Pervasive Data

Definition

According to the National Telecommunications and Information Administration:

“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.”

Dr. Jessica Vitak:

“Rich personal information generated through digital interaction and available for computation analysis.

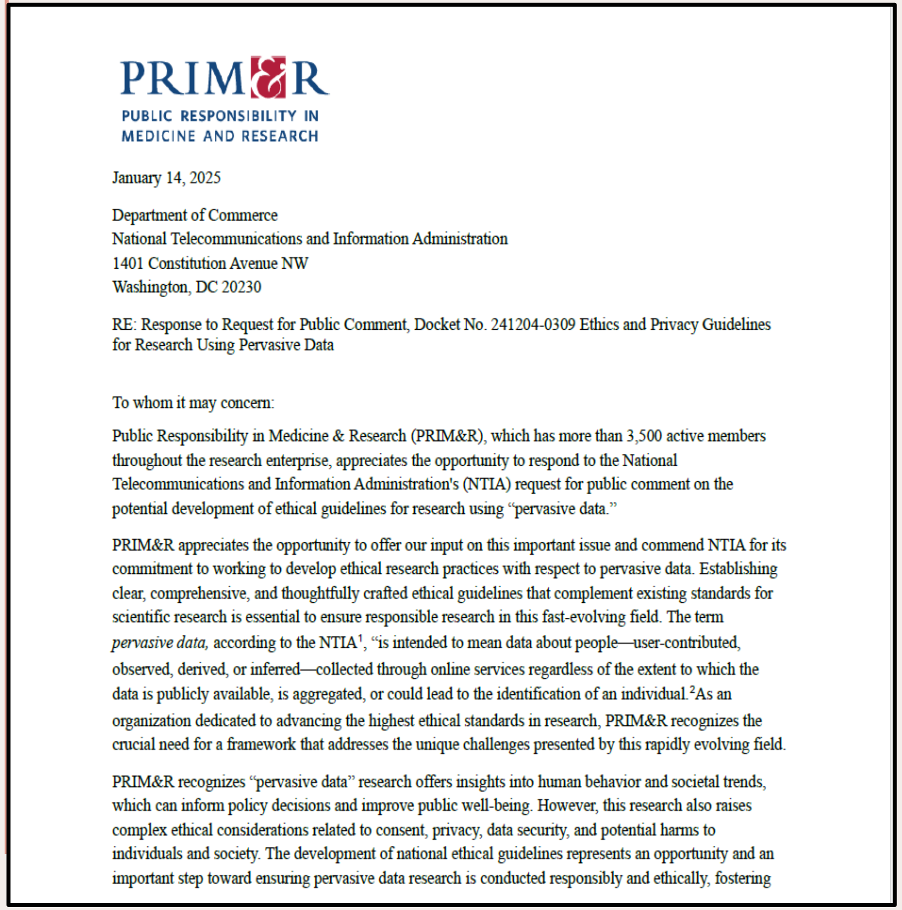
Pervasive data research is research that:

1. Gathers digital data about people;

2. Uses computational methods to understand individuals’ or groups’ health, habits, routines, or beliefs; and

3. Frequently collects data without the awareness of the studied population”

Synopsis of PRIM&R Response



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

Key points of document:

1. Current regulations are "difficult to relate consistently to research with pervasive data"
2. PRIM&R specifically recommends:
   - Building upon existing frameworks (Belmont Report, Menlo Report)
   - Considering a tiered risk structure that could parallel the Common Rule
   - Having clear categorization of data based on research questions
3. For consent and notification:
   - Data types should be clearly categorized based on risk
   - Some pervasive data used for non-research purposes may fit into excluded or exempt minimal risk categories
   - Research aimed at deducing identity should have planned consent mechanisms
4. The guidelines must not hamper progress of research with public information or research that doesn't differ from data collected in everyday life

January 17 Webinar Takeaways

- Meaningful informed consent
- Data quality
- Data sources/data access
- Social media considerations
- Privacy/security considerations
- Data & analysis & data-driven decision making

- AI model development requires large datasets like medical records, accessed via HIPAA waivers and waiver of consent
- Generative AI and new AI approaches make traditional data privacy and deidentification challenging
- Researchers and HRPPs/IRBs may need new risk minimization methods and review pathways for AI development
- For large datasets required to develop AI algorithms, there is a tension between the possible benefits of requiring consent and the potential for less representation and more bias and discrimination
- Public notification and education about personal/medical data use in research is critical for building trust