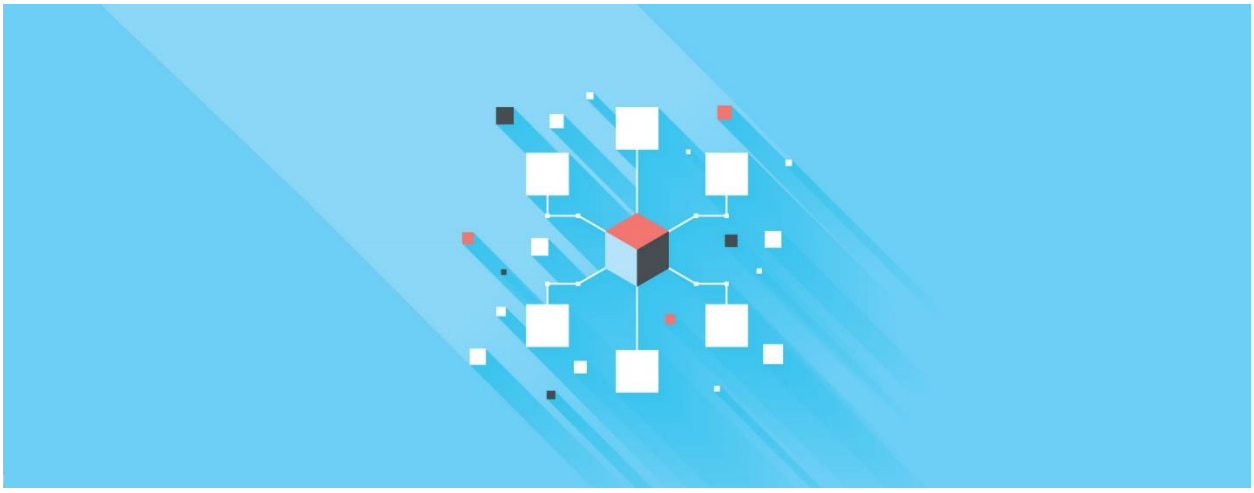
BEYOND IRBS: DESIGNING ETHICAL REVIEW PROCESSES FOR BIG DATA RESEARCH

CONFERENCE PROCEEDINGS



Thursday, December 10, 2015 • Future of Privacy Forum • Washington, DC

This material is based upon work supported by the National Science Foundation under Grant No. 1547506 and by the Alfred P. Sloan Foundation under Award No. 2015-14138.

Contents

The Future of Privacy Forum and FPF Education and Innovation Foundation gratefully acknowledge the support of the National Science Foundation and the Alfred P. Sloan Foundation to this project, with additional support provided by the Washington & Lee University School of Law.

Beyond IRBs: Designing Ethical Review Processes for Big Data Research

The ethical framework applying to human subject research in the biomedical and behavioral research fields dates back to the Belmont Report.[1] Drafted in 1976 and adopted by the United States government in1991 as the Common Rule,[2] the Belmont principles were geared towards a paradigmatic controlled scientific experiment with a limited population of human subjects interacting directly with researchers and manifesting their informed consent. These days, researchers in academic institutions as well as private sector businesses not subject to the Common Rule, conduct analysis of a wide array of data sources, from massive commercial or government databases to individual tweets or Facebook postings publicly available online, with little or no opportunity to directly engage human subjects to obtain their consent or even inform them of research activities.

Data analysis is now used in multiple contexts, such as combatting fraud in the payment card industry, reducing the time commuters spend on the road, detecting harmful drug interactions, improving marketing mechanisms, personalizing the delivery of education in K-12 schools, encouraging exercise and weight loss, and much more.[3] And companies deploy data research not only to maximize economic gain but also to test new products and services to ensure they are safe and effective.[4] These data uses promise tremendous societal benefits but at the same time create new risks to privacy, fairness, due process and other civil liberties.[5] Increasingly, corporate officers find themselves struggling to navigate unsettled social norms and make ethical choices that are more befitting of philosophers than business managers or even lawyers.[6] The ethical dilemmas arising from data analysis transcend privacy and trigger concerns about stigmatization, discrimination, human subject research, algorithmic decision making and filter bubbles.[7]

The challenge of fitting the round peg of data-focused research into the square hole of existing ethical and legal frameworks will determine whether society can reap the tremendous opportunities hidden in the data exhaust of governments and cities, health care institutions and schools, social networks and search engines, while at the same time protecting privacy, fairness, equality and the integrity of the scientific process. One commentator called this "the biggest civil rights issue of our time."[8]

These difficulties afflict the application of the Belmont Principles to even the academic research that is directly governed by the Common Rule. In many cases, the scoping definitions of the Common Rule are

---

[1] NATIONAL COMM'N FOR THE PROT. OF HUMAN SUBJECTS OF BIOMEDICAL AND BEHAVIORAL RESEARCH, BELMONT REPORT: ETHICAL PRINCIPLES AND GUIDELINES FOR THE PROTECTION OF HUMAN SUBJECTS OF RESEARCH (1979), *available at* http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html.

[2] HHS, FEDERAL POLICY FOR THE PROTECTION OF HUMAN SUBJECTS ('COMMON RULE'), http://www.hhs.gov/ohrp/humansubjects/commonrule/.

[3] EXECUTIVE OFFICE OF THE PRESIDENT, BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES, May 2014, http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf.

[4] *Cf.* Michelle N. Meyer, *Two Cheers for Corporate Experimentation: The A/B Illusion and the Virtues of Data-Driven Innovation*, 13 COLO. TECH. L. J. 274 (2015).

[5] For an analysis of big data regulatory challenges, *see* Jules Polonetsky & Omer Tene, *Privacy And Big Data: Making Ends Meet*, 66 STAN. L. REV. ONLINE 25 (2013); Omer Tene & Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 NW J. TECH & IP 239 (2013).

[6] Omer Tene & Jules Polonetsky, *A Theory of Creepy: Technology, Privacy and Shifting Social Norms*, 16 YALE J. L. & TECH. 59 (2013).

[7] Cynthia Dwork & Deirdre K. Mulligan, *It's Not Privacy, and It's Not Fair*, 66 STAN. L. REV. ONLINE 35 (2013).

[8] Alistair Croll, *Big data is our generation's civil rights issue, and we don't know it*, O'REILLY RADAR, Aug. 2, 2012, http://radar.oreilly.com/2012/08/big-data-is-our-generations-civil-rights-issue-and-we-dont-know-it.html.

strained by new data-focused research paradigms, which are often product-oriented and based on the analysis of preexisting datasets. For starters, it is not clear whether research of large datasets collected from public or semi-public sources even constitutes human subject research. "Human subject" is defined in the Common Rule as "a living individual about whom an investigator (whether professional or student) conducting research obtains (1) data through intervention or interaction with the individual, or (2) identifiable private information."[9] Yet, data driven research often leaves little or no footprint on individual subjects ("intervention or interaction"), such as in the case of automated testing for security flaws.[10] Moreover, the existence—or inexistence—of identifiable private information in a dataset has become a source of great contention, with de-identification "hawks" lamenting the demise of effective anonymization[11] even as de-identification "doves" herald it as effective risk mitigation.[12]

Not only the definitional contours of the Common Rule but also the Belmont principles themselves require reexamination. The first principle, *respect for persons*, is focused on individual autonomy and its derivative application, informed consent. While obtaining individuals' informed consent may be feasible in a controlled research setting involving a well-defined group of individuals, such as a clinical trial, it is untenable for researchers experimenting on a database that contains the footprints of millions, or indeed billions, of data subjects. danah boyd and Kate Crawford write, "It may be unreasonable to ask researchers to obtain consent from every person who posts a tweet, but it is problematic for researchers to justify their actions as ethical simply because the data are accessible."[13]

The second principle, *beneficence*, requires a delicate balance of risks and benefits to not only respect individuals' decisions and protect them from harm but also to secure their well-being. Difficult to deploy even in traditional research settings, such cost-benefit analysis is daunting in a data research environment where benefits could be probabilistic and incremental and the definition of harm subject to constant wrangling between minimalists who reduce privacy to pecuniary terms and maximalists who view any collection of data as a dignitary infringement.[14] In a recent White Paper titled *Benefit-Risk Analysis for Big Data Projects*, we offered decision-makers a framework for reasoned analysis balancing big data benefits against privacy risks.[15] We explained there that while some of the assessments proposed in that framework can be standardized and quantified, others require value judgments and input from experts other than privacy professionals or data regulators. For example, assessing the scientific likelihood of capturing a

---

[9] 45 CFR 46.102(f).

[10] *See, e.g.*, Arvind Narayanan & Bendert Zevenbergen, *No Encore for Encore? Ethical Questions for Web-Based Censorship Measurement*, WASH. & LEE L. REV. ONLINE (forthcoming 2016).

[11] *See, e.g.*, Arvind Narayanan & Ed Felten, *No Silver Bullet: De-Identification Still Doesn't Work*, July 9, 2014, http://randomwalker.info/publications/no-silver-bullet-de-identification.pdf; Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization* , 57 UCLA L. REV. 1701 (2010).

[12] See, e.g., Jules Polonetsky, Omer Tene & Kelsey Finch, *Shades of Gray: Seeing the Full Spectrum of Practical Data De-Identification*, SANTA CLARA L. REV. (forthcoming, 2016); Daniel Barth-Jones, *The Antidote for "Anecdata": A Little Science Can Separate Data Privacy Facts from Folklore*, Nov. 21, 2014, https://blogs.law.harvard.edu/infolaw/2014/11/21/the-antidote-for-anecdata-a-little-science-can-separate-data-privacy-facts-from-folklore/; Kathleen Benitez & Bradley K. Malin, *Evaluating Re-Identification Risks With Respect to the HIPAA Privacy Rule*, 17 J. AMER. MED INFORMATICS ASSOC. 169 (2010); Khaled El Emam et al, *A Systematic Review of Re-Identification Attacks on Health Data*, 6 *PLoS One* 1, December 2011.

[13] danah boyd & Kate Crawford, *Critical Questions for Big Data*, 15(5) INFO. COMM. & SOC. 662 (2012).

[14] Case C-362/14, Maximillian Schrems v. Data Protection Commissioner, 6 October 2015, http://curia.europa.eu/juris/document/document.jsf?docid=169195&doclang=EN; *also see* Ryan Calo, *The Boundaries of Privacy Harm*, 86 IND. L.J. 1131 (2011).

[15] Jules Polonetsky, Omer Tene & Joseph Jerome, *Benefit-Risk Analysis for Big Data Projects* (FUTURE OF PRIVACY WHITE PAPER, September 2014), http://www.futureofprivacy.org/wp-content/uploads/FPF_DataBenefitAnalysis_FINAL.pdf.

benefit in a specialized area, such as reduction of greenhouse emissions, cannot be made solely based on privacy expertise.

In response to these developments, the Department of Homeland Security commissioned a series of workshops in 2011-2012, leading to the publication of the *Menlo Report on Ethical Principles Guiding Information and Communication Technology Research*.[16] That report remains anchored in the Belmont Principles, which it interprets to adapt them to the domain of computer science and network engineering, in addition to introducing a fourth principle, *respect for law and public interest*, to reflect the "expansive and evolving yet often varied and discordant, legal controls relevant for communication privacy and information assurance."[17] In addition, on September 8, 2015, the U.S. Department of Health and Human Services and 15 other federal agencies sought public comments to proposed revisions to the Common Rule.[18] The revisions, which address various changes in the ecosystem, include simplification of informed consent notices and exclusion of online surveys and research of publicly available information as long as individual human subjects cannot be identified or harmed.[19]

For federally funded human subject research, the responsibility for evaluating whether a research project comports with the ethical framework lies with Institutional Review Boards (IRBs). Yet, one of the defining features of the data economy is that research is increasingly taking place outside of universities and traditional academic settings. With information becoming the raw material for production of products and services, more organizations are exposed to and closely examining vast amounts of often personal data about citizens, consumers, patients and employees. This includes not only companies in industries ranging from technology and education to financial services and healthcare, but also non-profit entities, which seek to advance societal causes, and even political campaigns.[20]

Whether the proposed revisions to the Common Rule address some of the new concerns or exacerbate them is hotly debated. But whatever the final scope of the rule, it seems clear that while raising challenging ethical questions, a broad swath of academic research will remain neither covered by the rules nor subject to IRB review. Currently, gatekeepers for ethical decisions range from private IRBs to journal publication standards, association guidelines and peer review.[21] A key question for further debate is whether there is a need for new principles as well as new structures for review of academic research that is not covered by the current or expanded version of the Common Rule.

In *Beyond the Common Rule: Ethical Structures for Data Research in Non-Academic Settings*, we noted that even research initiatives that are not governed by the existing ethical framework should be subject to clear principles and guidelines. Whether or not a research project is federally funded seems an arbitrary trigger for ethical review. To be sure, privacy and data protection laws provide an underlying framework governing commercial uses of data with boundaries like consent and avoidance of harms. But in many cases

---

[16] DAVID DITTRICH & ERIN KENNEALLY, THE MENLO REPORT: ETHICAL PRINCIPLES GUIDING INFORMATION AND COMMUNICATION TECHNOLOGY RESEARCH, U.S. Dept. of Homeland Sec., (Aug. 2012), *available at* https://www.predict.org/%5CPortals%5C0%5CDocuments%5CMenlo-Report.pdf.

[17] *Ibid*, at 5.

[18] HHS, NPRM for Revisions to the Common Rule, Sept. 8 , 2015, http://www.hhs.gov/ohrp/humansubjects/regulations/nprmhome.html.

[19] *Also see* Association of Internet Researchers, Ethical Decision-Making and Internet Research Recommendations from the AoIR Ethics Working Committee (Version 2.0), 2012, http://aoir.org/reports/ethics2.pdf (original version from 2002: http://aoir.org/reports/ethics.pdf).

[20] Ira S. Rubinstein, *Voter Privacy in the Age of Big Data*, 2014 WISC. L. REV. 861. .

[21] Katie Shilton, *Emerging Ethics Norms in Social Media Research*, WASH. & LEE L. REV. ONLINE (forthcoming 2016).

where informed consent is not feasible and where data uses create both benefits and risks, legal boundaries are more ambiguous and rest on vague concepts such as "unfairness" in the United States[22] or the "legitimate interests of the controller" in the European Union.[23] This uncertain regulatory terrain could jeopardize the value of important research that could be perceived as ethically tainted or become hidden from the public domain to prevent scrutiny.[24] Concerns over data ethics could diminish collaboration between researchers and private sector entities, restrict funding opportunities, and lock research projects in corporate coffers contributing to the development of new products without furthering generalizable knowledge.[25]

In a piece he wrote for a *Stanford Law Review Online* symposium we organized two years ago,[26] Ryan Calo foresaw the establishment of "Consumer Subject Review Boards" to address ethical questions about corporate data research.[27] Calo suggested that organizations should "take a page from biomedical and behavioral science" and create small committees with diverse expertise that could operate according to predetermined principles for ethical use of data. The idea resonated in the White House legislative initiative, the Consumer Privacy Bill of Rights Act of 2015, which requires the establishment of a Privacy Review Board to vet non-contextual data uses.[28] In Europe, the European Data Protection Supervisor has recently announced the creation of an Advisory Group to explore the relationships between human rights, technology, markets and business models from an ethical perspective, with particular attention to the implications for the rights to privacy and data protection in the digital environment.[29]

Alas, special challenges hinder the adaptation of existing ethical frameworks, which are strained even in their traditional scope of federally funded academic research, to the fast-paced world of corporate research. For example, the categorical non-appealable decision making of an academic IRB, which is staffed by tenured professors to ensure independence, will be difficult to reproduce in a corporate setting. And corporations face legitimate concerns about sharing trade secrets and intellectual property with external stakeholders who may serve on IRBs.

To address these important issues and set the stage for the introduction of IRB-like structures into corporate and non-profit entities, the Future of Privacy Forum (FPF) convened an interdisciplinary workshop in December 2015 titled "*Beyond IRBs: Designing Ethical Review Processes for Big Data*." The workshop aimed to identify processes and commonly accepted ethical principles for data research in academia, government and industry. It brought together researchers, including lawyers, computer scientists, ethicists

---

[22] FTC Policy Statement on Unfairness, Appended to International Harvester Co., 104 F.T.C. 949, 1070 (1984). *See* 15 U.S.C. § 45(n).

[23] Article 29 Working Party, WP 217, Op. 06/2014 on the Notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC, Apr. 9, 2014, http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf.

[24] The Common Rule's definition of "research" is "a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to *generalizable knowledge*." (Emphasis added).

[25] Jules Polonetsky, Omer Tene, & Joseph Jerome, *Beyond the Common Rule: Ethical Structures for Data Research in Non-Academic Settings*, 13 COLO. TECH. L. J. 333 (2015).

[26] Stan. L. Rev. Online Symposium Issue, *Privacy and Big Data: Making Ends Meet*, September, 2013, http://www.stanfordlawreview.org/online/privacy-and-big-data; *also see* stage setting piece, Jules Polonetsky & Omer Tene, *Privacy and Big Data: Making Ends Meet*, September 3, 2013 66 STAN. L. REV. ONLINE 25.

[27] Ryan Calo, *Consumer Subject Review Boards: A Thought Experiment*, 66 STAN. L. REV. ONLINE 97 (2013), http://www.stanfordlawreview.org/online/privacy-and-big-data/consumer-subject-review-boards.

[28] CONSUMER PRIVACY BILL OF RIGHTS §103(c) (Administration Discussion Draft 2015), *available at* https://www.whitehouse.gov/sites/default/files/omb/legislative/letters/cpbr-act-of-2015-discussion-draft.pdf.

[29] European Data Protection Supervisor, Ethics Advisory Group, Dec. 3, 2015, https://secure.edps.europa.eu/EDPSWEB/edps/site/mySite/Ethics.

and philosophers, as well as policymakers from government, industry and civil society, to discuss a blueprint for infusing ethical considerations into organizational processes in a data rich environment.

Workshop Theme: Defining the Problem

To set the stage for the workshop's diverse presentations and discussions, our early discussions were animated by a common theme: asking in what ways existing IRB structures may be inadequate or inappropriate for big data research activities.

This stage of the discussion highlighted two key research questions about ethics for innovative data use and lessons from institutional review boards. First, what are the biggest ethical challenges posed by potential new uses of data, and how can collective or societal benefits be weighed against privacy or other harms to an individual? Participants strove to identify when and which important ethical values are in tension with innovation and data research; how organizations can measure or evaluate abstract privacy and autonomy risks, particularly against the potential benefits of data use; how can and should context inform consumer expectations around innovative data uses and Big Data generally; and how subjective concepts like 'creepiness' can be used to inform ethical conversations around data.

The second research question asked: how can researchers work to promote trust and public confidence in research, as IRBs have in human subject testing, while being scalable to industry and avoiding the bureaucratic and administrative criticisms that have been directed at existing IRB structures? Participants sought to identify the primary benefits and drawbacks to existing IRB practices; how IRBs ensure both a variety of different viewpoints and the necessary expertise to evaluate a project; and how an IRB-like process might scale to meet business demands in markets. Participants also considered whether the ethical principles espoused by the Belmont Report and the Menlo Report are suitably represented by IRB practices, and which of their principles are most applicable to a more consumer-oriented review process.

By engaging the workshop's multidisciplinary participants in this critical scoping exercise, we were able to illuminate fundamental points of agreement across sectors and, equally important, points of disagreement about how to define the problem.

Keynote. The workshop began examining the role and structure of ethical reviews for big data research with a keynote presentation by **Ryan Calo, Assistant Professor, University of Washington School of Law**. Professor Calo highlighted the need for ethical review processes throughout the information economy, including both traditional academic institutions and corporate environments. Where information asymmetries and the potential for data-based manipulation or harms make it difficult for individuals to protect themselves, organizations need to confront ethical questions about what *should* or *should not* be done with data, even if those activities are not clearly illegal.

Professor Calo next highlighted similar struggles by both Facebook's internal ethical review and traditional academic IRBs to deal with aspects of big data research, including information asymmetries between researchers, research subjects, and IRB members; the impracticality of informed consent in many circumstances; and implementation of the Belmont principles in novel contexts. In order to help move both private sector review processes and IRB processes forward, he underlined the importance of formalized privacy/ethical review processes, transparent review criteria, and greater attention to the role of precedents by similarly situated review boards.

Finally, Professor Calo raised several fundamental questions about what values and functions we want private sector ethical reviews to embody: Should they be voluntary? Should they be substantive, or pro forma? Compliance-oriented or ethics-based? Is it necessary to have an outsider? How do we distinguish IRBs and ethical reviews from self-regulation? In discussion, workshop participants agreed with the identification interdisciplinary or cross-functional membership, transparent review criteria, and documented decisions and outcomes as key aspects of a corporate IRB.

Firestarters. Next, the workshop continued its efforts to define the problem through a series of "firestarter" presentations by leading experts. Their original research examined various aspects of existing ethical review processes and identified several core themes, including the value of transparency, widespread concerns about the utility of consent, the need for harmonization, and the need for practical solutions. These presentations and a group question-and-answer session afterwards were facilitated by two rapporteurs, **Joshua Fairfield**, **Professor of Law, Washington & Lee University School of Law,** and **Margaret Hu**, **Professor of Law, Washington & Lee University School of Law**.

First, **Katie Shilton, Assistant Professor, College of Information Studies, University of Maryland**, shared her research on *Emerging Ethics Norms in Social Computing Research*. Professor Shilton and her colleagues conducted a survey of 263 online data researchers and documented the beliefs and practices around which social computing researchers are converging, as well as areas of ongoing disagreement. Areas of growing consensus included: removing data subjects upon request; researchers remaining in conversation with both their colleagues and IRBs; sharing results with participants; and being cautious about outliers in the data. On the other hand, areas of disagreement among scholars included: whether it is permissible to ignore the terms of service of various platforms; deceiving participants; sharing raw data with key stakeholders; and whether consent is possible in large scale studies.

While many of the survey respondents indicated that they thought deeply about ethics when designing their own research processes, it also became clear that what each researcher considered ethical could differ across a wide set of principles. Nevertheless, Professor Shilton also identified several agreed-upon practices that researchers tended to agree should be utilized in this space, including: holding researchers to a higher ethical standard; notifying participants; sharing results with subjects; asking colleagues about their practices; asking IRBs for guidance; and removing individuals from studies upon request.

Micah Altman, Director of Research Program on Information Science, MIT and **Alexandra Wood, Berkman Center Fellow, Harvard University**, presented their work *Towards a New Ethical and Regulatory Framework for Big Data Research*. Along with their colleagues, Professors Altman and Wood sought to identify key gaps in the current regulatory framework for big data research, including: limits in the scope of coverage of the common rule; the inadequacy of informed consent requirements; reliance on narrow range of interventions such as notice consent and de-identification techniques; emphasis on the study design and collection stages of the information lifecycle; and limited oversight at other stages of the information lifecycle (such as storage, primary use, secondary use).

Accordingly, their recommendations for a new ethical framework are clustered around five main areas: 1) universal coverage for all human subject research, 2) conceptual clarity through revised definitions and guidance, 3) systematic risk-benefit assessment at each stage of the information stage cycle, 4) new procedural and technological solutions, and 5) tailored oversight with procedures calibrated to risk.

Arvind Narayanan, Assistant Professor, Department of Computer Science, Princeton University, presented his work on *Ethical Questions for Web-based Censorship Measurement.* Professor Narayanan underscored the divergent approaches to ethical reviews even within academia with a case study about the

Encore program and by reminding participants that many computer scientists conducting big data research rely on post hoc ethical reviews by program committees for computer science conferences, rather than traditional IRB research approvals. He pointed out the mismatch between computer science norms, including research disseminated via conference proceedings rather than journals and quickly evolving research methodologies, and those of IRBs, who often reject research outside their traditional definitions and may therefore exclude cutting edge computer science.

Some of the concerns raised by this sort of review include: lack of transparency; the fact that program committee reviewers may lack ethical training; and that by the time of review, the anticipated harm may have already occurred. On the other hand, advantages of these peer-driven review processes are the ability to adapt quickly to evolving norms and methodologies.

Neil Richards, Professor of Law, Washington University School of Law, shared his and Woodrow Hartzog's research on *Trusting Big Data Research*. Professor Richards began by identifying problems with the limits of consent and the procedural approaches and compliance mentality taken outside of the university context (e.g., FIPPs). If we can have the right set of rules to promote trust, he suggests, people will be more willing to expose their data in ways that can be beneficial for users, companies, and institutions more broadly.

Professor Richards identified four foundations for trust: protection, discretion, honesty, and loyalty. (1) Protection is the need to keep personal information securely against third parties outside the relationship. For this, Professor Richards discussed the need for industry standards and commitments by companies to go beyond just setting up a few technical safeguards by embracing comprehensive security programs. (2) In defining discretion, the intent is to capture a broader concept than pure confidentiality, as it recognizes that trust can be preserved even when the trustee shares information in a limited ways. In particular, discretion needs to be respected when considering secondary uses and disclosures. As to (3) honesty, this principle emphasizes that those entrusted with personal data must be honest and open with those who disclose personal information to them. These duties of candor and disclosure go beyond simple notice and choice or transparency efforts, requiring actual (not constructive) notice as an affirmative substantive principle sounding in fiduciary obligations. Finally, to earn trust by demonstrating (4) loyalty, companies must not exploit users' personal information for their own short-term and short-sighted gain at users' expense. Using data to manipulate unwitting users is also frequently disloyal; in drawing regulatory lines, regulators might look to consumer ability to understand the transaction, company representations, the nature of consumer vulnerabilities, and industry best practices for trust-promoting behavior.

Small and Large Group Discussions. After the firestarter session, workshop participants broke into four small group sections in order to tackle the issues more directly. These dialogues were facilitated by discussion leaders (1) **Jules Polonetsky, Executive Director, Future of Privacy Forum** and **Heng Xu**, **Program Director, National Science Foundation**; (2) **Mary Culnan, Professor Emeritus, Bentley University** and **Jeffrey Mantz, Program Director, National Science Foundation**; (3) **Omer Tene**, **Vice President, Research and Education, IAPP**; and (4) **Danielle Citron**, **Professor of Law, University of Maryland School of Law** and **Brendon Lynch, Chief Privacy Officer, Microsoft**.

Each group was tasked with answering the following questions:

1. What are the substantive problems with the current ethical review mechanisms (consent, applying principles, etc.)?
2. What are the structural problems with the current ethical review frameworks (scope, structure, consistency, etc.)?

3. Is the issue one of Research only or corporate analytics and product development?
4. Are the issues limited to informational privacy concerns? If not, how do we limit the scope?

After the breakout opportunity, all workshop participants reconvened to report back on their small group discussions and to raise any outstanding questions or issues related to the morning's goal of "defining the problem." Omer Tene, Vice President of Research and Education at the International Association of Privacy Professionals, led the discussion.

What are the substantive problems with the current ethical review mechanisms (consent, applying principles, etc.)?

One of the most common substantive issues identified by the small groups was uncertainty around the *scope* of reviews. One group focused on the definitional problem of what is "human" in terms of human-subject review, particularly when your research does not interact with any humans directly but explores their data in a very complex way. The absence of core criteria for evaluating research was another concern, which the group believed made the IRB process unsuitable for today's big data work. Another group of participants also wanted to specifically recognize that the general IRB approach has a problem of scalability that makes it unsuitable for day-to-day use.

What are the structural problems with the current ethical review frameworks (scope, structure, consistency, etc.)?

One group specifically identified issues with ethical review boards' ability to appreciate the differing risks to reusing data, which can vary widely depending on the context of the reuse. This has led to confusion between research that is exempt from review and data that is completely excluded from review under the Common Rule framework.

Another group wondered about overlaps and gaps among ethical frameworks from different business sectors, in addition to the social media platforms that have garnered the majority of research attention. For example, what ethical review process is used within the credit card industry to offer a particular interest rate to a potential client? How does Wikipedia and its global community handle ethical review? How can we make sure that these frameworks do not develop in inaccessible silos?

Another group also wanted to draw attentions to inconsistency among IRBs leading to "IRB laundering" among academics, wherein researchers will forum shop from IRB to IRB until they are given approval. Similarly, they noted that not all funding organizations conduct equally rigorous due diligence to ensure they are not funding unethical research.

Is the issue one of Research only or corporate analytics/product development?

Several groups agreed that there is a logical division between data-driven research in and for corporate environments and activities intended to create generalizable knowledge for the public and academic communities, but had difficulty identifying where to draw the line. Groups noted that both corporate and non-corporate spheres needed some form of ethical review, but that the structures need to be different to reflect the character of the research and of the researcher.

Another group recognized that as the line between academic and corporate research is increasingly blurry, triggering the threshold question of what conduct requires a review will become ever more important.

On the other hand, another commenter suggested that these should both be treated the same under a common framework for an ethics of information. This emphasizes that ethics is a substantive principle for which you can then develop frameworks for particular applications.

Are the issues limited to informational privacy concerns? If not, how do we limit the scope?

One small group reached a consensus that the issues animating this workshop go beyond just privacy and that ethical questions outside of informational privacy matter equally. At the same time, ethical review boards that intend to handle non-privacy related issues (such as safety) should be staffed by reviewers equipped to handle them.

Another group wanted to focus beyond privacy, given that privacy is "notoriously variable," and instead to focus on the use of data that may have harmful consequences for people. They would include within the scope of ethical review adverse and disparate impacts which may be the consequence of day-to-day business decisions within organizations using vast data sets.

From Problems to Solutions

Finally, as part of the morning's concluding discussion session participants began to pivot towards the afternoon's discussion topic: identifying solutions to the ethics and privacy issues they had surfaced during the first part of the workshop.

On the procedural front, several participants suggested building ethics awareness and literacy in various contexts, including introducing ethics awareness and oversight for general business activities into business school curricula or educating multidisciplinary groups within an institution or company on ethical decision-making. Other participants emphasized the need to balance process and substance, calling for "a more flexible process, but not a lighter ethics" while avoiding a "compliance mentality." Another participant identified the possibility of certifying certain methodologies as ethical, rather than relying on potentially inconsistent, case-by-case approvals for individual research projects.

Participants were also eager to jump into conversation about solutions to substantive ethics and privacy issues raised during previous sessions. For example, some participants wanted the group to consider looking to European models for balancing fundamental rights in the face of new technologies and business models, such as the legal analysis around "compatible use" tests under the Data Protection Directive and the General Data Protection Regulation or EU requirements for privacy impact assessments. Others suggested, however, that IRBs were designed with distinct enough goals—(i.e., IRBs are intended to protect research subjects, but not to evaluate externalities outside of that framework)—that the two models should be considered separately.

Finally, several contributors wanted to underscore the societal value of research. One additional commenter urged workshop participants to consider, throughout the day, not only the privacy or other individual harms that may arise during the research process, but also the societal harms that may arise from *not* conducting that research at all, whether in an academic institution or a corporate setting.

Workshop Theme: Paths to a Solution

Having explored and identified fissures in the existing IRBs model for ethical reviews throughout the morning, in the latter half of the workshop participants dedicated themselves to identifying and describing paths to a solution.

This next stage of the discussion highlighted additional research questions about existing approaches to privacy and ethical questions in industry and building ethical review mechanisms. First, the workshop explored what role privacy officers and other institutional review processes should play in setting expectations for ethical practices and considerations in firms. Participants discussed how possible solutions to protecting individual interests differ across various organizations (both industry and non-profits) and society at large; what role privacy and ethical committees play in establishing an organizational culture of respect for privacy; where are the primary stress points with existing ethical and social norms, and what questionable data practices could be adequately remedied by a review process; and how chief privacy officers and other compliance officials can promote ethical uses of data within organizations, and how can they encourage an ethical and privacy-protective culture within a firm.

Next, the workshop explored what processes and procedures an organization should follow to develop an ethical review process that can adequately address public and regulatory concern around data use. Participants considered how organizations can document a process for evaluating project benefits that is commensurate with traditional privacy impact assessments and how such processes can play an ongoing monitoring role, modifying their decisions as ethical boundaries become clearer and proposing appropriate accommodation or mitigation as circumstances change. Similarly, participants addressed the appropriate balance between the secrecy required to facilitate information sharing and open discussion and the transparency needed to enhance trust and promote accountability, and how principles from IRBs and other review mechanisms can be merged with existing privacy reviews.

These sessions retained the inclusive, open conversational structure of the previous discussions, where keynote and firestarter presentations laid a substantive foundation for the in-depth breakout and group discussions that were the crux of this workshop.

Keynote. To transition between the morning's examination of issues in the IRB and Big Data ethics spaces and the afternoon's search for solutions, participants heard from **Erin Kenneally, Portfolio Manager, Cyber Security Division, Science & Technology Directorate, U.S. Department of Homeland Security and lead author of the seminal *Menlo Report*.**[30] To transition between the morning's examination of issues in the IRB and Big Data ethics spaces and the afternoon's search for solutions, participants heard from Erin Kenneally, Portfolio Manager, Cyber Security Division, Science & Technology Directorate, U.S. Department of Homeland Security and lead author of the seminal Menlo Report. Ms. Kenneally led participants into a deep dive of types of data research using Information and Communication Technologies, the tremendous amounts of data they are able to garner, the privacy and ethical issues arising from these quickly evolving technologies, and how new frameworks like the Menlo Report have sought to adapt existing ethical approaches to modern technologies.

Ms. Kenneally started by laying out the differences between research and industry. Increasingly, both academic and commercial entities have the ability to collect and use new and existing online data without directly interacting with the data subject. While this poses potential for innovation, it also presents some risks given that, contrary to researchers whose main consideration is to benefit the public and who work within an Institutional Review Board structure, industry researchers are driven by profit and competition sans overarching oversight. Ms. Kenneally further highlighted the ethical issues arising from the collection of information from online "public" spaces. Finally, she shared her thoughts on potential solutions such as applying principles from sources like the Menlo Report and certain privacy and data protection laws,

[30] David Dittrich & Erin Kenneally, The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research (2012); David Dittrich & Erin Kenneally, Applying Ethical Principles to Information and Communication Technology Research: A Companion to the Menlo Report (2013).

orienting efforts on education and awareness, advancing innovative decision making tools, and implementing more effective oversight mechanisms.

Firestarters. The second half of the workshop aimed to identify and evaluate potential avenues for ethical review of big data research in a variety of contexts. Using the same approach as in the morning, the afternoon started with a series of "firestarter" presentations by leading experts. Each of them provided input on existing ethical review processes they had implemented or researched in various sectors, including health care, social networks, health and fitness, and environmental contexts. Their presentations and a group question-and-answer session afterwards were facilitated by **Kirsten Martin, Assistant Professor, Strategic Management & Public Policy, George Washington University School of Business**.

First, **Stan Crosley, Co-Director, Center for Law, Ethics, and Applied Research in Health Information, Indiana University School of Medicine**, shared his research on how to engender a new proxy for trust in the modern heathcare ecosystem. Professor Crosley started by addressing the evolution of the traditional healthcare ecosystem, noting that, a few decades ago, individuals trusted their physician, the FDA, and that prescription drugs were safe. The search for a relevant trust proxy, not only in healthcare, but also for industry at large, has now become a defining concern. He noted further that this is a two-sided issue, and that researchers need to consider both doing the right thing with data and using data to do the right thing. On this basis, he studied how to transpose the work that has been done in human research into corporate settings. He highlighted the need to have a broader sense of how ethics is set out in the context of big data research conducted by companies. He suggested that a core feature of ethical research was transparency, which involves consent and control by users/research subjects. This is at odds, he noted, with the fact that companies frequently collect more data than necessary and explains why traditional IRBs do not appear to be a viable solution in corporate contexts.

In an attempt to remediate this, Professor Crosley and his team at Indiana University have been working on creating an independent data review process, applicable to all industries, which would assess *data* risk for researchers (rather than just privacy risk). When one commenter asked about the composition of an IRB-like board for companies, Professor Crosley suggested participants consider dynamic IRB-like boards comprised of subject matter experts who are able to bring in their deep expertise, as well as consumers in order to align research practices as closely with consumers' reasonable expectations as possible.

Molly Jackman, **Public Policy Research Manager, Facebook**, continued next by describing the new ethical review process, inspired by the Belmont report, that Facebook has built to vet its research activities. She also emphasized the critical importance of ensuring any research conducted on the company's data is coherent and in conformity with the company's values.

Facebook's review process is divided into two independent branches: a research review and a privacy review. First, when joining the company, every employee receives privacy training, the nature and complexity of which varies depending on the employee's role. This specifically includes having any employee directly involved in research attending a thorough data and research "bootcamp." Employees involved as reviewers will also complete the National Institute of Health's human subjects training. Secondly, each research proposal is subject to a specific review process before it can begin. A board comprised of five Facebook employees hears the proposal and decides whether to authorize it or not. The review board considers several criteria, such as the sensitivity of the data or the community concerned by the research. Particularly sensitive research triggers a higher level of scrutiny. In order to grasp the subtleties and implications of the research, they involve experts in law, policy, communication and substantive research area experts both internally and externally. While typically the board deliberates and is able to come to an informed decision, additional veto points also exist throughout the process.

Ms. Jackman further highlighted the importance Facebook attaches to not having a "checkbox" system, but rather conducting a thorough review for each unique case brought before it. This flexible process is intended to allow the company to make faster and better decisions.

One commenter welcomed Facebook's attention to inclusiveness and asked if any particular criteria triggered the separate privacy review process. Ms. Jackman explained that any research proposal involving individual data would be directed to the privacy review process. In response to another comment, Ms. Jackman expressed Facebook's intention to increase transparency about their process, particularly towards their users.

Michelle de Mooy, Deputy Director, Consumer Privacy Project, Center for Democracy & Technology, presented the results of a partnership her organization, the Center for Democracy and Technology, had established with wearable company Fitbit. The project, called *Towards Privacy-Aware Research in Wearable Health*, aimed to produce guidance on how to do privacy-protective internal research at wearable companies that benefits customers and society at large. Fitbit, as a small company in the health and wellness market, recognized the need to maintain its users trust towards its products and in the brand itself. The company focused heavily on research and product development—its research team grew from five people at the beginning of this project to thirty by the time it came to an end.

Ms. De Mooy first described the project methodology, which started by investigating not only how researchers at Fitbit handle data but also about how some individuals can impact decisions. One of the major findings of her study was that most of Fibit's research subjects are in fact Fitbit employees. The research process at Fitbit was divided into "projects" and "hacks." Hacks consisted of researchers following their own curiosity and deciding whether or not a project is viable. Projects, on the other hand, were more formal and further divided into three categories based on what data was to be used: Fitbit employees' data, Fitbit user studies (which include all offices), or all Fitbit user data.

One of the main recommendations arising from the study was for Fitbit to treat all employee data as sensitive so that such data would receive a higher level of protection. Additional recommendations dealt with inculcating data stewardship as a crucial element of corporate ethical review processes and of a culture of privacy within a company.

When asked about the long-term use of data, Ms. DeMooy explained that CDT recommended having a set period where data would be automatically deleted every three months. Deleting historical datasets is crucial to preserve privacy and experience has shown that such datasets are not as useful as is often believed. Today's technology enables automatic deletion processes that companies should take advantage of as part of their efforts to implement good practices.

To close the afternoon firestarters, **Dennis Hirsh**, **Professor of Law, Capital University Law School** and **Jonathan King**, **Head of Cloud Strategy, Ericsson**, gave a presentation entitled "*Big Data Sustainability: An Environmental Management Systems Analogy*." Professor Hirsch and Mr. King explained how harm could be seen as a systems defect. On this basis, they wondered how defects could be reduced from the data analytics production process.

Professor Hirsch and Mr. King argued that leaders can learn from environmental management practices developed to manage the negative externalities of the industrial revolution. They observed that along with its many benefits, big data can create negative externalities that are structurally similar to environmental pollution which suggests that management strategies to enhance environmental performance could provide a useful model for businesses seeking to sustainably develop their personal data assets. They briefly chronicled environmental management's historical progression from a back-end, siloed approach to a more

collaborative and pro-active "environmental management system" approach. They then argued that an approach modeled after environmental management systems – a Big Data Management System approach – offers a more effective model for managing data analytics operations to prevent negative externalities. Finally, they discussed how a Big Data Management System approach aligns with: A) Agile software development and Dev Ops practices that companies use to develop and maintain big data applications, B) best practices in Privacy by Design and engineering and C) emerging trends in organizational management theory.

Small and Large Group Discussions. Following the afternoon firestarter session, workshop participants broke into four small group sections in order to tackle the issues raised by them more directly. These dialogues were again facilitated by discussion leaders **Jules Polonetsky, Executive Director, Future of Privacy Forum**; **Mary Culnan**, **Professor Emeritus, Bentley University**; **Omer Tene**, **Vice President, Research and Education, IAPP**; and **Danielle Citron**, **Professor of Law, University of Maryland School of Law**.

Each group was tasked with answering the following questions:

1. What would a new non-IRB structure look like? Consider consistency, confidentiality, expertise, diversity, composition, etc.
2. What are the elements of a new ethical framework? Are we updating Common Rule/Menlo Report guidance, doing an expanded version of FIPPS, or something else?
3. What is the feasibility of a formal structure in regards to corporate data uses (analytics, product development, new technology, etc.)?
4. How do ethics intersect with legal frameworks? Consider issues of legality, fairness, and benefit-risk analysis.

After the breakout opportunity, all workshop participants reconvened to report back on their small group discussions and to raise any outstanding questions or issues related to the afternoon's goal of "identifying and describing paths toward a solution." **Danielle Citron**, **Professor of Law, University of Maryland School of Law**, led the discussion.

What would a new non-IRB structure look like? Consider consistency, confidentiality, expertise, diversity, composition, etc.

One group oriented the discussion on what the process of a non-IRB structure would look like and agreed that such a structure should be internal to the company and comprised of insiders. The group also debated the issue of when such a process should be triggered, specifically whether at the data collection point or data use point. Consensus was reached that the ethical review process should apply when data use is intended to exceed the regular product improvement array, or when it may have an impact on people's lives.

Rather than having a completely new rule for the private sector, the group believed that our approach should focus on a structure that would be appropriate for companies of all sizes, or where an automated process could apply by default. If a particular situation proved to be more complex, the company should then implement and trigger a tailored and more thorough review process.

Participants also agreed that consistency of results could be supported by carefully calibrating the composition and expertise of the review body. To that end, particular attention should be directed to selecting the right combination of people.

What are the elements of a new ethical framework? Are we updating Common Rule/Menlo Report guidance, doing an expanded version of FIPPS, or something else?

Some of the participants agreed that existing principles were sound and reliable. Therefore, our focus should be on determining how to implement them and how they apply in corporate settings rather than redefining a whole new set of principles.

Participants also indicated that, depending on the question at stake, implementing ethical principles aligned with the company's culture and values may sometimes be more relevant than conforming to one or another particular process. General principles participants thought should be incorporated include concerns such as discrimination, privacy harm and economic harm. Participants also believed that companies should leverage the use of aggregated, anonymized data to the extent possible.

What is the feasibility of a formal structure in regards to corporate data uses (analytics, product development, new technology, etc.)?

One group acknowledged the difficulty of creating a core ethical culture within a company. Consequently, they believed that peer pressure within the industry is likely to play a key role in the implementation of formal structure applying to corporate data uses.

Another group raised the critical importance of training. Participants welcomed Facebook's model and the notion that a standardized structure would be an incredibly useful driver for ethics within organizations. Companies across sectors would need to review a particular project in the light of set principles and apply a relevant checkbox-equivalent system at the early stage of the process. While participants conceded it would represent a bigger challenge for larger companies where myriad of decisions are made on a daily basis, a majority believed it would still be reasonably feasible.

The group finally suggested that developing data-use and risk taxonomies could facilitate the implementation of a formal structure in regards to corporate data uses. Additionally, this would help identify where user choices and controls would do the most to mitigate potential harms.

How do ethics intersect with legal frameworks? Consider issues of legality, fairness, and benefit-risk analysis.

As several participants discussed, ethics is the foundation for law. It is defined as a set of moral and substantive aspirational principles relating to a specified group, field, or form of conduct. One group framed part of the problem as being that the corporate data uses discussed pertain to new human activity that did not exist before. As a result, ethic comes into play first and the desired ethical goals still need to be established in order to articulate the relevant principles that will be necessary before any law can be created.

During the full group discussion, one participant emphasized the need to motivate the private sector to move beyond consent and implement a benefit/risk assessment to achieve a balance. Participants agreed that laws constitute a negative incentive and recognized that corporations also need positive incentives to act.

To that end, another participant referenced the earlier firestarter presentation on environmental management systems, which do not offer a safe harbor *per se* but ensure some level of protection when companies have a formal process in place. In that context, when an issue arises at any stage of the process, a company that has a structured review system gets a lesser penalty.

Panel: Operationalizing Ethical Reviews. The last panel of the day convened corporate, academic and advocate leaders to discuss how solutions could be practically implemented within corporate structures.

Industry leaders from a diverse range of sectors shared their experience and input which resulted in an interactive conversation, moderated by **Susan Etlinger**, **Industry Analyst, Altimeter Group**.[31]

During this session, panelists described their organizations' approaches incorporating ethical review into corporate settings in a principled and practical manner. For example, **Hilary Wandall, Merck's Assistant Vice President, Compliance & Chief Privacy Officer** described how the company keeps the tenets of privacy and data protection aligned with company policy and values, which are publicly available on its website. The guiding principles for Merck employees include trust and the prevention of harm, which, in keeping with Professor Hirsch's presentation, they believe help make the company and its work more sustainable. Ms. Wandall also emphasized that, as a practical measure, Merck attempts to handle as many of its data and practice reviews as possible internally.

As background to his remarks, **David Hoffman, the Associate General Counsel and Global Privacy Officer of Intel,** submitted a 2015 white paper titled *Rethinking Privacy: Fair Information Practice Principles Reinterpreted*.[32] During the panel, as well as echoing Ms. Wandall's alignment of established company values with ethical data guidelines, Mr. Hoffman described Intel's long-standing use of a group of experts who are tapped for reviews of developing data practices. Given the similarity of this arrangement to more traditional IRB processes, he then raised the question of when and in what ways any new ethical review processes would be more appropriate or sufficient than the existing process for his organization. Mr. Hoffman also spoke to the practical challenges of ensuring the independence and anonymity of the reviewers, and sought suggestions from the other panelists (and audience) about additional ways to engage and use the team of reviewers.

Also during this panel, **Lauri Kanerva**, **Research Management Lead, Facebook,** spoke to the challenges of operating as a platform in both a commercial and research context. He described Facebook's commitment to finding ways to allow users to communicate without changing the outcome of those communications, including through processes such as its research and privacy reviews (see below, Molly Jackman and Lauri Kanerva, *Involving the IRB: Building Robust Review for Industry Research*, WASH. & LEE L. REV. ONLINE (forthcoming 2016)). Mr. Kanerva also discussed Facebook's decision to engage in multiple, internal review processes and the ways in which that approach best met the organization's unique needs and expertise. He proposed a framework of "ethics by design" to complement "privacy by design" and "security by design" principles already widely adopted around the world.

Marty Abrams, Executive Director, Information Accountability Foundation, provided participants with an overview of the IAF's work on a *Unified Ethical Frame for Big Data Analysis*,[33] an attempt to the satisfy the need for both a common ethical frame based on key values and an assessment framework. The latter consists of a set of key questions to be asked and answered to illuminate significant issues, both for industry and for those providing oversight to assess big data projects. Mr. Abrams also raised the question, both to the panel and to the workshop's participants, of how we should take into account a broader range of stakeholders and their interests. The ethical impact of a particular research path or data use should not be measured or considered in only the context of the direct participants, but also society at large.

---

[31] Author of THE TRUST IMPERATIVE: A FRAMEWORK FOR ETHICAL DATA USE (June 2015), https://bigdata.fpf.org/wp-content/uploads/2015/11/Etlinger-The-Trust-Imperative.pdf.
[32] David Hoffman & Paula Bruening, RETHINKING PRIVACY: FAIR INFORMATION PRACTICE PRINCIPLES REINTERPRETED (Nov. 2015), https://bigdata.fpf.org/wp-content/uploads/2015/11/Intel-Rethinking-Privacy.pdf.
[33] Information Accountability Foundation, UNIFIED ETHICAL FRAME FOR BIG DATA ANALYSIS (March 2015), http://informationaccountability.org/wp-content/uploads/IAF-Unified-Ethical-Frame.pdf.

Mr. Abrams also kicked off a discussion about the role of *dignity* in research, ethics, and privacy conversations. Several participants were concerned that if, increasingly, researchers will not need to give research subjects notice or get their consent for a particular study, this could lead to research subjects being dehumanized. Connecting U.S. thinking on this topic to similar EU efforts, Jules Polonetsky and others identified recent efforts by the European Data Protection Supervisor to address "a new digital ethics" for "data, dignity and technology."[34]

Camille Nebeker, **Assistant Professor, Department of Family Medicine and Public Health, University of California, San Diego,** presented to the panel and audience new challenges introduced by pervasive sensing methods and computational analytics when associated with human subjects research. Dr. Nebeker suggested novel operational models for ethical review processes that engage stakeholders more actively in shaping research ethics in the digital age. The Connected and Open Research Ethics (CORE) initiative based at the University of California, San Diego has developed a web-based platform to bring researchers, ethicists, privacy expert, technologists and participants together to develop standards and ethical practices. The CORE platform is being design using an iterative, participatory approach involving key stakeholders – at this time Institutional Review Board affiliates and researchers. The CORE's web-based platform and interdisciplinary, collaborative efforts were considered a useful model for private sector organizations to investigate further. Professor Nebeker also highlighted the foundational ethical challenges that will continue to arise as data collection technologies and research methodologies become increasingly sophisticated, including the difficulty of establishing contextual norms.

Finally, building from Professor Nebeker's discussion of contextualization, the panelists launched into conversation about how organizations can understand or help establish users' reasonable expectations for how their data will be collected, used, and shared in a Big Data environment. Several workshop participants joined in to echo the sentiment that ethical and privacy protective conduct is centered in users' expectations. Both company representatives and academics considered how to distinguish between when influencing people's behavior is an ethical decision (such as encouraging users to vote, or students to study harder) and when it is making public policy decisions (such as encouraging users to vote for a particular candidate, or for students to learn about one particular doctrine). One participant suggested a guiding principle that "When we use data beyond the user's imagination, it has to be for their benefit."

A Path Forward

At FPF's 2015 workshop "*Beyond IRBs: Designing Ethical Review Processes for Big Data*," 65 researchers, including lawyers, computer scientists, ethicists and philosophers, as well as policymakers from government, industry and civil society, came together to discuss a blueprint for infusing ethical considerations into organizational processes in a data rich environment.

The event, which revolved around a collection of interdisciplinary articles selected by a review committee following a call for papers, plenary discussions, and firestarter roundtables, brought the following insights and questions to light:

First and foremost, significant work is already being done on data ethics in different sectors, academic disciplines, industry initiatives and government and business organizations. For example, computer

[34] *See* TOWARDS A NEW DIGITAL ETHICS: DATA, DIGNITY AND TECHNOLOGY, September 2015, https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2015/15-09-11_Data_Ethics_EN.pdf.

scientists reported a variety of mechanisms for ethical review, including not only formal IRBs but also informal feedback from peer reviewers and selection decisions by conference program committees.[35] Companies such as Intel and Facebook have devised procedures and substantive guidelines to review new and sensitive uses of data.[36] Government initiatives, such as the Menlo Report in the United States and the European Data Protection Supervisor's Ethics Advisory Group in the European Union, seek to provide a roadmap for researchers and practitioners. Industry groups in healthcare related organizations and corporate accountability initiatives are making strides toward ethical best practices.[37] At the same time, the work on data ethics remains confined to separate silos, preventing cross-pollination and shared learning among disciplines, organizations and industry groups. Clearly, ethical principles should not be malleable and context dependent, nor should they mean different things to different people. Furthermore, absent interoperability, the transfer of knowledge between industry and the academic sector will be hampered. Efforts must be made to remove progress-impeding artificial barriers among discussion forums and to harmonize ethical processes and principles for the data economy.

Second, companies continue to struggle to define the contours of data research and the differences between day-to-day product testing and more ethically challenging projects and experiments.[38] In devising institutions for ethical review processes, companies must address common concerns about risk analysis, disclosure of intellectual property and trade secrets, and exposure to negative media and public reaction. As with environmental management systems, ethical reviews must not be relegated to entry or exit points of engineering or business cycles; rather they must be woven into organizational decision making at every stage of the development process.[39]

Third, existing legal frameworks in both the United States and European Union already provide strong grounds for ethical data reviews. The FTC's "unfairness jurisdiction" authorizes the agency to enforce against an act or practice that "causes or is likely to cause substantial injury to consumers, which is not reasonably avoidable by consumers themselves and is not outweighed by countervailing benefits to consumers or competition." In a recent report, the FTC highlighted risks integral to broad scale data analysis, including creating or reinforcing existing disparities, creating new justifications for exclusion, resulting in higher-priced goods and services for lower income communities, weakening the effectiveness of consumer choice, and more.[40] In Europe, the legitimate interest ground for processing personal data embraces a balancing of corporate interests with individuals' privacy rights. The newly reformed European privacy framework recognizes the significance of research and statistical analysis in the data economy. Article 83 of the General Data Protection Regulation (GDPR) provides a broad research exemption from various obligations; Recital 25aa acknowledges, "It is often not possible to fully identify the purpose of data processing for scientific research purposes at the time of data collection. Therefore data subjects should

---

[35] Narayanan & Zevenbergen, *supra* note 10.
[36] Molly Jackman and Lauri Kanerva, *Involving the IRB: Building Robust Review for Industry Research*, WASH. & LEE L. REV. ONLINE (forthcoming 2016).
[37] Camille Nebeker, Cinnamon Bloss & Nadir Weibel, *New Challenges for Research Ethics in the Digital Age*, WASH. & LEE L. REV. ONLINE (forthcoming 2016). *Also see* Information Accountability Foundation, http://informationaccountability.org/.
[38] Jules Polonetsky & Omer Tene, *The Facebook Experiment: Gambling? In This Casino?*, RE/CODE, July 2, 2014, http://recode.net/2014/07/02/the-facebook-experiment-is-there-gambling-in-this-casino/.
[39] Dennis D. Hirsch & Jonathan H. King, *Big Data Sustainability: An Environmental Management Systems Analogy*, WASH. & LEE L. REV. ONLINE (forthcoming 2016).
[40] FEDERAL TRADE COMMISSION, BIG DATA: A TOOL FOR INCLUSION OR EXCLUSION?, January 2016, https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf.

be allowed to give their consent to certain areas of scientific research when in keeping with recognized ethical standards for scientific research." To inform the implementation of these new rules, the European Data Protection Supervisor launched a process to explore data ethics and established an Ethics Advisory Group comprising experts in philosophy, law, economics and computer science.[41] As part of the new project, legal experts will interpret governing privacy and data protection legislation, jurisprudence and regulations to delineate the borders between ethical guidelines and legal obligations.

Fourth, additional work is needed to inform stakeholders on the definitions, nature and scope of the basic building blocks of an ethical framework. Professor Shilton has demonstrated that academic researchers lack coherent views on fundamental problems,[42] including guidance on acceptable methods of de-identification;[43] an understanding of individuals' sentiment and sensitivities around data use;[44] and delineation of new principles for ethical research. Establishing a common nomenclature and understanding of first principles will reduce the undesirable likelihood of different IRBs reaching conflicting conclusions about similar research projects.

Next Steps. When participants were asked to evaluate the workshop and recommend next steps, there was definitive and enthusiastic agreement that the presentations and papers helped prepare and advance discussions. Participants also agreed that the meeting included the right mix of participants, sectors, and disciplines. The mix of academics with industry leaders was noted as especially valuable toward a developing and continuing effort that addresses both academic and industry goals. As the workshop discussions made evident, problems and challenges around current ethical review mechanisms are clear and extend across different contexts and stakeholders. Paths to a solution are less clear and only slowly evolving. In the words of one participant, "everyone wants to see better options…[but] no one has it figured out."

While the workshop was lauded for appropriately identifying the range of topics and issues around ethical reviews from a privacy perspective and far beyond, its compacted timeframe limited the ability to synthesize discussion toward specific solutions. A shared view emerged, and represented above as the fourth of four key insights, that additional work is needed to achieve a greater sense of consensus on the best path forward for different sectors.

To this end, FPF reconvened many workshop participants at a follow-up event in spring 2016. In collaboration with the Ohio State University's Program on Data and Governance, FPF hosted a *Roundtable on Ethics, Privacy, and Research Reviews* to continue the discussion of ethics, privacy and practical research reviews in corporate settings. The event again brought together corporate and academic leaders to discuss how to integrate ethical and privacy considerations into innovative data projects and research. A major focus of the roundtable was on new papers by Facebook's Molly Jackman and Lauri Kanerva, titled *Evolving the IRB: Building Robust Review for Industry Research*, and by the Center for Democracy and

[41] Press Release, *EDPS starts work on a New Digital Ethics*, January 28, 2016, https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/EDPS/PressNews/Press/2016/EDPS-2016-05-EDPS_Ethics_Advisory_Group_EN.pdf; also see European Data Protection Supervisor, Opinion 4/2015, TOWARDS A NEW DIGITAL ETHICS: DATA, DIGNITY AND TECHNOLOGY, September 2015, https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2015/15-09-11_Data_Ethics_EN.pdf.

[42] Shilton, *supra* note 21.

[43] Polonetsky, Tene & Finch, supra note 12.

[44] Joseph Turow, Michael Hennessy & Nora Draper, *The Tradeoff Fallacy: How Marketers Are Misrepresenting American Consumers And Opening Them Up to Exploitation*, 2016, https://www.ftc.gov/system/files/documents/public_comments/2015/09/00012-97594.pdf.

Technology's Michelle De Mooy and Fitbit's Shelten Yuen, titled *Toward Privacy Aware Research and Development in Wearable Health*. Both papers were presented in draft form at the *Beyond IRBs* discussion.

Following this workshop, FPF has also sought to lay a foundation for organizations of all kinds to better operationalize their ethical considerations and review processes. In Appendix A, for example, FPF has consolidated and distilled leading ethical considerations for research review into a single operational questionnaire, including principles from the *Belmont* and *Menlo* reports, the Association of Internet Researchers' *Ethical Decision-Making and Internet Research* recommendations, the Information Accountability Foundation's *Big Data Ethical Framework*, and FPF's own published work on ethical review board operations, *Beyond the Common Rule: Ethical Structures for Data Research in Non-Academic Settings*.

Finally, FPF will continue to develop and advance common understandings and best practices for secure, privacy protective data sharing between businesses and researchers. Improving researchers' access to corporate data will help advance the state of research and scientific knowledge, enabling researchers and scientists to ask new questions and gain better insights into individual preferences, behaviors and trends.

It is our hope that the *Beyond IRBs* workshop and its associated collection of papers, thoughtful discussions, and active expansion plans will lead to widely acceptable principles, processes and best practices for ethical data reviews and contribute toward a unified data ethics agenda for government, industry, and civil society.

Appendix A: Considerations for Ethical Research Review

	Issue	Metrics		Yes	Maybe	No
Definitions	**Identifiable Private Information**	Are these data private? Consider reasonable expectations of privacy, given changing cultural and contextual circumstances.		○	○	○
			Is this information about behavior that occurs in a context in which an individual can reasonably expect that no observation or recording is taking place?	○	○	○
			Was this information provided for specific purposes by an individual and which the individual can reasonably expect will not be made public?	○	○	○
			Have these data previously been made public?	○	○	○
		Are these data identifiable? Consider the presence of direct identifiers and indirect identifiers in data sets and the presence of safeguards and controls covering the collection, use, and dissemination of the data. (See here for guidance on assessing the spectrum of data identifiability).		○	○	○
			Are data explicitly personal or potentially personally identifiable?	○	○	○
			Are data pseudonymized or key-coded?*	○	○	○
			Are data de-identified?*	○	○	○
			Are data very highly aggregated or anonymized?*	○	○	○
		Are these data persona? Consider local laws and regulations.		○	○	○

			Are these data reasonably linkable to a specific consumer, computer, or device? Consider that data may *not* be reasonably linkable if an organization (1) takes reasonable measures to ensure that the data are de-identified; (2) publicly commits not to try to re-identify the data; and (3) contractually prohibits downstream recipients from trying to re-identify the data. (See here for guidance on US FTC de-identification).	◯	◯	◯
			Are these data that identify an individual or with respect to which there is a reasonable basis to believe that the information can be used to identify an individual? (See here for guidance on HIPAA De-Identification),	◯	◯	◯
			Do these data relate to an identified or identifiable natural person? Consider that an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, online identifier, or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural, or social identity of that person. (See here for guidance on the EU General Data Protection Regulation anonymization standard).	◯	◯	◯
	Human Subjects	Are data obtained by an intervention or interaction with an individual human subject?		◯	◯	◯
			Consider also manipulations of the individuals' environments that are performed for research purposes, which could include manipulation of their computing devices, or automated appliances in the home.	◯	◯	◯
	Harm	Does this activity create a risk of harm to individuals? Consider that potential harms related to exposing the identity of research subjects engaging in sensitive behaviors, communications, or relationships, which they assume to be private, can extend beyond the direct research subject to family, friends or other community relations.		◯	◯	◯
			If yes, is the risk of harm no more than minimal? Consider whether the probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests	◯	◯	◯
		Does this activity target or affect vulnerable populations? Consider potential disparate impacts on groups such as minorities, children, the elderly, the disabled, or those suffering from health conditions.		◯	◯	◯

		Does this activity create a risk of societal harms? Consider potential harms related to: systems assurance; individual and organizational privacy; reputation, emotional well-being, or financial sensitivities; and infringement of legal rights (derived from constitution, contract, regulation, or common law).		◯	◯	◯
			Does this activity impact users' (including non-research subjects') ability to interact with data or technology without loss of privacy?	◯	◯	◯
			Does this activity disrupt users' (including non-research subjects') access to data or technology?	◯	◯	◯
			Does this activity create unreasonable constraints on protected speech or activity?	◯	◯	◯
	Research	Is this activity a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge?		◯	◯	◯
		Is this activity a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge, but not intended for publication or public dissemination?		◯	◯	◯
		Is this activity intended to improve, maintain, or add features to new or current products, including testing or analysis done for such purposes? Consider A/B market testing, for example.		◯	◯	◯
Process	**Triggers for Review**	Does this activity involve identifiable private information? (see above)		◯	◯	◯
		Does this activity involve human subjects? (see above)		◯	◯	◯
		Does this activity create a risk of harm (see above)		◯	◯	◯
		Is this activity research?		◯	◯	◯
		Does this activity involve using data for an unexpected or new purpose outside of the context in which it was initially created or collected? Consider cultural expectations and norms that individuals may attach to the venue in which they are interacting		◯	◯	◯

		Does this activity utilize sensitive data (e.g., health or mental conditions, intimate relations, political views, etc.)? Consider local laws and regulations, as well as social norms.		◯	◯	◯
		Does this activity target or affect special populations? Consider potential disparate impacts on groups such as minorities, children, the elderly, the disabled, or those suffering from health conditions.		◯	◯	◯
	Reviewing Body	Is the reviewing body internal to this organization?		◯	◯	◯
			Does the reviewing body's membership have the necessary expertise to understand the ethical and technical challenges of a particular data use or experimentation?	◯	◯	◯
			Does the reviewing body's membership reflect multiple viewpoints within an organization? Consider diversity of disciplines or professions, race, gender, cultural backgrounds, and sensitivity to issues such as community attitudes.	◯	◯	◯
			Does the reviewing body's membership include external viewpoints?	◯	◯	◯
			Is at least one member of the reviewing body not accountable to the organization's senior management?	◯	◯	◯
			Does the reviewing body have adequate resources to investigate data uses or experiments, obtain expert advice and counsel, and follow projects over time?	◯	◯	◯
			Does the reviewing body set forth transparent rules and procedures?	◯	◯	◯
			Does the reviewing body document the rationale for its decision-making?	◯	◯	◯
		Is the reviewing body external to this organization?	Does the reviewing body's membership have the necessary expertise to understand the ethical, business, and technical challenges of a particular data use or experimentation?	◯	◯	◯

			Does the reviewing body's membership reflect multiple viewpoints? Consider diversity of disciplines or professions, race, gender, cultural backgrounds, and sensitivity to issues such as community attitudes.	◯	◯	◯
			Does the reviewing body have safeguards to protect confidential intellectual property and trade secrets as necessary?	◯	◯	◯
			Does the reviewing body have the resources or capability to provide ongoing monitoring of and consultation about this activity over time?	◯	◯	◯
			Does the reviewing body set forth transparent rules and procedures?	◯	◯	◯
			Is the reviewing body subject to formal reporting requirements and regulatory oversight?	◯	◯	◯
	Time of Review	Does the reviewing body evaluate/approve activities prior to their commencement?		◯	◯	◯
		Does the reviewing body provide ongoing monitoring and consultation about activities?		◯	◯	◯
		Does the reviewing body evaluate/approve activities after they are completed but before publication?		◯	◯	◯
	Principles	Respect for Persons	Have you obtained informed consent from the subject of this activity?	◯	◯	◯
			Were individuals adequately notified about the research? Did they comprehend the notice? Was their consent voluntary?	◯	◯	◯
			Are individuals free to withhold consent to participation without negative consequences?	◯	◯	◯
			Will you provide ex post notification to affected individuals in the absence of consent?	◯	◯	◯
			Did you obtain informed consent for this specific research purpose or use, rather than rely on consent given for another research purpose?	◯	◯	◯
			Did you obtain consent from every individual in a group, rather than rely on individual consent to imply the consent of the group?	◯	◯	◯

		Beneficence	Have you systematically identified and weighed the risks and benefits of this activity for a range of stakeholders, including both affected individuals and society at large? Consider a cost-benefit or data benefit analysis. (See here for guidance on data risk-benefit analysis).	◯	◯	◯
			Have you diligently analyzed how to minimize harms and maximize benefits?	◯	◯	◯
			Have you preemptively planned to mitigate any realized harms, and implemented these evaluations into your activities?	◯	◯	◯
			Are the anticipated risks proportional to or outweighed by the anticipated benefits of this activity?	◯	◯	◯
		Justice	Have you identified all relevant stakeholders and their concerns? Consider individuals, organizations, political entities/governments, local communities, and society at large as well as their rights, interests, social norms, and cultural values.	◯	◯	◯
			Were affected individuals or groups selected for this activity in a fair manner? Consider that individuals or groups should not be arbitrarily targeted in ways that are not germane to legitimate research questions and activities based on attributes including (but not limited to): religion, political affiliation, sexual orientation, health, age, technical competency, national origin, race, or socioeconomic status.	◯	◯	◯
			Have you identified which stakeholders will accrue the benefits of this activity and ensured that benefits will be fairly distributed? Consider disparate impact analysis.	◯	◯	◯
			Have you identified which stakeholders will incur the costs of this activity, and ensured that the burdens will be fairly distributed? Consider disparate impact analysis. (See here for guidance on data risk-benefit analysis).	◯	◯	◯
			If this activity creates a risk of harm to its subjects, do the potential benefits to society outweigh those harms?	◯	◯	◯
		Respect for law and public interest	Did you engage in due diligence to identify applicable laws, regulations, contracts, industry codes, and other private agreements and then conduct your activities in ways that respect these restrictions? Consider the context of and conditions on data when they originates from others.	◯	◯	◯

			Have you clearly communicated the purposes of your activities (why data collection and/or direct interaction is required) and how research results will be used?	◯	◯	◯
			Have you clearly communicated the risk assessment and harm minimization related to your activities?	◯	◯	◯
			Have you determined whether the results of these activities will be published?	◯	◯	◯
			Can the anticipated result of this activity be achieved in a less data-intensive or intrusive manner?	◯	◯	◯
			Are the data and/or analysis used for this activity sustainable? Consider how long any insights achieved might endure once applied in practice, including factors such as whether the source data will be available for a period of time in the future, whether the data can be kept current, whether one has the legal permissions to process the data for the particular activity, and whether the analysis may need to be changed or refined to keep up with evolving trends and individual expectations.	◯	◯	◯
			Do you have procedures in place to document research methodology, ethical evaluations, data collected, and results generated and make them available responsibly in accordance with balancing risks and benefits?	◯	◯	◯
			Does the reviewing body have thorough procedural documentation, including board constitution, membership, reporting structure, and ongoing monitoring and review procedures?	◯	◯	◯

Appendix B: Workshop Participants

Marty Abrams, Executive Director and Chief Strategist, Information Accountability Foundation

Marty Abrams has over 35 years of experience as an information and consumer policy innovator. Multi-stakeholder collaboration has been a key for Abrams in developing practical solutions to dilemmas in information policy. His most recent work has been on big data governance and privacy compliance driven by demonstrable data stewardship. For the past five years, he has led the Global Accountability Project, which has refined the accountability principle that is part of various data protection laws and guidance documents. Prior to his work at the foundation, Abrams was the co-founder and President of the Centre for Information Policy Leadership at Hunton & Williams LLP, which he led for 13 years. Prior to that, he was Vice President of Information Policy at Experian and Director of Information Policy at TRW Information Systems, where he designed one of the earliest privacy impact assessment tools. He also chaired their Consumer Advisory Council.

Micah Altman, Director of Research and Head/Scientist, Program on Information Science for the MIT Libraries, Massachusetts Institute of Technology

Dr. Altman conducts work primarily in the fields of social science, information privacy, information science and research methods, and statistical computation — focusing on the intersections of information, technology, privacy, and politics; and on the dissemination, preservation, reliability and governance of scientific knowledge. Prior to arriving at MIT, Dr. Altman served at Harvard University for fifteen years as the Associate Director of the Harvard-MIT Data Center, Archival Director of the Henry A. Murray Archive, and Senior Research Scientist in the Institute for Quantitative Social Sciences. Dr. Altman is also a Non-Resident Senior Fellow at The Brookings Institution.

Solon Barocas, Postdoctoral Research Associate, Center for Information Technology Policy, Princeton University

Solon Barocas completed his doctorate in the department of Media, Culture, and Communication at New York University, where he remains an affiliate of the Information Law Institute. Dr. Barocas also works with the Data & Society Research Institute and serves on the National Science Foundation funded Council for Big Data, Ethics, and Society. His research explores issues of fairness in machine learning, methods for bringing accountability to automated decisions, the privacy implications of inference, the role that privacy plays in mitigating economic inequality, and the ethics of research involving repurposed data and online experimentation.

Lucy Bernholz, Ph.D., Senior Research Scholar, Stanford University

Lucy Bernholz is a Senior Researcher at Stanford University's Center on Philanthropy and Civil Society where she co-leads the Digital Civil Society Lab. She has been a Visiting Scholar at The David and Lucile Packard Foundation and a Fellow at the Rockefeller Foundation's Bellagio Center, with the Hybrid Reality Institute, and at the New America Foundation. She writes extensively on philanthropy, technology, information, and policy on her award winning blog, philanthropy2173.com. This work led *The Huffington Post* to hail her as a "game changer". She is the author of numerous articles and books about the philanthropy, policy, and technology, including the Blueprint Series of *Annual Industry Forecasts on Philanthropy and the Social Economy*, the 2010 publication *Disrupting Philanthropy*, and her 2004 book *Creating Philanthropic Capital Markets: The Deliberate Evolution.*

Marjory Blumenthal, Executive Director of the President's Council of Advisors on Science and Technology, White House

Marjory Blumenthal is a connector of people and ideas, translating technical concepts into plain English and leading discussions and collaborations across different kinds of natural scientists, engineers, social scientists, policy analysts, and civil society. She manages PCAST's analytical program, engaging its 20 distinguished scientists and engineers plus additional experts; speaks publicly about PCAST's work; and fosters the implementation of PCAST recommendations. Marjory's PCAST projects have addressed how systems engineering can improve the delivery of health care, the challenge of protecting privacy in the context of big data, new directions for cybersecurity, how information technology can improve education and training, and more—PCAST spans the landscape of science and technology. Previously she spent a decade as Associate Provost, Academic at Georgetown University, and was the founding Executive Director of the National Academies' Computer Science and Telecommunications Board (CSTB).

Geoffrey C. Bowker, Professor, School of Information and Computer Science, University of California, Irvine

Geoffrey C. Bowker is Professor at the School of Information and Computer Science, University of California at Irvine, where he directs a laboratory for Values in the Design of Information Systems and Technology. Recent positions include Professor of and Senior Scholar in Cyberscholarship at the University of Pittsburgh iSchool and Executive Director, Center for Science, Technology and Society, Santa Clara. Together with Leigh Star he wrote Sorting Things Out: Classification and its Consequences; his most recent book is Memory Practices in the Sciences

Jeff Brueggeman, Vice President, Global Public Policy, AT&T

Jeff Brueggeman is responsible for developing and advocating AT&T's global public policy positions on privacy, cybersecurity and human rights issues. He represents AT&T in a wide range of legislative, regulatory and policy development proceedings involving privacy and cybersecurity matters. In addition, he leads AT&T's engagement with various privacy and Internet policy organizations. Prior to assuming his current role, Mr. Brueggeman helped manage AT&T's privacy policies and coordinate the implementation of data privacy and security programs across the company. He has participated extensively in international Internet policy events and organizations, and served on the Internet Governance Forum's Multi-stakeholder Advisory Group.

Jill Caiazzo, Counsel, Government and Regulatory Affairs, IBM

Jill Caiazzo is Government and Regulatory Affairs Counsel at IBM Corporation. In this role, she advises on a variety of international trade regulatory issues with respect to IBM's worldwide operations, as well as advocates on public policy issues in the area of data privacy. Prior to joining IBM, Ms. Caiazzo was an attorney with Sidley Austin LLP in Washington, DC, where she was promoted to partner. Her practice focused on regulatory and policy matters affecting cross-border transactions. She also served as a judicial clerk to the Hon. Richard W. Goldberg of the U.S. Court of International Trade. While attending law school, Ms. Caiazzo was a management consultant in the area of government procurement at KAA Federal Solutions, Inc. Ms. Caiazzo is a magna cum laude graduate of the Georgetown University Law Center. She received her bachelor of science, cum laude, from Georgetown University's School of Foreign Service. Ms. Caiazzo also holds a certificate in comparative political and social studies from the Institut d'Etudes Politiques de Paris (Sciences Po). She is a member of the bars of the State of New York and the District of Columbia.

Ryan Calo, Assistant Professor, University of Washington School of Law

Ryan Calo is an assistant professor at the University of Washington School of Law and an assistant professor (by courtesy) at the Information School. He is a faculty co-director (with Batya Friedman and Tadayoshi Kohno) of the University of Washington Tech Policy Lab, a unique, interdisciplinary research unit that spans the School of Law, Information School, and Department of Computer Science and Engineering. Professor Calo is an affiliate scholar at the Stanford Law School Center for Internet and Society (CIS), where he was a research fellow, and the Yale Law School Information Society Project (ISP). He serves on numerous advisory boards, including the University of California's People and Robots Initiative, the Electronic Frontier Foundation (EFF), the Electronic Privacy Information Center (EPIC), Without My Consent, and the Future of Privacy Forum (FPF).

Subhashini (Shubha) Chandrasekharan, Ph.D.,AAAS Fellow Center for Global Solutions, Global Development Lab, USAID; Assistant Research Professor, Duke Global Health Institute, Duke University

Subhashini Chandrasekharan is global health researcher with a background in genetics and policy research encompassing, bioethics, law and the social sciences. Shubha received her PhD in genetics and molecular biology from the University of North Carolina at Chapel Hill and completed postdoctoral training in genetics there. She next completed a fellowship on ethical, legal and social (ELS) issues of genomics at the Center for Public Genomics at Duke University. She is currently Assistant Research Professor at the Duke Global Health Institute and studies ELS and policy issues surrounding implementation of new prenatal genomic testing technologies in the US and in low and middle-income countries (LMICs). In addition, she has strong interests in science diplomacy and social entrepreneurship for health services capacity building in LMICs.

She is currently an AAAS Science and Technology Policy Fellow at the Center for Global Solutions in the Global Development Lab, USAID. She works on projects focused on improving and strengthening health information systems capacity in West Africa as well as ethical issues in the use of digital technologies and data for development.

Danielle Citron, Lois K. Macht Research Professor & Professor of Law, University of Maryland; Senior Fellow, Future of Privacy Forum

Professor Danielle Citron is the Lois K. Macht Research Professor & Professor of Law at the University of Maryland Francis King Carey School of Law. Her work focuses on information privacy, cyber law, automated systems, and civil rights. In addition to 20 articles and essays in law reviews, Professor Citron is the author of *Hate Crimes in Cyberspace* published by Harvard University Press in 2014. Cosmopolitan and Harper's Bazaar nominated her book as one of the top 20 "Best Moments for Women" in 2014. She frequently writes for The Atlantic, New York Times, Time, CNN, The Guardian, the New Scientist, and Slate. She is a regular contributor at Forbes.com and Concurring Opinion. Professor Citron is an Affiliate Fellow at the Yale Information Society Project and an Affiliate Scholar at the Stanford Center on Internet and Society. She serves as an advisor to California Attorney General Kamala Harris's Task Force Against Cyber Exploitation and the American Law Institute's Restatement Third, Information Privacy Principles Project. She is on the advisory boards of the Cyber Civil Rights Initiative, Electronic Privacy Information Center, Future of Privacy Forum, Harvard Berkman Center's Initiative on Youth-Oriented Hate Speech, Without My Consent, and Teach Privacy.

Chris Clifton, Program Director, National Science Foundation; Professor of Computer Science (by courtesy), Purdue University

Dr. Clifton works on data privacy, particularly with respect to analysis of private data. This includes privacy-preserving data mining, data de-identification and anonymization, and limits on identifying individuals from data mining models. He also works more broadly in data mining, including data mining of text and data mining techniques applied to interoperation of heterogeneous information sources. Fundamental data mining challenges posed by these applications include extracting knowledge from noisy data, identifying knowledge in highly skewed data (few examples of "interesting" behavior), and limits on learning. He also works on database support for widely distributed and autonomously controlled information, particularly issues related to data privacy.

Stanley W. Crosley, Co-Director, Center for Law, Ethics, and Applied Research in Health Information, Indiana University

Stan Crosley chairs the Data Privacy and Health Information Governance team, a cross-disciplinary team of lawyers with health data privacy experience. Crosley is the former Chief Privacy Officer for Eli Lilly and Company, a position he held for 10 years, where he initiated Lilly's global privacy program and negotiated the company's compliance with FTC and State consent decrees, multiple European Data Protection Authority privacy inspections, and successful certification to the EU Safe Harbor program, gaining recognition for the program by receiving the 2007 Innovation Award from the International Association of Privacy Professionals. Crosley co-founded and served as chair of the International Pharmaceutical Privacy Consortium and was a member of the IOM Medical Research and Privacy Committee. He currently serves on the boards of the International Association of Privacy Professionals (IAPP), the Indiana Health Information Technology, Inc., the International Data Privacy Law Journal, and The Privacy Projects.

Mary Culnan, Professor Emeritus, Bentley University; Senior Fellow, Future of Privacy Forum

Mary Culnan is Professor Emeritus at Bentley University. She is a Senior Fellow at the Future of Privacy Forum where she also currently serves as President of the Board of Directors. Additionally, she serves as a member of the GAO's Executive Committee on Information Management and Technology and the Data Privacy Day Advisory Committee. Mary has testified before Congress and other government agencies on a range of privacy issues. Her current research interests include governance of privacy and security and improving online privacy notices. Her work has been published in a range of academic journals and has been used by the FTC to make recommendations to Congress.

Michelle De Mooy, Deputy Director, Consumer Privacy Project, Center for Democracy & Technology

Michelle De Mooy's work is focused on promoting strong consumer privacy rights through pro-privacy legislation and regulation, working with industry to build and implement good privacy practices, and analyzing emerging impact on data from new technology. Most recently, she has focused on personal health technology and related privacy concerns. De Mooy has been a panelist and featured speaker at many events related to digital privacy, including FTC workshops, the State of the Net, the Amsterdam Privacy Conference, and has testified before Congress on privacy and security issues. Her background is in advocacy for underserved populations, managing online media strategy for progressive causes, web product marketing, and software development.

Jana Diesner, Assistant Professor, iSchool, University of Illinois at Urbana-Champaign

Jana Diesner earned her PhD from Carnegie Mellon University, School of Computer Science, in the Computation, Organizations and Society (COS) Program. Diesner conducts research at the nexus of network science, natural language processing and machine learning. Her research mission is to contribute to the computational analysis and better understanding of the interplay and co-evolution of information and the structure and functioning of socio-technical networks. She develops and investigates methods and technologies for extracting information about networks from text corpora and considering the content of information for network analysis. In her empirical work, she studies networks from the business, science and geopolitical domain. She is particularly interested in covert information and covert networks.

Jeremy Epstein, Lead Program Officer for the Secure and Trustworthy Cyberspace (SaTC) Program, National Science Foundation

Jeremy Epstein is lead program officer for the National Science Foundation's Secure and Trustworthy Cyberspace (SaTC) program. SaTC is NSF's flagship cybersecurity research program, with about 700 active grants, 1000 researchers, and 2000 graduate students covering all aspects of cybersecurity & privacy. Jeremy is on loan to NSF from SRI International, where his research interests include voting system security and software security. He is associate editor in chief of IEEE Security & Privacy Magazine, founder of Scholarships for Women Studying Information Security, and a member of the Election Assistance Commission's Technical Guidelines Development Committee (TGDC) responsible for writing voting system standards. He holds an MS from Purdue University in Computer Sciences, and is ABD from George Mason University.

Susan Etlinger, Industry Analyst, Altimeter Group

Susan Etlinger is an industry analyst with Altimeter Group, where she promotes the smart, well-considered and ethical use of data. She conducts independent research on these topics and advises global executives on data and analytics strategy. Etlinger is on the board of The Big Boulder Initiative, an industry organization dedicated to promoting the successful and ethical use of data. She is a TED speaker, is regularly asked to speak on data strategy and best practices, and has been quoted in media outlets such as The Wall Street Journal, The New York Times, and BBC. Find Etlinger on Twitter at @setlinger, or on her blog at susanetlinger.com.

Joshua Fairfield, Professor, Washington and Lee University School of Law

Joshua Fairfield is an internationally recognized law and technology scholar, specializing in digital property, electronic contract, big data privacy, and virtual communities. His article *Privacy as a Public Good*, in the Duke Law Journal, is the first comprehensive treatment of privacy as a good subject to problems of social production. He is writing a book for Cambridge University Press, titled *ESCAPE: Property, Privacy, and the Internet of Things.* He has written on the law and regulation of e-commerce and online contracts and on the application of standard economic models to virtual environments. Professor Fairfield's other research focuses on big data privacy models and the next generation of legal applications for cryptocurrencies. Professor Fairfield consults with U.S. government agencies, including the White House Office of Technology and the Homeland Security Privacy Office, on national security, privacy, and law enforcement within online communities and as well as on strategies for protecting children online.

Kelsey Finch, Policy Counsel, Future of Privacy Forum

Kelsey Finch's projects at FPF include consumer wellness and wearables, big data, de-identification standards and privacy by design. Before coming to FPF, Kelsey was an inaugural Westin Fellow at the IAPP, where she produced practical research on a range of privacy topics and edited the FTC Privacy Casebook. She is a graduate of Smith College and the Benjamin N. Cardozo School of Law, with a concentration in Intellectual Property & Information Law.

Simson L. Garfinkel, Senior Advisor, National Institute of Standards and Technology

Simson L. Garfinkel is a Senior Advisor at the National Institute of Standards and Technology's Information Access Division. Garfinkel's research interests include digital forensics, usable security, data fusion, information policy and terrorism. He holds seven US patents for his computer-related research and has published dozens of research articles on security and digital forensics, including *Database Nation, The Death Of Privacy in the 21st Century* (O'Reilly, 2000).

Joshua M. Greenberg, Director of the Digital Information Technology Program, Alfred P. Sloan Foundation

Dr. Joshua M. Greenberg received his Bachelor of Arts in History of Science, Medicine and Technology from the Johns Hopkins University, and both Masters and Doctoral degrees from Cornell University's Department of Science & Technology Studies. His dissertation work on the early history of the consumer videocassette recorder and the invention of the video rental industry was published as "From Betamax to Blockbuster" by the MIT Press (2008). Prior to joining the Foundation, Dr. Greenberg was the New York Public Library's first Director of Digital Strategy and Scholarship, where he developed and led a digital strategy centered on building online visitors and deepening engagement through access to collections both on Library websites and third-party platforms and increased exposure to staff expertise via blogs and other social media. He is an active member of the broader digital library and digital humanities communities, and maintains active research and teaching interests in the history and sociology of information technology, the dynamics of public engagement with expert knowledge, and the methodological implications of new digital technologies for research.

Dennis D. Hirsch, Geraldine W. Howell Professor of Law, Capital University Law School

Dennis D. Hirsch is the Geraldine W. Howell Professor of Law at Capital University Law School where he teaches information privacy law, environmental law and administrative law. In 2010 he served as a Fulbright Senior Professor at the University of Amsterdam, Institute for Information Law (IViR), where he conducted research on Dutch data protection regulation and taught a course on Comparative Information Privacy Law. He is a faculty organizer of the University of Amsterdam's Summer Course on Privacy Law and Policy and teaches in that program each July. Professor Hirsch is the author of numerous articles and a prize-winning textbook and has lectured nationally and internationally at universities, law schools, bar associations and conferences. His current research focuses on governance theory, privacy law and policy, and big data's social and legal implications.

Anna Lauren Hoffman, Postdoctoral Researcher and Instructor, School of Information, University of California, Berkeley

Anna Lauren Hoffmann a trans woman and scholar working at the intersections of information, technology, culture, and ethics at the UC Berkeley School of Information. Generally, her work examines the ways in which the design and use of information technology can promote or hinder the pursuit of social justice, especially in data-intensive or "Big Data" related contexts. In addition, she employs discourse analysis to

explore the values and biases that underwrite understandings of technology, privacy, and research ethics as promoted by various stakeholders.

David A. Hoffman, Director of Security Associate General Counsel and Global Privacy Officer, Intel Corporation

David A. Hoffman is Director of Security Associate General Counsel and Global Privacy Officer at Intel Corporation, in which capacity he heads the organization that oversees Intel's privacy compliance activities, legal support for privacy and security, and all external privacy/security engagements. Mr. Hoffman joined Intel in 1998 as Intel's eBusiness attorney to manage the team providing legal support for Intel's Chief Information Officer. In 2005, Mr. Hoffman moved to Munich, Germany, as Group Counsel in the Intel European Legal Department, while leading Intel's Worldwide Privacy and Security Policy Team. Mr. Hoffman is the co-chair of the International Chamber of Commerce's Task Force on Data Protection and Privacy. Mr. Hoffman is also a Senior Lecturing Fellow at Duke University School of Law where he teaches a class on Information Privacy and Surveillance Law.

Margaret Hu, Assistant Professor, Washington and Lee University School of Law

Margaret Hu is an Assistant Professor of Law at Washington and Lee University School of Law. She currently teaches courses in Constitutional Law, Administrative Law, and Privacy Law. Her research interests include the intersection of immigration policy, national security, cybersurveillance, and civil rights. Previously, she served as senior policy advisor for the White House Initiative on Asian Americans and Pacific Islanders, and also served as special policy counsel in the Office of Special Counsel for Immigration-Related Unfair Employment Practices, Civil Rights Division, U.S. Department of Justice, in Washington, D.C. Professor Hu's recent articles include *Biometric ID Cybersurveillance*, *Small Data Surveillance v. Big Data Cybersurveillance*, *Big Data Blacklisting*, and *Taxonomy of the Snowden Disclosures*.

Molly Jackman, Public Policy Research Manager, Facebook

Molly Jackman is the Public Policy Research Manager at Facebook, where she is responsible for guiding the company's research agenda from a policy perspective. Her work includes prioritizing projects that will improve user experience, contribute to general knowledge, and provide insights that can lead to positive change in the world. Previously, she served as an Assistant Professor in political science at Vanderbilt University, and was a Fellow at the Brookings Institution. She received her PhD in political science from Stanford University in 2013, where she specialized in quantitative methods and political institutions.

Rey Junco, Associate Professor of Education and Human Computer Interaction, Iowa State University Faculty Associate, Berkman Center for Internet and Society, Harvard University

Rey Junco applies quantitative methods to analyze the effects of social media on youth psychosocial development, engagement, and learning. His research has focused on discerning better practices for using social technologies to enhance learning outcomes. Junco has found that technology, specifically social media like Facebook and Twitter, can be used in ways that improve engagement and academic performance. Junco is also interested in examining how social media affect interpersonal relationships, identity development, online discourse, and digital inequalities, and how the use of digital media promotes formal and informal learning. He is particularly interested in how online anonymity impacts youth identity formation. As part of his ongoing research program, he is investigating the ability to use trace data (seemingly irrelevant data collected through natural uses of technology) to provide real-time and unobtrusive prediction of student outcomes. Junco is the author of three books, with the latest *Engaging*

Students Through Social Media: Evidence-Based Practices for Use in Student Affairs focusing on translating interdisciplinary research findings to effective educational practices. Junco's work has been cited in major news outlets such as the New York Times, NPR, PBS, NBC, Time, US News & World Report, USA Today, The Guardian, The Atlantic, Huffington Post, and Mashable. Junco was also a regular guest on the NPR show, Tell Me More where he discussed how research informed the societal impact of social media use. His empirical work has been published in high-impact journals such as Computers & Education, Computers in Human Behavior, and the Journal of Computer Assisted Learning. Junco blogs at http://blog.reyjunco.com.

Lauri Kanerva, Research Management Lead, Facebook

Lauri Kanerva is the Research Management Lead at Facebook. In this role, he serves as the in-house adviser to executives and researchers regarding research oversight. Kanerva manages the Research Review process and participates in research and policy advisory committees, coordinates Facebook's research activities, particularly those involving partnerships with external organizations, and monitors compliance with Facebook's internal policies. Before joining Facebook, Kanerva spent 10 years running the non-medical IRB at Stanford University. Kanerva holds degrees in Business Administration, Sports Science and Physical Therapy.

Erin Kenneally, Portfolio Manager, Cyber Security Division, Science & Technology Directorate, U.S. Department of Homeland Security

Erin Kenneally is a licensed Attorney specializing in information technology law, including privacy technology, information risk, trusted information sharing, technology policy, cybercrime, ICT ethics, and emergent IT legal risk. Kenneally is currently the Portfolio Manager for cybersecurity research data sharing, privacy and ICT ethics in the Cybersecurity Division, Science & Technology Directorate at the U.S. Dept. of Homeland Security. Other positions included Founder and CEO at Elchemy, and Technology Law Specialist at the International Computer Science Institute (ICSI), the University of California San Diego, Center for Internet Data Analysis and the Center for Evidence-based Security Research. She holds Juris Doctorate and Master of Forensic Sciences degrees.

Jonathan H. King, Head of Cloud Strategy, Ericsson; Visiting Scholar, Washington University in St Louis School of Law

King's responsibilities include product area cloud strategy, business development, mergers and acquisitions, alliance development and go to market strategy. Prior to joining Ericsson, Jonathan was Vice President, Platform Strategy and Business Development for CenturyLink where he led the cloud and managed services platform transformation at CenturyLink. King received his J.D. from Loyola University Chicago, and completed a Master of Laws in Intellectual Property and Technology Law at Washington University School of Law. King completed his studies with Professor Neil Richards, and co-authored a Stanford Online Law Review Article called *The Three Paradoxes of Big Data* and a Wake Forest Law Review Article called *Big Data Ethics*. He is currently co-authoring with Professor Richards a book chapter called *Big Data and the Future for Privacy* to be published in a handbook of research on digital transformations.

Michael C. Lamb, Chief Counsel of Privacy and Information Governance, RELX Group

Michael C. Lamb, Esq. serves as Member of Advisory Board of The Future of Privacy Forum. Prior to his work in privacy for RELX Group, he served as VP and Lead Counsel (Privacy, Regulatory & Policy) for

LexisNexis Risk Solutions, and for nine years for AT&T as Chief Counsel, AT&T Internet, Chief Marketing Counsel, and AT&T Chief Privacy Officer.

Sagi Leizerov, Ph.D, Executive Director, EY

Sagi Leizerov leads the Ernst & Young's privacy practice, providing the firm's clients with privacy assurance and advisory services. Leizerov has over 20 years of experience in privacy, data protection, security, and crisis management. Leizerov has extensive experience working with both the public and private sectors and has served clients in various industries including healthcare, financial, pharmaceuticals, automotive, online, computer, and human resources. Leizerov holds a BA from the University of Maryland in behavioral sciences, an MBA with a marketing concentration from Johns Hopkins University, and a Ph.D. in conflict analysis and resolution from George Mason University.

Brenda Leong, Senior Counsel and Operations Manager, Future of Privacy Forum

Brenda Leong is Senior Counsel and Operations Manager at the Future of Privacy Forum, primarily supporting issues related to Education Privacy. She works on ed tech industry standards and collaboration on privacy concerns, as well as partnering with parent and educator advocates for practical solutions to the privacy challenges from the expansion of student data. Prior to working at FPF, Brenda served in the U.S. Air Force, including policy and legislative affairs work from the Pentagon and the U.S. Department of State. She is a 2014 graduate of George Mason University School of Law, and has her CIPP/US privacy certification.

Brendon Lynch, Chief Privacy Officer, Microsoft

Brendon Lynch has responsibilities for all aspects of Microsoft's privacy approach, including privacy policy creation and implementation across the company, influencing the creation of privacy and data protection technologies for customers and overseeing communication and engagement with all external audiences. Before joining Microsoft, Brendon led the privacy and risk solutions business at software maker Watchfire. Prior to entering the software industry in 2002, Brendon spent nine years in Europe and North America with PricewaterhouseCoopers, where he provided privacy and risk management consulting services. Brendon serves as Chairman of the International Association of Privacy Professionals (IAPP) Board of Directors, is a Certified Information Privacy Professional (CIPP) and holds a business degree from the University of Waikato, in his home country of New Zealand.

Mary Madden, Researcher, Data & Society Research Institute

Mary Madden is a veteran technology researcher, writer and public speaker, having worked to understand trends in American internet users' behaviors and attitudes for more than a decade. She is currently a Researcher for the Data & Society Research Institute where she is leading an initiative to understand the privacy and security experiences of low-SES populations. Supported by a grant from the Digital Trust Foundation, the project will provide freely accessible data to researchers working in this area and will seek to answer key questions that can help to ground current policy conversations and debates about privacy and security in the digital age. Mary is also an Affiliate at the Berkman Center for Internet and Society at Harvard University where she is part of a long-term collaboration with the Berkman Center's Youth and Media Project that combines quantitative and qualitative research methods to study adolescents' technology use and privacy management on social media. Prior to her role at Data & Society, Mary was a Senior Researcher for the Pew Research Center's Internet & American Life Project. She is a nationally recognized expert on privacy and technology, trends in social media use, and the impact of digital media on teens and parents.

Jeffrey Mantz, Program Director for the Interdisciplinary Behavioral and Social Science Research (IBSS) Program, National Science Foundation

Jeffrey Mantz received his PhD in Anthropology from the University of Chicago in 2003. From 2001-2003, he taught at Vassar College; from 2003-2007 at the California State University at Stanislaus; and from 2008-2012 at George Mason University. He was a Faculty Fellow at the Society for the Humanities at Cornell University, 2007-2008. While at George Mason University, he served as Director of the Anthropology Program, 2009-11, and Director of Graduate Studies, 2009-11. He came to NSF in August 2005 as Director of the Cultural Anthropology Program, first as a Visiting Scientist while on leave from George Mason University (2012-2014) and then as a NSF employee beginning in 2014. Since 2014, he has also served at the Human Subjects Research Officer for the NSF. He also serves as Program Director for the Interdisciplinary Behavioral and Social Science Research (IBSS) program.

Kirsten Martin, Assistant Professor of Strategic Management & Public Policy, George Washington University's School of Business

Kirsten Martin is the principle investigator on a three-year grant from the National Science Foundation to study online privacy. Martin is also a member of the advisory board of the *Future of Privacy Forum* and the Census Bureau's National Advisory Committee for her work on privacy and the ethics of "big data." Martin has published academic papers in *Journal of Business Ethics*, *First Monday*, *Business and Professional Ethics Journal*, and *Ethics and Information Technology* and is co-author of the textbook *Business Ethics: A managerial approach*. She has written teaching cases for the *Business Roundtable Institute for Corporate Ethics* including cases on Google in China as well as Bailouts and Bonuses on the financial crisis. She is regularly asked to speak on privacy and the ethics of big data.

Tim McGovern, Editor, O'Reilly Media

Tim McGovern works at O'Reilly Media, writing and publishing on the intersections of data, organizations, and human behavior. Before coming to O'Reilly, he worked at the University of Chicago Press on social science, particularly history, sociology, and sexuality studies. His background is in the history of ancient religions.

Michelle N. Meyer, Assistant Professor and Director of Bioethics Policy, Mount Sinai Bioethics Program, Union Graduate College–Icahn School of Medicine

Michelle N. Meyer is an Assistant Professor and Director of Bioethics Policy in the Union Graduate College–Icahn School of Medicine at Mount Sinai Bioethics Program, where she writes and teaches at the intersection of law, ethics, and science, with a focus on research ethics and regulation. She serves on an IRB and sits on the board of directors or advisors of three international academic research consortia, where she advises on matters pertaining to research ethics and regulation. Previously, Michelle was an Academic Fellow at the Petrie-Flom Center for Health Law Policy, Biotechnology, and Bioethics at Harvard Law School, a Greenwall Fellow in Bioethics and Health Policy at The Johns Hopkins and Georgetown Universities, and a Research Fellow at the John F. Kennedy School of Government at Harvard. She earned a Ph.D. in religious studies, with a focus on practical ethics, from the University of Virginia and a J.D. from Harvard Law School. Her writing on corporate experimentation for both scholarly and popular audiences can be found at www.michellenmeyer.com.

Kevin Miklasz, Assessment Specialist, **BrainPOP**

Kevin entered education as a trained scientist- he has a PhD in Biology from Stanford University. Both during and after his graduate studies, he has spent his time gaining a smattering of diverse experiences in education: designing science curriculum, teaching after-school science programs, designing science games, running a "cooking as science" blog, designing online learning platforms, running professional development for teachers and professional engineers, and analyzing educational assessment data. Through these experience, Kevin transitioned from scientific research to science education to educational technology to educational research and assessment. Kevin is currently the Assessment Specialist at BrainPOP where he is designing new, playful and meaningful assessments on BrainPOP's website, analyzing data from those assessments, and working on how those assessments can be displayed to teachers through dashboards.

Dr. John Murray, Program Director in the Computer Science Laboratory, SRI International

Dr. John Murray is a Program Director in the Computer Science Laboratory at SRI International. His research interests encompass interactive collaborative systems, software engineering, cognitive ergonomics, and human-machine systems. He has led many innovative interdisciplinary systems research and development projects both in academia and in industry, and has held technical leadership and executive management positions at several international corporations. His technical experience includes diagnostic modeling in complex networked systems, human behavior modeling in computer gaming studies, smart product design, and mobile wearable computer systems. Dr. Murray has received advanced degrees from Dublin Institute of Technology in Ireland, Stanford University, and the University of Michigan, where he was also an adjunct faculty member. He is also a Visiting Scientist in the College of Science at San Jose State University.

Arvind Narayanan, Assistant Professor of Computer Science, Princeton

Arvind Narayanan (Ph.D. 2009) studies information privacy and security and has a side-interest in technology policy. His research has shown that data anonymization is broken in fundamental ways, for which he jointly received the 2008 Privacy Enhancing Technologies Award. Narayanan leads the Princeton Web Transparency and Accountability project that aims to uncover how companies are collecting and using our personal information. He also studies the security and stability of Bitcoin and cryptocurrencies. Narayanan is an affiliated faculty member at the Center for Information Technology Policy at Princeton and an affiliate scholar at Stanford Law School's Center for Internet and Society.

Saira Nayak, Chief Privacy Officer, TUNE

Saira Nayak works to ensure internal compliance and leads TUNE’s external outreach to regulators, and other stakeholders on privacy and data security matters. Nayak, a San Francisco native, has over 15 years of legal experience in antitrust, intellectual property, privacy and data security matters. Before TUNE, she was Director of Policy at TRUSTe, where she defined the company’s external policy platform, and advised on TRUSTe’s privacy management solutions. Prior to TRUSTe, Nayak worked in-house at the Microsoft Corporation, where she counseled product groups on privacy and data security issues, and compliance under Microsoft’s antitrust consent decree with the US Department of Justice and several state AGs. Nayak also practiced at Dickstein Shapiro, (Washington, DC), where she advised clients on antitrust and consumer protection matters and served as Antitrust Counsel for the National Association of Attorneys General (“NAAG”).

Camille Nebeker, Assistant Professor in the Department of Family Medicine and Public Health, University of California, San Diego; Founder/Director, Collaborative for Advancing Professional and Research Integrity (CAPRI)

Prior to Camille Nebeker's appointment with UCSD, she served in an administrative leadership position where she directed academic research ethics and regulatory compliance programs. Her research has received support from the NIH, NSF and ORI and focuses on the design and testing of research/bioethics educational initiatives. Dr. Nebeker is also exploring the ethical dimensions (i.e., informed consent, risk assessment, data management) of socio/biomedical research that utilize emerging technologies. Nebeker refers to this project as MISST-E, which stands for **M**obile **I**maging, pervasive **S**ensing, **S**ocial media, and **T**racking - **E**thics (MISST-E). Dr. Nebeker is leading a study to develop MISST-E guiding principles with a goal of assisting researchers and ethics review boards navigate the ethical dimensions of research using these tools/methods. Using a stakeholder engagement approach, her team will design a web-based prototype called the **C**onnected and **O**pen **R**esearch **E**thics (CORE) to facilitate the development of a "learning ethics system" responsive to 21st century research.

Jules Polonetsky, Executive Director and Co-chair, Future of Privacy Forum

Jules Polonetsky serves as Executive Director and Co-chair of the Future of Privacy Forum. Polonetsky's previous roles have included serving as Chief Privacy Officer at AOL and before that at DoubleClick, as Consumer Affairs Commissioner for New York City, as an elected New York State Legislator and as a congressional staffer, and as an attorney. Polonetsky serves on the Advisory Board of the Center for Copyright Information. He has served on the boards of a number of privacy and consumer protection organizations including TRUSTe, the International Association of Privacy Professionals, and the Network Advertising Initiative. From 2011-2012, Jules served on the Department of Homeland Security Data Privacy and Integrity Advisory Committee.

Neil M. Richards, Professor of Law, Washington University, St Louis

Neil Richards is an internationally-recognized expert in privacy law, information law, and freedom of expression. He is a professor of law at Washington University School of Law, an affiliate scholar with the Stanford Center for Internet and Society and the Yale Information Society Project, and a consultant and expert in privacy cases. He serves on the boards of the Future of Privacy Forum, the Right to Read Foundation, and is a member of the American Law Institute. Professor Richards graduated in 1997 with degrees in law and history from the University of Virginia, and served as a law clerk to William H. Rehnquist, Chief Justice of the United States. Professor Richards is the author of *Intellectual Privacy* (Oxford Press 2015). His many writings on privacy and civil liberties have appeared in a variety of academic journals including the *Harvard Law Review*, the *Columbia Law Review*, the *Virginia Law Review*, and the *California Law Review*. He has written for a more general audience in *Time, Slate, Salon, Wired*, CNN.com, *Forbes*, the *Boston Review*, and the *Chronicle of Higher Education*.

Katie Shilton, Assistant Professor, College of Information Studies, University of Maryland

Katie Shilton's research focuses on ethics and policy for the design of information technologies, systems, and collections. She leads the Ethics & Values in Design (EViD) Lab at the UMD iSchool, and is the director of the CASCI research center. Her work has been supported by a Google Faculty Award and several awards from the U.S. National Science Foundation, including an NSF CAREER award. She received a B.A. from Oberlin College, a Master of Library and Information Science from UCLA, and a Ph.D. in Information Studies from UCLA. Shilton currently teaches courses in information policy, information and technology ethics, and digital curation.

Christine Task, Senior Computer Scientist, Knexus Research Corporation

In Spring 2015, Christine earned her PhD in Computer Science from Purdue University, with dissertation focus in Privacy-preserving Data Mining and Social Network Analysis. Her research centers on developing privatization techniques for processing data sets that contain sensitive personal information, to ensure that aggregate patterns can be shared for others to analyze while the privacy of individuals remains provably protected. In April 2012, she presented "*A Practical Beginner's Guide to Differential Privacy*" at the CERIAS Seminar lecture series; It has since been viewed online over 2,000 times, used in graduate and undergraduate courses across the globe, and linked on the Google Online Security research blog. She currently develops privatization technology solutions for Knexus Research Corporation.

Omer Tene, Vice President of Research and Education, International Association of Privacy Professionals; Senior Fellow, Future of Privacy Forum

Omer Tene is Vice President of Research and Education at the International Association of Privacy Professionals. He is an Affiliate Scholar at the Stanford Center for Internet and Society; and a Senior Fellow at the Future of Privacy Forum.

Mark Van Hollebeke, Privacy and Education Specialist, Microsoft

Mark Van Hollebeke joined Microsoft in January 2012, leveraging his years of experience in applied ethics and deep pragmatic commitments to develop educational and cultural change programs centered on building employee buy-in for privacy and online safety practices. His privacy trainings reach over 100,000 unique Microsoft employees each year, and help employees integrate meaningful privacy controls into the products and services they ship. The ideas Van Hollebeke values most are the ones that have positive practical results—allowing us to enhance the meaning and beauty of our everyday lives experience.

John Verdi, Director of Privacy Initiatives, National Telecommunications and Information Administration, U.S. Department of Commerce

John Verdi is Director of Privacy Initiatives at the National Telecommunications and Information Administration (NTIA). NTIA, located within the US Department of Commerce, is the principal advisor to the President on telecommunications and information policy issues. Mr. Verdi's work focuses on digital privacy and security issues; he leads NTIA's privacy multi-stakeholder process. Recently, his work has touched on unmanned aircraft systems (UAS), facial recognition technology, and mobile apps. Prior to joining NTIA, Mr. Verdi was General Counsel for the Electronic Privacy Information Center, where he supervised the organization's litigation program, pursued federal lawsuits regarding privacy issues, and authored Supreme Court briefs. He is the co-editor of *Litigation Under the Federal Open Government Laws* (25th Edition). Mr. Verdi earned his J.D. from Harvard Law School in 2002 and his B.A. in Philosophy, Politics, and Law from SUNY-Binghamton in 1998.

Jessica Vitak, Assistant Professor, College of Information Studies; Affiliate Professor, Department of Communication, University of Maryland

Jessica Vitak holds a B.A. in Communication and Journalism from Elon University, a M.A. in Communication, Culture & Technology from Georgetown University, and a Ph.D. in Media & Information Studies from Michigan State. Her research evaluates the benefits and drawbacks of mediated communication by focusing on the role that social and technical affordances shape interactions online. She is currently engaged in several research projects around how individuals and families navigate privacy online and how privacy concerns influence disclosures. In 2015, she received an ADVANCE Seed Grant with two journalism professors to study online misogyny, and serves as the faculty mentor for a Future of

Information Alliance (FIA) seed grant evaluating solutions to mitigate cyberbullying. More information on her research and teaching can be found on her website, http://jessicavitak.com.

Hilary Wandall, Chief Privacy Officer, Merck & Co., Inc.

Hilary Wandall has driven the development and global adoption of a values-based privacy program supported by management accountability and quantitative measurement of risk, effectiveness, and continuous improvement. She has broad multi-disciplinary experience in HIV research, genetic and cellular toxicology, internet marketing, corporate law, ethics and compliance, and privacy and data protection. Her career in healthcare spans 20 years, including 19 years at Merck. Wandall is also actively engaged in a broad range of industry and pan-industry outreach and advocacy efforts to address evolving health information and privacy and data protection policy issues. She is a member of the Board of Directors of the International Association of Privacy Professionals, the Board of Directors of the International Pharmaceutical Privacy Consortium, for which she recently served as Chair, and the Future of Privacy Forum Advisory Board. She recently served on the OECD Privacy Experts Group responsible for reviewing the 1980 OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data.

Anne L. Washington, Assistant Professor in the Organization Development and Knowledge Management Program, George Mason University School of Public Policy

Anne L. Washington's research investigates socio-technical aspects of transparency initiatives and electronic government projects. In 2012, she was the first U.S. citizen to be invited as a fellow with the Peter Pribilla Foundation at the Leipzig Graduate School of Management and Technical University of Munich (TUM). Political informatics, or poli-Informatics, is her current three-year National Science Foundation (NSF) grant that brings big data principles to the study of government and politics. She is leading a group of colleagues in using open government data to build research capacity for data intensive research. The topic of the 2012-2015 project is using computational analysis to better understand the financial crisis. She completed a PhD from The George Washington University School of Business, where her primary field was Information Systems and Technology Management and her secondary field was Organization Behavior.

Alexandra Wood, Berkman Center Fellow, Harvard University

Alexandra Wood's research involves exploring new and existing legal and regulatory frameworks for data privacy and developing legal instruments to facilitate the sharing and use of research data while preserving privacy, transparency, and accountability. Previously, Alexandra served as a legal fellow assisting with the technology, telecommunications, and intellectual property portfolio of U.S. Senator Barbara Boxer. As a law student, she worked with the Center for Democracy & Technology and the Electronic Privacy Information Center on privacy projects addressing emerging electronic surveillance, facial recognition, and mobile payments technologies. She was also a 2010 Google Policy Fellow with the Future of Music Coalition. Alexandra holds a law degree from George Washington University Law School, a master's degree in public policy from the University of Southern California, and a bachelor's degree in economics from Reed College.

Dr. Heng Xu, Program Director, Secure and Trustworthy Cyberspace, BIGDATA & RIDIR, Directorate for Social, Behavioral, and Economic Sciences, National Science Foundation

Dr. Heng Xu is currently on a temporary rotation as a Program Director for several interdisciplinary research programs at the National Science Foundation (NSF). Much of her work at NSF focused on bringing the social, behavioral and economic sciences to studies of major challenges in Big Data and Cybersecurity

& Privacy. Dr. Xu joined NSF from the Pennsylvania State University through the IPA agreement. At Penn State, she is a tenured associate professor in the College of Information Sciences and Technology (aka the iSchool). Her current research focus is on the interplay between social and technological issues associated with privacy and security. She approaches privacy and security issues through a combination of empirical, theoretical, and technical research efforts. She has authored and co-authored over 100 research papers on information privacy, security management, human-computer interaction, and technology innovation adoption. Her work has been published in premier outlets across various fields such as Information Systems, Law, and Human-Computer Interaction, including MIS Quarterly, Information Systems Research, University of Pennsylvania Journal of Constitutional Law, Proceedings of the International World Wide Web Conference (WWW), Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI), and many others. She was a recipient of an NSF Career award (2010) and the endowed PNC Technologies Career Development Professorship (2010-2013).

Michael Zimmer, Associate Professor, School of Information Studies, University of Wisconsin-Milwaukee

Michael Zimmer, PhD, is a privacy and Internet ethics scholar. He also serves as Director of the Center for Information Policy Research. With a background in new media and Internet studies, the philosophy of technology, and information policy & ethics, Zimmer's research focuses on the ethical dimensions of new media and information technologies, with particular interest in online privacy, social media & Web 2.0, big data, and internet research ethics.

Appendix C: Accepted Workshop Papers

Beyond IRBs: Ethical Guidelines for Data Research

by Omer Tene and Jules Polonetsky

Non-data related ethical concerns are not unique to big R research. As they develop products, companies frequently test and experiment in ways unrelated to the collection and use of personal information. They A/B test products, experiment with new drugs and closely examine the performance of new services. Some of these activities are governed specifically by a range of regulatory agencies handling safety issues, including the Food and Drug Administration, Department of Transportation, Consumer Product Safety Commission, Consumer Financial Protection Bureau and more generally, the FTC. This article focuses specifically on issues related to data-driven research, which is an area where the notion of harm is still hotly debated and both benefit and risk are typically intangible.

Research Ethics in the Big Data Era: Addressing Conceptual Gaps for Researchers and IRBs

by Michael Zimmer

Attempts to fill the policy vacuums and clarify the conceptual gaps created in the wake of big data have already begun, stemming largely from concerned computer scientists, social scientists, and legal scholars. This essay helps push forward the growing discourse on the ethics of big data research by disclosing critical conceptual gaps that often hamper how researchers and IRBs think about privacy, personal information, consent, and harm in the context of big data. In doing so, this essay follows a disclosive approach to computer ethics research, uncovering ethical issues in big data analytics and research that have not yet gained sufficient attention, and making visible the conceptual gaps that have emerged. Through such attempts to address and clarify these conceptual gaps we can strive to ensure that ethical guidelines and policies are properly formulated to support the central concerns of research ethics and human subject protection amid the rise of big data research.

New Challenges for Research Ethics in the Digital Age

by Camille Nebeker, Cinnamon Bloss, and Nadir Weibel

We have rapidly entered an era where researchers collect data ‘on-thefly,’ in real-time and, subsequently can design meaningful, personalized and adaptive health interventions. The technologies that we focus on in this paper include devices and apps that enable data collection via Mobile Imaging, pervasive Sensing, Social-media and location Tracking (MISST) methods. The MISST acronym describes the broad range of devices worn, deployed, carried or implanted to monitor or measure an individual’s behavior, activity, location and assorted biological indicators (e.g., sweat, heart rate). In addition to measurement and monitoring, MISST devices are programmed to interact with the research participant or patient to promote, for example, increased exercise or adherence to a medication schedule.

The IRB Sledge-Hammer, Freedom and Big-Data

by Curtis Naser

The application of the regulatory model of IRBs to research using big-data, particularly in the for-profit sector, raises a host of issues. I want to highlight three areas of concern where the regulatory framework of IRBs is challenged by this research: the power of the IRB, privacy, informed consent and the obligation of subjects to participate in research.

The practice of ethical and regulatory review of research involving human subjects has its roots in revelations of some very harmful research that came to light in the early 1970s. The Tuskeegee Syphilis Study is perhaps the most notable, in which a cohort of African American men from Alabama were denied treatment for syphilis with simple antibiotics over a period of over 20 years, resulting in many untimely deaths and the transmission of syphilis to their wives and children. Other studies that came to light involved infecting mental patients with hepatitis and injecting radioactive agents, all without consent, much less informed consent.

Architecting Global Ethics Awareness in Transnational Research Programs

by John Murray

Traditionally, the ethical principles that guide scientific studies involving people are primarily intended to cover direct human–-centered research. However, in the modern online world, cyber–-centric research is inherently data–-centered in nature, and researchers frequently operate with limited awareness of the potential human risks and effects of their activities. Indeed, the nature of their work is such that any organizational oversight of their research may be absent.

Recently, a series of updates to the U.S. policies and regulations governing Institutional Review Boards have been proposed, which are likely to have a significant impact on the online research community. However, since online studies inherently cross the boundaries of multiple jurisdictions, there is now an even greater need for harmonizing ethics observance regulations and guidelines in a global context.

Classification Standards for Health Information: Ethical and Practical Approaches

by Craig Konnoth

Secondary research using health information is on the rise. Not all informational research presents equal burdens. Yet, there has been little commentary on the distinction between different kinds of informational research. This paper helps remedy this problem. In so doing, it sets out the first step towards a blueprint for IRBs and other actors who must decide what kinds of constraints to apply to different kinds of information.

I first briefly explain what constitutes secondary research of health information, and outline the problem. Second, I point to analogous contexts in which health information is categorized. Third, relying on Helen Nissenbaum's approach to privacy as contextual integrity, I argue for a "scoring" methodology that IRBs should use in determining information sensitivity.

Selected Issues Concerning the Ethical Use of Big Data Health Analytics

by Lieke Jetten and Stephen Sharon

Privacy advocates have spent the better part of a decade teaching people that their data is precious and that once it's online, it's online forever. As this message finally takes hold, and users have finally started to limit the data they share online, Big Data initiatives are asking users to freely give up their data with no direct or immediate benefit. Without transparency of Big Data practices, users may continue reducing the data they share, especially health data, because they don't understand the collective value of their data. It is in this environment that the review boards of today and tomorrow must operate.

Traditionally, internationally accepted frameworks have guided decision makers and health professionals to decide what they should and should not do when dealing with health data. This was necessary because health data has customarily warranted special protections. However, in the era of rapid technology advancement, previously accepted frameworks are no longer sufficient for three reasons. First, innovation is outpacing the frameworks themselves, many of which reflect the world of data collection decades ago. Second, health data is increasingly being collected outside of traditional healthcare settings. Third, data are then shared with third parties not only for research, but also for commercial gain.

Beyond IRBs: Designing Ethical Review Processes for Big Data Research

by Simson L. Garfinkel

Big Data Research is something new, a practice rooted in large-scale data collection that frequently combines aspects of research, experimentation, and entertainment. Yet another mismatch between the worlds of Big Data research and IRB system are the different regulatory regimes experienced by researchers in academia, where work with human subject data is highly regulated, and many corporate researchers, where there are frequently no regulations because the work is not funded by the federal government.

One tempting approach is to erect a fence between big data research and IRBs, and provide data scientists with a new, market–-based, voluntary regulatory process. While such an approach might be attractive to some, self-regulated data science research is likely to be even less successful than self–-regulated standards for data collection, advertising, and a variety of other privacy–-intrusive business practices. Self-regulation is rarely effective in protecting one party when there is a significant power imbalance in a relationship. After all, it was the failure of self–-regulation in the 1950s and 60s that led to the Belmont Report, the Common Rule, and the current system of regulation for research involving human subjects.

Usable Ethics: Practical Considerations for Responsibly Conducting Research with Social Trace Data

by Jana Diesner and Chieh-Li Chin

Over the last decade, research on the privacy of user information has shown that often a) ordinary users pay little attention to privacy policies and b) when considering policies, people have a hard time understanding their meaning and practical implications. Usable computational solutions to this problem have been developed. We observe a similar trend with respect to the ethics, norms and regulations for using public digital data at any scale; big and small. By this we mean that researchers may have little awareness of the

different types of regulations beyond IRBs that might apply to their work, and difficulties to fully comprehend and implement applicable rules. This article focuses on practical issues with properly using social trace data for research and proposes solutions. For the purpose of this article, we define publicly available social trace data as information about people interacting with a) other social agents (e.g. social networks data from Facebook and Twitter), b) pieces of information (e.g. product review sites and discussion forums), and c) infrastructures (e.g. people checking in to places, geolocation services), and natural language text data (e.g. the content of posts and tweets) that all can be collected without intervention or interacting with users (also called passive measurement).

Ethics Review Process as a Foundation for Ethical Thinking

by Chris Clifton

The Institutional Review Board process, while not perfect, has been fairly effective in balancing the progress of research and protection to human subjects. I see two key benefits of the Institutional Review Board process in ensuring ethical human subjects research. The obvious one is the review – getting outside eyes who are not heavily invested in the project and outcome to look for issues. The problem is (hopefully) not a lack of ethics among researchers, but that the excitement over the project and outcomes blinds researchers to the potential risks. The second benefit is forcing researchers to think about issues before beginning research, and giving them the tools to do so.

Emerging Ethics Norms in Social Media Research

by Katie Shilton

Defining ethical practices for research using data from digital and social media communities is an ongoing challenge. This paper argues that we should learn from practice: that researchers working with open and online datasets are converging around norms for responsible research practice that can help guide IRBs or alternative arrangements interested in regulating research ethics. It uses descriptive ethics to suggest normative ethics.

Just because a community has come to agreement around particular practices does not mean that these practices are right. Outside deliberation is still needed; researchers will likely never be entirely self-regulating. But growing consensus among researchers provides guidance as to what researchers feel to be reasonable practice; a first step for understanding responsible conduct of research.

This essay draws on qualitative interviews with digital and social media researchers (Shilton & Sayles, 2016), as well as a survey of 263 social science, information science, and computer science researchers who use online data (Vitak, Shilton, & Ashktorab, 2016). The interviews investigated the challenges researchers experienced when collecting, managing, and analyzing online data. Analysis of the interviews reveals a diverse set of ethical challenges that push at the boundaries of existing research ethics guidance. The interview data also describes existing practices for navigating ethical quandaries, and documents resources that help researchers meet ethical challenges. The analysis of the data points to opportunities for review boards and ethics researchers as well as new debates to undertake as a community.

Trusting Big Data Research

by Neil Richards and Woodrow Hartzog

Although it might puzzle or even infuriate data scientists, suspicion about big data is understandable. The concept doesn't seem promising to most people. It seems scary. This is partly because big data research is shrouded in mystery. People are unsure about organizations' motives and methods. What do companies think they know about us? Are they keeping their insights safe from hackers? Are they selling their insights to unscrupulous parties? Most importantly, do organizations use our personal information against us? Big data research will only overcome its suspicious reputation when people can trust it.

Some scholars and commentators have proposed review processes as an answer to big data's credibility problem. It is possible that a review process for big data research could provide the oversight to ensure the ethical use of data we've been hoping for, applying sensible procedural rules to regularize data science. But procedure alone isn't enough. In this essay, we argue that to truly protect data subjects, organizations must embrace the notion of trust when they use data about or to affect their human users, employees, or customers. Promoting meaningful trust will involves structuring procedures around affirmative, substantive obligations designed to ensure organizations act as proper stewards of the data with which they are entrusted. To overcome the failures of a compliance mentality, companies must vow to be Protective, Discreet, Honest, and above all, Loyal to data subjects. Such commitments backed up by laws will help ensure that companies are as vulnerable to us as we are to them. When we know we can trust those using big data, the concept might not seem so scary after all. We will disclose more and more accurate information in safe, sustainable ways. And we will all be better off.

No Encore for Encore? Ethical questions for web-based censorship measurement

by Arvind Narayanan and Bendert Zevenbergen

A pair of computer scientists recently developed a clever way to measure Internet filtering and censorship worldwide, including countries such as China and Iran. Their system, named Encore, does this by executing a snippet of code on the web browsers of people who visit certain web pages — without the consent of those individuals. It caused a minor furor over research ethics in the computer networking and Internet measurement research communities.

We analyze this conundrum through the lens of established ethical principles, but keeping in mind the peculiarities of Internet and big data research: its global reach, large scale, and automated nature. We also comment on the unusual model that computer scientists use for ethical oversight. We hope that the questions we raise will be useful for researchers facing similar dilemmas in their own work, as well as for students of research ethics, both in technical disciplines and in fields such as law and philosophy.

Big Data Sustainability – An Environmental Management Systems Analogy

by Dennis D. Hirsch and Jonathan H. King

At this formative moment of mass big data adoption, we can learn from environmental management practices developed to manage negative externalities of the industrial revolution. Today, organizations globally wrestle with how to extract valuable insights from diverse data sets without invading privacy, causing discrimination, harming their brand or otherwise undermining the sustainability of their big data projects. Leaders in these organizations are thus asking: What is the right management approach for

achieving big data's many benefits while minimizing its potential pitfalls? Leveraging our analogy, we propose in this paper that adapting an Environmental Management System or "EMS" is a good reference model for organizations to consider for managing their big data developments.

We support our proposal by first further examining the utility of the analogy between the externalities of the information revolution and the industrial revolution. We then summarize the evolution of traditional environmental management from a siloed, command and control structure to a more collaborative, environmental management system approach. Finally, we argue why an environmental management system provides a good reference case for big data management.

Towards a New Ethical and Regulatory Framework for Big Data Research

by Effy Vayena, Urs Gasser, Alexandra Wood, David R. O'Brien, and Micah Altman

Emerging large-scale data sources hold tremendous potential for new scientific research into human biology, behaviors, and relationships. At the same time, big data research presents privacy and ethical challenges that the current regulatory framework is ill-suited to address. In light of the immense value of large-scale research data, the central question moving forward is not whether such data should be made available for research, but rather how the benefits can be captured in a way that respects fundamental principles of ethics and privacy.

In response, this Essay outlines elements of a new ethical framework for big data research. It argues that oversight should aim to provide universal coverage of human subjects research, regardless of funding source, across all stages of the information lifecycle. New definitions and standards should be developed based on a modern understanding of privacy science and the expectations of research subjects. In addition, researchers and review boards should be encouraged to incorporate systematic risk-benefit assessments and new procedural and technological solutions from the wide range of interventions that are available. Finally, oversight mechanisms and the safeguards implemented should be tailored to the intended uses, benefits, threats, harms, and vulnerabilities associated with a specific research activity.

Development of a new ethical framework with these elements should be the product of a dynamic multistakeholder process that is designed to capture the latest scientific understanding of privacy, analytical methods, available safeguards, community and social norms, and best practices for research ethics as they evolve over time. Such a framework would support big data utilization and help harness the value of big data in a sustainable and trust-building manner.