

2017

## Affordable Housing Law and Policy in an Era of Big Data

Nestor M. Davidson

Follow this and additional works at: <https://ir.lawnet.fordham.edu/ulj>

---

### Recommended Citation

Nestor M. Davidson, *Affordable Housing Law and Policy in an Era of Big Data*, 44 Fordham Urb. L.J. 277 (2017).  
Available at: <https://ir.lawnet.fordham.edu/ulj/vol44/iss2/2>

This Article is brought to you for free and open access by FLASH: The Fordham Law Archive of Scholarship and History. It has been accepted for inclusion in Fordham Urban Law Journal by an authorized editor of FLASH: The Fordham Law Archive of Scholarship and History. For more information, please contact [tmelnick@law.fordham.edu](mailto:tmelnick@law.fordham.edu).

# AFFORDABLE HOUSING LAW AND POLICY IN AN ERA OF BIG DATA

*Nestor M. Davidson\**

Introduction .....	277
I. Why Might Big Data Matter to Affordable Housing Policy? .....	281
A. Big Data and Policy .....	281
B. From Outputs to Outcomes in Affordable Housing .....	284
II. Big Data in Affordable Housing Law and Policy .....	287
A. Some Examples of Big Data’s Potential in Affordable Housing .....	288
1. Siting Decisions, Mobility, and Neighborhood Effects .....	288
2. Housing Portfolio Management .....	290
a. Subsidy Targeting and Market Conditions .....	290
b. Enforcement and Housing Quality .....	291
c. Unifying the Subsidized and Unsubsidized Housing Portfolio .....	293
3. Resident Relations and Services .....	293
B. Synthesis and Reflections on the Role of Law .....	295
III. Big Data’s Dark Side: Caveats and (Some) Responses .....	297
Conclusion .....	300

## INTRODUCTION

Every year, federal, state, and local governments invest more than \$50 billion to provide housing for people who cannot otherwise afford shelter.<sup>1</sup>

---

\* Professor and Associate Dean for Academic Affairs, Fordham University School of Law. This Article was prepared for the 2017 AALS Annual Conference and received helpful feedback from participants in workshops at the 2016 Property Works in Progress Conference and the Tulane Property Roundtable. I wish to thank Eva Schneider for her excellent research assistance, Clare Huntington for her key insights, and Raphael Bostic, well, just because.

1. See CONG. BUDGET OFF., FEDERAL HOUSING ASSISTANCE FOR LOW-INCOME HOUSEHOLDS (2015), <https://www.cbo.gov/publication/50782> [<https://perma.cc/MBG9-NRC8>]; Rachel Berquist et al., *State-Funded Housing Assistance Programs* (2014), <http://www.tacinc.org/media/43566/State%20Funded%20Housing%20Assistance%20Report.pdf> [<https://perma.cc/3MNQ-PFX7>].

In addition to this housing assistance, policymakers also make a variety of choices that impact the landscape of affordable housing, including in zoning, infrastructure, housing finance more broadly, and in a myriad of other policy domains.<sup>2</sup> These policies can make a profound difference for the millions of individuals and families helped, but are too often undertaken with only the vaguest, visceral sense of their consequences beyond the bare facts of putting roofs over people's heads.

However, affordable housing policy is beginning to experience a shift in perspective. To the extent that policymakers have collected data on impact, the focus traditionally has been primarily on outputs. These measures included the number of units built through a given investment, the number of people served under a given program, the number of construction or property management jobs created, and the like. But outputs are not always—indeed, not often—the same as *outcomes*, the actual short- and long-term consequences of policy interventions for those served by affordable housing programs and the communities at issue.<sup>3</sup>

In recent years, researchers and policymakers have begun to evaluate the results of policy interventions for people in subsidized housing on measures such as income, educational achievement, physical and mental health, and even subjective wellbeing.<sup>4</sup> Rather than merely track whether people have housing at a given level of affordability, this new focus understands that housing is a platform for a variety of life outcomes and that housing policy choices can meaningfully impact the arc of those outcomes.

This emphasis on outcome measures reflects a broader embrace of the use of data for decision making by managers and policymakers across the private and public sectors.<sup>5</sup> The ability to collect data in a more rigorous

---

2. In zoning for example, whether and where to allow multifamily housing can directly shape the availability of affordable housing in a community; whether to invest in transit can make the difference for the viability of affordable housing in many metropolitan areas; and in housing finance, everything from Federal Housing Administration insurance to the oversight of Fannie Mae and Freddie Mac shape the housing market.

3. As explored in Section I.B, *infra*, outputs can be a heuristically efficient proxy for certain basic outcomes. Thus, units provided give some rough sense of the impact of an investment, even if nothing is known about the conditions under which those units are provided, the location of those units, the resources available to people living in those units, or anything else. But those proxies have obvious limitations.

4. See *infra* Section I.B.

5. See generally Kristina McElheran & Erik Brynjolfsson, *The Rise of Data-Driven Decision Making is Real but Uneven*, HARV. BUS. REV. (Feb. 3, 2016), <https://hbr.org/2016/02/the-rise-of-data-driven-decision-making-is-real-but-uneven> [https://perma.cc/W86Q-UAF8] (exploring the range of sectors embracing data-driven decision making). The shift toward data is perhaps best known in popular culture through Michael Lewis' book *Moneyball*, which recounts how the Oakland A's found hidden value in players otherwise

and systematic way and the development of tools to make that information actionable—particularly to make transparent patterns that would otherwise be opaque—is beginning to change a range of decisional processes. This shift can be seen in everything from how professional baseball teams select players to how Facebook, Google, and other companies target their advertising to how investors seek value.<sup>6</sup> And data increasingly means “big data,” an admittedly fuzzy (and arguably hackneyed) concept that roughly refers to the use of relatively large data sets, often aggregated across previously disconnected areas, mined for the predictive value of underlying patterns and trends.<sup>7</sup>

Although much has been written about data-driven policymaking, the specific role of the legal system in improving program design and implementation deserves deeper exploration.<sup>8</sup> Across a range of affordable housing examples, explored below,<sup>9</sup> new data tools are starting to sharpen policy, but also creating a positive feedback loop in which agencies provide data to grantees to shape how they implement policy, gather more information about outcomes, and then share all of that information with other regulators or advocates to help advance other legal mandates, notably around enforcement and private rights of action. The confluence then between emerging analytic tools and a deepening understanding of the connection between inputs and outcomes makes affordable housing a particularly fruitful policy arena in which to explore law’s potential to generate as well as facilitate the deployment of data to improve policy.<sup>10</sup>

New data analytic tools are by no means a panacea. Incorporating big data into affordable housing law and policy raises serious concerns that are well rehearsed in the literature but worth reiterating in this context. On a structural level, data quality and integrity poses a basic challenge, as does the perennial risk that policies driven by what can be measured, such as bricks-and-mortar metrics, will inherently misdirect resources away from

---

underrated by traditional scouting methods. See MICHAEL LEWIS, *MONEYBALL: THE ART OF WINNING AN UNFAIR GAME* (2003).

6. McElheran & Brynjolfsson, *supra* note 5.

7. See James Manyika et al., *Big Data: The Next Frontier for Innovation, Competition, and Productivity*, MCKINSEY GLOBAL INST. 1 (2011), <http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation> [<https://perma.cc/3NTA-4UMH>].

8. Much of the legal literature on big data tends to focus on legal constraints on the misuse of emerging analytic tools, with a particular emphasis on privacy and threats to civil liberties. See Jonas Lerman, *Big Data and Its Exclusions*, 66 *STAN. L. REV. ONLINE* 55, 56 (2013).

9. See *infra* Section II.A.

10. A similar argument can be made in any number of other policy domains and some—such as health care—are much further along in terms of the centrality of data to decision making.

goals that are less quantifiable, such as policies that reflect the felt experience of people living in communities of concentrated poverty. Legal scholars have also raised important concerns about privacy and data security. And perhaps most importantly, there is a genuine risk that the very people whose lives should be at the center of affordable housing policy will be even more marginalized than they currently are when measurability is privileged over meaning.

There are potential answers to each of these concerns. Data must be taken in context and with appropriate skepticism; data should be anonymized and aggregated to the extent possible; and policymakers must find ways to be sensitive to the dignitary harms that attend quantification. But an alternative that fails to measure impact is hardly palatable and the potential benefits of enhancing the ability of policymakers to understand the consequences of their actions can actually advance the dignitary interests of those served by affordable housing policy.

This all may seem somewhat technical and dry—rarely has the heart fluttered for the ephemeral value of, well, analytics. Those motivated to engage in affordable housing—or other areas of poverty law and policy—by compassion may recoil at discussions of number crunching. That is entirely understandable, but data and compassion need not stand in opposition. For policy to be effective, we need both.

The Article is organized as follows. Part I describes big data’s potential for policy making and the emerging shift in affordable housing policy from outputs to understanding broader measures of outcomes. Part II then looks at several areas where data analytics are either currently changing approaches to affordable housing policy or where there is particularly strong potential for such a shift—by no means an exhaustive list, but illustrative. From these examples, the Part then synthesizes what this reveals about the role of law in generating data and deploying data. Finally, Part III highlights reasons for caution and some responses.

Before turning to the substance of this Article, I want to break the authorial fourth wall for a moment, to address a question about our current context. This Article grew in part out of my experience working in the U.S. Department of Housing and Urban Development (“HUD”) General Counsel’s office early in the first Obama Administration.<sup>11</sup> There, I worked with the team that developed the agency’s affirmatively furthering fair housing (“AFFH”) rules. I was also exposed to the agency’s work, discussed below, to embrace data-driven decision-making more generally.<sup>12</sup>

---

11. It is not just rote to include a reminder that nothing in this Article reflects the official views of HUD, nor relies on non-public information.

12. See *infra* text accompanying notes 35-41.

Since beginning this Article, however, the election of Donald Trump has injected uncertainty into affordable and fair housing policy, especially with Dr. Ben Carson's appointment as Secretary of Housing and Urban Development.<sup>13</sup> To the extent that Dr. Carson has a record with respect to HUD's mission, it consists primarily of voicing skepticism about the ability of the government to address segregation,<sup>14</sup> although Dr. Carson in his confirmation hearing seemed more open to the work of the agency.<sup>15</sup> This, together with the Trump Administration's general skepticism about regulation, might cast some of what this Article argues in doubt, although it is certainly too early to know. But it is not clear that data will be less important even if HUD moves away from promoting integration, economic opportunity, deconcentration of poverty, and other goals broader than simply subsidizing housing. Moreover, the longer-term trend will still favor data-driven decision making, state and local efforts will continue, and seeds planted in recent years may flower in other domains.

### I. WHY MIGHT BIG DATA MATTER TO AFFORDABLE HOUSING POLICY?

This section provides a brief primer on big data's potential for policy improvement. It then turns to the start of the embrace of outcome-driven policymaking in affordable housing. Big data and outcome-driven policymaking, of course, are not synonymous—the former has many applications and the latter does not require novel data tools—but their intersection raises intriguing possibilities.

#### A. Big Data and Policy

To state the obvious, policymakers have long relied on data to make decisions. Indeed, categorization, record-keeping, sorting, and similar bureaucratic informational management tools have so long been inherent to the administrative state that we rarely pause to note the ubiquity of the phenomenon.<sup>16</sup> In this sense, modern policy has always been data-driven,<sup>17</sup>

---

13. See Trip Gabriel, *Trump Chooses Carson to Lead HUD*, N.Y. TIMES (Dec. 6, 2016), <https://www.nytimes.com/2016/12/05/us/politics/ben-carson-housing-urban-development-trump.html> [https://perma.cc/676D-SJRG].

14. See Ben S. Carson, *Experimenting with Failed Socialism Again*, WASH. TIMES (July 23, 2015), <http://www.washingtontimes.com/news/2015/jul/23/ben-carson-obamas-housing-rules-try-to-accomplish/> [https://perma.cc/9AYV-GN5Y].

15. See Matt Flegenheimer & Yamiche Alcindorjan, *Ben Carson Urges Ending Reliance on Welfare in Bid to Be Housing Chief*, N.Y. TIMES (Jan. 12, 2017), <https://www.nytimes.com/2017/01/12/us/politics/ben-carsons-hud-housing-nominee-hearing.html> [https://perma.cc/2NMZ-UEDM].

16. See, e.g., Kristin A. Collins, *Administering Marriage: Marriage-Based Entitlements, Bureaucracy, and the Legal Construction of the Family*, 62 VAND. L. REV. 1085 (2009) (discussing the complex federal bureaucracy that developed to administer

and a broad array of policymakers have embraced the idea of evidence-based interventions.<sup>18</sup>

What is changing today is the availability of significantly more data and more powerful computing capabilities.<sup>19</sup> Big data is one of those clichéd terms that is hard to pin down precisely, and it is not necessary to do so here. Generally, though, the term encompasses a set of related phenomena.<sup>20</sup> First, big data refers to the capacity to collect and aggregate massive datasets and similar information—this is what is “big” about the phenomenon, as opposed to, well, just data.<sup>21</sup> Before I started research for this Article, I had no idea what an exabyte was (turns out to be one billion gigabytes and a gigabyte is about the amount of data it takes to digitally store a feature-length movie<sup>22</sup>), but the world is apparently currently generating two and a half times that much data every day.<sup>23</sup>

Second, big data also often involves the aggregation of information across heterogeneous sources. Billions of Twitter posts can reveal certain information, but Twitter posts correlated to other social media inputs adds new insights. Social media posts correlated to seemingly disconnected data

military pensions in the early nineteenth century); *see also* Uri Friedman, *Anthropology of an Idea: Big Data*, FOREIGN POL’Y (Nov. 2012), at 30 (noting that the advent of data analysis in policy can be traced to the use of punch-card technology to analyze the 1890 Census in a single year, rather than the eight years the task had previously taken).

17. *See* Meg Leta Ambrose, *Lessons from the Avalanche of Numbers: Big Data in Historical Perspective*, 11 I/S: J. L. & POL’Y FOR INFO. SOC’Y 201 (2015).

18. *See, e.g.*, Gordon L. Berlin & Rekha Balu, *Evidence at the Crossroads Pt. 3: Research-Practice Partnerships, the Future of the Evidence Movement*, WILLIAM T. GRANT FOUND. (Nov. 10, 2015), <http://wtgrantfoundation.org/evidence-at-the-crossroads-pt-3-research-practice-partnerships-the-future-of-the-evidence-movement> [<https://perma.cc/EE33-C8PU>].

19. *See generally* Viktor Mayer-Schönberger & Kenneth Cukier, *Big Data: A Revolution that Will Transform How We Live, Work, and Think* (2013); *see also* Jonathan Shaw, *Why “Big Data” is a Big Deal*, HARV. MAG. (Mar. - Apr. 2014), at 30.

20. These are shorthand descriptions of complex phenomena, to be sure. The computer-science literature on big data spells out these attributes in more detail. *See, e.g.*, Doug Laney, *3D Data Management: Controlling Data Volume, Velocity, and Variety*, APPLICATION DELIVERY STRATEGIES, META GROUP INC. (Feb. 2001), <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>, (coining the “3Vs,” which have become the mainstream definition of big data). However, the point here is to offer some general observations about what is most relevant for affordable housing law and policy.

21. Mayer-Schönberger & Cukier, *supra* note 19, at 8-9 (“The amount of stored information grows four times faster than the world economy, while the processing power of computers grows nine times faster.”).

22. *Id.* at 8.

23. *See, e.g.*, Matthew Wall, *Big Data: Are You Ready for Blast-off?*, BBC NEWS (Mar. 4, 2014), <http://www.bbc.com/news/business-26383058> [<https://perma.cc/9WQK-7Q2F>] (explaining that in 2012, 2.5 billion gigabytes of data were produced on a daily basis).

about, say, where people live and their spending habits may reveal even deeper patterns.

Third, big data generally refers to a set of analytic tools that can mine and activate that data. These new capabilities are facilitated not only by the digital footprints people increasingly leave as they navigate the online world, but also by the spread of new sensing technologies.<sup>24</sup> These analytics can be used for actionable prediction—not just to understand past trends, but to offer insights into how to act moving forward. The most important caveats to this somewhat idealized version of the promise of big data are discussed below,<sup>25</sup> but for policymakers, it is this last function—impact evaluation and prediction—that has the greatest relevance.

Examples of the use of big data are becoming increasingly common. For example, Google famously began real-time tracking of flu outbreaks in 2009 by linking common search terms with flu data from the Centers for Disease Control.<sup>26</sup> Google also transformed spell-checking from an educated guess by engineers at the most common errors and their corrections to a much more accurate approach that harnessed a massive database of actual errors and the way people corrected them.<sup>27</sup> Similar examples of information aggregation and new analytic capabilities abound in the private sector, and are becoming increasingly central to managerial planning and decision-making.<sup>28</sup>

---

24. See Neil M. Richards & Jonathan H. King, *Three Paradoxes of Big Data*, 66 STAN. L. REV. ONLINE 41, 42 (2013) (citing projections from Cisco that by 2020, thirty-seven billion intelligent devices, all collecting data, will be on line). In the housing context, this can be seen in the rise of “smart home” technology as well as building information systems, and other technologies to track the physical plant, energy usage, and similar information. See, e.g., Jessica Cocco, *Smart Home Technology for the Elderly and the Need for Regulation*, 6 PITT. J. ENVTL. & PUB. HEALTH L. 85, 91 (2011) (explaining that smart home technology, which collects physiological, locational, and movement data, can be used to effectively assist the elderly in health care and early intervention).

25. See *infra* Part III.

26. Mayer-Schönberger & Cukier, *supra* note 19, at 1-2; for a technical explanation, see generally Jeremy Ginsburg et al., *Detecting influenza epidemics using search engine query data*, 457 NATURE 1012 (2009).

27. See Gary Marcus, *The Web Gets Smarter*, NEW YORKER (May 23, 2012), <http://www.newyorker.com/culture/culture-desk/the-web-gets-smarter> [<https://perma.cc/K5DK-3AE4>].

28. See, e.g., Alistair R. Erskine et al., *How Geisinger Health System Uses Big Data to Save Lives*, HARV. BUS. REV. (Dec. 15, 2016) (“We use [big data] capabilities to track and analyze patient outcomes, to correlate their genomic sequences with clinical care, and to visualize healthcare data across cohorts of patients and networks of providers . . . with thousands of CPUs processing and delivering hundreds of terabytes of data every hour.”); Jeff Bertolucci, *Intel Cuts Manufacturing Costs With Big Data*, INFO. WEEK (Mar. 3, 2013) (“Processor giant uses big data to develop chips faster, identify manufacturing glitches and warn about security threats.”); see generally Kasturi E. et al., *Airline Route Profitability Analysis and Optimization Using Big Data Analytics on Aviation Data Sets under Heuristic Techniques*, 87 PROCEDIA COMPUTER SCI. 86 (2016) (explaining that airlines use big data to



Policymakers are also recognizing the value of these emerging tools. Perhaps most controversially, defense and national security agencies such as the CIA, NSA, and Department of Defense have embraced big data, raising significant civil liberties and privacy concerns.<sup>29</sup> More prosaically, big data is becoming increasingly important in fields such as public safety,<sup>30</sup> traffic management,<sup>31</sup> public health,<sup>32</sup> fraud prevention,<sup>33</sup> and many other areas.<sup>34</sup> To see the potential for affordable housing, we need to delve into another emerging trend, the shift from outputs to outcomes in that policy arena.

### B. From Outputs to Outcomes in Affordable Housing

Traditionally, affordable housing policymakers have focused on outputs but are beginning to focus on outcomes. This shift has the potential to open much broader horizons to improve how affordable housing is developed and delivered.

To the extent that policymakers have tracked the results of inputs such as subsidies or less direct interventions in housing, they have long focused on

---

calculate and predict everything from the route distance to seats, freight, or mails availability, and fuel to optimize profitability).

29. See Abraham R. Wagner & Paul Finkelman, *Security, Privacy, and Technology Development: The Impact on National Security*, 2 TEX. A&M L. REV. 597, 598 (2015); Adam R. Pearlman & Erick S. Lee, *National Security, Narcissism, Voyeurism, and Kyllo: How Intelligence Programs and Social Norms Are Affecting the Fourth Amendment*, 2 TEX. A&M L. REV. 719, 777 (2015).

30. See Helen Margetts, *The Promises and Threats of Big Data for Public Policy-Making*, OXFORD INTERNET INST.: POL'Y & INTERNET BLOG (Oct. 28, 2013), <http://blogs.oxi.ox.ac.uk/policy/promises-threats-big-data-for-public-policy-making/> [<https://perma.cc/3MDT-KVJL>] (discussing the example of California's use of predictive policing).

31. Los Angeles, for example, now not only tracks traffic in real time, but also calibrates every single traffic light in the downtown area. See Farnam Jahanian, *The Policy Infrastructure for Big Data: From Data to Knowledge to Action*, 10 I/S: J. L. & POL'Y FOR INFO. SOC'Y 865, 870 (2015).

32. See *supra* note 26.

33. See Helen Koh, *How Big Data Has Changed Public Policy*, DATAFLOQ, <https://datafloq.com/read/how-big-data-has-changed-public-policy-infographic/1880> [<https://perma.cc/FH3H-DKD4>] (noting that in 2014, the state of Indiana reported used algorithms to target tax audits based on anomalies in public records and tax-return information, identifying 75,000 fraudulent returns and saving approximately eighty-five million dollars).

34. The Obama Administration focused on big data across the public and private sectors. For example, the White House issued a report involving case studies in employment, education, criminal justice, and credit, to highlight how big data can be used to expand opportunities while potentially inadvertently introducing bias that could detrimentally affect marginalized groups and individuals. See EXECUTIVE OFF. OF THE PRESIDENT, *BIG RISKS BIG OPPORTUNITIES: THE INTERSECTION OF BIG DATA AND CIVIL RIGHTS* (May 4, 2016), <https://obamawhitehouse.archives.gov/blog/2016/05/04/big-risks-big-opportunities-intersection-big-data-and-civil-rights> [<https://perma.cc/9V89-C7TA>].

the immediate results of a given program or policy. For instance, HUD has long monitored and reported on basic measures relating to housing, such as the number of units built or supported, construction and management jobs created, as well as some indicia of the quality of the resulting housing, such as levels of overcrowding.<sup>35</sup>

Even as HUD has moved in recent years to broaden its planning and evaluation horizon, the Department still focuses to a large extent on metrics that concretely measure immediate outputs. For example, in its five-year Strategic Plan for the period from 2014 to 2018, the Department listed a comprehensive series of goals related to housing. Notable examples included:

- The number of households experiencing “Worst Case Housing Needs,” which HUD defines as having an income below fifty percent of Area Median Income (AMI), not receiving public assistance, and paying more than half of income on rent, living in severely inadequate conditions, or both;<sup>36</sup>
- The proportion of very low-income renters facing severe rent burdens;<sup>37</sup>
- The percentage of rental units built in the preceding four years that are affordable to very low-income renters;<sup>38</sup> and
- The production of rental units.<sup>39</sup>

These are very important core metrics and the fact that the Department has set clear, quantifiable goals around its central mission is commendable. Nevertheless, these output measures say relatively little about the lives of the people in the housing, or the larger community impacts of the housing investments and other policy interventions.

Other performance indicators in the HUD Strategic Plan do attempt to measure the difference federal investments are making for recipients beyond these immediate output indicators. For example, under the heading of “Economic Prosperity,” HUD now tracks the “[p]ercentage of participants enrolled in the FSS [Family Self Sufficiency] program who have increased wages.”<sup>40</sup> That begins to get at the connection between housing and consequences for residents, linking a housing-facilitated policy intervention to employment outcomes. Yet in other places the Department

---

35. See generally U.S. DEP’T OF HOUSING & URB. DEV., DATA SETS [https://data.hud.gov/data\\_sets.html](https://data.hud.gov/data_sets.html) [<https://perma.cc/8LTB-KFPD>].

36. See U.S. DEP’T OF HOUSING & URB. DEV., STRATEGIC PLAN 2014-18 (2014), at 32.

37. *Id.*

38. *Id.*

39. *Id.*

40. See *id.* at 25.

decries the lack of performance measures around topics such as health outcomes.<sup>41</sup>

As noted at the outset, however, researchers and policymakers are starting to link housing to longer-term life outcomes more thoroughly. For example, Raj Chetty, Nathaniel Hendren, and Lawrence Katz recently published the results of a study of HUD's "Moving to Opportunity" experimental voucher program. Under the experiment, HUD offered randomly selected families subsidies to move from high- to low-poverty neighborhoods, allowing researchers to compare relatively similar families who did not have the same opportunity. The researchers found that children who moved before age thirteen had a thirty-one percent higher average income, were sixteen percent more likely to attend college, lived in lower-poverty neighborhoods, and were less likely to be single parents as adults than peers who did not have the same exposure to low-poverty neighborhoods.<sup>42</sup>

Work in this vein has also tracked improvements from interventions that moved people to lower-poverty neighborhoods not only in terms of the immediate benefits of such neighborhoods, such as family safety, but also in terms of individuals' longer-term mental health, physical health, and subjective well-being.<sup>43</sup> This is not to over-read the data; much of the emerging neighborhood-effects literature remains contested.<sup>44</sup> But beginning to understand the consequences of the locational context of

---

41. *Id.* at 26-27 (setting as a goal that the agency would "[i]mprove performance management by enhancing HUD's collection and analysis of data pertaining to health-related outcomes across HUD-assisted housing programs. Also improve HUD's ability to integrate and/or conduct administrative data matches with other partner federal programs.").

42. See Raj Chetty, Nathaniel Hendren, & Lawrence Katz, *The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment*, 106 AM. ECON. REV., 855, 857 (2016).

43. See, e.g., Lawrence F. Katz, Jeffrey B. Liebman & Jeffrey R. Kling, *Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment*, 116 Q. J. OF ECON. 607 (2001); Jeffrey R. Kling, Jeffrey B. Liebman, & Lawrence F. Katz, *Experimental Analysis of Neighborhood Effects*, (Nat'l Bureau of Econ. Res., Working Paper No. 11577, 2005); Susan Clampet-Lundquist & Douglas S. Massey, *Neighborhood Effects on Economic Self-Sufficiency: A Reconsideration of the Moving to Opportunity Experiment*, 114 AM. J. SOC. 107 (2008); Jens Ludwig et al., *Long-Term Neighborhood Effects on Low-Income Families: Evidence from Moving to Opportunity*, 103 AM. ECON. REV. PAPERS AND PROC. 226 (2013).

44. Other analyses of the MTO program have not shown the positive results that the Chetty, Hendren, and Katz study has. See, e.g., Kling, Liebman & Katz, *supra* note 43 (finding no positive treatment effect for children in terms of educational or physical health outcomes, with the exception of mental health benefits and less risky behavior for girls). This is not surprising and it is important to note that seeking to understand the long-term consequences of policy interventions is hardly a definitive exercise. Over time, however, it is possible to discern correlations between inputs and outcomes, and the mixed MTO results underscore the need for more data.

affordable housing can help sharpen decisions about siting, mobility, transit connections, and more.

A shift from outputs to outcomes raises a number of questions, among which is an essentially functional one: how can policymakers realistically operationalize this change?<sup>45</sup> Among other things, doing so requires gathering and understanding significant amounts of information, which is most likely why agencies have often focused on much simpler metrics, such as the number of units built in a year. Given the rise of big data, however, there is potential to operationalize new approaches both to targeting resources and to linking related, but often functionally isolated, policy areas.

## II. BIG DATA IN AFFORDABLE HOUSING LAW AND POLICY

With that background, we can turn to the details of where the shift toward data-driven decision making is beginning to emerge in affordable housing as well as examples of areas where more sophisticated aggregation and analytical capacities seem to have particular promise. This section focuses on siting decisions and locational choice, resident relations and services, and portfolio management and property oversight, especially across the subsidized and unsubsidized stock of affordable housing. These examples are, of necessity, addressed below at a fairly high level of generality, but should give a sense of some primary areas of affordable housing policy and practice where new data tools have particular potential. The Part thus concludes with some reflections on what these examples

---

45. It is fair to ask whether affordable housing policy *should* focus on anything more than the simplest bricks-and-mortar outputs, to include a broader range of outcomes—in employment, education, integration, and the like. After all, it is challenging enough to manage the core tasks of building, maintaining, preserving, and making accessible decent, affordable housing given the paucity of resources devoted to subsidies compared to the need. *See e.g.*, John J. Infranca, *Housing Resource Bundles: Distributive Justice and Federal Low-Income Housing Policy*, 49 U. RICH. L. REV. 1071 (2015). But the reality is that housing inevitably and necessarily implicates health and education and employment and integration and equality and subjective well-being, as well as civic engagement and the health of the urban fabric. *See, e.g.*, LeighAnn M. Smith, *Affirmatively Further Fair Housing—and Potentially Further Fair Schooling*, 24 J. AFFORD. HOUS. & CMTY. DEV. L. 329, 330 (2015) (“The inherent connection between housing policy and education policy can be understood by an uncontroversial proposition: the makeup of our elementary and secondary schools is primarily drawn from the neighborhoods that surround the school building itself.”); *see also* Michael Diamond & J. Peter Byrne, *Affordable Housing, Land Tenure, and Urban Policy: The Matrix Revealed*, 34 FORDHAM URB. L.J. 527, 532-80 (2007) (tracing the broader goals of housing policy); Tim Iglesias, *Our Pluralist Housing Ethics and the Struggle for Affordability*, 42 WAKE FOREST L. REV. 511 (2007) (describing housing in terms of an economic good and home, as well as the locus of rights and social ordering). Ignoring that reality, however complex it renders policy making, risks unintended negative consequences—as when, for example, housing investments concentrate poverty on the basis of race.

suggest for the role of big data in improving affordable housing policy and law in improving the availability of that data.

### A. Some Examples of Big Data's Potential in Affordable Housing

#### 1. *Siting Decisions, Mobility, and Neighborhood Effects*

The first step beyond simple outcomes metrics in affordable housing involves the impact of neighborhood and regional context, which means that locational choice is particularly amenable to emerging analytics. HUD's new policy framework around AFFH provides a modest but promising attempt to use data to address—among other goals—the locational consequences of housing siting decisions.

The policy grows out of the federal Fair Housing Act's mandate that recipients of federal housing funding affirmatively further the purposes of the Act.<sup>46</sup> In a recently adopted rule,<sup>47</sup> HUD has created a national planning framework under which the Department will supply data to state and local governments and public housing agencies ("PHAs") across the country—almost all local governments get HUD funding and there are thousands of PHAs—with a range of uniform data on integration and segregation, housing needs, and indicia of economic opportunity.<sup>48</sup> The rule then requires that these program participants engage in their own analysis of fair housing based on that data and their own additional insights, to identify:

- "integration and segregation patterns and trends based on race, color, religion, sex, familial status, national origin, and disability within the jurisdiction and region;"

---

46. 42 U.S.C. § 3608(e)(5) (requiring that HUD programs and activities be administered in a manner to affirmatively further the policies of the Fair Housing Act). In addition to the Fair Housing Act's AFFH mandate, the Housing and Community Development Act requires local governments receiving state pass-through grants to certify that they will affirmatively further fair housing, 42 U.S.C. § 5306(d)(7)(B); the Cranston Gonzalez National Affordable Housing Act similarly requires, as part of a required comprehensive housing affordability strategy, that program participants certify that they will affirmatively further fair housing, 42 U.S.C. § 12705(b)(15); and for public housing authorities, the Quality Housing and Work Responsibility Act of 1998 ("QHWRA"), created a formal PHA planning process that similarly required certification that the PHA will affirmatively further fair housing. *Affirmatively Furthering Fair Housing*, 80 FED. REG. 42271, 42274 n.3 (July 16, 2015) (codified at 24 C.F.R. pts. 5, 91, 92, 570, 574, 576, 903) [hereinafter AFFH Final Rule].

47. AFFH Final Rule, *supra* note 46.

48. This HUD-supplied data comes from a range of sources, including general national sources such as the Decennial Census and the American Community Survey, but also more targeted information from HUD, the Census Bureau, the Department of Education, and the Environmental Protection Agency. *See* U.S. DEP'T OF HOUSING & URB. DEV., *AFFIRMATIVELY FURTHERING FAIR HOUSING (AFFH) DATA DOCUMENTATION* (July 2016), <https://www.hudexchange.info/resources/documents/AFFH-Data-Documentation.pdf> [<https://perma.cc/APF7-7KHU>].

- “racially or ethnically concentrated areas of poverty within the jurisdiction and region;”
- “significant disparities in access to opportunity for any protected class within the jurisdiction and region;” and
- “disproportionate housing needs for any protected class within the jurisdiction and region.”<sup>49</sup>

This data-driven exercise is a significant advance and has the potential to shed light on a wide range of decisions that local governments and PHAs make that are currently without a great deal of context. For example, understanding longitudinal patterns of the demographics of housing need can help inform siting decisions and shape affirmative marketing plans. It can also help policymakers understand questions such as concentration and agglomeration effects.<sup>50</sup> Each jurisdiction will use the data differently, but the underlying theory is that grantees will not only address obvious barriers to fair housing choice but will also deploy the planning tool data to guide housing siting, will also target enforcement resources and make other policy decisions in light of the patterns that the data reveals.

One challenge in this emerging framework is that the data tools that the agency and its grantees rely on are relatively broad-scale, in that they by necessity must be available for every jurisdiction in the country. Metrics such as neighborhood demographics, racial and ethnic concentration of poverty, school quality, job proximity, transit costs, and environmental health reveal a great deal about broad patterns and context for housing policy choices but require a significant amount of interpretation to shape decision making.<sup>51</sup>

---

49. AFFH Final Rule, *supra* note 46, at 42355 (to be codified at 24 C.F.R. § 5.154(d)(2)).

50. See Mindy Kao & Dan Immergluck, *AFFH Metrics for Affordable Housing Programs: An Approach to Assessing the Spatial Distribution of Housing Subsidies in Large Jurisdictions in the Assessment of Fair Housing* (May 25, 2016), [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2784583](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2784583) [<https://perma.cc/FS9A-GCEF>].

51. Scholars have called for similar kinds of analyses, such as human impact statements, *see, e.g.*, Marc L. Roark, *Human Impact Statements*, 54 WASHBURN L.J. 649, 653 (2015) (arguing that “cities and states should mandate that public projects and significant private projects that take advantage of city resources evaluate and report on the impact to low income populations, including the homeless populations, prior to approval of the project”) and community race audits, *see* R.A. Lenhardt, *Race Audits*, 62 HASTINGS L.J. 1527, 1534 (2011) (exploring the possibilities for a “voluntary tool called the “race audit” that can be utilized by localities interested in grappling with the inequalities that attend the color line”). For all its limitations, the Analysis of Fair Housing created under HUD’s AFFH rule can begin to serve some of these functions on a nation-wide basis. It is thus both a planning tool, but also a source of iterative information for policymakers that can be aggregated and mined for patterns.

Here, big data can produce a richer understanding of the link between inputs and outcomes in terms of the consequences of locational choice. New tools can do so by introducing much greater detail in terms of specific communities, but also by helping to tackle one of the greatest challenges in understanding impact: the longitudinal effects of policy interventions. The AFFH framework can give an accurate point-in-time picture of where the categories protected by the Fair Housing Act intersect with opportunity and housing investments, but it will take additional data and more sophisticated analytical tools to ferret out how that context changes in light of specific policy choices.

Ultimately, these tools could not only target community-level outcomes but potentially even city-wide and metro-level shifts, if aggregated across localities. They could also help policymakers focus on broader outcomes such as neighborhood change, tracking, for example, the pace of gentrification and displacement from various public investments. This will not make the decision to put a given housing development in site A versus site B necessarily easier (given the significant market and political constraints on all affordable housing), nor will it resolve the debate about place versus mobility.<sup>52</sup> Nevertheless, it can at least make the consequences of individual subsidy decisions and the larger structure of subsidies more transparent.

## 2. *Housing Portfolio Management*

A second general arena of affordable housing policy in which new data tools have particular promise involves market conditions as well as the oversight and management of housing assets. In these examples, the emphasis is less on data sharpening our understanding of where to deploy resources in terms of outcomes, and more on using technology to aggregate information to improve affordable housing practice and pragmatic policy implementation.

### a. *Subsidy Targeting and Market Conditions*

Big data tools can be useful in understanding how rent subsidy levels align with market conditions. As Matthew Desmond recently argued, better targeting of subsidies could free up substantial housing resources.<sup>53</sup>

---

52. See generally David Imbroscio, *Beyond Mobility: The Limits of Liberal Urban Policy*, 34 J. OF URB. AFF. 1 (2011); Gregory D. Squires, *Beyond the Mobility versus Place Debate*, 34 J. OF URB. AFF. 29, (2012).

53. See MATTHEW DESMOND, *EVICTED: POVERTY AND PROFIT IN AN AMERICAN CITY* 308-13 (2016) (proposing a broader, but more targeted housing voucher regime); see also David A. Dana, *An Invisible Crisis in Plain Sight: The Emergence of the "Eviction Economy," Its Causes, and The Possibilities for Reform in Legal Regulation and Education*,

The challenge, as Desmond points out, is that the broader the area for which rent-subsidy levels are set, the more over-and under-inclusive they are likely to be, compensating landlords in some neighborhoods above market while simultaneously preventing mobility to higher-income areas. HUD has begun to shift from metropolitan-level Housing Choice Voucher “fair market rents” to much more targeted zip code level measures in those metropolitan areas that are facing problems with voucher holders being concentrated in high-poverty neighborhoods.<sup>54</sup> But this is a pilot program and it only targets a subset of markets in a single program. By aggregating information about much more narrowly targeted market rents, as well a host of related transactional data, policymakers could get a much more fine-grained window into market conditions and thus appropriate subsidy levels across a range of programs.<sup>55</sup>

*b. Enforcement and Housing Quality*

New capacities to aggregate and analyze large data streams likewise hold promise to help policymakers more efficiently and effectively ensure the quality of housing and target interventions to remedy housing quality issues. Currently, local governments collect a great deal of property-level data from code enforcement, nuisance complaints, fire and police dispatches, and other sources, while states and the federal government collect information about housing quality through subsidy programs. But rarely is this information brought together, let alone used in a way to improve oversight.<sup>56</sup>

---

115 MICH. L. REV. 935 (2017) (reviewing *Evicted* and noting the political challenges of expanding voucher support).

54. See Establishing a More Effective Fair Market Rent System; Using Small Area Fair Market Rents in the Housing Choice Voucher Program Instead of the Current 50th Percentile FMRs, 81 FED. REG. 80567 (Nov. 16, 2016) (to be codified in scattered parts of 24 C.F.R. pts. 888, 982, 983, 985).

55. Cf. Norman Miller & Michael Sklarz, *Why Housing Price Indices Are Super Tools and How to Produce Them*, 45 REAL EST. REV. J. 2 (2016) (supporting big data valuation models using localized neighborhood price indices and data including terms of the sale, mortgage information, and property-specific details).

56. Some efforts have been made in this regard, showing the promise of aggregation and analysis to drive oversight. In South Bend, Indiana, for example, Mayor Peter Buttigieg in 2013 launched a Vacant and Abandoned Property Initiative with Code for America to bring code enforcement inspection data (which had been made on paper and had to be brought on line) with neighborhood input to craft a remediation strategy. See CODE FOR AMERICA, <https://www.codeforamerica.org/government-partners/south-bend-in> [<https://perma.cc/WR4R-L9Q3>]. Another example comes from how New York City revamped its approach to fire inspections under Mayor Bloomberg. A team of analysts aggregated disparate data from the building, housing preservation, tax, police, and fire departments to target illegal apartment conversions that posed particularly strong fire hazards. Prior to building this database, only thirteen percent of inspections were deemed



This kind of aggregation could be particularly important in the context of the link between housing quality and health outcomes. For example, the evidence linking indoor lead paint and problems with childhood development is overwhelmingly clear.<sup>57</sup> No comprehensive database of lead in housing exists, although some state and local governments do track lead-paint inspections and violations.<sup>58</sup> A big-data approach could consolidate multiple sources of information—home inspections, transactional data, lead-law citations, complaints—and provide a much more comprehensive picture not only of the landscape of risk, but also of more or less effective mitigation.<sup>59</sup> There are many other examples where data can improve policy around the nexus of housing and health outcomes, including indoor air quality, vectors of chronic illness related to damp, cold, and mold, and even injuries.<sup>60</sup>

Big data could also be used to track better and worse landlords, both to target enforcement resources and, ultimately, to influence choices about which housing developers or managers should receive assistance. HUD currently manages a process, known colloquially as the “2530 process”, to spot providers who pose particular financial or operational risks.<sup>61</sup> HUD recently revamped this process,<sup>62</sup> which until relatively recently involved paper submissions and individual review, but both HUD and other housing providers often have challenges assessing the risk of potential grantees (let alone using past performance to influence subsidy decisions in conditions of scarcity). Bringing information about providers, their management structures, and their past experience not only with specific subsidy programs but with other housing ventures, could not only streamline this

---

serious enough to warrant orders to vacate, wasting huge resources; after the analysis, that rate went up to seventy percent. Mayer-Schönberger & Cukier, *supra* note 19, at 186-88.

57. *See, e.g.*, Herbert Needleman, *Lead Poisoning*, 55 ANN. REV. MED. 209 (2004) (reviewing pediatric exposure effects not only from high-dose lead toxicity, but also from lower-dose exposure).

58. *See, e.g.*, MASS. HEALTH & HUM. SERVS., LEAD SAFE HOMES, <https://eohhs.ehs.state.ma.us/leadsafehomes/default.aspx> [<https://perma.cc/3TAL-RN6X>] (explaining the Massachusetts Childhood Lead Poisoning Prevention Program’s database for lead-inspected homes).

59. Some of this data is readily accessible, but much of it is not and it is rarely aggregated at any scale. Again, the cost of rendering data capable of interacting across disparate sources is not insignificant.

60. *Cf.* James Krieger & Donna L. Higgins, *Housing and Health: Time Again for Public Health Action*, 92 AM. J. PUB. HEALTH 758, 758-760 (2002) (reviewing the data relating housing quality to infectious disease vectors, chronic diseases, injuries, childhood development, and mental health).

61. *See* Retrospective Review—Improving the Previous Participation Reviews of Prospective Multifamily Housing and Healthcare Programs Participants, 81 FED. REG. 71244 (Oct. 14, 2016) (to be codified at 24 C.F.R. pt. 200).

62. *See id.*

kind of past-participation review, but could also help policymakers understand the operational and managerial factors that distinguish lower-risk partners from higher-risk partners.

*c. Unifying the Subsidized and Unsubsidized Housing Portfolio*

In this vein, new data tools may have particular promise in finding practical ways to bridge the subsidized and unsubsidized portfolios of affordable housing. For understandable reasons, affordable housing policy tends to focus on the portion of the housing stock available to low-income residents that is publicly subsidized. As Harvard's Joint Center for Housing Studies has noted, however, about three quarters of the stock of affordable housing is unsubsidized.<sup>63</sup>

The portion of the portfolio that is not part of any direct subsidy program is often relatively invisible to policymakers, tends not to be supervised in any meaningful way (except to the extent the portfolio intersects with nuisance claims and individual code enforcement), and is not integrated into holistic programs.<sup>64</sup> With improved data collection and analytics, it would be possible to start to unify this "shadow"<sup>65</sup> inventory with the much more transparent subsidized portfolio in a variety of ways, from gathering cost data related to development and operations to including outcomes related to the unsubsidized portfolio in locational analyses, to offering tools to assist with property management and tenant relations. New aggregation tools could also be used to better connect residents in the shadow portfolio to the kinds of services often more readily available to residents of the subsidized portfolio.

*3. Resident Relations and Services*

Moving, finally, from subsidy market conditions and portfolio management to individual residents, there is much that a big-data approach to affordable housing could facilitate in terms of individual residents. To begin, data aggregation tools can be used to track demand and even help with wait lists to foster more functional, even potentially regional, markets. In many cities, wait lists for available subsidized housing can be months

---

63. See JOINT CTR. FOR HOUS. STUD. OF HARV. U., *America's Rental Housing—Meeting Challenges, Building on Opportunities* 22 fig.18 (2011).

64. See, e.g. FAMILY HOUSING FUND, *The Space Between, Realities and Possibilities in Preserving Unsubsidized Affordable Rental Housing* (June 2013), [http://www.fhfund.org/wp-content/uploads/2013/06/Space\\_Between\\_Final\\_June-2013.pdf](http://www.fhfund.org/wp-content/uploads/2013/06/Space_Between_Final_June-2013.pdf) [<https://perma.cc/7EBK-UUJ8>] (discussing possible government involvement in the unsubsidized affordable housing market).

65. *Id.* at 18.

and, in some high-demand jurisdictions, even years long.<sup>66</sup> Given the mix of subsidies as well as the reality that much of the housing that might serve low-income communities is not subsidized, low-income housing markets are often fragmented and challenging to navigate.<sup>67</sup> Jurisdictions could benefit from clearinghouses or unified waiting lists, with new data tools greatly enhancing the ability of providers to match demand.

An even more significant way to use new data aggregation and analytical tools would be to build on the idea of housing as a platform to make that platform work better across social services. Housing is a particularly appropriate locus for a wide range of interventions, from educational support for children living in affordable housing, to job training for adults, to a range of wrap-around services for the formerly homeless in a “housing first” approach.<sup>68</sup>

Big data could be used here both to target services and to understand the consequences of those interventions. In terms of targeting, given that many residents of affordable housing encounter services across a range of areas, providers could access multiple sources of information to understand the needs of their clients without residents having to aggregate that information themselves. In terms of impacts, understanding the outcomes of various services could help providers over time focus on the relative efficacy of individual supports and the appropriate mix of support over large populations being served.

This kind of an approach is not distinctive to affordable housing, and social service agencies could use any other point of entry. If, however, policymakers could match education records, employment records, family

---

66. See, e.g., Mireya Navarro, *227,000 Names on List Vie for Rare Vacancies in City's Public Housing*, N.Y. TIMES, July 23, 2013, at A2, <http://www.nytimes.com/2013/07/24/nyregion/for-many-seeking-public-housing-the-wait-can-be-endless.html> [<https://perma.cc/H7MR-B93T>] (“There are now 227,000 individuals and families on the waiting list for Housing Authority apartments, totaling roughly half a million people, and the queue moves slowly. The apartments are so coveted that few leave them. Only 5,400 to 5,800 open up annually.”); see also Michael Twomey, *Legislation Proposed for Online List of All Affordable Apartments*, CITYLAND (Dec. 9, 2015), <http://www.citylandnyc.org/legislation-proposed-for-online-list-of-all-affordable-apartments/> [<https://perma.cc/39Q2-3R8K>] (noting proposal to aggregate data on affordable housing supply).

67. See, e.g., *Vacancy Clearinghouse*, TEX. DEP'T OF HOUSING & CMTY. AFF., <http://hrc-ic.tdhca.state.tx.us/hrc/VacancyClearinghouseSearch.m> [<https://perma.cc/WX9L-YMBN>]; see also FLA. HOUSING DATA CLEARINGHOUSE, <http://flhousingdata.shimberg.ufl.edu/> [<https://perma.cc/8ZPD-3P8T>].

68. Housing first programs take a harm-reduction approach to chronic homelessness, placing individuals facing substance abuse, mental illness, or dual diagnoses into housing and then using the fact that those individuals are in a safe, known place to craft services around that housing. See Nestor M. Davidson, “*Housing First*” for the Chronically Homeless: *Challenges of a New Service Model*, 15 J. AFFORDABLE HOUSING & CMTY. DEV. L. 125 (2006).

disruption data, and myriad other sources, they could potentially target assistance with much greater efficacy.

### B. Synthesis and Reflections on the Role of Law

To take a step back and synthesize the examples above, they generally point to several functions for big data to improve affordable housing policy. First, new data tools can improve policymakers' ability to target resources, whether in terms of direct subsidies, enforcement activities, or the larger policy context in which affordable housing exists. Second, and perhaps most promising in the short run, new data tools can help policymakers, providers, and advocates coordinate across silos. Vertically, technology can better integrate federal, state, and local operations. Horizontally, data tools can help bridge gaps across agencies at each level of government as well as across states and localities at the sub-federal level, where appropriate. If big data did nothing else but facilitate this kind of information sharing, it would still have great potential to improve the provision of services, respond to regional housing needs, and link housing data to other inputs such as environmental quality, public safety, employment opportunities, and education, just as the AFFH approach has begun to do. Finally, new data tools could add private information to the decisional matrix and to understanding outcomes. This is perhaps most challenging, legally, but also where law has the clearest role to play.

To be clear, in the immediate term policymakers may find the best value in what might be called "medium data."<sup>69</sup> Rather than attempting to connect the dots from billions of data points akin to searches on Google, policymakers in affordable housing may be able to utilize new analytic capabilities to make predictions about subsidies and related policy choices by aggregating relatively accessible information about market conditions, demographics, specific properties, and subsidies. The longer-term potential to increase this to a much more sophisticated look at, for example, dynamics of neighborhood change and the relative impacts of various policy levers is intriguing, but should not deter from shorter-term more manageable deployment.

This is not to be naïve about the genuine costs and other practical barriers to implementing any data effort that would truly yield new insights at scale.<sup>70</sup> The most obvious reason why affordable housing policymakers

---

69. See Jacob Harold, *Nonprofits: Master "Medium Data" Before Tackling Big Data*, HARV. BUS. REV. (March 20, 2013), <https://hbr.org/2013/03/nonprofits-master-medium-data-1> [<https://perma.cc/2APX-AJ8S>] (distinguishing "medium data" from "big data," and discussing possible uses for nonprofit organizations).

70. As a practical matter, some of the data that might be most relevant to a shift from outputs to outcomes is very difficult to collect and we won't know necessarily what we need until after the fact. Often data sources are difficult to link, making correlations (let alone

have defaulted to simple output measures, to the extent that they have tracked outcomes, is that it is just simply hard to do anything more. Perhaps if the value of tracking—and making predictions about—impact through new data tools were clearer, the cost-benefit might shift. But given the choice between investing in sophisticated analytics or investing in another unit of housing, most rational subsidy or housing providers will tend to prefer the immediate impact of a roof over the head of someone in need.

One answer may come by looking to HUD's approach to AFFH, which shows the potential role for federal investment in data tools. If local governments, public housing authorities, and non-profit housing providers do not have the resources to invest in analytics—and most do not—then the federal government can step in and do so at scale. Indeed, HUD has recently launched a broad-scale initiative as part of an interagency agreement with the Census Department to provide qualified researchers with access to certain HUD-sponsored evaluation datasets.<sup>71</sup> This agency “data democratization” effort is slated to expand to a broad range of HUD administrative and other information sources.<sup>72</sup> Ideally, information could flow not only to researchers, but from researchers to providers and policymakers.<sup>73</sup>

If affordable housing provides numerous examples of where new data tools can help shape policy, it is possible at this juncture to focus on how *law* can improve the flow of that information. Currently, statutory and administrative mandates play a relatively modest role in either generating or utilizing the potential of big data. Law is essentially mostly parasitic to data efforts, if not directly a hindrance.

But there can be a dialectic relationship between law and data for policy making. Legal mandates can help bring information to the surface and can create structures to capture and facilitate the analysis of that information. That information can then be used to sharpen policy. The new AFFH framework, for example, reflects an agency implementing a statutory

---

conclusions about causation for any given investment or intervention) challenging. Moreover, some data is easy to collect and hard to analyze and some data is hard to collect, and any effort to use law to generate new information must be sensitive to those dynamics.

71. See Katherine O'Regan, *Data Democratization and Evidence-Based Policy*, EDGE, <https://www.huduser.gov/portal/pdredge/pdr-edge-firm-asst-sec-010917.html> [<https://perma.cc/7SJM-SKM6>] (last visited Mar. 28, 2017).

72. *Id.*

73. There are many excellent intermediaries, such as the Urban Institute, that work to translate data into policy insights in the affordable housing arena. Highlighting the potential for federal investments in data is not to exclude the value of private and philanthropic efforts in this area as well. There is much symbiotic potential in public-private approaches to the data-policy nexus here.

mandate through planning tools that aggregate information across a variety of sources that mix government-generated data with information collected on a mandatory basis from grantees. Grantees then deploy this data, to analyze, plan, and act. The results of this nexus can then be collected and fed back to grantees to identify more or less effective policies. Thus, the legal system is facilitating the generation of data; data are aggregated and deployed to guide policy decisions; and the process is then iterated.

Equally important, patterns revealed through this cycle, which might otherwise be hard to see, can help inform how regulators and advocates advance other legal mandates, including enforcement and private rights of action. Data and deeper structural analysis can thus be used in tandem with substantive liability standards to target issues like discrimination. This could be the basis for litigation, whether individual or structural. The Home Mortgage Disclosure Act (“HMDA”), particularly after amendments in 1989 that made much more detailed information about the demographics of applicants and borrowers required,<sup>74</sup> is a model for this dynamic. Under HMDA, lenders are required to provide detailed information about lending practices, which advocates and regulators can then use to challenge patterns of discrimination.<sup>75</sup> Thus, the law requires private actors to disclose activity and that disclosure can reveal patterns that yield liability and, most importantly, change behavior.<sup>76</sup>

### III. BIG DATA’S DARK SIDE: CAVEATS AND (SOME) RESPONSES

To this point, it would be fair to read this Article as something of an unalloyed paean to big data’s potential value for affordable housing, but there are a number of reasons to be cautious in any exploration of the phenomenon’s potential. Several caveats emerge from the literature on big data worth attending to, both from a structural perspective, and in terms of the individuals that are the inevitable subjects—to use that term advisedly—of any big-data project.

To begin with structural concerns, many commentators have noted that big data, as much as traditional data (and perhaps more so), has the potential to suffer from basic challenges with informational quality.<sup>77</sup>

---

74. 12 U.S.C. § 2801-2810.

75. See generally Charles M. Lamb et al., *HMDA, Housing Segregation, and Racial Disparities in Mortgage Lending*, 12 STAN. J. C.R. & C.L. 249 (2016) (recounting the structure and function of HMDA in revealing patterns of segregation and discrimination).

76. Cf. Rigel C. Oliveri, *Disparate Impact and Integration: With TDCHA v. Inclusive Communities the Supreme Court Retains an Uneasy Status Quo*, 24 J. AFFORD. HOUS. & CMTY. DEV. L. 267 (2015).

77. See Solon Barocas & Andrew D. Selbst, *Big Data’s Disparate Impact*, 104 CAL. L. REV. 671, 684 (2016) (“In other words, the quality and representativeness of records might vary in ways that correlate with class membership (e.g., institutions might maintain

Moreover, it is not just that the quality of any analysis is inherently limited by the quality of the sources, and maintaining data governance can be a costly exercise. Measurability itself can be a misleading proxy for policy outcomes that might be harder to quantify but are no less (or even more) important.<sup>78</sup> These alternatives might be concrete but harder to measure (just as units produced are easier to measure than long-term subjective well-being), but there is also the risk that values that are not measurable—fairness, dignity—will be lost.

There are no definitive answers to these structural concerns and the best response is not so much legal, but practical. Policymakers, to the extent they are going to rely on data, especially the potentially unwieldy reams of information in a big data approach, must focus on continuous improvement and a feedback loop that can come from actually deploying the data in practice. Data quality and data management challenges also auger strongly for a measure of skepticism in not allowing the data to be overly determinative in decision making, as hard as it is to resist the allure of the concrete.

Shifting to the people whose lives and experiences generate the data on which big data lives—and whose lives will be influenced by decisions made based on that data—the rise of new analytic tools raises genuine concerns not just about privacy, but also about the loss of control of one’s identity.<sup>79</sup> Closely related to this is a concern with the risk that datafication will undermine the dignity and individual humanity of the people being quantified. Numbers can speak louder than individual voices and seemingly implacable data can obscure the very real life experiences. This affects policy as a structural matter but it also affects people because it reduces who they are to a series of numbers and quantifiable outcomes (graduation rates, health assessments, and the like). And this can all exacerbate existing power dynamics that risk minimizing the agency of people served by affordable housing programs.

These are important issues, and clearly policymakers must be sensitive to the perspectives of those served and power dynamics, but they are

---

systematically less accurate, precise, timely, and complete records for certain classes of people). Even a dataset with individual records of consistently high quality can suffer from statistical biases that fail to represent different groups in accurate proportions.”); *see also* Michael Mattioli, *Disclosing Big Data*, 99 MINN. L. REV. 535, 546 (2014) (“Data is often deeply infused with the subjective judgments of those who collect and organize it.”).

78. Moreover, data can improve the probabilities that a given intervention will tend toward a certain outcome, but the actual outcome will always be contingent. It is hard to imagine being able to overcome that contingency, however, no matter how much information informs decision making.

79. *See* Richards & King, *supra* note 24, at 43-44 (noting that big data threatens a corollary to the right of privacy, which is the right of identity—the right “to say ‘who I am’”).

ultimately manageable in the affordable housing context. Much of the data that would be a part of any effort to sharpen our understanding of outcomes is amenable to anonymization and must be treated on an aggregated basis. And privacy—for residents as well as providers—must be central to how data is collected, stored, and analyzed.

But privacy and the importance of preserving individual voice should not be an insuperable barrier to gaining practical value from new data tools. Consumers currently make choices on a daily basis to share information with technology companies in order to improve the collective information base available to all users. When someone accesses Google Maps on a mobile device, they are both receiving and sending information about location and traffic conditions and the like<sup>80</sup> and it is reasonable to trade privacy for the value of that information exchange. A similar dynamic can pertain with affordable housing and related areas of policy.

More fundamentally, it is worth asking “as compared to what?” For those appropriately concerned about individual voice and the dignity of residents, it is not so clear that our current affordable housing system is serving the most vulnerable right now as well as it could. Indeed, it is fair to argue that we do not pay enough attention to residents and clients in the current structure of housing policy.<sup>81</sup> Fundamental values can be advanced by better tools, and these are not ultimately incommensurate goals. We can focus on humanity and dignity in practice and use data to advance that. There is at least the potential that for all the problems with quantifiability, data can center housing policy choices on outcomes that put residents (rather than buildings or providers) more clearly at the center of policy focus. In many critical areas of public policy, including providing affordable housing for the most vulnerable, limited public resources could be used much more effectively. Data as an engine of decision making and the individual dignity of those served by public programs can and no doubt often do clash, but they can just as well interact.

Data, of course, cannot solve fundamental policy disputes that implicate not just comparative outcomes but also relative and at times

---

80. See Tim Stenovic, *Google Has Gotten Incredibly Good at Predicting Traffic—Here’s How*, BUS. INSIDER (Dec. 18, 2015), <http://www.businessinsider.com/how-google-maps-knows-about-traffic-2015-11> [<https://perma.cc/7U33-FQE7>] (“Here’s how it works: All iPhones that have Google Maps open and Android phones that have location services turned on send anonymous bits of data back to Google. This allows the company to analyze the total number of cars, and how fast they’re going, on a road at any given time.”).

81. It is striking, for example, that HUD, the national department dedicated to housing, contains no office or representative focused specifically on the residents of HUD-subsidized housing.



incommensurate priorities, as well as differing political economies.<sup>82</sup> Thus, no amount of impact assessment will necessarily conclusively resolve the perennial debate in housing policy about mobility versus community investment.<sup>83</sup> This debate will always reflect marginal trade-offs in an environment of scarcity and also interaction effects with other policy interventions and market forces. But beginning to understand the trajectory of the investments we make in the name of affordable housing can sharpen the focus of those investments. That might not be sufficiently ambitious, but would still be an improvement.

### CONCLUSION

This Article has sought to illuminate the potential benefits—and significant challenges—of deploying new data aggregation and analytic tools to advance affordable housing policy. Whether informing siting decisions, understanding the regional housing market consequences of local zoning policy, transforming management and resident services, or other areas of affordable housing, and recognizing the practical barriers to implementation, there is undeniable promise at a minimum to make better decisions.

In this arena, we are beginning to see a cycle in which law facilitates the generation of data that can be grouped across disparate areas to provide a clearer picture of the consequences of public investments and other policy choices. That data can also be used to drive other legal mandates, particularly around enforcement.

As noted, we must proceed cautiously in embracing these new tools—they will remain inherently limited in their ability to accurately capture ground-level reality and they must be used with a keen appreciation for the people whose lives are being measured and whose voice is too often ignored. But the alternative of continuing to muddle through is even less palatable.

---

82. Michael Diamond, *Affordable Housing and the Conflict of Competing Goods: A Policy Dilemma*, in AFFORDABLE HOUSING AND PUBLIC-PRIVATE PARTNERSHIPS 1 (Nestor M. Davidson & Robin Paul Malloy eds., Ashgate Press 2009).

83. See Nestor M. Davidson, *Reconciling People and Place in Housing and Community Development*, 16 GEO. J. ON POVERTY L. & POL'Y 1 (2009).