

Toward Metrics for Re(imagining) Governance:

The Promise and Challenge of Evaluating Innovations in How We Govern

By Aleise Barnett, David Dembo and Stefaan G. Verhulst

GovLab Working Paper

V. 1 Date April 18, 2013

Toward Metrics for Re(imagining) Governance:

The Promise and Challenge of Evaluating Innovations in How We Govern

By Aleise Barnett, David Dembo and Stefaan G. Verhulst

GovLab Working Paper

V. 1 Date April 18, 2013

MacArthur
Foundation

This paper was made possible with support from the
John D. and Catherine T. MacArthur Foundation

TABLE OF CONTENTS

I. Moving from faith-based to evidence-based interventions	4
II. Learning from evaluation: metrics and methods	5
A) Conceptual Framework	5
B) Metrics and indicators	6
C) Methodology	7
D) New trends and tools	8
III. Evaluating innovations in governance	9
A) Conceptual framework	9
B) Metrics and indicators	10
C) Methodology	11
IV. Conclusion and questions moving forward	12
Open bibliography	14

Please help us deepen, expand, and improve this version by
annotating this draft at <http://bit.ly/154CZQP>

TOWARD METRICS FOR RE(IMAGINING) GOVERNANCE:

The Promise and Challenge of Evaluating Innovations in How We Govern

By Aleise Barnett, David Dembo and Stefaan G. Verhulst ¹
GovLab Working Paper
V. 1 Date April 18, 2013

I. MOVING FROM FAITH-BASED TO EVIDENCE-BASED INTERVENTIONS

Governance—how institutions analyze information and make decisions to solve collective problems—is broken. Around the world, we face increasingly complex challenges ranging from widespread poverty to serious ecological crises that threaten our planet’s future. Yet trust in traditional institutions of governance is at an all-time low.

At the same time, we are living through the greatest era of disruptive innovation and rapid experimentation since the Industrial Revolution. Tremendous progress in information and communication technologies, including big data and social media, are empowering individuals to engage with one another—and with traditional institutions of governance—to tackle problems collectively. Groups of individuals with diverse social, intellectual, and professional backgrounds can now use technology to collaborate in new ways that can drive progress more rapidly and effectively than ever before. From local and federal governments to leading universities and Fortune 500 companies, institutions have an opportunity to reevaluate how they solve problems in the networked age.

While there is good reason to believe that breakthroughs may come from recent innovations such as community-based problem solving, behavioral economic insights about human behavior, or predictive analytic experiments, there are limited studies measuring exactly how productive it is to use these kinds of new governance techniques. Without a deeper understanding of whether, when, why and to what extent an intervention has made an impact, any initiative we design will be sub-optimal and will produce less than the desired results. If we are going to accelerate the rate of experimentation in governance and create more agile institutions capable of piloting new techniques and getting rid of ineffectual programs, we need research that will enable us to move away from “faith-based” engagement initiatives toward “evidence-based” ones.

This GovLab Working Paper is an attempt to describe the promise and challenge of evaluating new practices of problem solving in governance, specifically citizen-engagement interventions. The purpose is to inform those innovators who are eager to develop a deeper insight into what works for the key steps, opportunities and challenges involved when assessing impact. In addition, the paper aims to develop a foundation for further interdisciplinary research and debate.

This paper summarizes the recent lessons learned in the field of evaluation and examines how new methods could help assess the impacts of participatory governance and emerging social technologies used in governance. This work is not meant to be a comprehensive review of current approaches to evaluating new governance initiatives but is intended to frame the issues involved, and suggest where work needs to be done to develop better assessments. We are also providing an annotated and open bibliography for further reading, inviting others to expand upon it.

1. The authors are grateful to Joel Gurin for editorial support and Beth Noveck and Hania Farhan for their substantive review.

II. LEARNING FROM EVALUATION: METRICS AND METHODS

Before we try to change how we solve problems and plan new initiatives we need to know what has worked already and what hasn't. Over the past few decades there have been renewed calls to measure the impact of societal initiatives in order to allocate resources wisely and understand what to continue, scale up or cancel. With government and philanthropic funding shrinking, both government agencies and charities see outcome-based evaluation as an urgent need. There has also been an increased demand to understand with more precision and nuance whether programs are successful, how they are successful and what factors have led to that success. Even when an initiative demonstrates positive outcomes, it is still helpful to understand:

- Whether there are different impacts for some populations and objectives than for others.
- How much better off people are as a result of the initiative—only slightly, or significantly?
- Whether a program works differently in different geographical regions.
- Why a program has been effective and what factors have contributed to its success.²

It can also be helpful to understand what *doesn't* work, despite the incentive to report only successful elements of a program in order to maintain funding. Engineers Without Borders Canada, for instance, produces an annual Failure Report which shares learnings about less effective innovations in a step to “create a culture of creativity and calculated risk taking.”³ With a clear understanding of both successes and failures, more effective and more innovative initiatives can be designed and better opportunities for intervention can be identified.⁴

Evaluation is a complex and challenging task that requires judgment to choose how, what and when to measure along with the criteria for an intervention to be deemed successful. Our review of the existing literature on evaluation⁵ has highlighted several issues, challenges and approaches.

A) CONCEPTUAL FRAMEWORK

To determine an intervention's impact, an evaluation must be based upon a concept or theory of change. In order to understand the theory of how and why an initiative will work, it is helpful to devise a conceptual framework, or logic model. The logic model makes explicit the relationships among resources available to implement an intervention, activities planned, and sought-after results. It also theorizes how the results, or outputs, of the initiative will lead to both short-term beneficial outcomes and longer-term, fundamental impact.

The figure below depicts an example of a logic model for an initiative dedicated to reducing the incidence of diabetes. Part of the initiative's theory of change is that the morbidity of diabetes can be reduced by offering faith-based nutrition and exercise courses encouraging people at risk for diabetes to make healthier lifestyle choices. However, there are assumptions at each stage of the logic model: for example, the assumption that people will want to participate in the program or that participating in the program will lead to a change in knowledge or behavior. It is crucial to acknowledge the assumptions of the logic model and assess their accuracy, since flawed assumptions will lead to misleading conclusions.

2. Carol Weiss, *Evaluation: Methods for Studying Programs and Policies*, 2nd Edition (Englewood Cliffs, NJ: Prentice Hall, 1997).

3. “Failure Reports,” Engineers Without Borders Canada, accessed March 14, 2013, <http://legacy.ewb.ca/en/whowere/accountable/failure.html>.

4. Margaret C. Plantz, Martha Taylor Greenway, and Michael Hendricks, “Outcome Measurement: Showing Results in the Nonprofit Sector,” *New Directions for Evaluation* 1997 (1997): 15–30, accessed March 17, 2013, doi: 10.1002/ev.1077.

5. Julia Abelson and Francois-Pierre Gauvin, “Assessing the Impacts of Public Participation: Concepts, Evidence, and Policy Implications,” Canadian Policy Research Networks, March 2006; James Cutt and Vic Murray, *Accountability and Effectiveness Evaluation in Nonprofit Organizations* (New York: Routledge, 2000); and Weiss, *Evaluation: Methods for Studying*.

Figure 1: Bronx Health REACH ⁶ 2010 Logic Model ⁷



Funding	Training counselors	# of churches participating in peer counseling	Changes in knowledge, attitudes and behaviors of people at risk for diabetes	Decreased morbidity from diabetes in four zip code neighborhood in the South Bronx
Community volunteers	Recruiting participants	# of participants in peer counseling program		
Church leadership		# of counseling sessions		

B) METRICS AND INDICATORS

The selection of indicators and metrics for assessment is necessarily a value-based decision, since ultimately we measure what we deem important. A metric can be thought of as a numerical unit of measurement, such as ROI (return on investment), while an indicator is a metric tied to one or more targets, such as GDP (gross domestic product), so that it can measure the degree to which specific goals are being met.⁸ Indicators must be chosen with care: they become the feedback mechanism for allowing internal and external stakeholders to understand performance and how it is measured. Indicators also become points of reference that serve as part of the rationale for making decisions, both in the short- and long-term. They should be chosen in the context of the logic model for an initiative.

Indicators build on *outputs*, which are basic metrics of success in quantitative terms, such as number of trainings delivered or number of people participating in a program. Outputs themselves are basic indicators of a sort.

But more high-level indicators are also needed to measure how well an initiative achieves the short-term outcomes and long-term impact it is intended to.⁹

Different stakeholders may have different criteria for determining whether a program is successful. While some parties may be more interested in knowing whether certain outputs have been delivered, others may measure success only by the amount of impact an initiative made on the outcomes of interest.¹⁰ By consulting a variety of stakeholders in designing an evaluation strategy, researchers can confirm their agreement on a causal

6. Bronx Health REACH is a CDC-funded center to eliminate racial and ethnic disparities in diabetes and cardiovascular illness in southwest Bronx. The logic model portrayed here is for their Fine, Fit and Fabulous program, a 12 week diabetes prevention program teaching nutrition and fitness through group discussions and exercise sessions within a spiritual context.

See <http://www.bronxhealthreach.org/our-work/faith-based-outreach-initiative/fine-fit-and-fabulous/>

7. Carolyn Berry, lecture for “Explicating and Assessing Program Theory,” New York University, September 22, 2011.

8. Stefaan Verhulst, “ICT Indicators: Mapping Resources and Issues,” Markle Foundation, 2003.

9. Cutt and Murray, *Accountability*.

10. Abelson and Gauvin, “Assessing the Impacts.”

framework and establish agreed-upon indicators to measure success. The application of newer, more participatory and bottom-up design strategies provides new opportunities to fine-tune and improve the metrics used, reflecting what matters most for the targeted population.

C) METHODOLOGY

Once indicators have been established, data can be collected in a number of different ways. These methods include, for instance, direct observation, questionnaires, document review, focus groups, interviews, and examining administrative data. The best method or methods may depend on the program itself. For example, a questionnaire-based strategy won't work without access to the population participating in a program.

There are traditionally three broad categories of research designs to evaluate societal intervention programs:

- Experimental
- Quasi-experimental
- Non-experimental

Experimental designs use random assignment; researchers randomly assign one group of people to be program participants and another group to be non-participants, and compare results between the two groups to establish the effect of the program. Experimental designs are often commonly referred to as randomized controlled trials, or RCTs, and when done correctly are often thought of as the most rigorous research design.

For example, researchers may be interested in finding out whether participating in a microfinance program leads to higher savings on average. After obtaining a list of people who wish to participate in the program, they would randomly assign half of the people to the program group and the other half to the control group, who would not receive the program. The *only* difference between the control group and program group is access to the program. This means that researchers would be able to confidently conclude that it is the program which causes higher savings for the program group rather than some other characteristic or phenomenon.¹¹

Quasi-experimental designs may use a program group and comparison group which has not taken part in the program, or may simply consider one group, measuring the effects before and after that group receives the program. Quasi-experimental designs differ from experimental designs in that participants in the program and comparison group are NOT randomly assigned; rather a comparison group may be selected after a program group has been established. This technique is less rigorous because there is no certainty that the two groups are comparable across all characteristics; for example, those who choose to sign up for a microfinance program may be more motivated to save, and it may be this characteristic which ultimately explains their higher savings rather than the program.

Finally, *non-experimental* techniques, considered the least rigorous method for measuring program success, may involve a single survey delivered after a program in order to gain descriptive information.¹² Non-experimental techniques cannot accurately pinpoint the causes of the outcomes they measure. For example, if researchers using this design survey participants in a job skills development program and find that many of them have gotten jobs, they would not be able to determine whether this was simply due to a rise in the overall employment rate rather than the new skills this group developed.

Performance measurement is another assessment technique and can provide different insights than program evaluation. While outcome/impact evaluation measures the overall success of a program, performance measurement

11. Jonathan Bauchet et al., "Latest Findings from Randomized Evaluations of Microfinance," Consultative Group to Assist the Poor/The World Bank, December 2011, accessed April 1, 2013, <http://www.financialaccess.org/sites/default/files/publications/latest-findings-from-randomized-evaluations-of-microfinance.pdf>.

12. William Trochim, "Research Methods Knowledge Base," last modified October 20, 2006, accessed March 18, 2013, <http://www.socialresearchmethods.net/kb/>.

uses ongoing monitoring to establish whether operational goals are being met and delivering the desired outputs.¹³

Process evaluations are focused on determining whether a program has been implemented according to plan. Process evaluations consider one snapshot in time, while performance measurement is an ongoing monitoring process.¹⁴

D) NEW TRENDS AND TOOLS

Organizations looking to create social change are increasingly using more meticulous and comprehensive research methods to measure the impact of their work,¹⁵ using new devices and tools to gather evidence. The Impact Reporting and Investment Standards (IRIS), developed by the Global Impact Investing Network, for instance, can be used to describe an organization's performance from a social, environmental and financial perspective. IRIS allows organizations to understand their own performance. And by looking at different IRIS measures, stakeholders concerned about impact investing can estimate the aggregated impact of that sector.¹⁶

Another new device used for assessment in the development arena is a poverty scorecard, developed by Mark Schreiner, Director of Microfinance Risk Management, L.L.C. Individually created for target countries, these instruments are made up of 10 simple and observable indicators comprised of questions such as whether a household has a straw roof. Since many indicators of poverty are highly correlated, the scorecards are succinct because the answer to one question can imply the answer to another question. For example, in all likelihood a household with a straw roof will not have indoor plumbing.

In the poverty scorecard, indicators were selected to be objective, verifiable, quick to ask and answer, and liable to change over time so they can be used to track progress or negative trends. Each indicator was then assigned a point value, which can be used to calculate a poverty score. This is an easily implemented instrument which allows field workers to quickly assess poverty levels.¹⁷

Our review of the literature has also shown an increased call for qualitative methods rooted in ethnography. True ethnography focuses on understanding the perspective and culture of a group of people in order to understand their behavior. Quasi-ethnographic research methods intend to take into account people's experiences and perspectives to add a contextual dimension to research and evaluation.¹⁸ Small organizations that don't have the resources for traditional monitoring and evaluation may find it especially useful to solicit feedback and hear about the experiences of program participants. Ongoing feedback can also help organizations make changes and understand program accomplishments as they go.¹⁹

Finally, increased availability of real-time data along with enhanced analytical capabilities (often called big data) promises to radically change and improve how we assess outcomes and impact. When designed well, big data may allow practitioners to track progress and understand where existing interventions require adjustment much faster.

13. "Guide to Performance Measurement and Program Evaluation," Office for Victims of Crime Training and Technical Assistance Center, last modified January 2010, accessed March 17, 2013, <https://www.ovcttac.gov/taResources/OVCTAGuides/PerformanceMeasurement/pfv.html>; "Program Evaluation: Basic Information," United States Environmental Protection Agency, last modified June 14, 2012, accessed March 17, 2013, <http://www.epa.gov/evaluate/basicinfo/index.htm>.

14. Weiss, *Evaluation: Methods for Studying*.

15. Mark Benjamin, "The New Science of Measuring Impact," January 24, 2012, accessed March 17, 2013, http://poptech.org/e3_mark_benjamin.

16. "Impact Reporting and Investment Standards," accessed March 24, 2013, <http://iris.thegiin.org/>.

17. "Measuring Poverty and Poverty Scorecards," Unite for Sight, accessed March 24, 2013, <http://www.uniteforsight.org/global-health-university/poverty-scorecards>.

18. Alissa Gardenhire and Laura Nelson, "Intensive Qualitative Research: Challenges, Best Uses, and Opportunities," MDRC, April 2003, accessed March 24, 2013, <http://www.mdrc.org/intensive-qualitative-research>.

19. Suzie Boss, "Amplifying Local Voices," *Stanford Social Innovation Review*, Summer 2011, accessed March 24, 2013, http://www.ssireview.org/articles/entry/amplifying_local_voices1.

But in the words of a recent UN Global Pulse report on the use of big data to assess development goals, “With the promise come questions about the analytical value and thus policy relevance of this data—including concerns over the relevance of the data in developing country contexts, its representativeness, its reliability—as well as the overarching privacy issues of utilising personal data.”²⁰

III. EVALUATING INNOVATIONS IN GOVERNANCE

These basic considerations of evaluation—conceptual framework, metrics and indicators, and methodology—are critical in assessing any societal intervention. In addition, there are particular considerations in applying each of these to assessments of citizen engagement and data-sharing in governance. The following observations are presented to help establish best practices from the limited yet growing literature on assessing the impact of new governance initiatives.

A) CONCEPTUAL FRAMEWORK

In a recent paper, Fukuyama described the poor state of assessing governance worldwide, which he explains as resulting from the lack of any conceptual framework: “[W]e cannot measure what we cannot adequately conceptualize, we have to start with the concept first.”²¹

A related challenge in identifying a logic model or conceptual framework for governance innovation is linked with the diversity of goals underlying citizen engagement and data-sharing in governance. Goals may be as broad as enhancing democracy or as narrow as helping a specific agency operate more efficiently. Abelson and Gauvan, for instance, have this to say about goals of public participation and democracy generally:

*“Democratic theory tells us that public participation is undertaken for different purposes and with different underlying goals. Tensions exist between views of participation as an essential element of successful democracy (and inherently desirable in its own right) and participation as a means for achieving something else, be it a specific decision outcome, a desire for more informed, accountable or legitimate decision making, or perhaps to delay or share the blame for a difficult decision. Lying somewhere between is the desire for public participation to contribute to a more educated and engaged citizenry.”*²²

They go on to discuss how different parties to an intervention may have differing goals:

*“[D]ifferent evaluation perspectives... may exist among interested parties. For example, sponsors and taxpayers tend to be interested in value for money. But sponsors and organizers of public participation should also be interested in questions of efficacy and effectiveness (if the purpose is summative evaluation) and whether the public participation method implemented was successful as measured against its goals (to address a formative evaluation purpose). Participants themselves are increasingly interested in whether their involvement makes a difference (i.e., policy impact) and, as taxpayers, they also want to see that their involvement was meaningful given that investments in public participation are typically made at the expense of direct service and program delivery. These differing perspectives are integrally linked to the different underlying goals for public participation.”*²³

20. “Big Data for Development: Challenges & Opportunities,” UN Global Pulse, May, 2012, accessed April 1, 2013, <http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf>.

21. Fukuyama, Francis, “What Is Governance,” Working Paper 314, Center for Global Development, January 2013, accessed March 23, 2013, http://www.cgdev.org/files/1426906_file_Fukuyama_What_Is_Governance.pdf.

22. Julia Abelson and John Eyles, “Public Participation and Citizen Governance in the Canadian Health System,” in *Changing Health Care in Canada: The Romanow Papers*, vol. 2, ed. Pierre-Gerlier Forest, Tom McIntosh and Gregory Marchildon (Toronto: University of Toronto Press, 2004), 279-311.

23. Abelson and Gauvan, “Assessing the Impacts.”

Chess has identified three different approaches that are routinely taken to assess public participation:

“1) user-based evaluation, which assumes that different participants will have different goals and that the evaluation must take these different goals into account;

2) theory-based evaluation which is driven by theories and models of public participation and applies normative criteria universally to any public participation effort; and

3) goal-free evaluation which is not constrained by any stated goals and is conducted in the absence of any theory.”²⁴

B) METRICS AND INDICATORS

Selecting indicators and metrics of success for citizen engagement in governance is a value-laden process with abundant opportunity for misinterpretation.²⁵ It is not clear whether a universal framework for success can be achieved, but some have suggested that there is a need to develop a set of common standards for assessing citizen engagement initiatives.²⁶ Each initiative may be so different that the desired outcomes will vary widely depending on the program. However, it may be possible to develop thematic buckets within each subset of the participatory government space, defining a set of metrics for each area.

Some researchers assessing open government have suggested that rigid distinctions between quantitative and qualitative assessments are no longer helpful. The SUNY Albany Center for Technology in Government suggests that measures of the return on investment for government interventions should be judged only on whether they are “valid and useful in the relevant context of measurement.”²⁷

The Open Government initiative of the Obama administration provides a good example of the evolution in measurement and metrics needed in this field. To assess the success of this effort, the *Openthegovernment.org* organization developed agency evaluations using a scale of 1 to 3 to judge adherence to the directive. But that exercise is only able to determine whether agencies made their data available to the public. Missing from the directive itself, and from initial evaluation efforts, was any measurement of progress on the overall goal of the directive: to engage the public in governance. As these initiatives move forward, additional metrics are being developed to measure more substantive issues of participation. As the report on this study explained, “...directional policy without benchmarks or specific increments of improvement is akin to diagnosing fever without a thermometer.”²⁸

Several other organizations have been carrying out evaluations/assessments of open participatory government interventions, some of which already have case studies that give an indication of the effectiveness of various methods of evaluation. They include the following examples:

- The Coalition for Evidence-based Practice issued the report, “Which Study Designs Can Produce Rigorous Evidence of Program Effectiveness?” with an appendix that includes a number of case studies of interventions in government where RCTs have been effective.²⁹

24. Caron Chess, “Evaluating Environmental Public Participation: Methodological Questions,” *Journal of Environmental Planning and Management*, 43 (2000): 769-784, cited in Abelson and Gauvin, “Assessing the Impacts.”

25. Verhulst, “ICT Indicators.”

26. “Featured Conversation: Citizen Engagement,” Striking Poverty World Bank discussion forum, <https://strikingpoverty.worldbank.org/comment/360>.

27. Anthony Cresswell and Djoko Sigit Sayogo, “Developing Public Value Metrics for Returns to Government ICT Investments,” Center for Technology in Government, University at Albany, SUNY, June 2012, 3-4, accessed March 1, 2013, http://www.ctg.albany.edu/publications/reports/pubvalue_microsoft/pubvalue_microsoft.pdf.

28. Bertot et al., “Measurement of Open Government: Metrics and Process,” Paper presented at the 45th Hawaii International Conference on System Science (HICSS), January 4-7, 2012, accessed March 1, 2013, http://andromeda.rutgers.edu/~ncsds/spaa/images/stories/documents/Transparency_Research_Conference/Papers/Smith_Theodore.pdf.

29. Coalition for Evidence-Based Policy, “Which Study Designs Can Produce Rigorous Evidence of Program Effectiveness?” (working paper), January 2006, accessed March 20, 2013, http://coalition4evidence.org/wp-content/uploads/2012/12/RCTs_first_then_match_c-g_studies-FINAL.pdf.

- The Institute of Development Studies examined various evaluations in the open/participatory government arena. One of the reports looks at assessments of transparency and accountability initiatives in the fields of Service Delivery, Budget Processes, Freedom of Information, Natural Resource Governance, and Aid Transparency.³⁰
- The Center for Technology in Government at SUNY Albany has conducted a number of studies in related fields (see e.g., *The Dynamics of Opening Government Data*³¹ and *Developing Public Value Metrics for Returns to Government ICT Investments*³²). Taewoo Nam, also at SUNY Albany, reviewed a number of studies of evaluation specifically in the area of citizen sourcing using Web 2.0 technologies and discusses frameworks that researchers can use to evaluate the impacts of citizen sourcing.³³
- Exploring the Emerging Impacts of Open Data in Developing Countries (ODDC) is a recent project aimed at investigating ways in which open data improve governance, support citizens' rights and promote more inclusive development in developing countries. They are attempting to share information on evaluation and assessment of open data initiatives. Their website presents a number of case studies evaluating the impacts of open data.³⁴
- The World Bank, through its Striking Poverty program, is engaged in a debate on evaluating citizen involvement not only in transparency and accountability, but also in interventions that have an impact on people's lives.³⁵
- Nesta's initiative "Making Evidence Useful" includes a network of centres in the U.K. that will develop and evaluate assessment methods to identify what works.³⁶
- The Project on Government Oversight highlights best practices in open government in the U.S., including a list of resources on open government initiatives.³⁷

C) METHODOLOGY

Although costly and time-consuming, RCTs are considered helpful in rigorously establishing impact of relatively mature open government initiatives. There is a difference of opinion, however, as to the applicability of this approach to evaluating government. For this approach to work, experiments have to be set up ahead of time with a control group and a group that is part of the program. There are ethical issues involved, as providing government services to one part of the population and withholding them from another is tricky. One possibility that has been part of our federal "experiment" is to have individual states or localities test programs that then may be tried on a grander scale.

Michael Greenstone has written about the need for these kinds of experiments. In "Toward a Culture of Persistent Regulatory Experimentation,"³⁸ Greenstone points out some of the problems with randomized controlled trials and evaluating government programs in general. "Doing What Works," the project of the Center for American Progress, also calls for experimenting at the state and local level and adapting what works to the national level.³⁹

30. Rosemary McGee and John Gavanta, "Review of Impact and Effectiveness of Transparency and Accountability Initiatives: Synthesis Report," Institute of Development Studies, October 28, 2010, accessed March 20, 2013, <http://www.ids.ac.uk/files/dmfile/IETASynthesisReportMcGeeGavantaFinal28Oct2010.pdf>.

31. Natalie Helbig et al., "The Dynamics of Opening Government Data," Center for Technology in Government, University at Albany, SUNY, December 2012, accessed February 24, 2013, http://community.collaborativeforumonline.com/usercontent/site_143/s14/1000142873/56/The%2BDynamics%2Bof%2BOpening%2BGovernment%2BData.pdf.

32. Cresswell and Sayogo, "Developing Public Value."

33. Taewoo Nam, "Suggesting Frameworks of Citizen-Sourcing Via Government 2.0," *Government Information Quarterly*, 29 (2012): 12-20, accessed March 10, 2013, doi: 10.1016/j.giq.2011.07.005.

34. See <http://www.opendataresearch.org/emergingimpacts>

35. See <http://strikingpoverty.worldbank.org/comment/360>

36. See http://www.nesta.org.uk/home1/assets/features/making_evidence_useful

37. See <http://www.pogo.org/our-work/resources/2013/best-practices-for-open-and-accountable-government.html>

38. Michael Greenstone, "Toward a Culture of Persistent Regulatory Experimentation and Evaluation," in *New Perspectives on Regulation*, ed. David Moss and John Cisternino (Cambridge: The Tobin Project, 2009), accessed February 24, 2013, http://www.tobinproject.org/sites/tobinproject.org/files/assets/New_Perspectives_Ch5_Greenstone.pdf.

39. See <http://www.americanprogress.org/projects/doing-what-works/view/>

The Coalition for Evidence-Based Policy has a website devoted to “Top Tier Evidence”⁴⁰ that promotes the use of randomized controlled studies. They rely on a recent National Academy of Sciences report that includes “Recommendation on Criteria for Establishing Strong Evidence of Effectiveness.”⁴¹ The Coalition suggests, though, that the possibility of selecting a randomized intervention and control groups is itself a significant problem. The agency contemplating the intervention must be able to make assignments to those who will receive the intervention as well as to those who won’t, or who will receive a different intervention.⁴²

Quasi-experimental techniques can also be useful in evaluating outcomes while non-experimental methods can be a valuable starting point for new initiatives. Recent developments and new tools such as big data should also factor into how open government evaluations are designed.

And finally, as discussed in Section II, qualitative assessments are gaining renewed attention as ways of assessing open and participatory government initiatives. There are recent attempts to make such assessments more rigorous and also to mix qualitative and quantitative assessments, as advocated by Khargram and Thomas. These two argue for a new “Platinum” standard that incorporates RCTs with qualitative evidence to be used in assessing public administration and third party governance.⁴³

IV. CONCLUSION AND QUESTIONS MOVING FORWARD

We are witnessing broad experimentation in how we solve public problems through increased public participation in governance and the use of new information and communications technologies. But what do we actually know about the impacts of these changes? Given the budgetary conditions of most governments we can expect more calls for “doing what works” both for fiscal reasons and to have substantive impacts on social problems. But how do we know what works? And for whom? And exactly how? When we do discover that interventions “work,” or work to some extent? How do we replicate these efforts? Equally important, how do we adjust or end interventions that don’t work, or don’t work as well as we think they should?

The following four themes are important to incorporate as we move forward in evaluating interventions in citizen engagement in governance:

Different horses for different courses. “Citizen engagement” comprises many different “means” to achieve many different “ends.” Depending on the context, citizens can play different roles: as providers of ideas and expertise (think of crowd sourcing, predictive analytics, grand challenges, prize-induced innovation, brainstorming, etc.); or as representatives of specific interests (in the context of participatory budgeting, citizen juries and deliberative polling). And the contexts of engagement may differ substantially—from post-conflict zones to gentrified city blocks.

How do we provide answers to the questions we really should seek to answer: to achieve certain participatory objectives, what works, with whom, and under what conditions?

Improving people’s lives. Too often, the indicators used to measure citizen engagement are only meant to quantify the level of engagement—such as the number of people that participated or the volume of comments received. For citizen engagement to be meaningful and relevant, more effort is needed to answer the question:

How do we start determining what the impact is on on people’s lives?—the ultimate benchmark of success.

40. See <http://toptierevidence.org/>

41. National Research Council and Institute of Medicine, Youth and Families of the National Academy of Sciences. *Preventing Mental, Emotional, and Behavioral Disorders Among Young People: Progress and Possibilities*, ed. Mary Ellen O’Connell, Thomas Boat and Kenneth E. Warner (Washington DC: National Academies Press, 2009), accessed March 20, 2013, <http://www.whyy.org/news/sci20090302Mentalprepub.pdf>.

42. Coalition for Evidence-Based Policy, “Which Study Designs.”

43. Sanjev Khargram and Craig W. Thomas, “Toward a Platinum Standard for Evidence-Based Assessment by 2020,” *Public Administration Review*, 70, no. S1 (2010): S100-S106, accessed February 24, 2013, doi: 10.1111/j.1540-6210.2010.02251.

We measure what we value. Indicators (particularly of the statistical kind) are sometimes presented as neutral or scientific tools of measurement. In fact, though, they are inextricably linked with values—i.e., we measure what we care about. Indicators are socio-political constructs. As such, we should consider whether and how to involve citizens in determining how citizen engagement is measured (possibly using new on-line mechanisms such as rating and feedback tools).

How can we best engage with citizens to determine what success should look like and what to measure?

Experimentation in how we measure. While much experimentation in citizen engagement is taking place, experimentation in how we measure citizen participation is limited. Recent studies have shown that a mix of methodologies is needed to evaluate interventions in government. Reliance on the traditional gold standard of RCTs is not appropriate in many instances, and at the same time, relying solely on anecdotal evidence is no longer sufficient. But advances in both these areas hold out promise. In order to distill best practices and lessons learned, the field would benefit from increased experimentation using a variety of methods to understand the value that open government brings to people and the difference it makes in their lives. In particular, the use of big data to provide real time feedback that allows us to witness the impact of change to policy “as it happens” offers many opportunities for faster experimentation.

How can we improve evaluation of governance innovation through increased experimentation in methods and practice, including the use of big data?

OPEN BIBLIOGRAPHY

ON ASSESSING INTERVENTIONS IN OPEN AND PARTICIPATORY GOVERNMENT

Abelson, Julia and Francois-Pierre Gauvin. *Assessing the Impacts of Public Participation: Concepts, evidence, and policy implications*. Canadian Policy Research Networks. March 2006. Accessed March 20, 2013. http://www.cprn.org/documents/42669_en.pdf.

This study looks at goals of public participation efforts and assesses how interventions have met those goals. The study critiques different methods of measuring effectiveness and reviews case studies of evaluations.

Bertot, John Carlo, Patrice McDermott and Ted Smith. "Measurement of Open Government: metrics and process." Paper presented at the 45th Hawaii International Conference on System Science (HICSS), January 4-7, 2012. Accessed March 20, 2013. http://spaa.newark.rutgers.edu/images/stories/documents/Transparency_Research_Conference/Papers/Smith_Theodore.pdf.

The paper examines the Obama administration directive on open government and ways of assessing agency compliance. It describes the development of a measurement framework, assessment tool, and methodology that is being used by the OpenTheGovernment.org organization and partners and recommends ways of moving forward in assessing open and transparent government.

Board on Children, Youth and Families of the National Academy of Sciences. *Preventing Mental, Emotional, and Behavioral Disorders Among Young People: Progress and possibilities*. National Academies Press. 2009. Accessed March 20, 2013. http://www.nap.edu/catalog.php?record_id=12480.

Beginning on page 371 of this report, the National Academy of Sciences outlines its recommendations for evidence-based assessments.

Bornstein, David. "The Dawn of the Evidence-based Budget." *New York Times*, May 30, 2012.

This Times op ed piece, prompted by the OMB's call for agencies to indicate how they will evaluate their activities is used by the Pew Charitable Trusts Results First website (<http://www.pewstates.org/projects/results-first-328069>) as backing for their efforts to help states use evidence-based assessments in deciding on interventions. Bornstein briefly reviews the history of evidence-based calls by Bush and Obama administrations and mentions a few efforts to rely on such assessments.

Boss, Suzie. "Amplifying Local Voices." *Stanford Social Innovation Review*, Summer 2011.

A call for using story-telling to develop data for project evaluation. Based on Global Giving's storytelling project.

Chess, C. "Evaluating Environmental Public Participation: Methodological questions." *Journal of Environmental Planning and Management*, 43(6): 769-784, cited in Abelson and Gauvin.

Chess examines evaluation of public participation in environmental programs and discusses why, what and how to evaluate. Different methodologies are reviewed and recommendations made to improve evaluations.

Cresswell, Anthony and Djoko Sigit Sayogo. *Developing Public Value Metrics for Returns to Government ICT Investments*. A report to the Microsoft Corporation. SUNY Albany Center for Technology in Government. June 2012.

This paper looks at ways to assess investment in public programs, including both quantitative and qualitative assessment tools particularly for ICT, but also with broader applications. There is a lengthy discussion of public values and measurement and an analytical framework is developed.

Coalition for Evidence-based Policy. *Which Study Designs Can Produce Rigorous Evidence of Program Effectiveness?* Working Paper. The Coalition. January 2006. Accessed March 20, 2013. http://coalition4evidence.org/wp-content/uploads/2012/12/RCTs_first_then_match_c-g_studies-FINAL.pdf

The study recommends RCTs and the best way to evaluate programs and suggests why other methods are lacking. Also discusses quasi-experimental trials but recommends the former if at all possible. Then goes into detail on how best to set up RCTs.

Fukuyama, Francis. "What Is Governance." Working Paper 314. Center for Global Development. January 2013. Accessed March 23, 2013. http://www.cgdev.org/files/1426906_file_Fukuyama_What_Is_Governance.pdf

Fukuyama looks at different types of country governments and ways of assessing their performance and suggests four types of measures: (1) procedural; (2) capacity measures; (3) output measures; (4) measures of bureaucratic autonomy. He rejects output measures in favor of using capacity and autonomy as measures of executive branch quality.

Greenstone, Michael. "Toward a Culture of Persistent Regulatory Experimentation and Evaluation." Chapter in Moss, David Moss and John Sinternino. *New Perspectives on Regulation*. The Tobin Project. 2009.

As the title suggests, Greenstone argues that regulations should be constantly assessed and reassessed and experimented with to enable better cost-benefit decisions based on the best way to provide the greatest benefit to the public. He argues for both experimental (e.g., CRTs) and quasi-experimental assessments and constant experimentation. Chapter also looks at criticisms of these approaches.

Heusser, Felipe I. *Understanding Open Government Data and Addressing Its Impact* (draft version). World Wide Web Foundation.

The World Wide Web Foundation, in collaboration with IDRC has begun a research network to explore the impacts of open data in developing countries. In addition to the Web Foundation and IDRC, the network includes the Berkman Center for Internet and Society at Harvard, the Open Development Technology Alliance and Practical Participation. More information on the network is available here: <http://www.opendatasearch.org/about>.

Jukic, Tina and Mirko Vintar. "Ex-Ante Evaluation of E-Government Projects: Which Method to Use." *NI-SPAcee Annual Conference*. 2009. Accessed April 1, 2013. <http://www.nispa.org/files/conferences/2009/papers/200906301344320.JukicVintar.doc>

This paper discusses traditional evaluation methods as applied to e-government programs (including cost-benefit, payback period, return on investment, internal rate of return and net present value). The researchers find these methods wanting and suggest an alternative "multi-attribute model for ex-ante evaluation.

Kannan, P.K. and Ai-Mei Chang. Beyond Citizen Engagement: Involving the public in co-delivering government services. IBM Center for the Business of Government. 2013. Accessed April 2, 2013. www.businessofgovernment.org/sites/default/files/Beyond%20Citizen%20Engagement.pdf

This report focuses on the opportunities emerging technologies provide in the open government arena for involving the public in co-design (involving public in initial design of programs), co-producing and co-delivering of government services. The report concludes with a brief section on “How Do You Know If It’s Working?”

Khargram, Sanjev and Craig W. Thomas. “Toward a Platinum Standard for Evidence-based Assessment by 2020.” *Public Administration Review*. 70. December 2010.

The authors review the positive attributes of both quantitative and qualitative (Gold Standard I and II) assessment methods and argue for an integration of the two into a “Platinum Standard” by 2020.

Levy, Deb and Kathy Brennan. *Measuring Social Impact*. Skoll World Forum. 2005. <http://skollworldforum.org/2006/09/25/measuring-social-impact/>.

The authors discuss the difficulties in measuring social impact of interventions by non-profits and others and the differences in measuring outcome versus performance and choosing indicators accordingly.

Liu, Jianwei, et al. “eGovernment Project Evaluation. An Integrated Framework.” EGOV 2008, 2008. Accessed April 1, 2013. http://130.37.52.27/Jliu/Paper/Egov%2008_E-government%20assessment%20framework_published.pdf

In this paper, as with the paper by Jukic and Vintar (“Ex-Ante Evaluation of E-Government Projects: Which Method to Use?”) the authors demonstrate how traditional economic evaluation methods are not effective in demonstrating the value or lack thereof of e-government projects. The researchers propose a new model for determining the value of e-governance called the “Value Cube” which takes into account different value categories [strategic, operational, social and financial] and the specific goal areas of different stakeholders.

McGee, Rosie and John Gavanta. *Review of Impact and Effectiveness of Transparency and Accountability Initiatives: IETA Synthesis Report*. Institute of Development Studies. October 28, 2010. Accessed March 20, 2013. <http://www.ids.ac.uk/files/dmfile/IETASynthesisReportMcGeeGavantaFinal28Oct2010.pdf>.

This report looks at transparency and accountability initiatives and efforts to assess them. Specifically, they look at the cases of Service Delivery; Budget Processes, Freedom of Information, Natural Resource Governance and Aid Transparency. It examines how we know what we know, and importantly suggests gaps in our knowledge and ways of moving forward to improve evaluation efforts. The authors have completed a number of articles and reports on these issues.

Nam, Taewoo. “Suggesting Frameworks of Citizen-sourcing Via Government 2.0.” *Government Information Quarterly*. 2012, 29.

The paper examines the Obama directive on Open Government and how to assess the record of citizen participation in these initiatives. It presents a framework for evaluating citizen sourcing and how that can be evaluated in terms of the Open Government’s directive related to transparency, participation and collaboration.

Office of Management and Budget. *Use of Evidence and Evaluation in the 2014 Budget*. May 18, 2012. Accessed March 20, 2013. <http://www.whitehouse.gov/sites/default/files/omb/memoranda/2012/m-12-14.pdf>.

Memo by Jeffrey Zients based on Obama's call for evidence-based policy. "Where evidence is strong, we should act on it. Where evidence is suggestive, we should consider it. Where evidence is weak, we should build the knowledge to support better decisions in the future."

Organization for Economic Cooperation and Development. *Citizens as Partners: Information, Consultation and Public Participation in Policy-making*. OECD. 2001.

This early report discusses the benefits of public participation in governance and the need for evaluation of public information, consultation and participation. The EU developed a common assessment framework (CAF) to aid in reviewing actions taken to improve openness and understanding, public access to services and actions taken to empower customers/citizens.

Evaluating Public Participation in Policy Making. OECD. 2005.

This book builds on the 2001 *Citizens and Partners* report and attempts to address the "evaluation gap" identified in that study: the fact that there were many attempts to increase participation in government and very few attempts to assess the effectiveness of such efforts. It includes case studies of a number of participatory endeavors in several countries. The OECD also has a website that links to evaluation guidelines of a number of organizations and countries who are members of the DAC Network on Development Evaluation: <http://www.oecd.org/derec/guidelines.htm>.

Podesta, John and Rushing Reece. *Doing What Works: Building a government that delivers greater value and results to the American people*. Center for American Progress. February 2010. Accessed March 20, 2013. http://www.americanprogress.org/wp-content/uploads/issues/2010/02/pdf/dww_framing.pdf.

This report by the Center for American Progress calls for challenging the status quo in government by measuring what works, experimenting and innovating, and coordinating and consolidating federal programs to increase efficiency and effectiveness of government programs. It calls for use of new technologies to engage the public and action based on evaluations of what works.

Rowe, Gene and Lynne Frewer. "Public Participation Methods: A framework for evaluation." *Science, Technology and Human Values*. 2000, 25(1): 3-29.

These three articles by Rowe and Frewer lay some of the groundwork for the Abelson and Gauvin report, especially in the area on methodology. The authors have much experience in public participation in food safety issues.

"A typology of public engagement mechanisms." *Science, Technology, and Human Values*. 2005, 30(2): 251-290.

"Evaluating public participation exercises: A research agenda." *Science, Technology, and Human Values*. 2004, 29(4): 512-556.

Winkler, Mary Kopczynski et al. Building a Common Outcome Framework to Measure Nonprofit Performance. The Urban Institute. 2006.

This report develops a framework to provide indicators for measuring success of nonprofit programs.

World Bank. Striking Poverty: Inspiring innovation, accelerating development. Discussion forum hosted by the World Bank available at: <https://strikingpoverty.worldbank.org/c130311>.

This discussion forum seeks to find out what works and what doesn't in public participation in government. It is hosted by Tiago Peixoto of the Bank and includes Joanne Caddy of the OECD, Vera Schattan Coelho of the Centro Brasileiro de Analise e Planejamento and Edward Anderson of Involve.

WE WOULD WELCOME YOUR SUGGESTIONS FOR ADDITIONAL SOURCES THAT WE CAN INCLUDE IN THE NEXT VERSION AND ON THEGOVLAB.ORG WEBSITE.

Please send your recommendations to Stefaan Verhulst (sverhulst@nyu.edu) or use the url <http://bit.ly/150j0Ct>,