with the inevitable limits of knowledge in a constructive manner is fundamentally different from setting limits to data generation in the presumed “best interest” of the patient or as a way to mask uncertainty and limit the doctor's liability. The very terms used in discussions about when and how to best return findings to participants illustrate where the core agency lies: It lies with those who return and not with those who receive. Shifting the focus of the debate to access increases the scope of agency of research participants. The outcomes of that have shown to be without harm so far (5, 6, 10, 11).

Access to raw data is independent from the prospective delivery of interpreted information at a later point in time. Even the most thoughtful delivery of comprehensive information does not render obsolete a prior right (12) to access the raw data set.

References and Notes

1. R. C. Green *et al.*, *Genet. Med.* **15**, 565–574 (2013).
2. W. Burke *et al.*, *Genet. Med.* **15**, 854–859 (2013).
3. C. G. van El *et al.*, *Eur. J. Hum. Genet.* **21**, 580–584 (2013).
4. Presidential Commission for the Study of Bioethical Issues, *Anticipate and Communicate: Ethical Management of Incidental and Secondary Findings in the Clinical, Research, and Direct-to-Consumer Contexts* (U.S. Department of Health and Human Services, Washington, DC, 2013); www.bioethics.gov.
5. M. P. Ball *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 11920–11927 (2012).
6. B. Prainsack, A. Buyx, *Med. Law Rev.* **21**, 71–91 (2013).
7. Levett Committee, Noort Committee, Drenth Committee, *Flawed Science: The Fraudulent Research Practices of Social Psychologist Diederik Stapel* (Tilburg, Groningen, Amsterdam, 2012); <https://www.commissielevelt.nl>.
8. The National Commission for the Protection of Human Subjects in Biomedical and Behavioral Research, *Ethical Principles and Guidelines for the Protection of Human Subjects in Research* (U.S. Department of Health, Education, and Welfare, Washington, DC, 1979); www.hhs.gov/ohrp/humansubjects/guidance/belmont.html.
9. U.S. Food and Drug Administration, A. Gutierrez, Warning Letter to 23andMe, Inc., 22 November 2013; www.fda.gov/ICECI/EnforcementActions/WarningLetters/2013/ucm376296.htm.
10. C. S. Bloss *et al.*, *J. Med. Genet.* **50**, 393–400 (2013).
11. K. Wasson *et al.*, *J. Community Genet.* **4**, 495–505 (2013).
12. Moral rights may or may not be reflected in enforceable legal rights at a given place or time. An in-depth analysis of this aspect falls outside the scope of this article.

Acknowledgments: J.E.L. receives funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013; REA grant no. 298698). The funding body and institutions had no role in the writing of the manuscript or in the decision to submit the manuscript for publication. The views expressed are entirely the authors' own. J.E.L. and G.M.C. thank the staff of the Personal Genome Project (PGP-Harvard) for discussion. All the authors thank J. Aach for helpful comments on an earlier version of this manuscript.

10.1126/science.1249382

Science

Raw Personal Data: Providing Access

Jeanine E. Lunshof, George M. Church and Barbara Prainsack

Science **343** (6169), 373-374.
DOI: 10.1126/science.1249382

ARTICLE TOOLS

<http://science.sciencemag.org/content/343/6169/373>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. 2017 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. The title *Science* is a registered trademark of AAAS.