

Breakout Discussion Notes: Policy Development

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

This document was produced based on notes taken during the Policy Development 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

This breakout session raised a wide variety of questions and explored issues on the themes of Big Data policy in relation to justice and discrimination, transparency and access, and building common vocabulary for the analysis of these themes across various contexts. The group focused on identifying issues rather than discussing possible solutions to particular problems.

Themes and Discussion Topics

Definitions and A Common Language

The group conversation centered initially on the tensions between Big Data considered holistically, and Big Data in particular contexts, such as health, education, finance, marketing, etc. Early in the session, some participants stressed that there should be a common language and set of issues across the broad data community that could be applied to its respective groups. Transparency and accuracy were among the core issues that emerged. Yet, other participants suggested that the banner of Big Data needs to be reassessed and considered in specific contexts (such as housing, education, finance, etc.). There may be a need to think about those particular cases in isolation rather than under the aegis of “Big Data” applied broadly. It was also pointed out that there is a need to identify the data “haves” and “have nots” with respect to collection and use in order to illuminate instances of discrimination.

Some suggested creating a hierarchy of benefits and harms that Big Data Civil Rights groups can prioritize. Others asked, how do we identify Big Data problems? And yet, what are the biggest concerns about processing and using algorithms? The important question here seems to be, where do the respective goals of Big Data and Civil Rights align and where do they conflict?

Justice and Discrimination/Bias

There was also some discussion about the role of data to respond to discrimination. How and when do we need data in order to respond to discrimination? It was stressed that in the healthcare context, app developers need guidance about the laws and about discrimination issues.

There was also a push to map out occasions where implicit bias infects data and to quantify those points. This goal is made challenging by the fact that it is difficult to determine whether data is predictive simply because it's accurate or because it is using biased data. In other words, statistical bias is one thing, but is probably easier to identify than issues of overly-specific inferences.

Transparency, Accuracy, Access, and Revision (or Resistance?)

One big issue that surfaced during the conversation was that of transparency. Generally, participants felt that transparency needs to be increased across the board. Yet, transparency alone is not enough because it needs to be coupled with actual access and needs to be subject to revision. Reforms should, therefore focus on access, as well. Individuals need to know when data has been collected about them, to be able to access it, and to have the power to correct any mistakes or inaccuracies presented. People need to be given the tools to be able to respond and revise transparent public disclosure; otherwise, that “transparency” can slip into an invasion of privacy or the publication of misleading information.

Transparency about how the algorithms are applied needs to be improved, as well. Some suggested that the Federal government, when subcontracting with institutions that capitalize on information, could devise a means for individuals to be given the opportunity to access and correct the information that the subcontractors have. (For example, the Federal government doesn't have information on Pell grantees, but it does subcontract with Pierson, which has a lot of information these grantees). Consumers also need to be educated about potential harms and benefits of data collection across disabilities, race, language, and other categories.

Areas for Further Exploration

Additional knowledge needed in order to formulate policy on Big Data

One of the biggest challenges will be to advance any specific programs in these areas under the banner of ‘big data’, primarily because both the technologies and the language/definitions that circulate around this phenomenon are not refined enough to identify that ‘big data’ is the root cause of a singular or thematic series of harms that cut across multiple domains, or is available as a remedy to problems that occur across sectors. In other words, one big data solution is not necessarily a proscription across domains, or issues. However, in order to devise a set of standards, or conversational and programmatic tools, for framing and discussing issues around the big data phenomenon, the participants suggested three thematic areas for further exploration, namely: identifying and protecting best practices; transparency; and regulation. In other words, what frameworks could produce a baseline for approaching issues that do come up in common across sectors, or for identifying discontinuities between them?

Identifying and Protecting Best Practices

How can safe harbors for best practices be created wherein transparency is increased across the board? How can one get more insight into the algorithms, what inputs are used and how are the algorithms applied? Where might we find examples of good analytics? Who is conducting empirical analyses that practitioners and policy-makers can refer to?

Transparency

What requirements exist about disclosing when data is collected and used to make decisions in different arenas? How can we identify when big data practices are functioning in the background, when there is an algorithmic function, and when there is collection?

And how can we identify where data is collected but not analyzed? Can we identify when data is not being collected? What are the goals and incentives of the people using big data analytics? Further, how can individuals be given the tools to challenge inaccurate or unfair decisions and information collected about them?

Regulation

What are the current regulatory frameworks and what agencies that are administering those laws? Second, are those agencies active in those areas and looking for new methods of regulating big data practices? Third, what internal assessments are occurring, and where do those internal assessments line up with regulatory frameworks?