PRIM&R
PUBLIC RESPONSIBILITY IN MEDICINE AND RESEARCH

January 14, 2025

Department of Commerce
National Telecommunications and Information Administration
1401 Constitution Avenue NW
Washington, DC 20230

RE: Response to Request for Public Comment, Docket No. 241204-0309 Ethics and Privacy Guidelines for Research Using Pervasive Data

To whom it may concern:

Public Responsibility in Medicine & Research (PRIM&R), which has more than 3,500 active members throughout the research enterprise, appreciates the opportunity to respond to the National Telecommunications and Information Administration's (NTIA) request for public comment on the potential development of ethical guidelines for research using "pervasive data."

PRIM&R appreciates the opportunity to offer our input on this important issue and commend NTIA for its commitment to working to develop ethical research practices with respect to pervasive data. Establishing clear, comprehensive, and thoughtfully crafted ethical guidelines that complement existing standards for scientific research is essential to ensure responsible research in this fast-evolving field. The term *pervasive data,* according to the NTIA[1], "is intended to mean data about people—user-contributed, observed, derived, or inferred—collected through online services regardless of the extent to which the data is publicly available, is aggregated, or could lead to the identification of an individual.[2]As an organization dedicated to advancing the highest ethical standards in research, PRIM&R recognizes the crucial need for a framework that addresses the unique challenges presented by this rapidly evolving field.

PRIM&R recognizes "pervasive data" research offers insights into human behavior and societal trends, which can inform policy decisions and improve public well-being. However, this research also raises complex ethical considerations related to consent, privacy, data security, and potential harms to individuals and society. The development of national ethical guidelines represents an opportunity and an important step toward ensuring pervasive data research is conducted responsibly and ethically, fostering

[1] https://www.federalregister.gov/documents/2024/12/11/2024-29064/ethical-guidelines-for-research-using-pervasive-data#footnote-1-p99844

[2] The NTIA states that the definition currently under consideration "... may include text, images, videos, biometric information, information about a data subject's behavior (purchases, financial standing, media consumption, search history, medical conditions, location, etc.), and other information that makes up a person's digital footprint." Further, the NTIA states, "Online services may include a wide range of information technologies throughout the technology stack/technical infrastructure, including but not limited to web-based monitoring tools, content delivery networks, blockchain technology, digital labor platforms, education technology, Internet of Things devices, connected cars, wearable devices, mobile sensors, data brokers, streaming services, search engines, online marketplaces, social media platforms, and AI systems."

public trust while promoting valuable scientific inquiry. These guidelines should build upon existing ethical principles, such as those articulated in the Belmont Report, to offer clear and comprehensive guidance for researchers navigating the complexities of pervasive data. Ideas noted in the 2012 Menlo Report[3] should preempt new discussion and expand understanding. Notably, the Menlo Report includes specific context for Respect for Persons, Beneficence and Justice with applied relevance to pervasive data. These include voluntary participation, continued efforts to do no harm, and approaches to research with balanced justice. The Menlo Report also broadens these tenets to mention legal due diligence, transparency, and accountability. Indeed, all are vital for the continuation of ethical research with pervasive data. PRIM&R offers our comments regarding parallels to existing regulation and applicability of the Menlo Report to AI in today's Common Rule framework.

PRIM&R will focus its comments on the following key areas identified by the NTIA:

<u>Question 1</u>. What are the potential benefits of developing national-level ethical guidelines for researchers collecting, analyzing, and sharing pervasive data?

<u>Comment:</u>

The development of national-level ethical guidelines for researchers collecting, analyzing, and sharing pervasive data would fill gaps not addressed in current federal regulations governing human subjects research. Current regulations are difficult to relate consistently to research with pervasive data. A challenge becomes evident when attempting to apply regulations designed either for prospective data collection or existing data that rely on unspecified deidentification methodologies[4] and consent practicability[5] standards. The responsibility of reviewing data in these categories is vague and must be addressed in the digital space. This is especially pertinent if the guidelines address cases where research using pervasive data does not meet the definition of human subjects research. Ethical guidelines can provide guiding principles that empower researchers to navigate the complexities of pervasive data while respecting ethical standards including:

Enhanced Trust and Transparency: Clear and consistent ethical guidelines can foster trust between researchers, data subjects, and the public by promoting responsible and transparent data practices. Transparency can also improve public understanding and confidence in research.

Protection of Data Subjects: Guidelines can help mitigate the risks of privacy violations, re-identification, and other harms to individuals by establishing standards for data handling and use.

Addressing Societal Risks: Ethical guidelines can help researchers anticipate and mitigate potential societal-level risks, such as discrimination, bias, and the erosion of trust in research.

---

[3] Erin Kenneally & David Dittrich, *The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research*, SSRN JOURNAL (2012), http://www.ssrn.com/abstract=2445102 (last visited Nov 7, 2024).
[4] 45 CFR 46.104 (d)(4)
[5] 45 CFR 46.116(f)(3)

Ethical Guidance for Researchers: National guidelines would offer researchers a framework for understanding ethical complexities, promoting best practices, and addressing potential harms throughout the research process.

Complementing Existing Ethical Frameworks: National guidelines can build upon existing ethical frameworks, such as the Belmont Principles, to provide a comprehensive approach to pervasive data research. They can also complement existing regulatory frameworks like the Common Rule.

These principles enhance education and awareness about ethical responsibilities, aligning with PRIM&R's commitment to advancing ethical research through education and professional development.

As part of the ethical guidelines developed, the NTIA should consider making recommendations to organizations regarding the type of policies or review processes they might need beyond the IRB process to assess the ethics of research that might not meet the definition of human subjects research promulgated by the Common Rule (e.g., because the identity of the subject may not be readily ascertained by the investigator or associated with the data) or is considered exempt (e.g., due to variations in the interpretation of how "publicly available" is defined), but nonetheless may pose harms to groups or individuals or appear to be a violation of privacy expectations.

Additionally, providing recommendations for ethics training for researchers using pervasive data would strengthen ethics concerning pervasive data. As Zimmer notes[6], researchers in big data fields often do not undergo the more comprehensive training as other researchers who work with data or biospecimens derived from humans.

<u>Question 2.</u> What are the potential drawbacks of developing national-level ethical guidelines for researchers collecting, analyzing, and sharing pervasive data?

<u>Comment:</u>

PRIM&R understands the federal government may envision a role in setting legal and ethical standards to avoid egregious or malicious use of pervasive data. The pace of technology and the pace of oversight must operate similarly. The current bandwidth of the U.S. Government (USG) may not suffice to support routine and pointed regulatory efforts. Further, other data classifications may overlap and apply to multiple data types creating difficulty implementing a "one size fits all" model and contribute to redundancy. Proposed national-level guidelines must not hamper the progress of research with public information or research that does not differ from data collected in everyday life. PRIM&R cautions that an overreach of rigidity could inadvertently limit flexibility in innovative research methodologies. These factors converge with data ownership, retention, and inclusivity of diverse disciplines. There are challenges to setting effective boundaries that deter abuse and actually would warrant enforcement.

The responsibility to define terms and understand the field must not fall to one body alone. First, a consensus on subcategories associated with pervasive data must be clearly defined for researchers, IRBs, scientific organizations, and participant advocacy groups in a collaborative consortium with the federal government. This must exist prior to codifying any rules. PRIM&R also encourages consideration of all

[6] Michael Zimmer, *Addressing Conceptual Gaps in Big Data Research Ethics: An Application of Contextual Integrity*, 4 SOCIAL MEDIA + SOCIETY 2056305118768300 (2018), https://journals.sagepub.com/doi/10.1177/2056305118768300 (last visited Jan 8, 2025).

research institutions to ensure those with lower resources are not driven out of research by an overabundance of regulation.

Principles that are nebulous can be either purposely avoided or inadvertently missed. Oftentimes, harm can be done without intent. Thus, clear guidelines and understandable definitions can prevent harm from occurring.

Government appointed advisory groups requiring relevant expertise, similar to the HHS Secretary's Advisory Committee on Human Research Protections (SACHRP), may be beneficial in creating guidance and modernized governance. This could also allow transparent, non-technical language to serve as a discoverable resource to the public, while allowing any enforcement activity to reside within the USG.

Additionally, an effective solution includes building partnerships to work with government entities to continuously fund and support shared responsibilities through demonstration projects or coalitions. As a research compliance organization, committed to advancing the highest ethical standards, PRIM&R welcomes the opportunity to assist with these efforts.

Question 6. Consent and autonomy are key principles in human subjects research ethics. However, users of online services may be required to divulge certain personal information and/or have no ability to freely make decisions about its use. How should researchers working with pervasive data consider consent and autonomy?

Comment:

Data are not equal from a risk perspective and must be clearly categorized based on the research questions are being explored. Regulations in research often employ phased and/or tiered risk or development structures. For example, FDA Clinical Research Phases[7] offer strategies and clear steps from feasibility to marketing and beyond. This standard could help set precedent for a systematic way to transition broad models to pointed, intentional constructs.

Further, ethical committees are organized to first classify review processes on the basis of research risk criteria. This model establishes prioritization and allows the field to narrow the types of research that require higher level oversight (e.g., mental health, genetics, illegal activities), while excluding or exempting clearly lower-risk research. Given the everyday use of pervasive data already encountered for non-research purposes, the risk of use for research may look similar to existing categories in human subjects research.[8]A tiered risk structure could overlie the Common Rule and other associated frameworks around risk. For example, a multi-tiered structure could involve developing a color-coded labeling system to flag different types of data collection. Therefore, the level of identifiable data use required to inform affected participants and allow consent options could parallel those required in the Common Rule.

Question 7(h). How can researchers best conduct research with pervasive data in a way that engages the community, users, and data subjects? What are the best practices for such

[7] 21 CFR 312.21 also outlined by FDA at https://www.fda.gov/patients/drug-development-process/step-3-clinical-research#Clinical_Research_Phase_Studies

[8] 45 CFR 46 et seq.

participatory research that uses pervasive data? What are the challenges and/or barriers to conducting participatory research? What important research questions cannot be answered using participatory mechanisms, and why?

Comment:

In line with PRIM&R's commitment to diversity, equity, inclusion, and justice, best practices for participatory research include engaging data subjects and communities early, fostering shared ownership of the research process. Complexity exists when these data are to be made generalizable and are susceptible to concerns of security, misuse, and bias. Given the known importance of public trust in science and innovation, opportunities must exist for participatory research with pervasive data. Community based participatory research (CBPR) sets a clear framework for future movements.

Transparency in goals and outcomes helps ensure meaningful engagement and trust. PRIM&R emphasizes the importance of applying these principles in the digital world, especially when community members are known and accessible to researchers (e.g., social media groups, patient advocacy connections). Research resonates within communities with place-based efforts that highlight community awareness and the long-lasting relationships that can be built with researchers based on conversations about needs and societal or health challenges.[9]

Indeed, this model of CBPR is useful and promotes scientific literacy and builds trust and education in science for those that participate or encounter the results. Clear benefits must accompany clear risk descriptions. Standardized format and plain language along with verification of the participant's understanding must be considered as risk increases. Communities should contribute to the design and development in pervasive research and define in advance what boundaries must remain during reporting and future research. Regulation could assist in this way to ensure that both researchers and ethics bodies are consistent in their application of community-based collaborations.

Notification

8e. Under what conditions should data subjects be notified that their data is used for research?

Comment:

Researchers and regulators together must first define the risk levels associated with pervasive data. Categorization aligned with the Common Rule could be established to expand existing understanding. Some pervasive data already used for non-research purposes may fit into excluded or exempt minimal risk research designations not requiring signed or documented informed consent under current standards. Whereas research completed for a purpose to deduce identity, or that inadvertently engineers methodologies that unexpectedly expose or reverse engineers identity for research purposes, should contain a planned mechanism for consent; this is especially true when the topics of exposure do not qualify for exemption under the Common Rule.

Ethical data use should prioritize:

[9] Barbara L. Brush et al., *Success in Long-Standing Community-Based Participatory Research (CBPR) Partnerships: A Scoping Literature Review*, 47 HEALTH EDUC BEHAV 556 (2020), https://journals.sagepub.com/doi/10.1177/1090198119882989 (last visited Jan 3, 2025).

- Transparent communication with data subjects, emphasizing accessibility and cultural sensitivity.
- Scenarios where notification may be impractical (e.g., anonymized or aggregated datasets) should be addressed through clear policies, ensuring researchers balance ethical obligations with feasibility.
- Innovative tools, like dynamic consent platforms, can help researchers achieve transparency without creating unnecessary administrative burdens.
- Researchers should also be encouraged to broadly share results of research conducted with pervasive data. Sharing research findings with the public bolsters trustworthiness in science.

What are necessary and/or best practices for communicating with data subjects when their data is used for research?

Data owners are also responsible to safeguard and protect its use. When organizations, their suppliers, subcontractors, subsidiaries, etc. agree to share data for research purposes, many require the use of agreements or assertations from data researchers or third-party processors. Though not all data hold the same high disclosure risk, it could be useful to borrow some basic tenets from HIPAA. For example, the principles behind a formal Data Use Agreement (45 CFR § 164.514(e)(2)) used for a "Limited Data Set" apply to limiting data to specific, defined categories. Importantly, prior to using these data, recipients assert that they will not purposefully reidentify or contact participants and will store data at a level of an agreed standard. These concepts could be tailored to situations when specific highly-personal or greater than minimal risk pervasive data (e.g., internet browsing histories, genetic information, or location data) are utilized for research purposes.

What barriers exist to notifying data subjects?

A large barrier exists in attention to privacy and research practices. While internet users are aware of tracking, scamming, spam, phishing, and other malicious attempts of data use, a more effective model would provide more transparency and public engagement around research using pervasive data. When pervasive data are collected, data subjects generally do not understand the technical terms and potential purpose of widespread sharing. PRIM&R believes in trustworthy and transparent practices for research data use.

Pop-up and acceptance of cookies can be an unclear and ineffective form of messaging. As noted above, a multi-tiered structure could be developed with a color-coded labeling system to flag different types of data collection. A responsibility exists for data owners to provide plain language and ADA accessible descriptions of how data are used and how to preserve one's right to privacy. Further, the use of public messaging campaigns by NTIA or other areas of the federal government may contribute to a better understanding of data use.

10a. What steps should researchers take to protect data subjects or against societal-level harms prior to the dissemination of research outputs (publications, presentation slides, data visualization, datasets, AI/ML models, etc.)?

Comment:

Historic research atrocities often point to alienation or exploitation of groups without their knowledge or consent. Therefore, both the magnitude of potential risk along with the probability of such risks must

factor into any definitions or framework around pervasive data. It is important to note that any limitations on the use of pervasive data should have clear reasoning. Standards set too stringently may have ramifications on development. This supports the use of a prioritization structure to allow instances with the highest risk levels to be expertly evaluated.

As stated in the Menlo Report,[10] "Taken as a whole, the intent of the Common Rule is to protect persons who might be harmed from involvement in research, not simply with whether humans are participating in research." This illustrates the caution that must be revisited while researchers and IRBs contemplate the interactions between pervasive data and their study. Use of deidentification algorithms and "build rules," designed to avoid reidentification, could be the norm, with exceptions requiring consent and/or ethical approval by a neutral group. Data sources must be reliable and analyzed within positionality and availability contexts. Further, ensuring the integrity and accuracy of any data also would help to minimize the risk of bias and misinterpretation.

11. What existing ethical frameworks, such as those from professional organizations or government agencies, should be considered when drafting national-level ethical guidelines for research with pervasive data?

Comment:

Rather than creating new governing bodies, researchers should build upon existing ethical standards such as the Belmont Report, Menlo Report and the Common Rule. Further, we must leverage professional guidelines and principles, including PRIM&R's focus on inclusivity, to ensure guidelines are not duplicative. Instead, new guidelines should complement existing systems.

For example, some research concerns about pervasive data parallel the long-term storage and identifiability concerns outlined with early high-throughput or next-generation genetic research. Ethicists contemplate the potentials for propagation of bias and machine learning in next generation sequencing and note that results must include analysis and interpretation to be beneficial to a patient/participant.[11] These considerations of public benefit and community understanding are of utmost importance to PRIM&R.

Further, in culturally specific research, IRBs often rely on expertise from a person familiar with the community in the form of a cultural appropriateness statement. While this mechanism merely creates a spokesperson for the culture, it is still a relevant start to acknowledge formally. This framework is already familiar to researchers and IRBs. Teaching principles of cultural humility and empathy in research contexts is recognized and PRIM&R supports embedding this into required training curriculum with a clear assessment and continuous development as the pervasive data research field expands.

[10] Kenneally and Dittrich, *supra* note 2.

[11] Nicole Martinez-Martin & David Magnus, *Privacy and Ethical Challenges in Next-Generation Sequencing*, 4 EXPERT REVIEW OF PRECISION MEDICINE AND DRUG DEVELOPMENT 95 (2019), https://www.tandfonline.com/doi/full/10.1080/23808993.2019.1599685 (last visited Jan 3, 2025).

12. What are the existing requirements and legal obligations that impact research with pervasive data?

a. What are the risks around research that uses pervasive data, if any, that currently fall beyond the usual considerations of IRBs operating under the Common Rule or FDA regulations.

Comment:

The task of developing guidance for ethical use of pervasive data is not insurmountable and must mirror past and present advancements in medical, behavioral, and computational science. As discussed in the RFC, the use of the Belmont Report and Menlo Report must serve as key guidance tools for any regulation or framework. Additionally, structure could be borrowed from existing state and international frameworks like CCP and GDPR. PRIM&R continues to promote harmonization of the policies that the federal government issues around research. Though the efforts around the broad application of these privacy laws may present too much overall burden to implement large scale, there are notable considerations that allow an amalgamation of relevant existing effective rules. Several examples are provided for context.

Rule/Framework/Law	Relevant Concept	Potential Applicability to Pervasive Data Research
HIPAA 45 CFR 164.512(i)	Honest Broker	Standardized efforts exist that allow neutral intermediaries to deidentify data.
HIPAA 45 CFR § 164.502	Minimum Necessary	Researchers looking to utilize pervasive data define why each datapoint is necessary to the research question.
Declaration of Helsinki (v.2024)	Defining Participants	The recent changes to the Declaration of Helsinki utilize terminology to further address and recognize autonomy as a participant rather than a human subject.
Export Administration Regulations (EAR) at 15 CFR § 764.2.	Willful Violation	Consequences of deliberate, unauthorized disclosure are prioritized and enforceable.
The Common Rule 45 CFR §46	Key Information	Short, concise explanations of what might allow a person to decide to participate data.
General Data Protection Regulation (GDPR): Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016	Records of Processing Activities	Defined logs and records exist to track compliance and breech potential.
California Consumer Privacy Act (CCPA), California Civil Code at §§1798.100-199.	Right to be forgotten	Clear parameters must exist to opt-out
FDA Clinical Trial Framework 21 CFR 312.21	Phased design	While the effort would not necessarily need to be focused on marketing. Using a phased or tiered framework to define broad efforts versus focused efforts would assist in a categorization framework understood by existing IRBs.

NIH Guidelines for Research Involving Recombinant or Synth etic Nucleic Acids, Section IV-B-4 through IV-B-6.	Board Expertise Requirements	Assurance that an ethics board contains data security and/or pervasive data protections expertise when review occurs in this field.
Certificate of Confidentiality (NIH) 21st Century Cures Act P.L. 114-255 Privacy Certificate (NIJ) Protocols 28 CFR § 22.23	Protection from compelled disclosure	Certain limitations must exist when research data are applied beyond typical contexts. Currently, these protections only exist for NIH and NIJ.

b. What steps can be taken to ensure that potential new guidelines for research with pervasive data complement the existing regulatory framework for human subjects research?

Comment:

Ethical guidelines should:

- Address emerging risks beyond the scope of traditional IRBs, such as AI/ML-driven re-identification risks, without duplicating existing regulatory efforts.
- Emphasize the need for collaborative approaches among researchers, institutions, and communities to bridge gaps in current frameworks while maintaining simplicity and clarity.

PRIM&R supports the importance of the need for deep understanding and application of the Belmont Report and aligned Menlo Report. We will continue to provide accessible and understandable programming in this discipline to our research ethics population, researchers, and participants. The basic tenets of ethical research must underscore any technological advancement. We must spotlight and combine the needs of public health, safety, and social justice with information that undeniably supports research discovery.

14. How should ethical guidelines take into account future technological advances around research with pervasive data?

Comment:

PRIM&R considers adherence to the Belmont Report to be the most understood and reliable way to frame the future of pervasive data research. Guidelines should be principle-driven and flexible, avoiding prescriptive regulations that may quickly become outdated.

Respect for Persons: Whenever possible, individuals should possess the autonomy to decide how their interactions with the internet and company services are used in line with the original intention. Any expansion beyond these boundaries must clearly contemplate consent possibility and practicability.

Beneficence: Benefits and risks with pervasive data must be appropriately balanced. If generalizability of data is intended, it must be considered against a gauge of direct benefit first. For example, targeting to better advertise could be viewed as a benefit to conduct research on behalf of a company. But the use of those data may be considered less of a benefit to the participant or the public; especially if a risk of unexpected disclosure is involved. We must be transparent and decisive about what true benefit entails.

Justice: The use of pervasive data may cause inequity in the consumer or internet user data. Those who have the privilege of internet access and literacy to interact may be used to collect data that could be perceived as a norm or standard. This may restrict those persons based on factors of socioeconomic status, intellectual ability, and location to be less represented in research. This must be considered when the use of pervasive data makes claims related to biomedical and social behavioral research.

We must also encourage and allow opportunities for periodic community and stakeholder input to adapt practices in response to new challenges and in line with collective respect. This aligns with PRIM&R's commitment to ongoing dialogue and inclusivity.

Once again, thank you for the opportunity to provide feedback on this critical issue and for your commitment to developing ethical research practices which would apply to pervasive data. Clear, comprehensive, and thoughtfully developed ethical guidelines, which complement existing scientific research ethical guidelines, will be crucial to ensuring the responsible advancement of research in this rapidly evolving area.

Sincerely,

Ivy R. Tillman, EdD, CIP
Executive Director
Public Responsibility in Medicine and Research (PRIM&R)