# Breakout Discussion Notes: New Research

Data & Civil Rights
October 30, 2014 – Washington, D.C.
http://www.datacivilrights.org/

*This document was produced based on notes taken during New Research breakout session of the Data & Civil Rights conference. This document represents a general summary of the discussion that took place. Not all attendees were involved in every part of the conversation, nor does this document necessarily reflect the views and beliefs of individual attendees. For an overview of the breakout sessions, including a description of the questions participants were asked to consider, see:*
*http://www.datacivilrights.org/pubs/2014-1030/BreakoutOverview.pdf*

## Overview

In coming together to discuss new research, the group had an opportunity to (1) reflect on the conclusions from other sections, (2) re-iterate the common threads between different areas of concern, (3) imagine future possible actions, and (4) address what concerns they thought had not been included in the course of discussions. The result was therefore a wide-ranging discussion, taking input from all involved. However, as areas of concern were re-visited the common themes running through all of them became clearer. In the end, this session focused on what the group could begin to do to produce positive interventions. What are the most productive steps for researchers who examine data to take to further a variety of social causes across domains?

## Themes and Discussion Topics

There are aspects of the data-inflected landscape of public life that the group agreed require some form of intervention. There needs to be a greater awareness among the public about how the constellation of different data systems work together. Ideally, there would be a structural map of how and where data exist in these systems, but such a mapping is not likely to arise spontaneously; it will require deliberate work. Furthermore, the ambivalence between data technology's capacity for efficacy and capacity for harm (illustrated vividly in the case of health care provision) needs to be reconciled. How do we balance competing values, and to what extent should we prioritize privacy *over* other values?

In each of the spheres of concern discussed in other sessions, the group summarized the most pressing concerns:

- Education — what is meaningful consent and data ownership; what is useful educational data?
- Housing — housing data has a deep capacity for positive and negative impacts in a system that is remarkably hard to reverse engineer. In particular, this is exacerbated by the prevalence of very large housing companies, who are responsible for a range of decisions distinct from those even 20 years ago.
- Employment — how are algorithmic processes and different types of data brought to bear on hiring decisions?
- Criminal Justice — much of the useful data currently exists in limited forms.
- Health — public health authorities might have an obligation to collect even more data, but there

might be significant barriers to how they use and distribute that data, particularly sensitive data surrounding stigmatizing medical conditions like H.I.V. However, this is one area where privacy is a diminishing value compared to the overall goals of improving public health.

When discussing these different areas of concern, however, it became clear to the group that similar themes stretched across all of these different domains. Data is becoming integrated and interoperable, and therefore we need a strong basis for the analysis of power in relationship to what controls on data availability, accessibility, or usage exist in a Big Data environment. Many of the proceeding conversations worked on the ambivalence of good and bad effects from data technologies. In every case, however, good and bad had to be contextualized by *for whom* it was good or bad. This is why an analytical language of power (one beyond "good" and "bad") is required to advance the discourse on data. There is a need to better describe the structural character of commercial, scientific, governmental, and private individual uses of data, and data analytic capabilities. The ethical complexity of the good and bad effects from data technologies is further exacerbated by the technical complexity of data collection and analyses technologies. This includes non-human networks, machine learning networks, and robotics. This complexity is also rooted in the social infrastructure from which data technologies draw their data: different races/ethnicities are more accurately represented in different contexts. And new types of data (e.g. public Twitter data) are adding further dimensions to the conversation.

## Areas for Further Exploration

Finally, the group assembled what they thought were the gaps left that still needed attention. This included asking what individuals can do with their own data, and what role data literacy plays in issues of social justice. This led to a self-reflexive consideration: what can a community of researchers (the group participants and their larger network of collaborators) do to advance the status of data in society? Some concerns:

- What tasks can only be accomplished through collaboration?
- What are the consequences of studying non-decision makers?
- What happens when processes are opaque even to decision makers?
- What are the researchers own ethics on data use?
- What kind of framework is sufficiently effective to bring to Institutional Review Boards (IRBs), and beyond IRBs, what frameworks should inform best practices for users who are not obliged to go through IRB review?
- What kinds of authoritative footing are needed to intervene in such a fast-changing domain?
- Who *are* the people to talk to based on their power and influence and/or needs? What *are* the already existing coalitions?
- What are models and best practices that can structure future research?
- What is the effect of the wide divide between those studying the implications of data, and those using it to maximize individual applications?
- How can researchers discover/reveal/evaluate the causal processes behind the correlations provided by data analysis?
- What is the history of the current moment of data technology? How did we transition from the periods of technological development and initial legislation, to now when legal applications are made to previously un-imagined scenarios?