

# Data&Society

Meeting Minutes: March 2, 2015 Advisors Telemeeting

## Attendees

Geoffrey C. Bowker, University of California, Irvine (co-PI)  
danah boyd, Data & Society / Microsoft Research (co-PI)  
Kate Crawford, Microsoft Research / New York University (co-PI)  
Rachelle Hollander, National Academy of Engineering  
Barbara Koenig, University of California, San Francisco  
Eric Meslin, Indiana University Center for Bioethics  
Alondra Nelson, Columbia University  
Paul Ohm, University of Colorado Law School  
Seeta Peña Gangadharan, New America / Data & Society

Jacob Metcalf, postdoc

*Council members unable to attend:*

Alessandro Acquisti, Heinz College, Carnegie Mellon University  
Mark Andrejevic, Pomona College  
Solon Barocas, Princeton University  
Edward Felten, Princeton University  
Alyssa Goodman, Harvard University  
Arvind Narayanan, Princeton University  
Helen Nissenbaum, New York University (co-PI)  
Frank Pasquale, University of Maryland  
Latanya Sweeney, Harvard University  
Sharon Traweek, University of California at Los Angeles  
Matt Zook, University of Kentucky

## 1. Introduction

This telemeeting focused on two primary topics: whether to pursue Institutional Review Boards (IRBs) as a plausible regulatory option for big data research and consideration of the “10 Simple Rules” series at PLoS as a venue for a collective publication.

## 2. IRBs

Multiple members reiterated complaints about IRBs as a regulatory apparatus. Historically, IRBs have been oriented toward the epistemology and risk profile of biomedical research. However, the foundational assumptions behind the Common Rule (which governs IRBs for research funded by the US Federal government) are a poor fit for many other research methods. The complaints from social scientists about inconsistent application of opaque rules and regulatory over-reach may signal what data scientists would experience under IRB regulation. There is some diversity in structure and purpose for IRBs, including private IRBs and community IRBs.

In contrast, Eric Meslin suggested that IRBs are currently an unexpectedly fertile area of debate and innovation. He stated that we shouldn't assume the historical problems with IRBs will always apply, and big data is a perfect opportunity to consider the governance structures, as they are being newly opened up and looked at thoroughly. He suggested that many of the most burdensome problems with IRBs are historical quirks rather than issues inherent to research ethics regulations. For example, cancer research cluster groups have struggled to deal with the peculiar needs of each institution for multi-sited research, so why not allow the multiple sites of research to use one IRB? The historical reason is that legal counsel didn't want to defer approval to another group, but that can be addressed relatively easily. It is still an open question as to whether big data should be treated as human subjects

research requiring review, but there is a movement, particularly in Europe and also in India and Africa, that wants to treat all big data as human-subjects research.

Geof Bowker pointed out that computer science has considered some things human-subjects and others not. There's not a lot of precedence of computer science practitioners treating their work as though it impacts human subjects. Barbara Koenig pointed to the genomic repository BDGap, where researchers are required to share data; prohibiting a requirement that individuals need to go to the IRB before using data that has already been cleared. She asked if there is a division we need to apply between "human" data and other data.

Kate Crawford said this is a big question that the Council should think about publicly, especially whether big data is human subjects research. Koenig replied that Health and Human Services (HHS) recently released an Advance Notice of Proposed Rulemaking announcing upcoming revisions to the Common Rule. She also pointed to a consensus study from the National Academies, "Strategies for Responsible Sharing of Clinical Data," which may be of interest. Rachelle Hollander pointed to another recent report from the National Research Council about the proposed changes to the regulation of social and behavioral sciences.

Paul Ohm stated that the question at hand is not just a matter of who thinks their data is human subjects research, but also clarifying what research should not be covered. There are many disciplines and types of data that do not properly fit the criteria for human subjects research but keep falling under its purview. The current attempt to revise the Common Rule is the latest in a long list of revision efforts, and anything the Council can do to figure this out collectively would be helpful. Value that the Council could add may be to disabuse funders from trying to tweak old systems rather than attempting something creative.

Jacob Metcalf asked what practical steps might be taken to pursue such public input. Members offered suggestions including approaching the HHS with offers to support their rule making process, addressing foundations and non-profits (such as Robert Wood Johnson Foundation and Creative Commons), addressing industry associations and corporate entities, addressing academic organizations that are already coping with these issues (such as American Society of Human Genetics), and highlighting possibilities for industry to engage in in-house review of business practices (as suggested by the Consumer Privacy Bill of Rights). Outputs from the Council on these issues would likely be seized upon by any number of these bodies and organizations, especially Chief Privacy Officers at tech companies.

Kate Crawford suggested that a major research question that could be addressed by Council members is what counts as a human subject for human-subjects research ethics and regulations in the context of big data research and practice.

danah boyd raised the matter of what sorts of public engagements are missing and should be cultivated. She said that there is not much evidence of cross-pollinating across domains. We can offer lessons learned from one field, but how can they transfer to another? How should we structure that effort?

Barbara Koenig responded by asking how we can accomplish public engagement when public means more than particularly situated audiences? In particular, how do we deal with the dynamic that it feels like anyone can get access to all data at any time if they have the right tools? It is hard to deal with big data ethics issues in a subtle manner when so much of the public's experience is with the extreme examples that foster a sense of having no control over what happens with one's data. Alondra Nelson noted that it is easy to have a sense of nihilism around data protection when you look at mainstream interventions, op-eds, and letters. We are often stuck in our own disciplinary silos, so we should go on tour and hold panels at major conferences. Paul Ohm agreed that we all spend too much time in disciplinary silos and that the Council is important because it is not about one domain alone. He noted that he suspects there are many publics we are not even aware of.

Rachelle Hollander pointed to the Association of Science and Technology Centers, the professional association for science museums. It is probably the case that there are nearby science museums wherever we are located, and they are all good at developing programs that interest a public audience.

Kate Crawford stated that the Council should pursue the maximum, broadest impact possible with minimum cost. She suggested that the discussions the Council is having could be turned into white papers that can have a life well beyond the group if done well. One example of a low-cost, high-impact output that was suggested is the "10 Simple Rules" series of short papers at PLoS.

Seeta Peña Gangadharan asked if low-lift efforts would include making ourselves available as consultants for writers outside of academia. danah boyd said that Data & Society could support this and asked what narrative the Council wants to impart in such engagements. Geof Bowker suggested that the Council should build a responsive role for big scandals so that we could produce something relevant easily. Kate Crawford noted that a group of bioethicists responded to the Facebook contagion study in *Science*. She said that developing more long-range thinking collectively would enable easier responses to shorter-term issues. Rachelle Hollander noted that there is a policy and research agency audience that needs a certain kind of information and engagement that is quite different from the more media-oriented audience. That kind of engagement could be better facilitated if we had space to share papers and reactions via the website.

*\* Funding for the Council for Big Data, Ethics, and Society was provided by the National Science Foundation (#IIS-1413864).*