PERVADE
PERVASIVE DATA ETHICS
for computational research

UNIVERSITY OF MARYLAND 18 56
COLLEGE OF INFORMATION

Ethical Considerations for Pervasive Data Research

Dr. Jessica Vitak
Professor, College of Information
University of Maryland
jvitak@umd.edu

My background

Who am I?

Privacy and ethics scholar evaluating risks of new technologies for more than a decade.

Who & what is PERVADE?

PERVADE (Pervasive Data Ethics for Computational Research) is an interdisciplinary collaboration between seven researchers at six institutions, funded by a six-year grant from the NSF(2017-2023).







New PERVADE focus

In 2024, Michael Zimmer, Casey Fiesler, and I received a new round of funding focused on developing a suite of ethics education resources for computer science/data science students and educators (including RECR training).

What do we mean by pervasive data?

Rich personal information generated through digital interaction and available for computational 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.

Notice

Ethical Guidelines for Research Using Pervasive Data

A Notice by the National Telecommunications and Information Administration on 12/11/2024

PUBLISHED DOCUMENT: 2024-29064 (89 FR 99844)

PDF
Document Details
Document Dates
Table of Contents
Public Comments
Regulations.gov Data
Sharing
Print
Document Statistics
Other Formats

DOCUMENT HEADINGS

Department of Commerce
National Telecommunications and Information Administration
[Docket No. 241204-0309]
RIN 0660-XC064

AGENCY:

National Telecommunications and Information Administration (NTIA), Department of Commerce.

ACTION:

Notice, request for public comments.

SUMMARY:

The National Telecommunications and Information Administration (NTIA) is seeking public input on the potential writing of ethical guidelines for the use of "pervasive data" in research. "Pervasive data" refers to data about people gathered through online services. NTIA will rely on these comments, along with stakeholder engagements, in considering whether to draft and issue non-binding guidelines to assist researchers working with pervasive data. Such guidelines, if warranted, would detail how researchers can work with pervasive data while meeting ethical expectations of research and protecting individuals' privacy and other rights.

Request for public comments just closed (Jan 15) on ethical guidelines for pervasive research!

What are we studying?

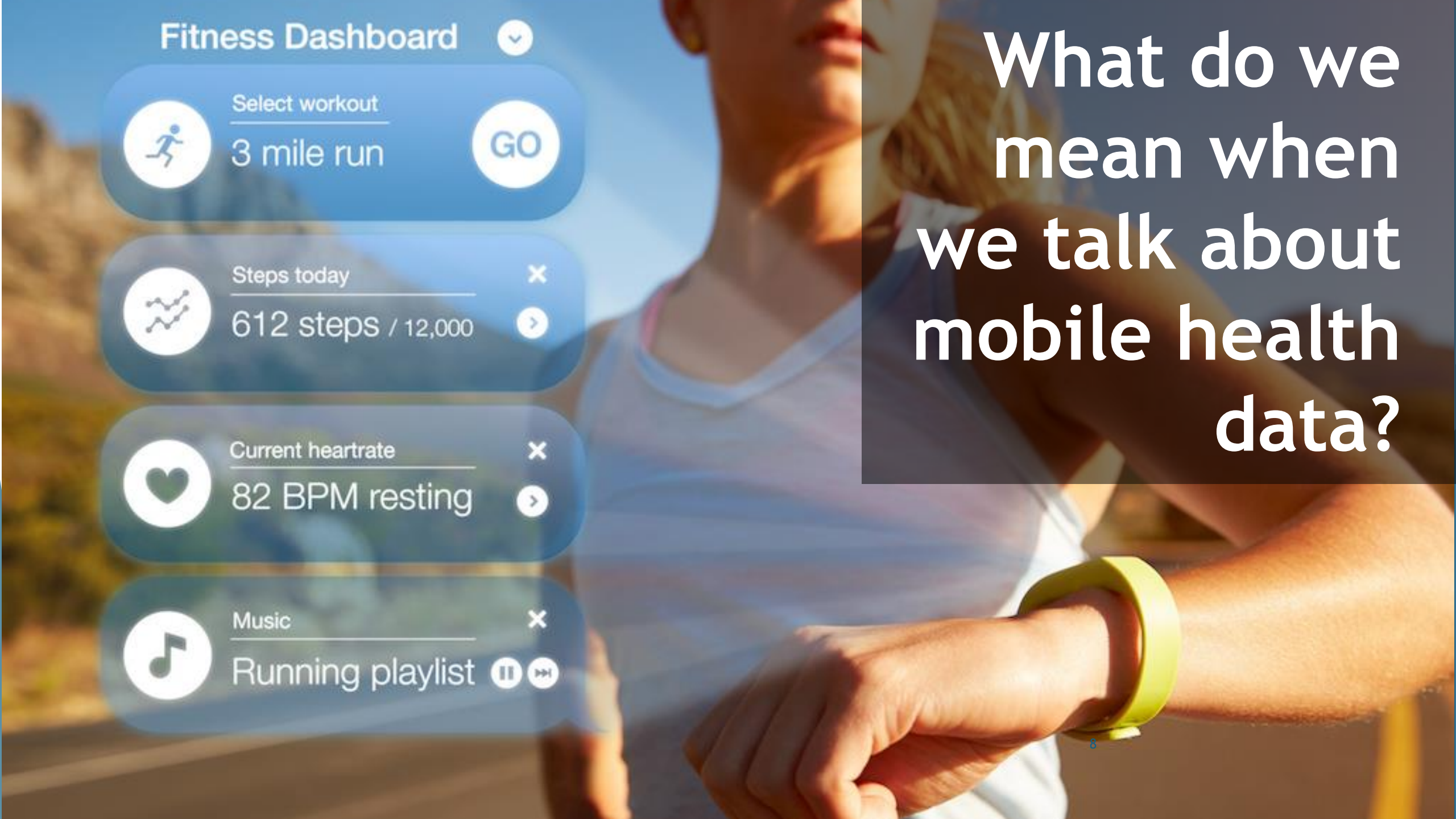
We believe the growth in the scale, scope, speed, and depth of human data research requires reconsideration of fundamental ethical assumptions, including:

1. risks within data itself;
2. beliefs and opinions in user communities;
3. **beliefs and practices in computing research communities;**
4. practices among regulators and policymakers such as IRBs.

See https://pervade.umd.edu for more on our work.

Pervasive data creates a vast gray area regarding (un)ethical research practices

- Is it feasible to collect informed consent?
- Isn’t public data public?
- Should you be more transparent about your research?
- Who is being left out by your data collection strategies?
- Is it possible to truly anonymize a dataset?
- When should AI be used—and when should it not?



What do we mean when we talk about mobile health data?

Tracking everyday health data is expanding thanks to technology

The Quantified Self (QS) uses technology for data acquisition on aspects of a person's daily life.

- **Inputs:** food consumed, symptom tracking, steps taken
- **States:** mood, location, blood oxygen levels
- **Performance:** mental and physical

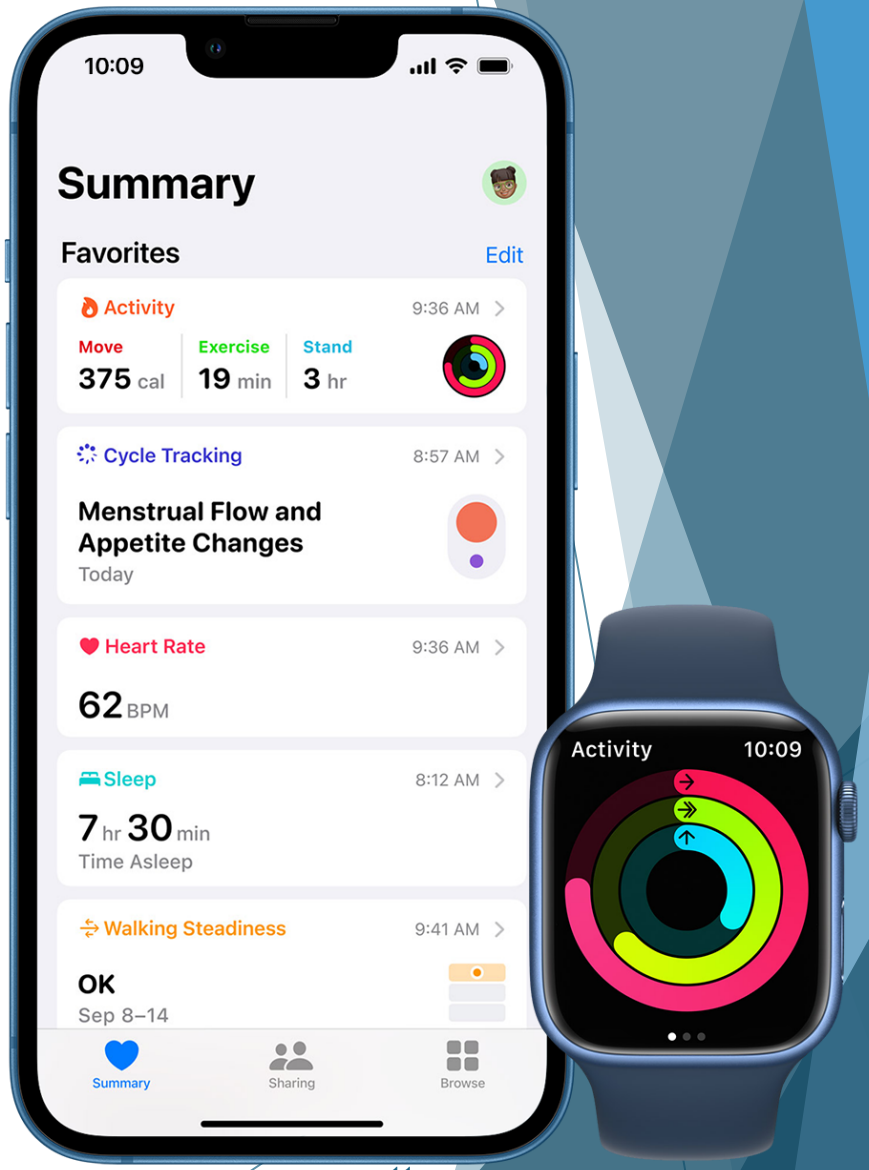
Mobile & wearable technologies facilitate tremendous opportunities for health-related data collection.

- Wearables collect continuous, real-time health inputs.
- Mobile apps can synthesize this data and/or allow user input.
- Mobile apps paired with smart technologies can provide rich and detailed data in situ.

Example: Apple Watch

Apple can collect/track:

- Activity
- Sleep
- Nutrition
- Mindfulness
- Body measurements
- Health records
- Vitals (blood pressure, heart rate, ECG, body temperature, breathing rate)



Example: Disease management

Diabetes tracking (many apps pair with an on-body glucose monitor)

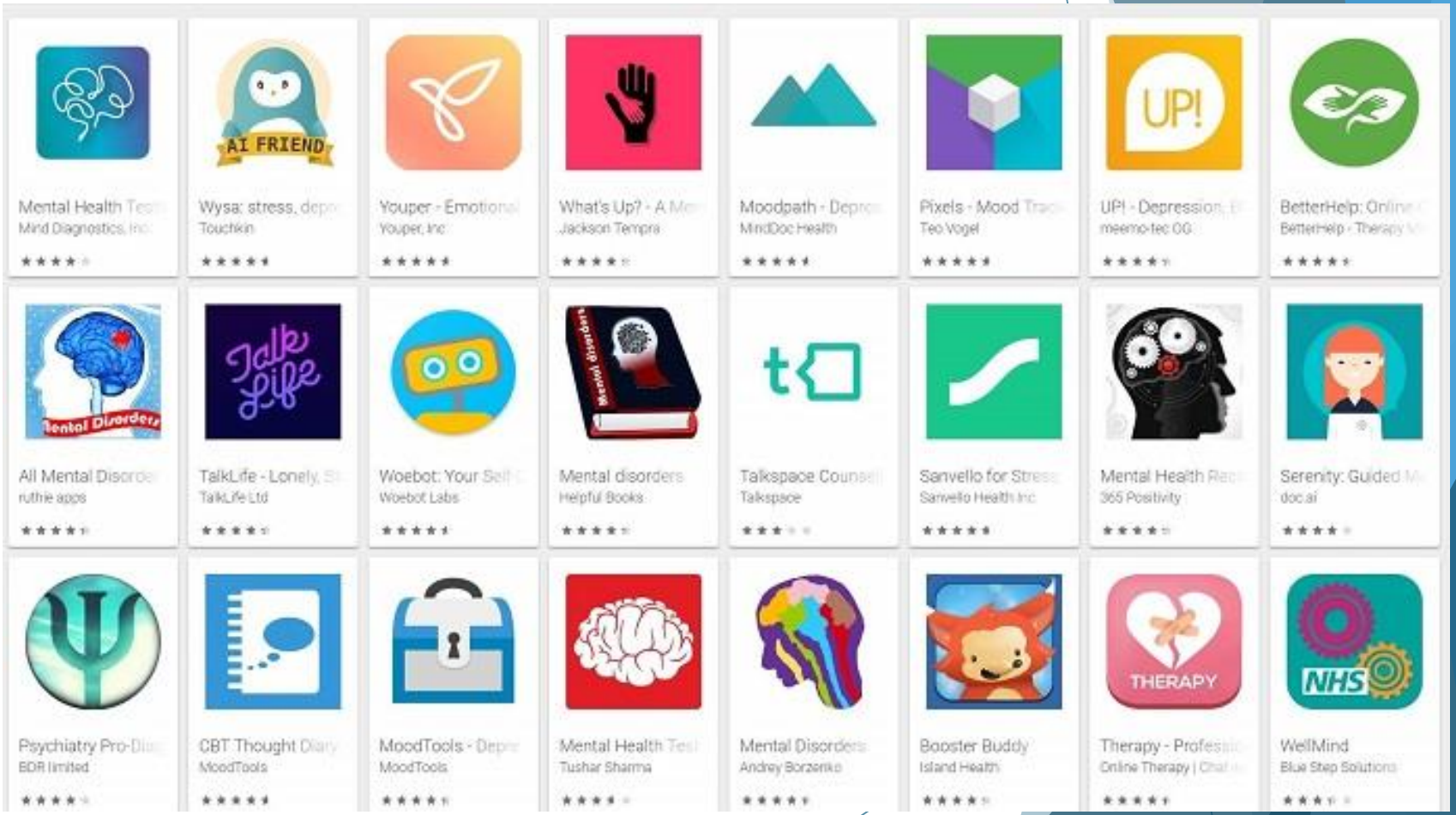
Migraine Buddy tracks triggers, symptoms, and treatment

Example: Smart home integration & data collection

- Smart home technology is increasingly used to enhance health and well-being.
- This is significant interest in how this technology can help older adults age in place and stay connected to loved ones.
- Amazon Echo Show (right) has a drop-in feature to make it easy to stay connected.

Example: Mental Health management

There is tremendous interest in mental health and well-being mobile apps, with 10,000+ existing apps and significant investment from Silicon Valley.



Ethical considerations for mobile health research

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

Ethical consideration #1: Challenges with meaningful consent

- Consent documents are often long and hard to parse.
- Consent via a pop-up on a mobile app may be ignored.
- “Set it and forget it” approach means users may quickly forget that they are sharing their data with researchers.
- Consider mechanisms for reminding participants about data collection.
- Give participants easy methods to pause or fully stop data collection.
- Give participants mechanisms for reviewing and deleting part (or all) of their data.

Ethical consideration #2: Collecting/analyzing platform data

- Public social media data is typically characterized as “non-human subjects” data.
- Social media can provide rich and large datasets on a wide range of health topics.
- Research with social media users finds they have numerous concerns with their data being used for unexpected purposes.

Big Data & Society
Volume 10, Issue 1, January-June 2023

https://doi.org/10.1177/20539517231164108

SAGE journals

Original Research Article

When research is the context: Cross-platform user expectations for social media data reuse

Sarah Gilbert, Katie Shilton, and Jessica Vitak

Abstract

Social media provides unique opportunities for researchers to learn about a variety of phenomena—it is often publicly available, highly accessible, and affords more naturalistic observation. However, as research using social media data has increased, so too has public scrutiny, highlighting the need to develop ethical approaches to social media data use. Prior work in this area has explored users' perceptions of researchers' use of social media data in the context of a single platform. In this paper, we expand on that work, exploring how platforms and their affordances impact how users feel about social media data reuse. We present results from three factorial vignette surveys, each focusing on a different platform—dating apps, Instagram, and Reddit—to assess users' comfort with research data use scenarios across a variety of contexts. Although our results highlight different expectations between platforms depending on the research domain, purpose of research, and content collected, we find that the factor with the greatest impact across all platforms is consent—a finding which presents challenges for big data researchers. We conclude by offering a sociotechnical approach to ethical decision-making. This approach provides recommendations on how researchers can interpret and respond to platform norms and affordances to predict potential data use sensitivities. The approach also recommends that researchers respond to the predominant expectation of notification and consent for research participation by bolstering awareness of data collection on digital platforms.

Ethical consideration #3: Privacy & security challenges

- Mobile devices & wearables collect a lot of (sometimes sensitive) data from users.
- Informed consent may not sufficiently capture potential harms.
- Recognize that public data may not be viewed as public by user.
- Practice **data minimization** whenever possible.
- Consider practices like data donation, where people voluntarily choose to share data with researchers.

Ethical consideration #4: Building predictive models

- Mobile health data offers tremendous opportunities for predictive modeling of disease and other health outcomes.
- Researchers must be very careful in considering how their models may be biased.
- Look at data sources, data accuracy, who is excluded, bias in output, etc.

How can we help researchers address these ethical challenges?

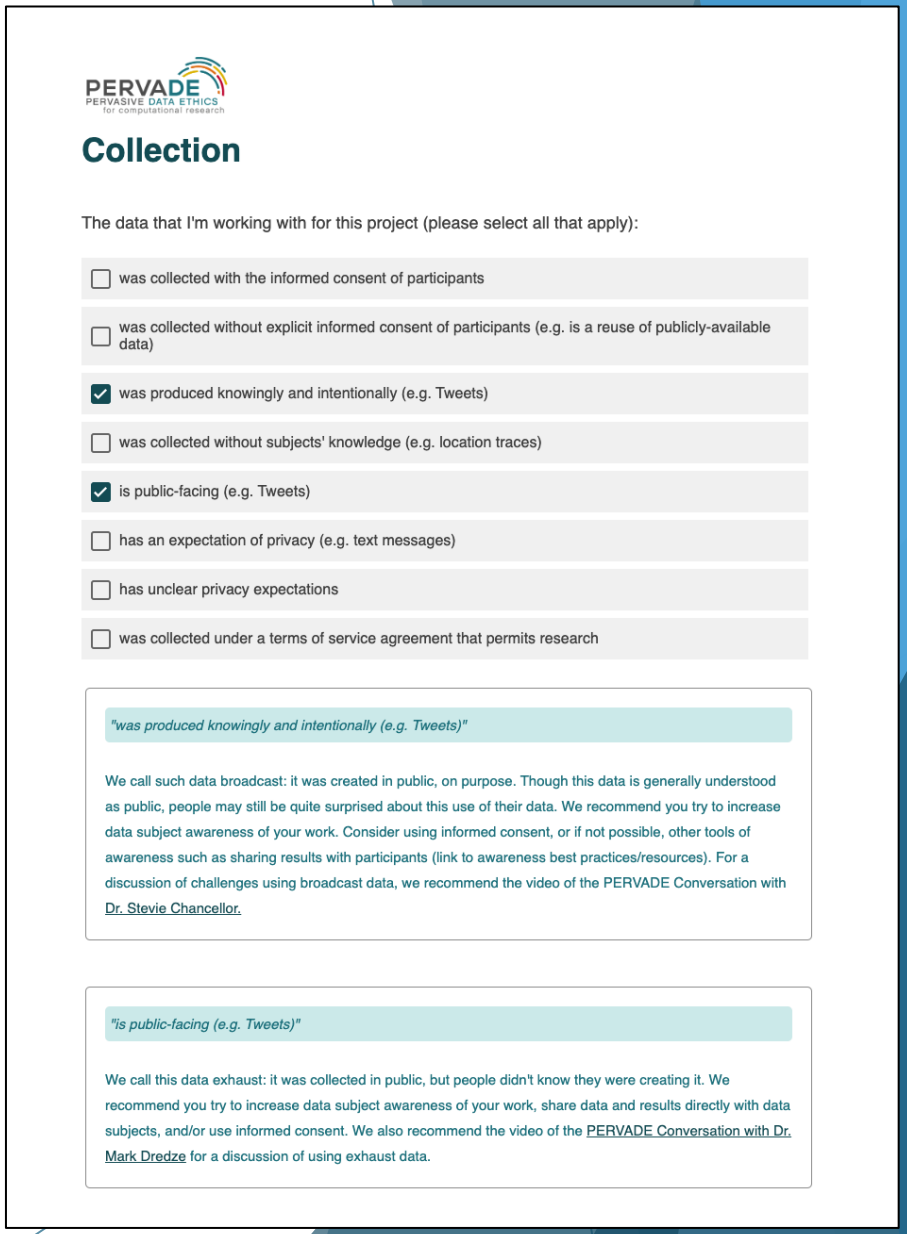
Scenario-based training* [example]

A researcher did a study examining public tweets about use of PrEP, a medicine that people who are at risk for HIV take to reduce their risk of infection. He found thousands of public tweets in which people were talking about their experiences with PrEP. In the final manuscript of the paper, he reported the types of tweets he found in a table that included illustrative examples of the tweets. While he reported the text of illustrative tweets verbatim, he did not include user's Twitter handles to protect their privacy.

*This research is in progress and being led by Dr. Sherry Pagoto at Uconn.

How can we help researchers address these ethical challenges?

- PERVADE researchers have been building a decision support tool to help data science researchers reflect on awareness and power.
- Questions capture information on research goals, data being collected, issues of power, data sources & contextual expectations, as well as data storage, processing, analysis, and sharing.
- The tool provides recommended resources to help researchers reflect on and (hopefully) resolve ethical issues identified through this process.



Thank you!

For more information, visit https://pervade.umd.edu

Jessica Vitak

jvitak@umd.edu

https://jessicavitak.com

PERVADE
PERVASIVE DATA ETHICS
for computational research

This research is supported by National Science Foundation (Awards #2419006 & #1704315).