

Workshop Discussion Notes: Housing

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

This document was produced based on notes taken during the Housing workshop 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. All workshop participants received workshop materials prior to the event to spark discussion. The primer can be found at: <http://www.datacivilrights.org/pubs/2014-1030/Housing.pdf>

Overview

The housing workshop focused on the use of data in housing discrimination. The discussion centered on the civil rights implications of new online housing platforms, new practices by lending institutions, and emerging issues in the rental market. There was a large discussion between housing advocates and technical experts on new methods of data collection and analysis. A recurrent theme was the need for transparency and accountability in data profiling and data collection, particularly given the increased capacity for private companies to collect, aggregate, and sell unprecedented quantities of personal data; and their ability to use it to profile and predict behavior in ways that could be harmful. Data-derived predictions necessarily incorporate past history of discrimination, and thereby help to reproduce its impact.

The motives for data collection and analyses, and the transparency with which they are presented to the public, were key and recurrent concerns. There was a tension between the ways that data can be used, particularly for profiling purposes, that can benefit parties with good and bad intentions, and how we sort out acceptable data use in the context of these intentions. Outside of the discussion on motivations and intentions, third-party datasets were asserted to be generally unreliable, at a practical level. However, the promise of big data analyses to create strong predictive values has been accepted by many industries, despite concerns about data quality.

In addition to the themes described below, the discussion explored the following questions: (1) Will our new digital environment mitigate discrimination, or will we simply transport segregation into this new space? (2) How can big data sets be leveraged to identify new forms of discrimination against particular groups? (3) Can we identify data gathering and analysis practices in our new digital environment that are consistent with civil rights?

Themes and Discussion Topics

Data: Reinforcing Discrimination or Reversing It?

At the beginning of the workshop, several participants suggested an examination of the positive uses of new datasets. Participants wanted to balance the critical view that big data may lead to further data-driven discrimination with an examination of how big data could be used to identify

discriminatory patterns. They suggested that just as data plays an important role in monitoring the practices of home mortgage lenders, new datasets could help identify emerging forms of discrimination online and offline.

Stepping back to understand positive and negative uses, one participant described how data can generally be used for two purposes: (1) descriptively, to research and understand the current environment or playing field, and (2) to reduce information asymmetries that can exacerbate inequality. Descriptive data could be used to demonstrate the potential discriminatory harms of using criminal background checks in employment and housing decisions. In terms of information asymmetries, we could imagine that with the availability of online services and ubiquitous information about products and prices, the problems of geographic-based pricing disparities could be reduced.

In response, participants noted that as companies seek to maximize their profits and collect more personalized information, we know that some of the data points will correlate with existing patterns of segregation, and will therefore, reproduce patterns of disadvantage. Additionally, reducing of information asymmetries requires that people have equal access to information and technology, which is not the case. Otherwise, the current information asymmetries will only grow. The group also made references to differences in capability to understand/make sense of this information. Increasing supply does not automatically reduce asymmetries.

Some participants suggested that the motivations of the data-collector and data-analyzer are important to understanding potential harms. Advocates succeeded in requiring mortgage lenders to report on their loans, and that data has been used to identify discrimination against protected classes. However, when private companies are collecting data there is generally an interest in leveraging that data for profit. Participants noted that in each specific case, one can ask if the data mining is for money or to gain greater social understanding. Later in the discussion, one advocate voiced concern that advocacy groups themselves use profiling and market segmentation to reach targeted groups (e.g. in mailing out advocacy materials). Participants agreed that each action must be evaluated by its consequences, and possible disparate impact.

Turning to the question of profiling, one participant noted that in many cases race is not explicitly used by companies to target people, but instead, companies use proxies. They noted that companies examine cookies to determine if a person has visited Telemundo, for example. With that information, the company can then infer the ethnicity of the person. Or for example, a company might analyze who shopped at Home Depot in order to guess gender, or Toys-R-Us site visits to identify people with children. This level of profiling represents a new world of challenges for civil rights.

Access and Transparency

As a metric for identifying good and bad data practices, participants discussed access and transparency as important characteristics. For example, the Census is transparent about their data collection methods and makes their data available publicly. Third-party datasets rarely provide this level of access and transparency, and so it is difficult to know whether their data collection and analysis practices might violate civil rights. It was noted that while the Census has tended to undercount communities of color, they provided those error rates to the public and worked to address those issues transparently.

In the workshop, the question over transparency stood out as a distressing and acute concern. Particularly with big data derived online, participants raised the concern that we do not know what is collected, how it is collected and who has access to it. This is a key difference between the data collected by agencies like Census, which goes through a careful review process, and the data collected by private companies.

Data Vendors and Data Quality

Participants that work with third-party datasets for demographic information explained that in their experience, the data was unreliable. One participant noted that even retailers question the validity of many third-party datasets. The overarching concern was that these less reliable datasets are most error-prone in communities of color and low-income communities. While industries may also have a good sense that their data is unreliable, this quality may be irrelevant; at the corporate level, what matters is that the data are "good enough" to provide a short term competitive advantage.

With these errors in mind, participants wondered if new big data offers real predictive value, or instead, if these companies were selling "magic spells." Participants asked if there were empirical studies demonstrating that the data was predictive or the data-driven campaigns effective? One participant argued that industry, at least, was convinced of these new data-driven campaigns and that data-mining algorithms were designed to test their own efficacy in many cases. However, this again might indicate that an algorithm is predictive enough to improve sales, rather than to meet imagined standards of accuracy or precision.

Defining Big Data

Participants moved to a discussion about how big data might be different from other, more familiar datasets. One participant suggested that big data has two different analytic methods: machine learning, which builds new predictive models with each new piece of data; and, inference, which seeks to understand how two things are connected.

Participants pointed to a distinction between big datasets that do not identify a particular person, but are used for predictions about behavior, and, privacy-invasive individual-level datasets. Both pose civil rights challenges. For decades, civil rights advocates have demonstrated how patterns of segregation skew aggregated geographic data, which in turn reinforce patterns of segregation. The civil rights community must now turn to the current trend of companies collecting mass individual data. Several participants noted that this new individual data continues to be impacted by the footprint of existing patterns of segregation, as many characteristics tend to be correlated. As noted above, participants pointed to the fact that so little is known about these big data sets. The public does not know what is collected, how it is collected, or how it will be used.

Methods for Detecting Discrimination Online

In order to understand how data is used online and how it could be used to detect discrimination, one participant suggested that big data offers the opportunity to conduct true random experiments to test discrimination claims. Participants were interested in understanding what would be possible and how that could be done.

In order to test for discrimination online, some participants wondered if it was possible to know, for example, what products and price marketing an older white man receives versus a young black man. Other participants responded that automated, large-scale tests about differential treatment on the web are possible. However, the vast number of factors involved in online advertising make it difficult to tell why a person would be profiled in a particular manner. The discussion pointed to the need to do more analysis on online profiling to understand if and how it is having negative consequences for particular communities.

New Online Housing Platforms

Participants discussed the potential impacts of new online housing platforms (e.g. Trulia, Airbnb, Couchsurfing, HomeAway, Craigslist, etc.). Overall, there was general concern that civil rights lessons learned about the traditional housing market were not being transferred to these new platforms. For instance, the group discussed one study¹ that suggests that users of Airbnb (and, presumably other similar platforms) are discriminated against based on their skin color, as made evident in their profile photos, which did not surprise, but did concern, fair housing advocates.

On the other hand, a few participants saw the potential for these new platforms to address longstanding patterns of discrimination. For example, the explicit rating and implicit ranking (search order results) of buyers and sellers, which is commonplace on new platforms, provides additional transparency that could help balance the traditional power dynamic.

A participant noted that non-traditional sites may be able to help break down existing social barriers. For example, a person using Airbnb might choose to stay in neighborhood they may not otherwise stay in, or have an experience with a person from another cultural background that they would otherwise not have had. Alternatively, it's possible for landlords "test" potential tenants using Airbnb's longer-term rental option, which could be an option for potential tenants whom landlords have valid reasons to turn away. However, many participants were skeptical that patterns of racism and discrimination online would not simply mirror the offline realities.

There was not sufficient time in the workshop to discuss recent concerns that real estate sites promote self-segregation, leading prospective buyers or renters to particular neighborhood based on factors that correlate strongly with race. A participant in the industry noted that these sites use a rich array of private and public datasets to profile neighborhoods. These profiles may undermine fair housing laws that are in place to prevent steering of customers into particular neighborhoods by licensed brokers. It will be challenging to address the extent to which individual consumers are making use of these geo-demographic websites to do their own assessments of neighborhoods.

Banks and Profiling: Changes Since the Subprime Mortgage Crisis

One participant raised a question about how banks are targeting new and existing customers since the Sub-prime Mortgage crisis. Particularly since the crisis, banks have re-focused on targeting new products towards existing customers, no longer reaching into "emerging" markets.

¹ Benjamin Edelman and Michael Luca, "Digital Discrimination: The Case of Airbnb.com," HBS Working Paper No. 14-054, January 10, 2014, http://www.hbs.edu/faculty/Publication%20Files/14-054_e3c04a43-c0cf-4ed8-91bf-cb0ea4ba59c6.pdf, accessed November 2014.

Banks are collecting ("hunting") huge quantities of data about where current clients are keeping their money outside of the bank, in hopes of luring more of their clients' money to the bank.

Participants noted that the pendulum has swung back from subprime lending to redlining. Financial institutions are de-banking lower-end customers and prioritizing their traditional customers, who tend to be higher-income and white. One participant asked that if banks were collecting so much data, why had they not predicted the subprime crisis?

Discrimination in the Rental Market

The discussion turned toward the rental market. Participants acknowledged that more work needs to be done to understand and fight discrimination in the rental market. They raised the need to understand new patterns of rental discrimination and fine-grained practices, such as how people are treated when the need repairs, or when they ask to move units.

One participant explained that large rental companies have an increasing share of the housing stock. As one would expect, those large companies collect data on their renters. However, there is much less transparency than found with mortgage lenders, who are more regulated. For example, participants asked how a renter would know why their rental application was declined? Or, what rent you will be charged? There was an open question about how data brokers are contributing to rental company analysis. What datasets are rental companies using outside of credit scores? What data is legal to use?

One participant noted that rental companies keep evictions lists. There was concern that as more rental properties are being aggregated by large companies, more people will be excluded from traditional rentals and pushed into a predatory rental market. It is unclear what redress individuals have around evictions lists and how they can be removed from such lists. This discussion raised the question of, how have changes in industry structure affected the ability of local organizations to influence corporate practices?

Areas for Further Exploration

The workshop raised several areas for further investigation. First, the civil rights community needs to continue to investigate how new data sets and data tools can help detect discrimination online and offline. Second, there is a need to investigate the civil rights consequences of new trends, such as new methods people use to find housing, and changes to the rental and mortgage markets following the Sub-prime Mortgage Crisis. One participant suggested the re-establishment of something like a United States Office of Technology Assessment that would be responsible for examining the potential harms of algorithms, data analysis methods, data collection schemes and uses of data for profiling or marketing. Finally, throughout many of the conversations, participants raised the issue of discrimination against the poor, and the need to focus on socioeconomic discrimination. One participant noted that the poor should be included as a legally protected class.