

Current Oversight Approaches and Research involving “Big Data”

Jeffrey Kahn, PhD, MPH

Andreas C Dracopoulos Director

Johns Hopkins Berman Institute of Bioethics

jeffkahn@jhu.edu

@KahnEthx

The growth of data available for research purposes

- **Personal health data online has grown exponentially**
 - much “created” or at least added by individuals themselves
- **Evolving functionality and applications of web, mobile and social media have created a new research environment**
 - Uses of data are increasingly different than researcher-participant interactions

Health-related data

- **Information ‘actively’ supplied by individual users**
 - medical histories, genomic data, web posts
- **Personal information collected while users interact online, social media, increasingly via mobile, and passively (quantified life)**
 - Location, content, behavior data
- **Disclosures to users of the potential uses of personal data vary dramatically**

Collecting Big Data

- **What is the right data to collect?**
- **How to collect it?**
- **How much to collect?**
 - From where?
 - How to determine what is relevant?
- **What does it mean?**
 - and how to validate what we think it means?
- **BUT,**
- **What are conditions or limitations of use?**
- **What is the relevance of public health vs. other uses? and**
- **What about ethics?**

How have we come to research ethics protections?

- 1970s approaches to research *protection* being employed in 2017ff contexts
 - Regulations in substantial part driven by reaction to scandal and desire to prevent exploitation of subjects
 - Consent conceptualized as between researchers and subjects
 - Are these concerns relevant today?
 - Are they relevant for research using Big Data?
 - Web-oriented “consent” standards are de facto practice
 - » Different than research consent
 - Consumer platforms being used for research purposes
 - » Terms of service, etc. on websites, phones, smart devices
 - Regulatory or contractual standards vs. ethics
 - IRBs are applying rules crafted for a different species of research

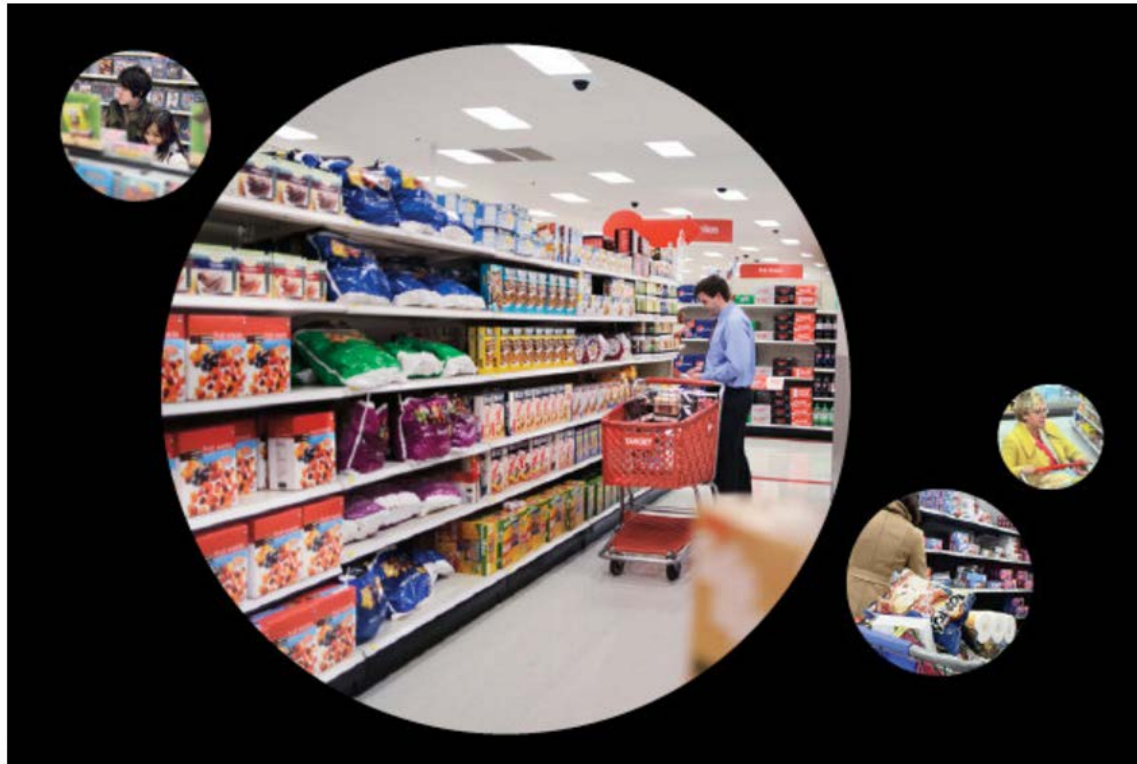
Consent in an evolving research environment

- **What do we hope to achieve in the consent process?**
 - Disclosure of information
 - Understanding
 - Of uses, by whom, for how long, possibility of secondary disclosures, etc.
 - Of risks and potential benefits
 - Voluntary participation
 - The evolving concept of control of information
- **Collection of information for research purposes as a condition of use**
 - Three concerns
 - General consent rather than consent to specific research use
 - Disclosure is boilerplate, which calls into question meaningfulness or even awareness
 - Based on consumer agreement rather than informed consent to research
- **Opt-in to research**
 - Seems closest to satisfying conventional criteria of informed consent
- **Opt-out of research**
 - Not clear how consistent these approaches are with informed consent for research
- **These are all carryovers from more consumer-oriented web environment**

Magazine

How Companies Learn Your Secrets

By CHARLES DUHIGG FEB. 16, 2012



Issues outside of the the “traditional” research environment

- Social media content as research data
 - Are terms of service enough?
 - What do we mean by the public nature of social media content?
 - For all to see may be different than for all to use
 - Among the required protections for traditional research participation is opportunity to opt out
 - How to accommodate when terms of service effectively *require* participation?
 - Legal standards may be met, but not the sprit of how we understand the ethics of consent
- What criteria are important in determining whether and under what conditions consent may be required?
 - Identified vs. anonymous?
 - Is there a threshold of metadata collection before identifiability?
- Should the purpose of research be a factor in determining the levels of protection necessary?
 - public health vs research for marketing, recruiting, or other business-related motives
 - Individual rights are trumped by public health; not so in other areas



OPINION



OPINION

Opinion: Learning as we go: Lessons from the publication of Facebook's social-computing research

Jeffrey P. Kahn^{a,1}, Effy Vayena^b, and Anna C. Mastroianni^c

^a*Berman Institute of Bioethics, The Johns Hopkins University, Baltimore, MD 21218;* ^b*Institute of Biomedical Ethics, University of Zurich, 8006 Zurich, Switzerland;* and ^c*University of Washington School of Law, Seattle, WA 98195*

and application of regulations continue to evolve as a result (13, 14). As Fiske and Hauser recently argued in PNAS, research involving human participants in social-computing environments suffers from a similar mismatch of the realities of research and the policies gov-

www.pnas.org/cgi/doi/10.1073/pnas.1416405111

PNAS | September 23, 2014 | vol. 111 | no. 38 | 13677–13679

The shortcomings of existing approaches

- **Regulatory fit**
 - What counts as research on human participants?
 - What ethics oversight applies to private sector and collaborative research?
- **Informed consent and the meaning of protection of participants**
- **Confusion over relevance and applicability of state and international jurisdictions**
- **Rules for publication**

What to do about them?

- **New thinking about consent to research in data-rich contexts**
 - At a minimum, modify disclosures
 - Committing to levels of privacy protection
 - Maximally, modifying consent to more dynamic, context specific process
- **Allowing individuals to manage use of about them (vs. from them)**
 - Privacy, control, access
- **No research stds => credibility suffers**
- **Few ethics stds => credibility may suffer more widely**
- **Time for the research community to work to create standards for ethically acceptable social media research**

Proposals for a new framework

- **Drawing on Vayena et al.**
 - **Closing old and new gaps in required oversight**
 - **Clarity**
 - **Definitions**
 - **What and who counts as research?**
 - **Standards for privacy protection**
 - **Learn from evolving best practices**
 - **Create and offer new process and technological solutions**
 - **Beyond consent and de-identification**
 - **Safe harbor for use of endorsed solutions**
 - **Calibrated oversight**
 - **Tiered access to data**
 - **Variable access based on criteria of risk-benefit**
 - **Wider stakeholder involvement in development of approaches**
 - **Researchers**
 - **IRB professionals and members**
 - **Industry**
 - **Regulators**
 - **Ethics and privacy experts**
 - **Journal editors**
 - **Research participants**