American Psychological Association

American Psychologist

0003-066X/18/$12.00

2018, Vol. 73, No. 2, 138–145
http://dx.doi.org/10.1037/amp0000240

Ethical Aspects of Data Sharing and Research Participant Protections

Michael W. Ross
University of Minnesota Medical School, Minneapolis, Minnesota

Martin Y. Iguchi
RAND Corporation, Santa Monica, California

Sangeeta Panicker
American Psychological Association, Washington, DC

Open access is fast becoming the norm across science. Sharing research data broadly has the potential to accelerate scientific progress, optimize the value of data, and promote scientific integrity. However, data sharing also poses new practical and ethical challenges to the conduct of research with human participants. This article provides an overview of how open access to research data has impacted the core principles of research ethics—respect for persons, beneficence, and justice—and, in turn, how a reinterpretation of these principles translates to procedures for the protection of the rights and wellbeing of human research participants.

Keywords: data sharing, research participant protections, data access, group harms

Open access and data sharing, practices that are long-standing norms in many of the physical sciences, are rapidly becoming implemented within the biomedical and behavioral sciences (Boulton, Rawlins, Vallance, & Walport, 2011; Friedlander & Adler, 2006; Holdren, 2013; Margolis et al., 2014; Tenopir et al., 2015). Worldwide, funding entities are increasingly asking that data be made available for use by others, breaking with traditional norms of data ownership, maintenance, and storage, typically archived in private repositories for personal use by academics (Brewer, Potterat, & Muth, 2010; Cragin, Palmer, Carlson, & Witt, 2010; Koslow, 2000; Nelson, 2009; Pisani, Whitworth, Zaba, & Abou-Zahr, 2010; Pisani, Whitworth, Zaba, & Abou-Zahr, 2011; Savage & Vickers, 2009; Tenopir et al., 2011; Walport & Brest, 2011). Indigenous communities have also proposed the ownership, control, access, and possession model for data management (Schnarch, 2004), and a redefinition of the term "researcher" to include communities.

The rationale for data sharing is straightforward. Providing broad access to data holds potential for maximizing the utility of available information, allows access to those with different skill sets and perspectives, increases the likelihood of data merging and transformation, and decreases the cost of knowing. Data sharing also holds potential for preventing fraud and/or detecting error, as transparency allows for ease of verification, replication, refutation, and/or refinement.

The path to open data is not, however, straightforward. A complex array of issues must be considered, from human participant protections, to data ownership and relative power in access, to standards and definitions, to regulations and technical constraints, to costs and benefits, and of course, exceptions to every rule. This article is a review of the impact of data sharing on human research participant protections and related ethics.

Ethical Framework

The ethical underpinnings of human research participant protections in the United States are firmly grounded in occidental philosophical traditions that hold individual rights as paramount. Thus, the authors of the seminal document *The Belmont Report* (National Commission for the Protection of Human Subjects of Biomedical & Behavioral Research, 1978), which provides the ethical framework for research with human participants, delineated three basic principles: (a) respect for persons, (b) beneficence, and (c) justice. *Respect for persons* requires recognizing individuals

Editor's note. This article is part of a special section titled "Data Sharing in Psychology," published in the February–March 2018 issue of *American Psychologist*. Jennifer Crocker and Leah Light served as editors of the special section.

Authors' note. Michael W. Ross, Program in Human Sexuality, Department of Family Medicine and Community Health, University of Minnesota Medical School, Minneapolis, Minnesota; Martin Y. Iguchi, Drug Policy Research Center, RAND Corporation, Santa Monica, California; Sangeeta Panicker, Science Directorate, American Psychological Association, Washington, DC.

Correspondence concerning this article should be addressed to Sangeeta Panicker, Science Directorate, American Psychological Association, 750 First Street North East, Washington, DC 20002. E-mail: spanicker@apa.org

as autonomous agents with unalienable rights for self-determination. This principle extends further to provide additional protection to those less capable of self-determination, whether from decreased mental capacity, illness, or circumstances of vulnerability (e.g., prison inmates). Abiding by this principle in the research setting requires researchers to present prospective participants with the opportunity to provide voluntary informed consent. *Beneficence* addresses the researcher's obligation to protect the well-being of human participants, not only by minimizing risks of harm to the participant, but also by maximizing any potential benefits to the participant. Typically, this entails ensuring that potential benefits outweigh risks of harm to individual participants; however, this principle can also extend to considerations beyond a particular research participant if the potential for loss of benefits to society as a whole, is taken into account. In addition to the proscription to do no harm to the research participant(s), the principle of beneficence encompasses a prescriptive ethical mandate to promote good. Thus, a lack of explicit benefits to the individual participant can be acceptable should broader (i.e., societal) benefits exist, so long as every effort is made to protect the participant from risks of harm. *Justice* addresses the need for equitable distribution of the benefits and burdens of research on humans. This requires a fair selection of research participants based on the purposes and anticipated outcomes of the research, and not based on their easy availability, manipulability, or otherwise compromised position. Given the ubiquity of societal injustice to vulnerable groups of citizens, justice in human research requires careful consideration of the rationale for the exclusive selection of participants from certain groups.

In addition to the impetus for open access and data sharing, other events have substantially impacted the research landscape which have challenged the interpretation of the prevailing research ethics framework as exclusively focused on the individual. Advances in digital and communication technologies have made international research and international collaborations more common. As a result, U.S.-based researchers must contend with norms and conventions of different societies, including those that have less of an individual-focused culture. Furthermore, the ubiquity of the Internet, digital technologies, and social media has empowered the public to become more engaged in research, which has fostered nontraditional research collaborations, such as community-academic partnerships. These collaborations also have implications for data sharing and the protection of individual participants as well as participants' communities.

Impact of Data Sharing on the Belmont Principles and Research Participant Protections

Researchers and entities charged with the oversight of research, such as institutional review boards (IRBs) and research ethics boards, are required to attend to and balance the dictates of these basic principles. However, it is the requirement for voluntary, informed consent from potential research participants—based on the principle of respect for persons—that is considered the bedrock of ethically responsible research with human participants. Therefore, it has been standard practice during the process of obtaining informed consent to assure participants that all the information they provide during the research study will be kept confidential to the extent permitted by law, with access restricted to members of the research team. Furthermore, depending on the sensitivity of the data, researchers might provide additional assurances regarding how the data will be published or otherwise shared with the scientific community; for example, by deidentifying the data before they are shared. The increasingly universal requirement for data sharing, however, means that data will be available more broadly to researchers in general, beyond just the original research team and original study (and the original country, regulatory/legal system, and culture) for which the participant gave consent. In addition, for some types of data, deidentification may not be an option, because it would render the data useless. For example, audio and video recordings may lose scientific value if substantially altered, as may heavily redacted transcripts. Finally, while the ability to aggregate different types of data from numerous sources has the potential to increase the societal benefits, the nature and scale of risks of harm is also compounded by broad access to big data. A recent study by de Montjoye, Radaelli, Singh, and Pentland (2015) on the reidentifiability of credit card metadata found that four spatiotemporal points are enough to uniquely reidentify 90% of individuals, with the dollar amount of a card transaction increasing that risk of reidentification by 22% on average. de Montjoye et al. noted that finding the right balance between privacy and utility is crucial to realizing the great reidentification potential of metadata because even such apparently anonymous data as credit card metadata—such as economic, behavioral, temporal, and spatial information—when aggregated, render the data identifiable. Similarly, Sweeney (1997, 2000) demonstrated the ease with which deidentified demographic information can be combined with other deidentified data such as medical/health information, to reidentify individuals.

Thus, the move toward open access and open data as the norm across the biomedical and behavioral sciences has had a profound impact on the practical interpretation of the principle of *respect for persons* in the research setting. It has forced a reassessment of the fundamental concept of informed consent in research with human participants so that it truly protects the autonomy interests of individual participants. Data deposited in repositories will likely be accessed by researchers other than the research team that collected the data, and for purposes that might significantly differ from the focus of the original study. Consequently, the

participant will not have control over who might access the data in the future and for what purposes, as these are unknowns during the initial consent process. Thus, the consent they provide will not be "informed" in the traditional sense, but rather general, in one of two ways: (a) open-ended permission/blanket consent (i.e., consenting to participate in the present study well knowing that the nature and intent of future uses of the data are unknown) or (b) broad consent as defined by Grady et al. (2015; i.e., consent for an unspecified range of future research subject to a few content and/or process restrictions). Broad sharing of research data could have an effect on the integrity of the science if individuals and communities are more willing to trust and provide data to researchers from known institutions, but less likely to do so if the identity and motivations of a future user are unknown. While this is conjecture, it is a fruitful avenue for research.

Open access and data sharing practices result in so-called secondary use, wherein previously archived data are mined by other investigators to study different questions. Such secondary uses of data also require weighing of risks of harm versus potential benefits, similar to the assessment for the original study. Thus, the impact of data sharing on the principle of *beneficence* itself is no different from the ethical and scientific justification for the original study. However, from a practical perspective, the manner in which the principle is interpreted through regulations has ramifications for participants and/or their communities. For example, under current regulations some secondary uses of deidentified data are not considered research with human participants; consequently, regulatory protections do not apply (Protection of Human Subjects, 2009). However, given the ease with which reidentification is possible with broad data sharing and current and future technologies, there is need for a more cautious approach such as calibrated levels of access and prospective review requirements depending on the sensitivity of the original data, probability and magnitude of risks of harm, and/or the intent and scope of the secondary use.

In addition to the proscription to do no harm, *beneficence* also provides a prescriptive ethical mandate to maximize potential good from existing data. Thus, data sharing must be considered a beneficent act, unless the risks of harms overwhelmingly outweigh the potential benefits associated with it. Making data sharing with appropriate safeguards the default position not only promotes beneficence, but also reshapes the view of data sharing not as a necessary evil, but an ethical mandate and potential good.

Open access to data that engenders secondary uses also has implications for the third Belmont principle of *justice*, just as it does for the original study. In other words, individuals and/or groups that contributed to the initial study should also benefit from the results, if relevant, emanating from secondary uses of a data set.

Group Harms

It is perhaps not surprising that individualistic, as opposed to communitarian, cultures, focus on more individualistic interpretations of ethics. However, we argue that harm occurs in collectivistic as well as individual contexts, and that research data have the potential to harm both individual participants and their communities.

Thus, each of the three Belmont principles may have a social and community as well as a personal dimension, which can impact data sharing. In a collectivistic/communitarian culture/society, abiding by the principle of respect for persons might entail obtaining permission from an entity other than the potential research participant (e.g., village elder, council, matriarch/patriarch, spouse), and potentially assent, in the case of a minor, and consent from an adult participant. Consequently, regardless of the individual participant's choice, ability to share data might be constrained by community norms and decisions or choices. The ethical dilemma posed by such a situation, which pits respect for the individual's autonomy against the community's decision, may suggest the need to reframe the ethical principles underlying research. Similarly, the collectivist interpretation of the principle of beneficence might require assessing the risks and potential benefits of sharing data to both individuals and communities. Furthermore, the ethical mandate to do good and maximize benefits inherent in the principle of beneficence provides a strong ethical foundation for sharing of research data when such sharing capitalizes on the value of the data in the public's interest, so long as appropriate safeguards are in place to protect participants and their communities from risks of harm. Finally, the principle of justice requires that research participants are fairly selected with regard to the purpose and expected outcome of the research, not only when considering the participant as an individual, but also as *a member of a collective group or society*.

As an example, Drabiak-Syed (2010) reported on a settlement agreement between the Havasupai tribe and Arizona State University (ASU) Board of Regents. Researcher from ASU had collected blood samples from members of the Havasupai tribe, for what was believed to be a diabetes study. Later, it was discovered that the samples had been shared by ASU with other researchers to study other issues including inbreeding, human migration, and schizophrenia. Regardless of whether individual participants were identified, sharing of the data resulted in social and psychological harms to the tribe as a whole. Tribal members sued ASU when the university refused to return the blood samples, alleging cultural, dignitary, and group harm. The eventual settlement involved monetary compensation, return of blood samples, return of all research derived from the samples, termination of IRB approvals, and disclosure of all researchers outside of ASU receiving transferred blood

samples. Drabiak-Syed argued that the return of blood samples constituted recognition of the rights of both the individual and the tribe, with an emphasis on the importance of blood samples to the spiritual wellbeing of both the individual and the tribe. Drabiak-Syed also made the case that many indigenous groups have a well-formed sense of group identity, that spiritual beliefs and cultural practices must be taken into consideration when defining harms to the individual and to the larger group, and that consideration of consent must also consider consent from group representatives, particularly when the group is defined as a distinct sovereign entity.

While regulations and policies for research with human participants in the United States seldom, if ever, explicitly consider potential harms to individuals or groups not involved in the research, other countries have addressed such issues more comprehensively. For instance, harms to nonparticipants are considered in the Australian Government National Health and Medical Research Council (2015) statement on ethical conduct in human research, where risks and potential benefits are considered for individuals, their families, and groups with which they identify

> Examples of risks to non-participants include the risk of distress for a participant's family member identified with a serious genetic disorder, the possible effects of a biography on family or friends, or infectious disease risks to the community. Some social research may carry wider social or economic risks; for example, research in a small community into attitudes to specific subpopulations may lead to unfair discrimination or have effects on social cohesion, property values, or business investment.

The statement notes that

> Within some communities, decisions about participation in research may involve not only individuals but also properly interested parties such as formally constituted bodies, institutions, families or community elders. Researchers need to engage with all properly interested parties in planning the research.

Consent includes consideration of the "requirements of the codes, laws, ethics and cultural sensitivities of the community in which the research is to be conducted". Consideration of the impact on nonparticipants arose from research that negatively affected indigenous communities. Aboriginal Australians consider their DNA to be collective cultural property, and have expressed concerns over its research use including studies on racial classification and on the inferiority of indigenous peoples, and possible patenting of genetic sequences. In a new process known as "dynamic consent," participants are given full control over their genome. The world's first dynamic consent model allows DNA donors to provide or revoke consent for specific projects even after they have consented to their sequenced genome being held on file (Wahlquist, 2016). Every application to access the data would be decided upon by the Indigenous Governance Board, which is chaired by the Indigenous human rights commissioner. Such close consultation with the potentially affected communities before releasing research data could help thwart potential lapses in privacy protection or other harms.

This consideration of group harms is also addressed in *The Belmont Report*, which notes that when groups are socially defined (e.g., by race or ethnicity), research might be used to rationalize prejudice or perpetuate discrimination against participants and nonparticipants. As noted by Khoury, Little, and Burke (2004), such "injustice arises from social, racial sexual and cultural biases institutionalized in society" (p. 66). The question thus also arises as to whether consent by the individual participant in research that may have potential to harm the group is sufficient, and points to the need to consider research ethics beyond participant harm. Here, the existence of a Community Advisory Board at an early stage in the research development through to analysis and communication of data is an important consideration in managing potential group harms.

Special Considerations in International Research

Data sharing poses additional challenges to international research. Local sociopolitical and cultural factors, in addition to local laws and regulations, significantly affect the risks of harm to research participants and their communities. Often, the potential harms stem from unauthorized access to the data or compelled disclosure of the data and/or participant identities to state authorities. Therefore, in addition to ensuring that participants are fully informed and comprehend the risks of harm inherent in participating in the study, researchers are obligated to be judicious in selecting data repositories and, depending on the nature of the data and the risks of harm to individual participants, groups, or communities, employ mechanisms such as delayed access, restricted access, and varying levels of access. Furthermore, it is the researchers' responsibility to ensure that data are deposited in a manner that precludes misinterpretation by legitimate secondary users of the data, which could harm individual research participants and/or their communities. However, beyond the point of deposit, the researchers are unable to guarantee the use or misuse of the data.

Data sharing can have damaging consequences for individuals and communities, such as when there are political or religious motives to discredit particular communities or to justify repressive laws and sanctions. "Mining" data to make a specific case against a community, which would be considered unethical if carried out by a scientist in the United States, cannot always be prevented when data are accessed outside of the United States. Protection of communities, therefore, has to be part of the equation in considering data sharing, particularly where data are from outside the United States or being used outside the United States. It is conceivable that such data may be used to justify violence against communities, vigilante activity, or legal repression, or to fuel discrimination. It is here that the

researcher needs to "above all, do no harm." Close consultation with potentially affected communities (here, the importance of a well-consulted Community Advisory Board cannot be overemphasized) before releasing data could help to thwart potential lapses in privacy protection through unintended reidentification, or harms against communities through selective malicious use of data which may fuel stigma or discrimination. It might be argued that the Hippocratic maxim "above all, do no harm" is an impossible standard as one might envisage almost any data doing harm. The appropriate standard, we believe, is that the researcher needs to actively consider reasonably foreseeable harms, in consultation with communities in weighing potential benefits against risks of harms.

It must be borne in mind that once research data are in the public realm they are out of the control of the original researcher—regardless of the conditions set by those individual researchers—unless a legally binding agreement is signed. What is legally binding in the United States, however, might not be binding in other countries, and scientists and participants must be aware that the ethical conventions for use of research data may also vary in other countries. Thus, data protection mechanisms, such as Certificates of Confidentiality from the National Institutes of Health or Privacy Certificates from the National Institute of Justice, may be inadequate for protecting research participants and their communities from risks of harm.

The following example illustrates complexities of sharing data when the research involves academic-community partnerships in international settings.

> Dr. A had completed the first study of gay and bisexual men in a low-income country, a country where a number of prominent members of the government had called for the death penalty for homosexual men. Subsequently, Dr. A was contacted by Dr. B, asking for access to the data set. When Dr. A delved into Dr. B's research background, it became apparent that Dr. B was employed by the government in that country. On inquiring about the nature of the secondary use of the data, Dr. B was initially evasive and indicated that he was interested in the data to identify the proportion of the gay men who had been sexually abused as children or had sex with another man as children or teens. Furthermore, he did not seem to understand that there was a difference between homosexual men and transsexual men and also indicated that he wanted to look at the data to see "why these men want to become women." Dr. A indicated that the data set was coowned by a colleague in the country (Dr. C), and that both investigators as well as the gay community partner would need to assent to the transfer of data to Dr. B (a transfer the community partner was uncomfortable about on hearing the facts). Subsequently, local police in the country contacted Dr. C's university administration with an order to seize the data, which unbeknownst to them were stored behind a firewall and password protected on the computer at Dr. A's home institution in the United States. Soon thereafter, the government moved to legislate for the death penalty for homosexual activities used as justification that homosexual men "recruited" minors by sexual seduction and abuse, which they argued "made" the minors homosexual.

While details of the case have been modified to avoid identifying details, and highlight the risks (it is a combination of several situations), this hypothetical case illustrates potential dangers of data access, including (a) researchers who are unqualified or poorly qualified in the area of research asking for access to data; (b) the use of selective data to push agendas, which may threaten human rights or lead to community harm and harm to individuals in the community; (c) the necessity for protection of colleagues (Dr. C) and data in states where seizure of jointly owned data is possible; (d) the need to foresee possible harms arising from data misuse by governments because it was not clear that even if Dr. B was qualified, others in his department with other intentions could have access. Outside of the United States, U.S. law, and other rights and U.S. IRBs have no authority. Indeed, local law and IRBs have primacy.

Participant Protections for Sharing Different Types of Data

As mentioned, open access to research data might pose different risks of harm to research participants, depending on the nature of the data (qualitative vs. quantitative), its sensitivity (risky health behaviors vs. language development), as well as other factors such as characteristics unique to the study sample (e.g., location as suburb or reservation, race or ethnicity, religion, conviction or diagnosis, or physical characteristics), location of the study population, and so forth. For example, the site of Humphreys' (1970) study of homosexual encounters in public toilets was quickly identified despite attempts to anonymize it, leading to police surveillance and arrests. Thus, different types of data will require different types of restrictions to protect the research participants.

Qualitative data (including clinical interviews). By their nature, qualitative data are narrative and describe situations. Qualitative data cannot be shared without being heavily redacted, with regards to names, places, times, dates, and events. Indeed, even a series of qualitative or clinical interviews can be placed together to give clues that might reveal identities, in conjunction with other public data, including criminal records. Reidentification of individual participants or groups from such data is a matter of great concern. While the risk of reidentification from quantitative data stems from the increasing ease with which data from various sources can be aggregated, qualitative data are by their very nature vulnerable to reidentification. Anonymization techniques may not only obscure identities, but also crucial interpretive details (or substitute ones that could be real and are not). Qualitative data often tell a complex story in which identifiers (person, context, relationship,

intent, time, and place) are a key to interpretation. Thus, great care needs to be taken when preparing qualitative data for sharing including controlling access to the data by instituting mechanisms for identifying qualified secondary users and stringent data use agreements. This is particularly critical for certain types of research, for example, in studies of dyads such as spouses, employer–employee, teacher–student, or clinician–patient that involve matched data; unfettered access to the data could allow one member of the dyad to identify themselves in the data and therefore the responses of their partners, which in turn might negatively affect the relationship and/or cause harm.

Clinical records themselves are data, as is information abstracted from them. A clinical interview is, essentially, qualitative data collection. Sharing qualitative clinical records may also be covered by protections stipulated in the Health Information Portability and Accountability Act of 1996 (HIPAA; Pub. L. 104–191, §110 Stat. 1936). Data abstracted from clinical records may also be covered if they include "individually identifiable health information," including demographic data.

Anthropological data. Field notes are a form of data that are rich in detail and context and it may not be possible to fully redact them without losing important information. In small communities, it is particularly difficult to anonymize data because of the limited number of individuals, and the identification of individuals with specific roles or situations. Further, interpretation of data using the researcher as an "instrument" cannot be exactly replicated with other researchers with less experience and knowledge of the research area and/or population. Thus, IRBs or a discipline's norms may require that field notes that could be subject to criminal investigation subpoena be destroyed, thereby precluding sharing with other researchers. For example, sociologist Alice Goffman destroyed her field notes of her ethnographic study that became the basis for her book *On the Run: Fugitive Life in an American City* prior to its publication, to ensure that the subjects of the study were protected from the criminal justice system (Lewis-Kraus, 2016; Parry, 2015).

Geographic Information System (GIS) data. GIS data can be particularly dangerous given their identification of specific locales, especially if stigmatized or illegal activities are the focus of the study. Even if the scale of the data is changed to be sufficiently broad as to location (e.g., suburb or city level), identification of locales of illegal activity by law enforcement agencies, journalists, or vigilante groups may be possible. If the location is misinterpreted, individuals may be mistakenly targeted and subject to harassment, arrest, or attack. Representations of GIS data should be at a low enough scale (probably not below a suburb level) that the published data cannot be used to target individuals or locales if stigmatizing or criminal activity data are the subject of the study. Scale level is often dependent on the frequency of the phenomenon under study: where a phenomenon is rare and visible, for example some dermatological or genetic conditions where only one or two people in a suburb are represented, the scale may need to be lowered. HIPAA regulations, for example, consider city, state, and five-or-more-digit zip code as protected health information. These are not deemed deidentified information and as such are subject to the requirements of the privacy regulations.

Photographic or other image data. The increasing use of images, whether "photovoice" related, images of neighborhoods or specific locales, or of individuals, pose risks of harm to individuals or communities when shared if the data include information about stigmatizing attributes or criminal behavior. If data are to be shared, it is best research practice to have signed consent forms for each image when individuals in the image are identifiable. Images of body parts (especially if there are identifying characteristics such as tattoos) must be treated with similar care. Health-related images such as MRIs, x-rays, or other scans must have any identifying data, including practitioner, code number, or date, removed before sharing. Health-related images often come under the additional protection provisions of HIPAA.

Quantitative data. Data sharing where there are no individual identifiers usually does not present a major problem, although in small communities, small subsamples, or when specific cell frequencies are small, individual participants could be identified by motivated persons with some knowledge of a participating individual or protocol. Similar considerations apply to biospecimens. Sharing the results of analyses may be less controversial than the actual biosamples, where additional analyses may be performed beyond the original IRB approval. Such analyses may include DNA analyses, which could establish identities (Strand, 2016). A study by Gymrek, McGuire, Golan, Halperin, & Erlich (2013) illustrates this problem. The research team was able to recover surnames from "deidentified" genomes using short tandem repeats on the Y chromosome and querying publicly available genetic genealogy databases. As genealogy databases continue to grow, genetic material becomes easier to identify for the individual and for those genetically proximal to the identified individual.

A recent report in the news highlights both the advantages of data access to facilitate transparency, trust, and accuracy, while also highlighting an important concern. *The New York Times* ("Decoding the N.F.L.," 2016) was able to determine that a database of concussion reports compiled by the National Football League was incomplete, by matching dates, locations, and public injury reports, allowing them to determine that over 100 player concussions had not been included. The newspaper named individuals who had been left out of the database, thereby demonstrating their ability to determine who was in the data set. While steps had been taken to make identities anonymous, the steps were inade-

quate to protect participants from disclosure. Clearly, access to data was good for monitoring of scientific integrity, but at the expense in this case of participant anonymity.

While we assume that all professional researchers in the United States are well trained and ethically educated, deposited data can be accessed by other kinds of researchers—employees of foreign governments, law enforcement agencies or lawyers, journalists from sensational media, blackmailers, and people with agendas against particular communities. We believe that it is only a matter of time before research participants suffering tangible harm from being identified sue researchers and data depositories. Perhaps, as in medicine, we should "hope for the best but prepare for the worst" in protecting and deidentifying data, and implementing mechanisms, such as restricted access, that protect the research participants from risks of harm.

We highlight the potential for problems not to argue against data sharing, as we see great potential value, but rather because implementation will require considerable foresight with regard to potential harms and revised training for researchers in data management, documentation, and disclosure. It will also require a rethinking of grant funding to account for additional expenses associated with yet-to-be-determined best practices for data sharing, as well as development of curricula to ensure proper preparation for data safeguarding by those conducting human research. Future directions should include more attention to the handling of other kinds of data beyond quantitative data in IRB training, including case histories, and dialog with data repositories regarding appropriate protections for nonquantitative data. In addition, data ownership issues involving institutions, researchers, participants, communities, and society at large must be resolved. Involvement of major funding agencies in these endeavors is crucial as the mandates on data sharing seem to have focused on quantitative data sets without significant consideration of the risks of nonquantitative data and the potential for harm to individuals, to communities, and to the reputation of science. Involvement of scientific societies representing researchers and clinicians most impacted by sharing data is also crucial. Finally, social media companies such as Google, Facebook, and Twitter must also be part of the debate as they are engaged in both the applied and commercial collection of data, and because their data are increasingly being used for scientific research. And finally, in the absence of a universal policy or mandate to share research data, disciplines, subdisciplines, funding agencies, and individual researchers are obtaining broad consent for secondary uses of data, as described by Grady et al. (2015). However, when data sharing becomes the norm, consent might well have to be more open-ended, with restricted access to protect the research participants. Regardless, both broad consent and blanket consent pose threats to the integrity of the science by potentially skewing the study population and results—that is, only the data of those who provide broad/blanket consent will be available for secondary research. Thus, to ensure that scientific discoveries and progress are balanced with respecting the rights and well-being of research participants, it is critical to garner public appreciation and support for how open science and the widespread sharing of data have impacted traditional norms of research ethics.

References

Australian Government National Health and Medical Research Council. (2015). *National statement on ethical conduct in human research*. Retrieved from https://www.nhmrc.gov.au/book/national-statement-ethical-conduct-human-research

Boulton, G., Rawlins, M., Vallance, P., & Walport, M. (2011). Science as a public enterprise: The case for open data. *Lancet, 377,* 1633–1635. http://dx.doi.org/10.1016/S0140-6736(11)60647-8

Brewer, D. D., Potterat, J. J., & Muth, S. Q. (2010). Withholding access to research data. *Lancet, 375,* 1872. http://dx.doi.org/10.1016/S0140-6736(10)60870-7

Cragin, M. H., Palmer, C. L., Carlson, J. R., & Witt, M. (2010). Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 368,* 4023–4038. http://dx.doi.org/10.1098/rsta.2010.0165

Decoding the N.F.L. database to find 100 missing concussions. (2016, March 24). *The New York Times*. Retrieved from https://www.nytimes.com/2016/03/25/sports/football/at-least-100-concussions-left-out-of-nfl-studies.html

de Montjoye, Y.-A., Radaelli, L., Singh, V. K., & Pentland, A. S. (2015). Identity and privacy. Unique in the shopping mall: On the reidentifiability of credit card metadata. *Science, 347,* 536–539. http://dx.doi.org/10.1126/science.1256297

Drabiak-Syed, K. (2010). Lessons from Havasupai Tribe v. Arizona State University Board of Regents: Recognizing group, cultural, and dignitary harms as legitimate risks warranting integration into research practice. *Journal of Health & Biomedical Law, VI,* 175–225.

Friedlander, A., & Adler, P. (2006). *To stand the test of time: Long-term stewardship of digital data sets in science and engineering: A report to the National Science Foundation from the ARL Workshop on New Collaborative Relationships–The Role of Academic Libraries in the Digital Data Universe*. Retrieved from http://files.eric.ed.gov/fulltext/ED528649.pdf

Grady, C., Eckstein, L., Berkman, B., Brock, D., Cook-Deegan, R., Fullerton, S. M., . . . Wendler, D. (2015). Broad consent for research with biological samples: Workshop conclusions. *The American Journal of Bioethics, 15,* 34–42. http://dx.doi.org/10.1080/15265161.2015.1062162

Gymrek, M., McGuire, A. L., Golan, D., Halperin, E., & Erlich, Y. (2013). Identifying personal genomes by surname inference. *Science, 339,* 321–324. http://dx.doi.org/10.1126/science.1229566

Health Information Portability and Accountability Act (1996). Pub. L. 104–191, § 110 Stat. 1936.

Holdren, J. P. (2013, February 22). *Memorandum for the heads of executive departments and agencies: Increasing access to the results of federally funded scientific research*. Retrieved from https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

Humphreys, R. A. L. (1970). *Tearoom trade: A study of homosexual encounters in public places*. London, United Kingdom: Duckworth.

Khoury, M. J., Little, J., & Burke, W. (2004). *Human genome epidemiology: A scientific foundation using genetic information to improve health and prevent disease*. Oxford, United Kingdom: Oxford University Press.

Koslow, S. H. (2000). Should the neuroscience community make a paradigm shift to sharing primary data? *Nature Neuroscience, 3,* 863–865. http://dx.doi.org/10.1038/78760

Lewis-Kraus, G. (2016, January 12). The trials of Alice Goffman. *The New York Times*. Retrieved from https://www.nytimes.com/2016/01/17/magazine/the-trials-of-alice-goffman.html?_r=0

Margolis, R., Derr, L., Dunn, M., Huerta, M., Larkin, J., Sheehan, J., . . . Green, E. D. (2014). The National Institutes of Health's Big Data to Knowledge (BD2K) initiative: Capitalizing on biomedical big data. *Journal of the American Medical Informatics Association, 21,* 957–958. http://dx.doi.org/10.1136/amiajnl-2014-002974

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, Department of Health, Education and Welfare. (1978). *The Belmont Report: Ethical principles and guidelines for the protection of human subjects of research*. Washington, DC: U. S. Government Printing Office.

Nelson, B. (2009). Data sharing: Empty archives. *Nature, 461,* 160–163. http://dx.doi.org/10.1038/461160a

Parry, M. (2015, June 12). Conflict over sociologist's narrative puts spotlight on ethnography. *The Chronicle of Higher Education*. Retrieved from http://www.chronicle.com/article/Conflict-Over-Sociologists/230883

Pisani, E., Whitworth, J., Zaba, B., & Abou-Zahr, C. (2010). Time for fair trade in research data. *Lancet, 375,* 703–705. http://dx.doi.org/10.1016/S0140-6736(09)61486-0

Pisani, E., Whitworth, J., Zaba, B., & Abou-Zahr, C. (2011). Withholding access to research data—Authors' reply. *Lancet, 375,* 1873. http://dx.doi.org/10.1016/S0140-6736(10)60871-9

Protection of Human Subjects, 45 C. F. R. 46, Subpart A (2009).

Savage, C. J., & Vickers, A. J. (2009). Empirical study of data sharing by authors publishing in PLoS journals. *PLoS ONE, 4,* e7078. http://dx.doi.org/10.1371/journal.pone.0007078

Schnarch, B. (2004). Ownership, control, access, and possession (OCAP) or self-determination applied to research: A critical analysis of contemporary First Nations research and some options for First Nations communities. *Journal of Aboriginal Health, 1,* 80–95.

Strand, N. K. (2016). Shedding privacy along with our genetic material: What constitutes adequate legal protection against surreptitious genetic testing? *American Medical Association Journal of Ethics, 18,* 264–271. http://dx.doi.org/10.1001/journalofethics.2016.18.3.pfor2-1603

Sweeney, L. (1997). Weaving technology and policy together to maintain confidentiality. *The Journal of Law, Medicine & Ethics, 25,* 98–110. http://dx.doi.org/10.1111/j.1748-720X.1997.tb01885.x

Sweeney, L. (2000). *Simple demographics often identify people uniquely* (Data Privacy Working Paper 3). Retrieved from https://dataprivacylab.org/projects/identifiability/paper1.pdf

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., . . . Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLoS ONE, 6,* e21101. http://dx.doi.org/10.1371/journal.pone.0021101

Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., . . . Dorsett, K. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLoS ONE, 10,* e0134826. http://dx.doi.org/10.1371/journal.pone.0134826

Wahlquist, C. (2016, August 17). Indigenous DNA at centre of ethical furore could help reconnect stolen generations. *The Guardian*. Retrieved from https://www.theguardian.com/australia-news/2016/aug/18/indigenous-dna-at-centre-of-ethical-furore-could-help-reconnect-stolen-generations

Walport, M., & Brest, P. (2011). Sharing research data to improve public health. *Lancet, 377,* 537–539. http://dx.doi.org/10.1016/S0140-6736(10)62234-9

Received May 4, 2017
Revision received October 10, 2017
Accepted October 13, 2017 ■