

Stanford  
LawSchool | Stanford Criminal  
Justice Center

# Rate My District Attorney: Toward a Scorecard for Prosecutors' Offices

Katherine K. Moy, Dennis M. Martin,  
and David Alan Sklansky

2018



# **Rate My District Attorney: Toward a Scorecard for Prosecutors' Offices**

**A Report of the Stanford Criminal Justice Center**

**January 2018**

**Katherine K. Moy, Dennis M. Martin, and David Alan Sklansky\***

---

\* Katherine K. Moy and Dennis M. Martin are third-year students at Stanford Law School. David Alan Sklansky is Stanley Morrison Professor of Law at Stanford Law School and Faculty Co-Director of the Stanford Criminal Justice Center.

The **Stanford Criminal Justice Center** (SCJC), led by faculty co-directors Joan Petersilia, David Sklansky, and Robert Weisberg and executive director Debbie Mukamal, serves as a research and policy institute on matters related to the criminal justice system. For more information about our current and past projects, please visit our website: <http://law.stanford.edu/criminal-justice-center>.

Copyright © 2018 Stanford Criminal Justice Center  
All rights reserved.  
Stanford, CA  
Stanford Criminal Justice Center  
Stanford Law School  
559 Nathan Abbott Way  
Stanford, CA 94305

# Table of Contents

Executive Summary .....	4
Introduction .....	5
Prior Efforts at Developing Metrics for Prosecutors’ Offices .....	6
Fig. 1: NDAA Chart Depicting Prosecution Goals, Objectives, and Performance Measures .....	8
Developing a Prosecutorial Scorecard.....	9
Stage One: Assembling Personnel and Logistical Planning .....	11
Stage Two: Defining the Index’s Goals and Audience .....	12
Stage Three: Designing the Index .....	15
Fig. 2: Excerpted Summary of State Scores for 2016 ACEEE.....	19
Stage Four: Gathering Data .....	20
Fig. 3: Excerpted Chart of Proxies for Measuring Voter Registration Performance .	23
Stage Five: Constructing the Index .....	25
Stage Six: Disseminating and Building Support for the Index .....	27
Conclusion .....	28

# Executive Summary

Local prosecutor elections can have uniquely consequential results for the American criminal justice system. Paradoxically, however, these elections attract much less voter engagement than other races, and incumbents are repeatedly re-elected. As a result, activists seeking to convince prosecutors to pursue reforms, or to elect new reform-minded prosecutors, have a hard time communicating just how well a given office is performing.

A prosecutorial rating system is one approach to remedying this information gap. Much like indices used in other public policy areas, such a rating system could be a critical way of communicating to voters and potential electoral challengers whether a prosecutors' office has effectively pursued the electorate's policy priorities.

This report begins to chart a path toward building such a rating system. Drawing on the expertise of experienced public policy index developers, the report outlines a procedure that developers can follow to design and build their own scorecard. The process described in the report involves several stages, during which developers will need to grapple with key policy and logistical issues.

Although the contours of the process are flexible, the report lays out the following steps to developing a prosecutorial rating system:

- 1) **Gather key personnel and experts and set project benchmarks.**
- 2) **Define the index's goals and target audience**, including any intermediaries that might be enlisted to convey the index's message.
- 3) **Select the variables** the index will use to measure performance and **decide how much weight** to attribute to each variable.
- 4) **Gather data** for each variable, including any proxy measurements to use where direct data is unavailable.
- 5) **Aggregate and normalize the data** in a coherent, rigorous, digestible format.
- 6) **Disseminate and build support** for the index.

Each of these stages involves complex decisions, many of which may need to be revisited throughout the development process. But walking through each of the stages methodically can help highlight areas of dispute and place in a broader procedural context. By keeping the index's overall goals in mind as they work through the minute details of each stage, developers are more likely to be able to create a successful index to help meet their reform objectives.

# Introduction

Prosecutors play a central role in American criminal justice. They decide who is charged with crimes, which crimes they are charged with, and what kind of a plea agreement to offer. They supervise law enforcement agencies and are responsible for ensuring that exculpatory information is turned over to the defense. It is widely believed that mandatory sentencing laws have greatly magnified the power of prosecutors over the past several decades; in many cases, prosecutors effectively decide not only who will be charged, and what they will be charged with, but how much time they will spend in prison.

Most prosecutors in the United States work in county-level offices headed by a locally elected official, commonly known as a “District Attorney” or “State’s Attorney.” There are more than 2,400 elected prosecutors across the country.<sup>1</sup> Historically, the ballot box has served as a weak tool for oversight of American prosecutors. Incumbents running for reelection rarely faced serious challenges, and campaigns focused on personalities, not on policies.<sup>2</sup> Recently, however, in a string of prosecutorial races across the United States, incumbents have been defeated by challengers calling for greater moderation in criminal punishment, greater vigilance against wrongful convictions, and greater scrutiny of police misconduct. A growing number of criminal justice reformers are focusing on electing prosecutors committed to change.<sup>3</sup>

One challenge these reformers and the prosecutors they help to elect must confront is the difficulty of assessing the day-to-day operations of a prosecutorial office. Part of the problem is opacity: Prosecutors’ offices do not have a strong tradition of collecting and sharing data. Another part of the problem, though, is conceptual: There is no consensus about how prosecutors’ offices should be judged. Elected prosecutors wear many hats. They are politicians, managers, policymakers, advocates, and adjudicators. We expect prosecutors to be strong and effective litigators, but we also want them to be impartial and objective ministers of justice; we want them to uphold the rule of law, but also to exercise mercy.<sup>4</sup> No one thinks that conviction rates alone are a good measure of how well a prosecutorial agency is functioning—there can be good reasons and bad reasons for conviction rates to rise. But voters looking for a better measure of prosecutorial performance find themselves at a loss. So do many prosecutors themselves.

One solution to this problem is to develop a method of rating or ranking prosecutors’ offices. In a range of other areas—from education and environmental protection to

---

<sup>1</sup> *OLR Research Report: States that Elect Their Chief Prosecutors*, OFFICE OF LEGISLATIVE RESEARCH: CONNECTICUT GENERAL ASSEMBLY (Feb. 24, 2003), [www.cga.ct.gov/2003/rpt/2003-R-0231.htm](http://www.cga.ct.gov/2003/rpt/2003-R-0231.htm).

<sup>2</sup> See Ronald F. Wright, *How Prosecutor Elections Fail Us*, 6 OHIO ST. J. CRIM. L. 581 (2009).

<sup>3</sup> See David Alan Sklansky, *The Changing Political Landscape for Elected Prosecutors*, 14 OHIO ST. J. CRIM. L. 647 (2017).

<sup>4</sup> See David Alan Sklansky, *The Nature and Function of Prosecutorial Power*, 106 J. CRIM. L. & CRIMINOLOGY 473 (2016) [hereinafter Sklansky, *Nature and Function*].

election oversight—public policy indices have shown promise in empowering democratic oversight. In many ways, those who seek to develop a public policy index in this field are working on a blank slate; there is almost no information available to voters about prosecutorial performance to compete with or incorporate. By the same token, the vagaries of local prosecutor elections pose unique challenges to developers wading into the field. Developers of a prosecutorial scorecard would need to answer some hard questions: What should the index measure, ideally? What can it realistically attempt to measure? How can it effectively get its message across?

This report does not try to answer those questions; instead, it charts a path toward answering them. It sets forth a framework for developers of a prosecutorial rating system to follow. Fortunately, the developers of a number of well-regarded public policy indices have published their own reflections on the best practices of index development. This report builds upon the recommendations of those experts.

The report has two parts. The first part discusses past efforts toward developing ratings systems for prosecutors' offices—efforts that future developers will want to build on. The second provides a road map for the development process, breaking the process into a series of stages: assembling personnel and planning, defining the index's goals and audience, designing the index, gathering data, compiling and analyzing the data, and disseminating the results. The report identifies key questions that will need to be addressed at each of these stages.

## **Prior Efforts at Developing Metrics for Prosecutors' Offices**

At least two major projects have focused on using performance measurements to spur reform of local prosecutors' offices. Although they have not been implemented at a large scale, these projects' findings can help developers of a prosecutorial rating system get a sense of what ground experts and reformers have already tread.

The National District Attorneys Association (NDAA), through its research branch the American Prosecutors Research Institute, published a series of four papers in 2007 that recommended and tested performance measures for prosecutors' offices. The NDAA began by stating that simple measurements like crime rates, conviction rates, and even recidivism fail to capture much of what prosecutors do. In particular, “such traditional measures often do not adequately address the interests of victims and the community, nor do they adequately explain prosecutorial discretion and decision-making.”<sup>5</sup>

In the place of those metrics, the NDAA outlined the three ideal functions of prosecutors: (1) “To promote the fair, impartial, [and] expeditious pursuit of justice”;

---

<sup>5</sup> NAT'L DIST. ATTORNEY'S ASS'N, PROSECUTION IN THE 21ST CENTURY: GOALS, OBJECTIVES, AND PERFORMANCE MEASURES 1 (2007), [http://www.ndaa.org/pdf/prosecution\\_21st\\_century.pdf](http://www.ndaa.org/pdf/prosecution_21st_century.pdf) [hereinafter NDAA REPORT].

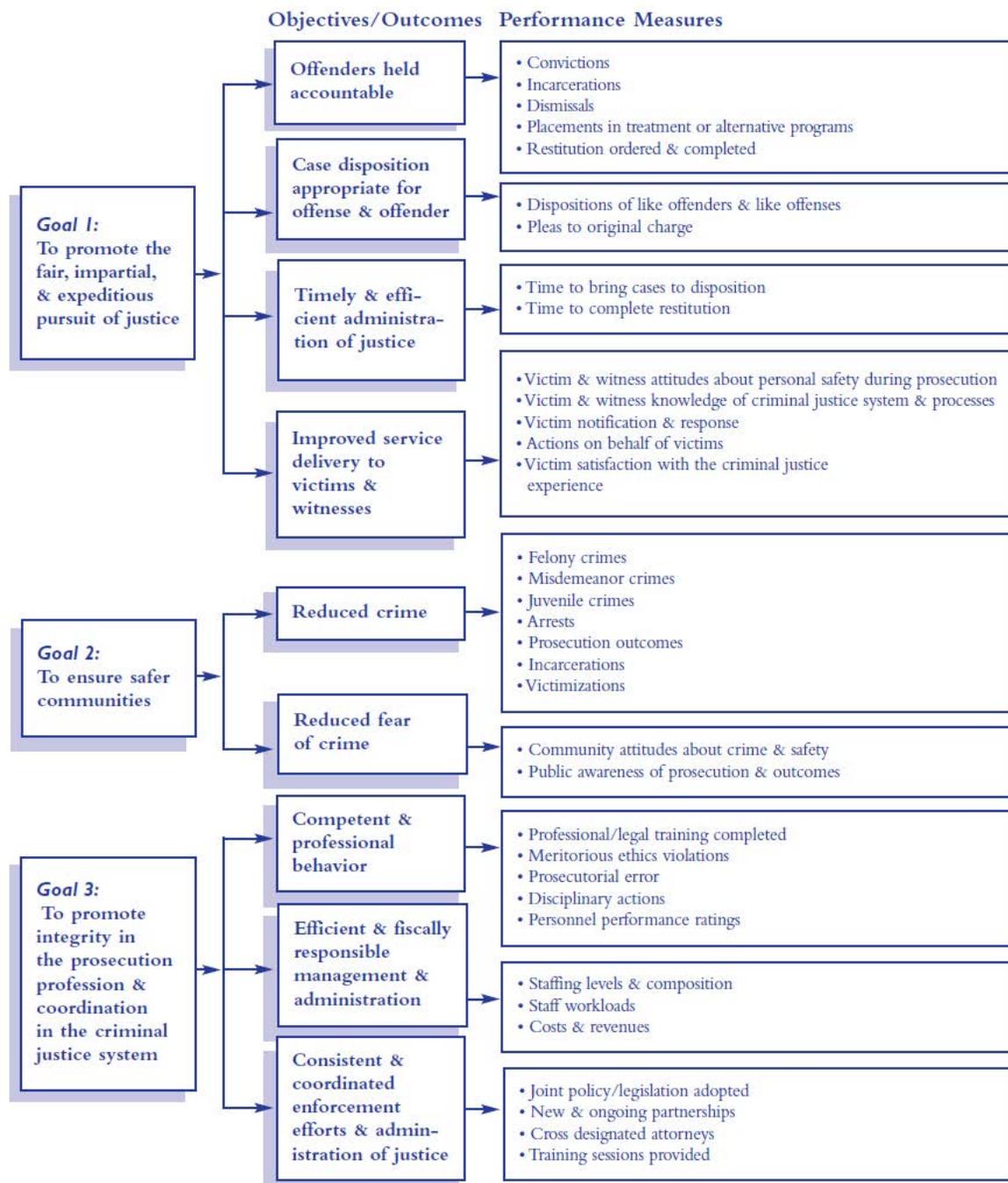
(2) “To ensure safer communities”; and (3) “To promote integrity in the prosecution profession [and] coordination in the criminal justice system.” For each of these goals, the NDAA recommended discrete objectives that an office should strive to implement. In turn, it provided several measurements that might be available to track prosecutors’ progress on the objectives.<sup>6</sup> For example, under its goal of promoting the fair, impartial, and expeditious pursuit of justice, the NDAA listed “[i]mproved service delivery to victims [and] witnesses” as an objective. The NDAA proposed that experts and observers measure this service delivery with a variety of metrics, including how safe victims and witnesses feel during prosecution, how well a prosecutor’s office notifies victims of important developments, and how satisfied victims feel with the criminal justice experience overall. The interplay between these core functions, objectives, and performance measurements is depicted in Figure 1.7

---

<sup>6</sup> *Id.* at 19.

<sup>7</sup> *Id.*

**Fig. 1: NDAA Chart Depicting Prosecution Goals, Objectives, and Performance Measures**



**Source:** Nat'l Dist. Attorney's Ass'n, *Prosecution in the 21st Century: Goals, Objectives, and Performance Measures* 19 (2007)

The second major effort at developing a series of metrics for prosecutors' offices was carried out by the Brennan Center for Justice at New York University School of Law. In 2014, the Brennan Center convened a panel of prominent jurists (including former judges, U.S. Attorneys, and cabinet members) to determine what goals federal

prosecutors should set for themselves and their offices.<sup>8</sup> They proposed these goals as an alternative to traditional measures of success such as number of convictions, sentence lengths, and conviction rates.

The panel agreed on three core priorities that all federal prosecutors' offices should target: (1) reducing violence and serious crime; (2) reducing prison populations; and (3) reducing recidivism. The panel also identified three "optional" priorities: (4) reducing pre-trial detention; (5) reducing public corruption; and (6) increasing coordination. And they proposed measures of success for each priority. The experts suggested that "reducing violence and serious crime," for example, can be tracked by the change in violent-crime rate, change in the percent of violent-crime cases on docket, and the percentage of the community that feels safe.

While the NDAA's project provided insights as to how prosecutors themselves view the issue (even they are concerned about being judged by conviction rates, plea-bargain rates, or the outcome of a single high-profile case), the Brennan Center's recommendations highlighted reformers' ideas about the need for more descriptive measurements of prosecutorial performance.

Both the Brennan Center and the NDAA have advanced the public policy conversation about what good prosecutors' offices look like. The NDAA in particular began the critical step of defining discrete metrics that might be used to measure performance. But both the Brennan Center's panel<sup>9</sup> and the NDAA<sup>10</sup> intended their recommendations as internal reforms—priorities that individual offices, or the DOJ as a whole, might adopt. They did not intend them to be used for external evaluation, or for office-to-office comparisons. Still, they provide guidance in thinking about possible components of a prosecutorial scorecard.

## Developing a Prosecutorial Scorecard

The remainder of this report attempts to outline the various stages of developing a public policy index. Each stage includes key questions developers will likely have to wrestle with before they can move on to the subsequent stage. Some of the questions are noticeably more complex than others; with luck, those questions that are less easily answered can guide future research.

---

<sup>8</sup> BRENNAN CENTER FOR JUSTICE, FEDERAL PROSECUTION FOR THE 21ST CENTURY 9-11 (2014), [http://www.brennancenter.org/sites/default/files/analysis/Buying\\_Time/Federal%20Prosecution%20for%20the%2021st%20Century.pdf](http://www.brennancenter.org/sites/default/files/analysis/Buying_Time/Federal%20Prosecution%20for%20the%2021st%20Century.pdf) [hereinafter BRENNAN CENTER REPORT].

<sup>9</sup> *Id.* at 20. The Brennan Center suggests measures be tracked from year to year so that offices' progress toward goals over time can be seen. In this way, "offices are primarily measured against themselves and not against other offices, which may have more resources or different criminal justice challenges." *Id.*

<sup>10</sup> *See, e.g.*, NAT'L DIST. ATTORNEY'S ASS'N, PERFORMANCE MEASURES FOR PROSECUTORS: FINDINGS FROM THE APPLICATION OF PERFORMANCE MEASURES IN TWO PROSECUTORS' OFFICES 14-15 (2007), [http://www.ndaa.org/pdf/performance\\_measures\\_findings\\_07.pdf](http://www.ndaa.org/pdf/performance_measures_findings_07.pdf).

The development process described below is drawn largely from an extensive guide published by the Yale Center for Environmental Law and Policy (YCELP), which developed and maintains the Environmental Performance Index (EPI).<sup>11</sup> YCELP's guide serves as a model because it is comprehensive; it describes almost every phase of the development process in detail. It is also readily adaptable to different development scenarios—although YCELP cautions that the guide is “meant for those who are specifically seeking guidance on how to develop *environmental* performance indicators or aggregated, composite indices,”<sup>12</sup> it describes the process in general terms that can apply outside the environmental policy context.

The below analysis is thus more an exercise in supplementing YCELP's robust framework with lessons from other indices than in deriving a procedure from the ground up. The report supplements YCELP's recommendations with comparisons to other guides to suggest possible variations in the process, as well as specific examples of how other indices have tackled particular issues in each development stage. Thinking about what other indices do well, and comparing them to existing funding and data constraints, will likely help developers assess and critique potential proposals for a prosecutorial rating system.

A word of caution is warranted: The variation in the processes other developers have used makes clear that even breaking the process down into stages is somewhat misleading. That is, depending on the nature of the index and the available data, it may be impossible to answer some of the questions in Stage Three, for example, without also wrestling with those in Stage Five. Developers, therefore, should begin with a sense of the entire process before diving into any given stage.

Finally, it is important to note at the outset that the order of the stages and the emphasis placed on each key question will depend on the goals of the index and the obstacles unique to its development. Although the order of the stages is largely borrowed from YCELP's guide, there is no literature suggesting this is the only viable sequence; on the contrary, other experts have placed some tasks in different phases of the development process, or omitted stages altogether. Ultimately, the “stages” themselves are more of a conceptual tool than a rigid system. They need not be followed sequentially, and chances are that they will overlap significantly. The stages are meant to be a helpful reference point to return to when developers are in the weeds of developing the index.

---

<sup>11</sup> YALE CTR. FOR ENVTL. LAW & POLICY, MEASURING PROGRESS: A PRACTICAL GUIDE FROM THE DEVELOPERS OF THE ENVIRONMENTAL PERFORMANCE INDEX (EPI) (2013), [http://archive.epi.yale.edu/files/ycelp\\_measuring\\_progress\\_manual.pdf](http://archive.epi.yale.edu/files/ycelp_measuring_progress_manual.pdf) [hereinafter MEASURING PROGRESS].

<sup>12</sup> *Id.* at 9.

## *Stage One: Assembling Personnel and Logistical Planning*

*Who should be involved in developing the index?  
What is our timeline?  
What will the ultimate work product be?*

The first step in developing an index is to assemble a team and set management benchmarks and deadlines. The key decisions to make at this stage are largely divorced from the substance of the index. There are thus few best practices to report; each development team’s timeline and ultimate work product will depend on goals of the index as well as the resources the team can bring to bear. YCELP lists this planning step as its second stage, but for reasons elaborated below, in a field with little consensus and no prior indices to build on, it will likely be necessary to wrestle with the *project’s* parameters (as distinct from those of the index) first.

The first key element of this stage is personnel. Regardless of exactly when in the process it occurs, developers should set aside time to consider who ought to be involved in the development of the index, and in what capacity. YCELP envisions two categories of personnel: the “core team,” and everyone else. The core team—also known as the “management team”—will take the lead in managing the construction of the index.<sup>13</sup> Importantly, the core team should ideally be staffed with individuals whose primary job is developing and running the index. The ultimate composition of any full-time management team for a policy index is flexible,<sup>14</sup> but this attribute of the core team is worth replicating to the extent resources allow.

To supplement the core team, YCELP recommends bringing in additional experts at critical junctures “to provide feedback and guidance on methodology, indicator and data selection, and preliminary results.”<sup>15</sup> YCELP breaks the experts down into three types, all of whom have valuable insight to share with the index’s developers:

*Subject area experts*, who have extensive background in the field and an awareness of the dynamics the index is trying to measure. In the world of prosecutorial performance, these could include academics, researchers, criminal justice reformers, judges, current or former prosecutors, defense attorneys, and formerly incarcerated people.

*Methodological experts*, who can provide the most current knowledge on index construction. Data scientists or statisticians are solid candidates for this category.

---

<sup>13</sup> *Id.* at 21.

<sup>14</sup> The core team running the EPI—staffed jointly between YCELP and Columbia’s Center for International Earth Science Information Network—consists of “vision leaders,” project managers, research staff, analytical staff, and administrative staff. *Id.*

<sup>15</sup> *Id.* at 22.

*Target audience experts*, who are willing to serve as bellwethers for initial results, and/or have knowledge of target audience populations such that they can gauge the resonance and relevance of the index.<sup>16</sup> These might be election fundraisers, academics, community advocates, or even voters themselves.

The optimal role of outside experts will depend on the index's goals, the kinds of data it will incorporate, and the availability of data, but YCELP suggests soliciting expert feedback on both the goals of the index as well as the indicators (discussed in detail in Stage Three).<sup>17</sup> Because expert feedback will likely be necessary throughout the process, it may be wise to form a standing (or at least standby) panel of experts from each of these categories who can follow the project through its various development stages. In addition to continuity, establishing a standing panel likely would foster a more transparent working environment than would seeking input from experts on an ad hoc basis.

In addition to personnel decisions, YCELP recommends dedicating energy at this stage to laying out a more definite process and timeline for the development of the index.<sup>18</sup> Developers should set preliminary goals for the kinds of work products they hope to produce and when they hope to complete them. Of course, these projections will change based on the ultimate structure of the index and any obstacles that arise during development, so the exact timing of this stage is highly flexible. Still, setting a timeline can help keep the project from spiraling out of control, and likely will help uncover any obstacles sooner rather than later. Ultimately, the primary takeaway from this stage is the importance of setting aside time in the midst of substantive decision making to contemplate logistical issues and make a realistic plan.

## ***Stage Two: Defining the Index's Goals and Audience***

<p><i>What are we trying to accomplish with this index?</i> <i>What is the public policy knowledge gap we are trying to fill?</i> <i>Who do we hope to reach? How do we hope to influence them?</i></p>
---

The first substantive step in developing a public policy index is to set the foundation for the entire process by defining the index's goals and audience. This is the big-picture stage, and the answers to the key questions here will set the parameters for much of the index-development process.

Perhaps the most intuitive—and daunting—task in the development process is to lay down the goals of the index. Developers must first ask themselves what they hope to accomplish. Potential goals might include measuring a public policy phenomenon; raising awareness about an issue; comparing the performance of two or more entities; framing an issue along certain terms; providing “decision making and management

---

<sup>16</sup> *Id.*

<sup>17</sup> *Id.* at 25 (diagram of timing of expert consultations).

<sup>18</sup> *Id.* at 25-28.

tools” to policymakers; or some combination thereof.<sup>19</sup> The goals might also be external to the index itself; perhaps developers hope to use the index to spur policy reform or expand a research institution’s notoriety.<sup>20</sup> A single index might be designed to accomplish all of these goals, but zeroing in on a few specific ideas for what the index will add to a public policy conversation can help guide developers later in the design process as they confront more complex, contentious questions.

Of course, the goal of a given index may change over time. For example, what begins as a way to inform observers can morph into a tool designed to help authorities craft policy solutions. Before the EPI, for example, YCELP had designed and managed the Economic Sustainability Index (ESI). When they transitioned from the ESI to the EPI in the early 2000s, their goal changed: “Rather than providing a relative measure of sustainability for countries . . . , we changed the focus of the EPI to measure country performance against absolute targets for which countries could take policy action.”<sup>21</sup>

Notwithstanding the potential for change over time, writers reflecting on index development agree that clarity is critical in this goal-setting stage. YCELP emphasizes the “importance of a clearly defined overarching goal for the design and communication of an index.”<sup>22</sup> Likewise, the developers of the Commitment to Development Index (CDI) learned from their experience that “[c]larity of concept and purpose is essential.”<sup>23</sup>

A prosecutorial rating system could have one of several goals. Developers might seek to ensure that underperforming (or otherwise unsatisfactory) prosecutors are voted out of office and replaced by better ones. At the same time, they might prioritize using their index to shift popular understanding of what “good” prosecution looks like. Alternatively, developers might endeavor simply to hold prosecutors accountable to voters’ idiosyncratic policy preferences, whatever those preferences may be (for example, “toughness on crime”).

A second key element of this stage is defining the index’s audience. Developers should clearly establish to whom they hope to communicate the index’s information. Just as an index can have multiple goals, it can have multiple audiences. Take, for example, the World Wildlife Fund’s Environmental Paper Company Index (EPCI). The biannual index gives individual paper producers a score out of 100 for the sustainability of their production practices.<sup>24</sup> The EPCI describes itself as a “tool for companies who are striving for continual improvement of their environmental performance for pulp and paper

---

<sup>19</sup> *See id.* at 14-15.

<sup>20</sup> DAVID ROODMAN, CENTER FOR GLOBAL DEVELOPMENT, BUILDING AND RUNNING AN EFFECTIVE POLICY INDEX: LESSONS FROM THE COMMITMENT TO DEVELOPMENT INDEX 12 (2006), <http://www.cgdev.org/content/publications/detail/6661>.

<sup>21</sup> MEASURING PROGRESS, *supra* note 11, at 15.

<sup>22</sup> *Id.*

<sup>23</sup> ROODMAN, *supra* note 20, at 12.

<sup>24</sup> *See Environmental Paper Company Index 2015: Resolute Forest Products*, WORLD WILDLIFE FEDERATION, <https://epci.panda.org/results/resolute-forest-products> (last visited Sept. 5, 2017).

production and want to be recognized publicly for their leadership in transparency.”<sup>25</sup> But at the same time that it “[a]ims to motivate and reward company efforts towards sustainability,” it also serves to “inform[] buyers of paper, investors and other stakeholders that [a] company takes environmental responsibility seriously.”<sup>26</sup> The EPCI thus has at least two audiences: paper producers and—perhaps secondarily—paper consumers. To accomplish its goals, the EPCI has to be a useful source of information to both groups.

The diverse potential goals of a prosecutorial rating system lend themselves to different target audiences. An index that focuses on giving voters the best information possible implies a limited audience—voters. By contrast, an index that aims to get poorly rated prosecutors voted out could be targeted at anyone who might be involved in prosecutorial elections—from voters to candidates to interest groups. Where reformers’ ability to influence voters directly is uncertain, it may be necessary to consider a broader audience.

An element of pragmatism is likely required at this stage; it may be necessary to consider an intermediate audience (or “transmission” audience) to communicate the index to the target audience if the developers cannot connect to them directly. For example, voters and potential candidates might not frequently consult the Stanford Criminal Justice Center for guidance, but they may be receptive to media outlets, academics, donors, or interest groups. Which transmission audience is best will likely depend on an index’s target audience and the types of data available. Developers should also recognize that the index may garner secondary or tertiary audiences: A rating system that aims to motivate voters to unseat poorly rated prosecutors might also be used as a roadmap and impetus for prosecutors to improve their offices.

At any rate, developers should aim to be as specific as possible at this stage. Just as with defining the index’s goal, having a clear picture of the audience in mind early on can help guide the rest of the development process. YCELP instructs that “[s]etting a clear boundary for an index’s audience can help to act as a filter for what is pertinent in the selection of data and indicators.”<sup>27</sup> Developers will want to identify as precisely as they can what the target audience does not know, and what they hope it will do with the information the index communicates.

---

<sup>25</sup> *Id.*

<sup>26</sup> *Id.*

<sup>27</sup> MEASURING PROGRESS, *supra* note 11, at 17.

## *Stage Three: Designing the Index*

*What exactly will the index measure? What inputs will we include?  
Are we concerned primarily with outcomes? Policies? Both?  
Do we have a set of criteria each input must meet before it can be included?  
How much weight should each input get?  
How does the nature of the inputs limit how we can display the results?*

Perhaps the most difficult and contentious stage of index development is the stage YCELP places third: actually doing the work of designing the index. Although in practice it may not come at precisely this juncture, it can be recognized as the stage at which the most critical policy decisions behind an index will need to be made.

*Identifying Indicators and Measuring Amorphous Concepts.* The first, and perhaps the most daunting, task of the stage is to determine what the index will measure; that is, what inputs will go into it. YCELP refers to these inputs as “indicators.”<sup>28</sup>

This stage is daunting because the universe of possible indicators is extremely broad. Developers can measure particular *outcomes* (e.g., conviction or crime rates), or *input variables* (e.g., whether an office has adopted a particularly favorable policy). To make the process as rigorous as possible, developers should take care to challenge pre-conceived notions of what belongs in a policy index. Even if the inquiry leads to the conclusion that the universe of legitimate indicators is so confused that it makes more sense to measure only one dimension (e.g., how punitive a prosecutor’s office is) rather than overall performance, it is still worth understanding the full range of options available.

The simplest way to start selecting indicators is to use an established policy framework or guiding document for reference. For example, the Freedom House’s Freedom in the World Index rates each country based on the political rights and civil liberties available to its citizens. Those two umbrella categories include seven subcategories, which themselves are derived from the United Nations’ Universal Declaration of Human Rights.<sup>29</sup> Along these lines, developers of a prosecutorial rating system might choose to model their indicators off the policies set forth by the Brennan Center or the NDAA.<sup>30</sup>

---

<sup>28</sup> *Id.* at 31. The term is somewhat self-explanatory: At this stage, index designers must ask themselves what would *indicate* that a particular entity is succeeding at whatever the index is evaluating. To name a few: What would indicate that a country is having a positive impact on the environment? What would indicate that a developed country is committed to global development? What would indicate that a particular U.S. State has the freest, most accessible election process possible? Or, in the case of a prosecutorial rating system, what would indicate that a particular prosecutor’s office is performing well?

<sup>29</sup> FREEDOM HOUSE, FREEDOM IN THE WORLD: AGGREGATE AND SUBCATEGORY SCORES, <https://freedomhouse.org/report/freedom-world-aggregate-and-subcategory-scores> [hereinafter FREEDOM IN THE WORLD].

<sup>30</sup> See *supra* notes 5-10 and accompanying text.

In the absence of a guiding document or policy framework, index developers must themselves decide which policy priorities to include as indicators. YCELP provides a number of established analytical frameworks that can guide the process of deciding which indicators to include and how to include them.<sup>31</sup> In addition, to be included in the EPI, each indicator must meet several criteria.<sup>32</sup>

The developers of other indices have taken a more organic approach to indicator selection, resulting in several approachable models.

The developers of the Commitment to Development Index (CDI) chose their indicators by committee, according to a few guiding principles: Each indicator proposed had to be (1) available for all the countries covered in the index; (2) reasonable to collect data on with a relatively low budget (for the whole project, about \$200,000 per year); and (3) defensible as an indicator of commitment to development.<sup>33</sup> The developers and their team of experts wrote up rationales for their list of indicators, and after internal review, the core team made final design decisions.

Reconciling the sometimes opposing views about what to include in the CDI was a major stumbling block for its developers. The CDI seeks to evaluate developed countries' policies related to "development in poorer countries," an admittedly "broad and poorly defined outcome"<sup>34</sup>—not unlike prosecutorial performance. At the outset, they knew their goal would be to "do something that is analytically impossible in a way that is analytically credible."<sup>35</sup>

The major data gaps the CDI developers faced forced them to make difficult decisions about which indicators to include. And in wrestling with those decisions, they found guidance in focusing on their audience and ultimate goals. For example, the developers knew that global North-South migration was having an effect on some developing countries, but there was no evidence to show exactly *how* migration policies were impacting development.<sup>36</sup> The CDI developers reasoned that "as a matter of public education," it was important to include something about migration in the index.<sup>37</sup> By the time the developers had debated the issue and consulted with experts in the field, the specific measure of migration policy that ended up in the index was "reasonable . . . but

---

<sup>31</sup> MEASURING PROGRESS, *supra* note 11, at 31-33.

<sup>32</sup> The EPI criteria for indicators are (1) Relevance: "The indicator tracks the environmental issue in a manner applicable to countries under a wide range of circumstances"; (2) Performance orientation: "The indicator provides empirical data on ambient conditions or on-the-ground results for the issue of concern, or it is a "best available data" proxy for the outcome measures"; and (3) Established scientific methodology: "The indicator is based on peer reviewed scientific data, data from the United Nations or other institutions charged with data collection." *Id.* at 39. The guide also lists criteria for including a particular set of *data* in the EPI. *Id.*

<sup>33</sup> ROODMAN, *supra* note 20, at 4.

<sup>34</sup> *Id.* at 2.

<sup>35</sup> *Id.*

<sup>36</sup> *Id.* at 3.

<sup>37</sup> *Id.*

debatable.”<sup>38</sup> Rather than preserving the pure accuracy the index by omitting an imperfect measure of migration—a key dynamic in global development—the developers opted to hew toward their goal of educating their audience by providing a more complete, if imperfect, picture.

In the end, embracing the notion that “every design choice in the CDI implies a stand,” and being transparent about their choices, helped developers move forward.<sup>39</sup>

Another way to determine which inputs to include is to ask the audience directly. This tactic can be especially useful when either data is lacking or there are legitimate debates among experts about how to measure an amorphous concept

Heather Gerken recommends that the developers of her proposed Democracy Index engage in “politicking” in the process of indicator selection. As a way to preempt challenges to the utility and resonance of the completed Index, she suggests seeking input directly from the audience. This might include using “polling and focus groups to be sure that they are including metrics that appeal to most voters”; asking “seasoned politicians whether including this or that metric in the Index will trip it up politically”; and “focusing on measures that [election] administrators can live with.”<sup>40</sup>

Gerken recognizes that some developers may be hesitant to subject the development of an index to popular perceptions. Indeed, it may give some experts pause to select the indicators of prosecutorial performance based on what a bare majority thinks prosecutors’ offices should be doing. But it is at least one strategy for breaking any impasses among experts or shedding new light about what the audience might actually value in an index. Regardless of exactly how audience impact is solicited, Gerken’s recommendations serve as a reminder that indices are not developed in a vacuum.

A final approach to selecting an index’s indicators relies on radically narrowing the index’s scope. Where there is dissent among experts about which indicators are most important, or where developers determine that several indicators—though meaningful measurements on their own—cannot be aggregated coherently, it may be informative to tell the audience about an entity’s performance along a single metric.

This approach could be especially fitting in the prosecutorial performance context. Despite the variety of proposals legal scholars have put forth to measure prosecutorial performance, “[t]here is nothing remotely approaching a consensus. And even if scholars could agree about what information prosecutors’ offices should be disclosing, it is not clear the public would care about it.”<sup>41</sup> When faced with such an impasse, it may make

---

<sup>38</sup> *Id.*

<sup>39</sup> *See id.* at 5.

<sup>40</sup> HEATHER K. GERKEN, *THE DEMOCRACY INDEX* 122 (2009). Note that this list of potential stakeholders includes but extends beyond the “audience experts” that YCELP recommended be included as regular consulting experts. *See supra* note 16 and accompanying text.

<sup>41</sup> Sklansky, *Nature and Function*, *supra* note 4, at 516.

sense to avoid making a normative statement about what prosecutors' offices should do, and instead provide a descriptive assessment of one facet of their performance. And if experts cannot agree on which of the proposed single metrics is best, developers might also choose to report several individual metrics, without attempting to compile or weigh them against one another.

Several legal scholars have argued for such an approach. Ronald Wright and Marc Miller have suggested using offices' as-charged conviction rate.<sup>42</sup> Tracy Meares has similarly recommended measuring prosecutors' charge-to-conviction ratio.<sup>43</sup> Russell Gold has suggested using a jurisdiction's total expenditures, including prosecution, public defense, and incarceration.<sup>44</sup> And Stephanos Bibas has suggested a survey system based on evaluations of prosecutors by "[f]ellow prosecutors, judges, defense counsel, defendants, victims, and jurors."<sup>45</sup>

*Weighting Indicators.* In addition to deciding what exactly an index will measure, developers must determine what weight each indicator deserves. Put simply, this task involves laying out how much a certain quality matters to a jurisdiction's overall performance.

Developers have two options for determining the weight to give each indicator. They can make each indicator count equally or they can decide to weight certain indicators more than others. The decision ultimately rests on what developers want to communicate with the index. And lest it seem from this discussion that giving indicators equal weight is the more rational, objective option, it is important to remember that even that decision would require input from experts.

The developers of the CDI ultimately decided to give each of their seven factors equal weight. A primary motivator for that decision was the fact that those who opposed equal weighting could not reach a consensus on the appropriate weights to give each indicator, if not to weight them equally.<sup>46</sup> Outside scholars addressed this problem by surveying development experts from around the world to solicit their preferred weights. Ultimately, the correlation between the equally weighted CDI and the one the scholars developed was more than 99%.<sup>47</sup> In addition, the developers of the CDI reasoned that weighting the indicators equally would be more intuitive, and allow for more striking visual representations.<sup>48</sup> The CDI's experience suggests that where there is no consensus

---

<sup>42</sup> Ronald Wright & Marc Miller, *The Screening/Bargaining Tradeoff*, 55 STAN. L. REV. 29, 33 (2002).

<sup>43</sup> Tracey L. Meares, *Rewards for Good Behavior: Influencing Prosecutorial Discretion and Conduct with Financial Incentives*, 64 FORDHAM L. REV. 851, 873-75 (1995) (suggesting the use of monetary rewards to minimize tactical overcharging and charge bargaining).

<sup>44</sup> Russell Gold, *Promoting Democracy in Prosecution*, 86 WASH. L. REV. 69, 108 (2011).

<sup>45</sup> Stephanos Bibas, *Prosecutorial Regulation Versus Prosecutorial Accountability*, 157 U. PA. L. REV. 959, 989 (2009).

<sup>46</sup> ROODMAN, *supra* note 20, at 6.

<sup>47</sup> *Id.*

<sup>48</sup> *See id.* at 4 (depicting a stacked bar graph with each indicator in a different color).

around which indicators are most important to performance, equal weighting may be the most natural option.

But it is not the only option. The American Council for an Energy-Efficient Economy (ACEEE) produces a State Policy Scorecard for Energy Efficiency, which “assesses state policies and programs that improve energy efficiency in our homes, businesses, industries, and transportation systems.”<sup>49</sup> Experts give each state a score based on its performance in six policy areas, and the scores are totaled to form an aggregate score. Utility and public benefits policies are worth 20 out of the 50 points possible, while building energy codes are worth 7 points, and appliance efficiency standards are worth only 2 points.<sup>50</sup> This organization is depicted in Figure 2. The ACEEE clearly determined that whether a state is succeeding at implementing energy efficient policies should depend largely on their utilities and public benefits programs. Unfortunately, the ACEEE does not explain why it gives so much weight to that policy priority, but it is not a stretch to assume that either data or expert judgment indicated that utilities and public benefits policies make the most difference in energy efficiency outcomes.

**Fig. 2: Excerpted Summary of State Scores for 2016 ACEEE**

Rank	State	Utility & public benefits programs & policies (20 pts.)	Transportation policies (10 pts.)	Building energy codes (7 pts.)	Combined heat & power (4 pts.)	State government initiatives (7 pts.)	Appliance efficiency standards (2 pts.)	TOTAL SCORE (50 pts.)	Change in rank from 2015	Change in score from 2015
1	California	15	10	7	4	7	2	45	1	1.5
1	Massachusetts	19.5	8.5	7	4	6	0	45	0	1
3	Vermont	19	7	7	2	5	0	40	0	0.5
4	Rhode Island	20	6	5	3.5	5	0	39.5	0	3
5	Connecticut	14.5	6.5	5.5	2.5	6	0.5	35.5	1	0
5	New York	10.5	8.5	7	3.5	6	0	35.5	4	3
7	Oregon	11.5	8	6.5	2.5	5.5	1	35	-3	-1.5
8	Washington	10.5	8	7	2.5	6.5	0	34.5	0	1
9	Maryland	9.5	6.5	6.5	4	5.5	0	32	-2	-3
10	Minnesota	12.5	4	6	2.5	6	0	31	0	0

**Source:** American Council for an Energy-Efficient Economy, Executive Summary: The 2016 State Energy Efficiency Scorecard 6 (2016), <http://aceee.org/sites/default/files/pdf/summary/u1606-summary.pdf>.

Stage Three is likely the point at which developers are most likely to get tangled up in difficult decisions. In considering what a rating system *should* measure, it is natural to think about what it realistically *can* measure, an inquiry which is discussed in detail in Stage Four. To the extent possible, it is important at this point to try to disentangle these two questions. Having an open (if someone idealistic) conversation about the message an index seeks to communicate to its audience, and reserving for later decisions about measurements that can communicate that message, will create a more robust index than focusing on the limitations of available data. Of course, data and measurement constraints are intimately related to what an index will communicate; but keeping the key

<sup>49</sup> AMERICAN COUNSEL FOR AN ENERGY-EFFICIENT ECONOMY, EXECUTIVE SUMMARY: THE 2016 STATE ENERGY EFFICIENCY SCORECARD 1 (2016) [hereinafter 2016 ENERGY EFFICIENCY SCORECARD], <http://aceee.org/sites/default/files/pdf/summary/u1606-summary.pdf>.

<sup>50</sup> *Id.* at 6.

questions distinct in each stage should help developers move forward when the input selection process is overwhelming.

### ***Stage Four: Gathering Data***

*Is the information we need already being collected and compiled?  
If so, can we access it? Is it reliable?  
If not, can we convince anyone to begin collecting the data? Can we collect  
it ourselves?  
Are there any other sources of information that allow us to measure our  
indicators indirectly?*

Once the indicators and weights of an index have been decided, developers can select the data that will populate the index. In an ideal world, there would be plenty of data available about the issues the index seeks to measure, and developers would simply need to pick the very best data sources to mine. YCELP emphasizes the importance of evaluating the quality of available data, and instructs index developers to make sure any data that goes into the index is both “valid (i.e., corresponds to the “real world”) and reliable (i.e., consistent over time and across observations).”<sup>51</sup> They provide a number of strategies to guide selection of data in order to produce a rigorous and reliable index. For example, actual observational data is preferable to that generated by algorithmic modeling.<sup>52</sup> Likewise, datasets that are regularly updated are better than those that are allowed to stagnate.<sup>53</sup>

But the extent to which *data evaluation* requires its own stage depends on how much data is available, and it is at this point that the EPI model becomes less helpful to the project of creating a prosecutorial rating system. The study of the environment spans several scientific disciplines, and technology allows scientists to measure physical occurrences in the environment on a near constant basis. In a field with comparatively little data—like the performance of local prosecutors’ offices—some proactive efforts may be necessary to generate the data required to populate an index.

*Existing Data.* The primary federal agencies that compile data on state prosecutions are the FBI and the Department of Justice’s Bureau of Justice Statistics (BJS). The FBI tracks the volume and rates of violent crimes and property crimes through its Uniform Crime Reporting program.<sup>54</sup> This data is available at the county level and is updated annually, but with a roughly two-year lag ; the latest full report available is for 2014. The BJS, meanwhile, has a wealth of meaningful data that could be mined. Unfortunately, that data is not consistently updated.<sup>55</sup> For example, BJS produces a dataset called State Court

---

<sup>51</sup> MEASURING PROGRESS, *supra* note 11, at 43.

<sup>52</sup> *Id.*

<sup>53</sup> *Id.*

<sup>54</sup> *Uniform Crime Reporting*, FBI: UCR, <https://ucr.fbi.gov> (last visited Nov. 14, 2017).

<sup>55</sup> BJS does, however, routinely update its Annual Survey of Jails and the National Corrections Reporting Program. See *Data Collection: Annual Survey of Jails*, BUREAU OF JUSTICE STATISTICS,

Processing Statistics that tracks criminal defendants from charging through disposition, but this dataset has not been updated since 2009.<sup>56</sup> BJS also produces a National Census of State Court Prosecutors that has lain dormant since 2007.<sup>57</sup>

Private organizations occupy some space in the field, but they rely primarily on data published by government entities and released in judicial opinions. Watchdogs' observations of prosecutorial misconduct in particular rely on published opinions, so they miss most of the misconduct that occurs at the trial level.<sup>58</sup> Misconduct records may be a suboptimal data source because if misconduct were to come to light en masse, it might produce enough of a scandal to get an elected prosecutor voted out of office even without rigorous data collection.<sup>59</sup> Data on exonerations is of higher quality, but too low density to be useful: The National Registry of Exonerations counts only 2,124 exonerations ever.<sup>60</sup>

Finally, some states have begun collecting their own data, to mixed results. The available data for prosecutorial performance varies dramatically from state to state. North Carolina, for example, releases very granular data through its Office of the Courts.<sup>61</sup> Researchers can discover, for example, the number of cases per felony that produced pleas to a lesser charge, pleas to no contest, convictions at trial, a finding of no probable cause, etc.<sup>62</sup>

California's data is much lower quality than North Carolina's. The California Attorney General's Office's Criminal Justice Statistics Center records primarily crime-centric data.<sup>63</sup> California does maintain an Offender Based Transaction Statistics (OBTS)

---

<https://www.bjs.gov/index.cfm?ty=dcdetail&iid=261> (last visited Nov. 14, 2017); *Data Collection: National Corrections Reporting Program (NCRP)*, BUREAU OF JUSTICE STATISTICS, <https://www.bjs.gov/index.cfm?ty=dcdetail&iid=268> (last visited Nov. 14, 2017). From these datasets, developers can derive information about incarceration rates, whether admission rates are trending up or down, and how that information breaks down along demographic lines.

<sup>56</sup> *Data Collection: State Court Proceeding Statistics*, BUREAU OF JUSTICE STATISTICS, <http://www.bjs.gov/index.cfm?ty=dcdetail&iid=282> (last visited Nov. 14, 2017).

<sup>57</sup> *Data Collection: National Census of State Court Prosecutors*, BUREAU OF JUSTICE STATISTICS, <http://www.bjs.gov/index.cfm?ty=dcdetail&iid=265> (last visited Nov. 14, 2017).

<sup>58</sup> See, e.g., *Registry Database*, CENTER FOR PROSECUTOR INTEGRITY, <http://www.prosecutorintegrity.org/registry/database> (last visited Nov. 14, 2017) (providing a searchable database of published opinions on prosecutorial misconduct).

<sup>59</sup> For an example of such a large-scale exposure, see Jordan Smith, *Anatomy of a Snitch Scandal: How Orange County Prosecutors Covered Up Rampant Misuse of Jailhouse Informants*, THE INTERCEPT (May 14, 2016, 6:57 AM), <https://theintercept.com/2016/05/14/orange-county-scandal-jailhouse-informants>.

<sup>60</sup> *Detailed View*, NATIONAL REGISTRY OF EXONERATIONS, <http://www.law.umich.edu/special/exoneration/Pages/detaillist.aspx> (last visited Nov. 14, 2017).

<sup>61</sup> See, e.g., *Case Activity Reports, Fiscal Year 2015-2016*, THE NORTH CAROLINA COURT SYSTEM, [http://www.nccourts.org/Citizens/SRPlanning/Statistics/CARReports\\_fy15-16.asp](http://www.nccourts.org/Citizens/SRPlanning/Statistics/CARReports_fy15-16.asp) (last visited Dec. 16, 2017).

<sup>62</sup> *Id.* [click on "Felony Case Activity Report" to access data].

<sup>63</sup> See *Criminal Justice Statistics Center*, STATE OF CALIFORNIA DEPARTMENT OF JUSTICE, <https://oag.ca.gov/cjsc/spereq> (last visited Dec. 16, 2017) (listing categories of information collected by the Criminal Justice Statistics Center).

program, which is overseen by DOJ, but the data collected by that system does not seem to be publicly available. The data is used in an annual report, published by the Judicial Council’s Office of Court Research, on whether there is any systematic sentencing bias based on race or ethnicity.<sup>64</sup> Unfortunately, the report only answers that question on a statewide (not county-to-county) basis, and does not make granular data available.<sup>65</sup> One bright spot in California’s data-collection efforts is the California Sentencing Institute’s data on incarceration<sup>66</sup>; it is more detailed and clearer than the information available through BJS.

*Devising Proxies.* In the absence of data directly measuring their indicators, developers must rely on their creativity to devise alternative ways of measuring the index’s indicators. Experts refer to these alternatives as “proxies.” Proxies are ways of getting at roughly the same information that direct observational data would provide.

Heather Gerken’s Democracy Index contains a detailed discussion of how proxies function in index development. The Democracy Index aims to measure the performance of local election administration in terms of three goals: whether (1) “every eligible voter who wants to register can do so”; (2) “every registered voter who wants to cast a ballot can do so”; and (3) “every ballot cast is counted properly.”<sup>67</sup> Gerken prefers to rely on hard data rather than qualitative assessments (more on this dichotomy below), but she acknowledges that “we don’t yet have the data we need.”<sup>68</sup> Indeed, Gerken argues that “[t]he most significant obstacle to creating an Index is the shocking dearth of data on how well our system is working.”<sup>69</sup> In 2007, political scientists, law professors, and election administrators held a brainstorming session at the Democracy Index conference to identify reasonable metrics that were already available.<sup>70</sup> Figure 3 provides a summary of the brainstorming session for just one of the proposed goals of the Index. It also provides a look into how proxies function:

---

<sup>64</sup> Letter from Martin Hoshino, Admin. Dir., Judicial Council of California, to Diane f. Boyer-Vine, Legislative Counsel, California State Assembly, et al, *Disposition of Criminal Cases According to the Race and Ethnicity of the Defendant: 2015 Report to the California Legislature as Required by Penal Code Section 1170. 45* (Nov. 20, 2015), available at [http://www.courts.ca.gov/documents/lr-2015-raceandethnicity-pc1170\\_45.pdf](http://www.courts.ca.gov/documents/lr-2015-raceandethnicity-pc1170_45.pdf).

<sup>65</sup> See CRIME IN CALIFORNIA, CAL. DEPT. OF JUSTICE 64 (2015) <https://oag.ca.gov/sites/all/files/agweb/pdfs/cjsc/publications/candd/cd15/cd15.pdf>? (describing the OBTS data).

<sup>66</sup> See *Adults*, CALIFORNIA SENTENCING INSTITUTE, <http://casi.cjcg.org/Adult> (last visited Dec. 16, 2017).

<sup>67</sup> GERKEN, *supra* note 40, at 29.

<sup>68</sup> *Id.* at 26.

<sup>69</sup> *Id.* at 111.

<sup>70</sup> *Id.* at 130-31.

**Fig. 3: Excerpted Chart of Proxies for Measuring Voter Registration Performance**

	[Voter] Registration
Performance goal	Every eligible voter who wants to register can do so
Relevant values	Convenience, integrity, accuracy
Possible proxies	<p>Survey data on how many voters encountered a problem registering</p> <p>Time it took to register (and success rate) based on testers or “Neilsen voters” or randomly tagged registration cards</p> <p>Number of voters who appeared to cast a ballot on Election Day but were told they were not registered</p> <p>Number of provisional ballots counted because registration was verified</p> <p>[ . . . ]</p>

**Source:** Heather K. Gerken, *The Democracy Index* 130-31 (2009)

Proxies understandably require deviating from a preference for hard data. Gerken notes that because of the dearth of data currently available to populate the Democracy Index, developers may need to rely on *subjective evaluations* (whether voters felt that they wasted a lot of time going to the polls) or *policy inputs* (what policies a locality has in place to minimize wait times) rather than hard numbers (data for how long each voter actually waited in line). But these imperfect measurements are not only more readily available than hard data, they may be more cost effective than generating data from scratch, at least in the near term. And they may be the best way to move forward and generate momentum around the development and widespread use of an index.

Indeed, it may be worth considering whether a fully subjective measurement might be a viable way forward. Freedom House provides a model for the whole-hog subjective evaluation approach. Its Freedom in the World Index relies on the expertise of more than 80 analysts, “who prepare the draft reports and scores, use a broad range of sources, including news articles, academic analyses, reports from nongovernmental organizations,

and individual professional contacts. The analysts score countries based on the conditions and events within their borders during the coverage period.”<sup>71</sup>

An expert-evaluation approach may not be as opaque as it sounds; although the ultimate determinants of a country’s score are scholars inside a conference room, the detailed questionnaire they use to reach their scores is available online.<sup>72</sup> Freedom House also prioritizes consistency and contemplation in their approach, which helps insulate it from accusations of bias: From year to year, a country’s score will typically change only if a disruptive event occurs which “warrants a decline or improvement (e.g., a crackdown on the media, the country’s first free and fair elections).”<sup>73</sup>

*Building Buy-In for Data Collection.* In describing how to implement the Democracy Index, Gerken confronts the problem of insufficient data head-on. She devotes an entire section of her book to possible strategies for obtaining the necessary data, or, as she calls it, “smoothing the way for data collection.” One critical step is getting buy-in from election administrators, especially secretaries of state. To achieve that buy-in, administrators must believe that collecting and providing better data is in their own interest. To that end, the developers of the Democracy Index, with the help of Pew, were working as of 2009 to develop a *preliminary* index to rank how well the states are doing on data collection.<sup>74</sup> The goal there is to raise awareness of the need for better data.<sup>75</sup> But, as Gerken acknowledges, awareness is unlikely to compel election administrators to hand over data about how their own systems are performing: “No one likes to be evaluated.”<sup>76</sup>

Gerken posits several ways in which harvesting and handing over additional data might be in administrators’ best interest, all of which may apply to prosecutors’ offices. First, data can serve as a shield for accusations of partisanship or bias within an office. Second, a ranked index (based on the data provided) can help show constituents that other jurisdictions are facing the same challenges. Finally, data-driven policymaking is “likely to be an irresistible force,” suggesting that “[w]hether or not the Democracy Index is assembled, we will see a push for data.” It is in administrators’ interest, Gerken argues, to be at the forefront of that push.<sup>77</sup> Emphasizing similar ways in which increased

---

<sup>71</sup> *Methodology: Freedom in the World 2016*, FREEDOM HOUSE, <https://freedomhouse.org/report/freedom-world-2016/methodology>.

<sup>72</sup> *Id.*

<sup>73</sup> *Id.*

<sup>74</sup> *Id.* at 112.

<sup>75</sup> The developers of the Quality Counts education assessment took a similar tack. At its inception, very few school systems collected the kind of information Quality Counts sought. The organization released a “necessarily imperfect” inaugural report anyway, thus setting a standard for the kind of information a system should be collecting. *Id.* at 32. Over time, “[the index] has prompted many school systems to collect precisely the information that the first version of Quality Counts had identified as important.” *Id.*

<sup>76</sup> *Id.* at 114.

<sup>77</sup> *Id.*

transparency and monitoring can help prosecutors do their jobs may help generate more data collection in the long term.

There is reason to be optimistic that prosecutors may be open—or at least not openly hostile—to aiding in the development of a prosecutorial rating system. Both the NDAA and the Brennan Center’s existing studies on prosecutorial virtues were driven by former and current prosecutors.<sup>78</sup> They agree that conviction rates are a bad measure of performance, and advocate much more reformist measures than might have been expected. Many of the measurements that reform-minded scholars have advocated are echoed in the NDAA’s reports. While the research branch of the NDAA is likely to be more progressive than many local prosecutors,<sup>79</sup> there is reason to hope that resistance from prosecutors would at least not be uniform.

Ultimately, the actual selection of data is a stage at which it would be especially important to bring in data scientists and others with quantitative experience. They will likely bring to bear several established schools of thought about how best to use what data are available and identify gaps that need to be filled.

### ***Stage Five: Constructing the Index***

*Do our data sources have uniform measurement scales? If not, how do we normalize them?*

*What are the targets we should expect jurisdictions to shoot for? If there are none, how can we compare jurisdictions to one another?*

*How do we aggregate our various data inputs?*

*How do we categorize our various indicators and ultimately display our result?*

Once developers have selected all the indicators they want to include, and all the sources of data to populate them, the work of actually calculating and building the index can begin. In a sense this is the most straightforward step; in the words of high school math teachers everywhere, all that’s left to do is “plug and chug.” YCELP cautions, however, that “the methodology of indices is far from mechanical.”<sup>80</sup> Statistical results, after all, require interpretation by experts. And, as mentioned in Stage Three, benchmarks for success are often debatable.

One of the deceptively simple tasks to perform at this stage is normalizing or aggregating data across the indicators. For example, imagine that developers have data for how often

<sup>78</sup> BRENNAN CENTER REPORT, *supra* note 8, at 9-11; NDAA REPORT, *supra* note 5, at 21-22 .

<sup>79</sup> See, e.g., Josh Keller & Adam Pearce, *This Small Indiana County Sends More People to Prison than San Francisco and Durham, N.C., Combined. Why?*, N.Y. TIMES (Sept. 2, 2016), [https://www.nytimes.com/2016/09/02/upshot/new-geography-of-prisons.html?\\_r=0](https://www.nytimes.com/2016/09/02/upshot/new-geography-of-prisons.html?_r=0) (quoting one local prosecutor, who is responsible for an exorbitantly high incarceration rate in rural Indiana, as saying, “My constituents are the people who decide whether I keep doing my job. The governor can’t make me. The legislature can’t make me”).

<sup>80</sup> MEASURING PROGRESS, *supra* note 11, at 55.

in a given year a prosecutor’s office sought the harshest sentence possible. Suppose they also have data for how often cases resulted in plea agreements. Those two metrics are not automatically comparable. Assuming both measurements should be included as indicators of prosecutorial performance, developers have to give the office a uniform score for each value to be able to aggregate them into the larger score.

One way to normalize scores across measurements—and the method YCELP uses for the EPI—is to set an ideal or “target” value for each indicator and measure an entity’s performance as a percentage of that target. YCELP thus assigns scores between 0 and 100 for each country based on how close they come to the target for each indicator.<sup>81</sup> With all the indicators then normalized to a 100-point scale, they can be more readily combined or averaged.

The developers of the CDI decided not to normalize most of their indicators against an external ideal value.<sup>82</sup> Instead, they build normalized scores around the *average* value for each indicator of a developed country’s commitment to development.<sup>83</sup> Those countries that are performing at the average are assigned a score of 5 out of 10, those that perform at twice the average get a 10, and those that perform even better score higher than 10. Correspondingly, countries doing twice as bad as average get a 0 out of 10.

The task is substantially more difficult when there is no expert consensus on the ideal value for a given indicator, or the scale is not linear. (For example, developers may not want prosecutors to seek the highest possible sentence in 100% of cases, but they may not want them to do so in 0%, either.) In such cases, giving jurisdictions a more subjective score, at least for certain indicators, may make more sense. Developers might thus make a policy call as to the “optimal” value of an indicator, and score jurisdictions on a bell curve around it. Of course, transparency about such policy judgments would be key to communicating the index effectively (as discussed in Step Six).

Where jurisdictions are given more subjective scores, normalization can be even simpler. For example, for the State Policy Scorecard for Energy Efficiency, experts give each state a score based on its performance in each of six policy areas, and the scores are totaled to form an aggregate score.<sup>84</sup> Importantly, although the ACEEE displays a map with each state’s overall ranking, it does not list the state’s total score without a breakdown of its scores in each policy area. This transparency helps observers understand not only how each state is performing in each area, but also how much weight the index gives to the various policy areas. Similarly, Freedom House’s Freedom in the World index also

---

<sup>81</sup> See *id.* at 56 (“For the EPI, the worst-performing country usually establishes the low performance benchmark for a particular indicator.”).

<sup>82</sup> ROODMAN, *supra* note 20, at 6 (“[T]he only indicators that are normalized relative to ideal maxima are binary ones on whether countries have ratified the Kyoto Protocol and certain other international treaties.”).

<sup>83</sup> Specifically, the CDI’s developers calculate the 21-country average and multiply by five. For a detailed discussion of the technical principles governing the CDI’s methodology, see *id.* at 5.

<sup>84</sup> 2016 ENERGY-EFFICIENCY SCORECARD, *supra* note 49, at 1.

publishes each country's scores for its seven subcategories "in order to generate debate and discussions within countries as to areas that are most in need of reform."<sup>85</sup>

### ***Stage Six: Disseminating and Building Support for the Index***

*How do we communicate the results of the index in a way that is credible and useful?*

This final stage of index development does not appear in YCELP's guide. It received such significant attention in other sources, however, that it deserves its own discussion. The primary lesson for this stage is that even an index with flawless methodology is only as good as its ability to actually influence people. Developers thus will have to work hard to communicate to their audience and other stakeholders that the index is useful, thorough, and fair. Unfortunately, there is no real blueprint for how to execute such a communication strategy.

Ultimately, the most effective communication and presentation method will depend on an index's audience. First, developers have to wrestle with where to reach their audience. If the audience is primarily industry actors—recall the Environmental Paper Company Index—a presence at trade events may be critical. Attending (or better yet, presenting at) academic conferences may be one way to engage experts. Op-eds are often a great way to bring the issue to a broad audience. And where voters are the primary audience, mass mailings or their online equivalent may be helpful.

Assuming the audience can be reached, developers then must decide how to frame the index's conclusions to have the most impact with their target audience. Some of this thinking likely will have been covered in the process of building the index (e.g., whether it makes sense to give a comparative title like "more free" or "less free"). But at this stage, even the wording of headlines may matter. Developers will want to consider whether leading with numbers will be more effective than a simple, powerful bumper sticker like "Prosecutor A is the worst in the state."

Regardless of the ultimate means of selling the index, a number of sources emphasize the importance of transparency in this stage. After all, the goal is not only that people be aware that the index exists; they must believe it is a useful tool for them to use. As Gerken notes, "[p]erhaps the most important thing that designers of the [Democracy] Index can do to sell the Index after it's assembled is to spend lot of time talking about the choices that go into the Index *before* it's assembled."<sup>86</sup> It should be clear by this point that the design and construction of an index will involve layer upon layer of value judgments.

---

<sup>85</sup> See FREEDOM IN THE WORLD, *supra* note 29.

<sup>86</sup> GERKEN, *supra* note 40, at 121.

Commentators agree that being transparent about those value judgments is critical to gaining buy-in for the index.<sup>87</sup>

## Conclusion

Local prosecutors wield exceptional influence in the American criminal justice system. Their decisions can shape everything from the outcome of an individual sentencing hearing to the way law enforcement officers engage with their community. Yet despite prosecutors' central role, the elections in which they are chosen historically have attracted much less voter attention than other races, and incumbents are seldom challenged.

In recent year, however, criminal justice reformers have increasingly sought to take advantage of prosecutors' influence, and have focused their efforts on recruiting and electing reform-minded prosecutors.<sup>88</sup> A number of organizations are likewise dedicated to supporting newly elected prosecutors as they pursue reforms within their own offices.<sup>89</sup>

Modeled on indices developed for other public policy fields, a prosecutorial rating system is one way for these reformers to break through the opacity surrounding prosecutorial elections and enable voters to better hold their elected prosecutors accountable. At the same time, a prosecutorial scorecard would enable sitting prosecutors to better understand where their offices need the most work.

This report aims to support reform efforts by outlining a process for developing such a rating system. The best practices identified above are ripe for adaptation to the process of building a prosecutorial scorecard. At each stage in the development process, developers will have to wrestle with key questions about how to measure prosecutorial performance and how best to convey the index's findings. Fortunately, when it comes to the difficult policy choices that underlie an index, there are very few wrong answers; developers have the ability to tailor the index to their unique objectives. Relying on the lessons other index developers have learned will help guide developers of a prosecutorial scorecard as they seek build the most effective index to meet their reform goals.

---

<sup>87</sup> See *id.*; ROODMAN, *supra* note 20, at 5 (“The principle driving our decisions was that in the face of ignorance about the “true” parameters, not to mention the functional form of the development “production function,” it is best to be transparent and minimally arbitrary to avoid seeming biased.”).

<sup>88</sup> See, e.g., Maryclaire Dale, *Civil Rights Champion Poised To Be Philadelphia DA*, ASSOCIATED PRESS (May 17, 2017), <https://apnews.com/839d1da7f40f441abf5859cd526b1858> (noting reform organization Color for Change's support of Larry Krasner's candidacy Philadelphia District Attorney).

<sup>89</sup> See, e.g., *Our Work and Vision*, FAIR AND JUST PROSECUTION, <https://fairandjustprosecution.org/about-fjp/our-work-and-vision> (last visited Dec. 14, 2017).