Data Ownership, Reciprocity, and Research
Meeting Summary

Background

Governments, companies, academic research institutions, and the public are increasingly aware of the value of personal data, prompting a reevaluation of the question of who “owns” and controls data access and use. Privacy laws, such as the General Data Protection Regulation (GDPR)\(^1\) and the California Consumer Privacy Act,\(^2\) aim to increase transparency around the collection and use of data, establish legal bases for the processing of consumer data, and/or establish consumer rights to decide how their data are used. Pharmaceutical and life sciences companies and academic research institutions seek to leverage, and sometimes sell, data for secondary research. While the specifics differ, other companies (e.g., Consuli,\(^3\) Private Access,\(^4\) Hugo Health,\(^5\) Hu-manity.co,\(^6\) and others) aim to empower individuals to control who accesses their health data and how their data are used for research and other purposes.

Discussion at the April 5 Bioethics Collaborative focused on two key questions:

- Who should **own, control, and benefit from** the use of personal data?
- What would a **model of stewardship** for research data look like?

The meeting scope included data collected during clinical research and health data from the medical record that are used for research.

*Issues With Current Approaches to Data Management*

---

\(^1\) Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119).


After an introductory presentation by the MRCT Center, discussion turned to identifying issues with the current state of data management in research. Bioethics Collaborative attendees agreed that data are a public good but noted that ethics boards often restrict data sharing for research due to the perceived risks to individual privacy and confidentiality, while undervaluing the benefits to the public good. Institutional Review Boards (IRBs) are aware of but not responsible for interpreting and applying state, national, and international privacy laws and regulations (e.g., the Health Insurance Portability and Accountability Act, GDPR). To minimize institutional risk, IRBs may approach these laws and regulations conservatively, limiting data sharing to the extent possible. In addition, the definition and valuation of “public good” are ambiguous and obscure, while risks to privacy and confidentiality, on the other hand, are well-defined and familiar.

The process of requiring a participant to provide consent to share/use their health or research data appears to respect autonomy and the right to informational self-determination (i.e., “a person’s ability to freely decide whether and how personal data and information about her are collected, stored, multiplied, processed and transferred by third parties”). Some Bioethics Collaborative attendees, however, questioned whether a consent-based model is the best approach.

First, there are unintended and problematic consequences of requiring consent. Marginalized populations who experience or have experienced negative interactions in the healthcare system—individually or collectively—may be less likely to provide consent for data sharing and use. Issues of distrust may contribute to hesitancy. Requiring consent, then, might result in a lack of representativeness in datasets used for secondary research. Genomics and other databases already lack diversity by ancestry, race, ethnicity, and socioeconomic status, among other factors. The results of secondary research will not include information from groups left out of the data, resulting in increasing health disparities. Too strong an emphasis on consent may inadvertently end up harming the very individuals and groups it seeks to protect.

Second, for some research participants, consent itself does not do enough to enable autonomy, even when that consent is informed. These participants argue that individuals should have control over their research data and the ability to share it with whomever they wish, including their physician(s) and other researchers, rather than their data being controlled by researchers. The stakes are particularly high for research participants who have unmet medical needs and enroll in research partly to learn about their condition and/or receive possible treatments. In the United States (U.S.), participants have been denied access to their data secondary to interpretations of the Clinical Laboratory Improvement Amendments (CLIA) that prevent non-

---

7 Christoph Schickhardt et al., Do patients and research subjects have a right to receive their genomic raw data? An ethical and legal analysis, BMC Med. Ethics, 7 (2020).
CLIA certified laboratories from returning individual results that could be used to make clinical decisions or have personal utility.

Finally, whether or the extent to which consent is “informed,” that is, whether participants understand the nature of the risks and benefits of sharing their personal data, is not always clear. Indeed, the process of consent fails to enable autonomy when individuals consent to data sharing without being informed or understanding key features of the choice. Research consent forms are, however, lengthy and complex, posing barriers to participant comprehension. In clinical settings, the consent forms that authorize data and biospecimen sharing for research are increasingly presented in ways that do not facilitate understanding. Patients, for example, may be asked to provide an electronic signature without accessing or reading the appended consent and/or may be led to believe that receiving medical care is conditional upon consenting to data and biospecimen sharing. Rarely do patients understand the technical differences between identifiable, de-identified, and anonymous data or the differences in risk between these categories.

Potential Approaches

Bioethics Collaborative attendees proposed a number of approaches for consideration, grounded in the contextual nature of assessing the benefits and risks of data sharing and use. Ethics boards should perform balanced assessments of the benefits of data sharing to the public good, compared with the risks of informational privacy. To this end, they should consider concrete examples of public good made possible by data sharing.

It was not lost on attendees that identifiable information is routinely shared outside of healthcare, and shared either without permission or based on terms of use in an agreement that almost no one reads. Every credit card purchase and Google search is or can be tracked without specific consent for its use. But a Google search is not always equivalent to sharing personal, often sensitive, health information. Attendees voiced affirmation of the potential benefit of third-party, external review for proposals involving the sharing of participant data. That external party need not be an IRB or research ethics committee; a separate board or a

---

8 National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Health Sciences Policy, Committee on the Return of Individual-Specific Research Results Generated in Research Laboratories, Returning Individual Research Results to Participants: Guidance for a New Research Paradigm (Autumn Downey et al. eds., 2018).

9 Federal regulations apply “to all research involving human subjects,” in which human subject means a living individual about whom an investigator (whether professional or student) conducting research:
   (i) Obtains information or biospecimens through intervention or interaction with the individual, and uses, studies, or analyzes the information or biospecimens; or
   (ii) Obtains, uses, studies, analyzes, or generates identifiable private information or identifiable biospecimens. (45 CFR 46.101 and §46.102(e).

Federal regulations do not, therefore, apply to some contexts of data sharing and use discussed in this Bioethics Collaborative.
data use committee might be empaneled to serve this role, ensuring that the risks are minimized and balanced by the potential benefits. Board members would ideally have expertise in data ethics, and the board’s review would not be constrained by federal regulations written to protect identifiable data and biospecimens.

Another, complementary approach would be to enact laws that implement fiduciary duties of stewardship and beneficence upon the parties that control clinical research data and health data used for research. Implementing a data review board and/or stewardship laws may obviate the need for—and may be preferable to—consent-based models of data management.

Nevertheless, consent will remain important even if under less expansive conditions. Bioethics Collaborative attendees considered the potential for “user-centered design”10 to improve the consent form and process. Notably, user-centered design would need to consider cultural and linguistic differences among different populations, with some expressing skepticism about its ultimate promise.

Paying research participants and patients “data dividends” for the use of their data for research was suggested. Compensation respects individuals who contribute their data and may encourage other individuals to authorize the use of their data for research. However, the prospect of payment, even in small amounts, could raise concerns that research participants and patients are being unduly influenced to share their data. At the same time, we lack adequate data on whether payment in research contexts distorts decision-making in any concerning way. Additionally, compensation may be less valuable to some participants than returning results.

Bioethics Collaborative attendees suggested other solutions to encourage the use of data for the public good, respect the individuals contributing data to research, and/or build trust in the parties that control the use of research data. The importance of educating individuals and the public on the importance of health and research data and the tenets of ethical data management was emphasized. The ideal setting for this education in the U.S. may be high school health classes; students are responsive to new information, and health classes are mandated in many states. In addition, data controllers should be accountable for their actions and consequences for misuse defined.

---