

## PRESIDENT'S STATEMENT 2007: THE IMPORTANCE OF DATA

*Paul Brest\**

“Numbers matter. Through counting and accounting, data can improve the well-being of people by informing the decisions that affect them; poor, missing, and old data contribute to bad decisions, lack of trust, and lower societal well-being.”

– *The Center for Global Development*<sup>1</sup>

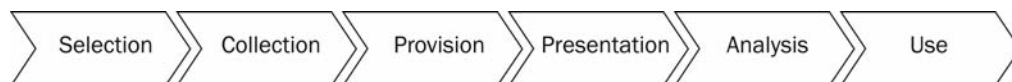
“It is easy to lie with statistics, but easier to lie without them.”

– *Frederick Mosteller*<sup>2</sup>

Foundations support the development and dissemination of myriad forms of knowledge, ranging from scholarship of no immediately foreseeable use to applied research designed to improve health, education, and the environment. Data, whether quantitative or qualitative, are the atoms and molecules of knowledge. No program to promote or assess progress in any sector of society can succeed without a solid grounding in data.



We think of a data system as involving six stages:



At the end of this essay, I will say a few general words about the role of philanthropy in supporting data systems. First, though, some examples, drawing mainly on the work of the Hewlett Foundation's grantees.

### I. DATA SELECTION

The first step in promoting a data system is ascertaining what data policymakers need to make better decisions. Since societies—especially those in the developing world—have limited resources, we cannot hope for exhaustive knowledge. Setting priorities is crucial.

One class of data is universally necessary: governments must know who their citizens and residents are and where and how long they live. In sub-Saharan Africa, fundamental limitations of resources have prevented national governments from collecting reliable demographic data on births, deaths, and migration, let alone data concerning education, government expenditures, and air and water quality. A new Hewlett initiative called Demographic Dynamics for Development (3D) is assessing the state of data in sub-Saharan Africa and the potential role of private philanthropy in helping improve data systems.

Beyond basic demographic information, policymakers must assess where resources devoted to data collection are likely to produce commensurate social improvements. Californians lack reliable data about students' test scores, course completion, and dropout rates. For instance, current data on dropout rates exclude students who leave between middle and high school. But figuring out just which data will best aid educators in improving our schools is a challenge. The Hewlett and Gates Foundations are supporting efforts by state education agencies to determine what information

needs to be collected. The foundations have offered to help pay to develop statewide data on students' educational histories that can follow them from one school to another. Like a medical chart, this will eventually eliminate the costs and inaccuracies of re-collecting information and will ensure that California has a complete picture of each of its students. When combined with other information, it should allow researchers to identify what educational strategies work—and don't work—with the promise of improving children's education outcomes.

From time to time, social scientists step back to question the adequacy of data that were once assumed to be the primary measures of a society's progress. In the last century, economists, including Amartya Sen and Partha Dasgupta, helped establish human-development indicators that captured multidimensional aspects of well-being. Since the turn of the new century, the Hewlett Foundation has supported work by the distinguished labor economist Alan Krueger, Nobel laureate Daniel Kahneman, and other economists and psychologists to study the ways individuals evaluate their well-being, including satisfaction in life, work, and relationships, and the absence of chronic negative emotions. Krueger is seeking to have questions on these issues included in the Bureau of Labor Statistics' American Time Use Survey with the ultimate goal of presenting a more complete picture of American well-being.

With a grant from the Hewlett Foundation, the Organization for Economic Cooperation and Development (OECD) is working on similar efforts at an international level. The OECD aims to complement its current collection of data on quantitative indicators with citizens' qualitative assessments of their well-being. It remains to be seen how useful these sorts of subjective data will be in assessing societal progress and designing policies to promote it. But the Kahneman-Krueger and OECD efforts promise to deepen our understanding of how societies can measure progress.

## II. DATA COLLECTION

Essential data are often collected sporadically and improperly, omitting responses from key demographic groups. Non-random sampling biases data and reduces their value for making the generalizations necessary for prudent policymaking. For example, censuses in many developing countries exclude people living in informal slums, even though that's where a large majority of urban residents live. The African Population and Health Research Center in Kenya is working to address this problem by collecting data on health and education in Nairobi's slums; it has the potential to serve as a model for other parts of Africa.

Even data sets that are pulled from a random sample can be distorted in other ways. Agricultural production data from sub-Saharan Africa on supply, demand, projections, and trade are sometimes biased to overstate exports because governments fear domestic political backlash if they report net food imports. And even accurate agricultural data are often out-of-date, making them useless in dynamic markets. Our grantees, including Manobi, the International Fertilizer Development Center, and a Mali-based Michigan State University project are working to collect supply and demand and market and input price data that will help farmers improve crop yields and governments improve agricultural policy.

Like the drunk looking for the lost coin under a streetlight rather than in the dark corner where he lost it, policymakers often favor those data that are easy to collect rather than the most useful. This is true of many assessments of education in developing countries, where the number of students enrolled has been treated as a proxy of success. But (to use an education colloquialism) just getting "butts in seats" hardly ensures that children come out of school better equipped to succeed. That calls

for collecting different information. Pratham, a grantee of a collaborative initiative with the Gates Foundation, is addressing this problem in India through a nationwide survey of children's basic math and reading skills. Pratham's data is collated into an Annual Status of Education Report, which highlights learning outcomes in primary school and aims to stimulate nationwide interest in more effective schooling methodologies.

### III. DATA PROVISION

Once data are collected, they must be made available to policymakers and the public.

In some cases, cash-strapped governments lack the resources to preserve data and make them available. For example, longitudinal census data in many sub-Saharan African countries exist only in paper archives and are not only difficult to access but in danger of being lost forever. The 3D initiative endeavors to make such data available for policymakers and researchers.

Costs aside, politicians are sometimes motivated to keep embarrassing information from reaching public scrutiny. Philanthropy can play an important role in supporting systems that make governments more transparent and accountable.

Indeed, this is a key objective of our Global Development Program. Hewlett grantees such as the International Budget Project (IBP) create tools that citizens can use to ensure that their governments are providing them with adequate information. IBP's Open Budget Index rates countries on how accessible domestic budgets are to their citizens. Initially compiled for fifty-nine countries, the Index is expanding to include more countries and to include subnational budgets as well. Armed with comprehensive and practical budget information, and provided the training to interpret them, citizens' organizations can identify reforms needed to strengthen the delivery of public services.

Another grantee, the University of San Diego's Justice in Mexico Project, is working to revamp government-provided data on crime rates. At present the government collects considerable data, but charges the public—and sometimes other sectors of the government—hundreds of dollars to access them. The Project's database will present information on crime, victimization, police, and the judicial system that is timely, accessible, and free—all disaggregated by state and locality. The Project will ultimately use these data to advise Mexican officials on the optimal use of resources to improve the administration of justice. This would be the first time in Mexican history that such comprehensive information will be available to professionals working to protect the public.

The Foundation's support for transparency and accountability in Mexico has already had some unanticipated benefits. Thanks in part to our grantees' efforts, all states and the federal government now have access to information (ATI) laws—the equivalent of U.S. freedom of information laws—that require making data available in a broad range of areas, from government spending to welfare lists to prison files. The Mexican Federal Access to Information Institute piloted a program to make these new laws relevant to poor communities, with some dramatic results:

- ▶ Poor women in the state of Veracruz learned that their names appeared on the lists of beneficiaries for health and housing programs—but they had never received the benefits. Indeed, some people on the list of beneficiaries for Pap smears and mammograms were men. The women entitled to benefits are now pressing for the benefits they supposedly received.
- ▶ Prisoners in a federal penitentiary in the state of Nuevo León—the majority of whom are too poor to have lawyers and are behind bars for petty offenses—used

the law to gain access to their own files. Though the prisoners were initially denied the information, they appealed and, in a precedent-setting ruling, won the right to access information for all prisoners. Once they exercised their right to information, 36 percent were able to show that, under the terms of their sentencing, they were eligible for early release. This has set a new standard of openness for other prisons nationwide, and has enormous potential for reuniting poor families and substantially lowering prison costs.

Unfortunately, despite (or perhaps because of) Mexico's impressive strides toward greater access to information rights, Hewlett grantees and other access to information advocates in the country are facing increasing resistance—and even harassment—from some government authorities, especially at the state and local levels.

The need to increase government transparency is by no means limited to developing countries. For example, until recently, the public could only learn about the recipients of European farm subsidies by piecing together disparate pieces of information obtained through freedom of information requests. Thanks in part to the hard work of the Foundation's grantee, EU Transparency, the European Commission has committed to making farm subsidy data readily available and free online.

I have focused on efforts to bring to light information that governments and businesses would often prefer to keep in the dark. But foundations also help willing participants use technology to overcome barriers to aggregating data. This is the case of the California Cultural Data Project, which seeks to provide comprehensive information about the state's cultural sectors. The Project's website will include information about workers who make their living through the arts, annual arts and culture activities, and revenue generated by arts events. A large enough data set will allow analysis of how the cultural sector affects state and local economies and their residents' quality of life, with the goal of giving philanthropists, policymakers, and citizens a better sense of the arts sector's assets, value, and needs.

#### **IV. DATA PRESENTATION**

To be most useful, data must be not only available, but presented in ways that enable citizens, policymakers, and analysts to apply them to the problems at hand. Data presentation has two major aspects: linking and coordinating data sets, and putting information into a platform that's easy to navigate.

Coordinating data sets is a pervasive problem in public education systems. Whether because of cost, negligence, or intention, student and teacher data for California's K–12 education system are not interconnected. Moreover, students' high school records are not linked with their performance in the state's higher education systems, because each institution maintains its own separate database. Indeed, many community colleges do not link data on students' initial test scores with data on their advancement through college, and nothing is known about their progress if they transfer to four-year institutions.

To improve California's schools, we need to know how different educational approaches affect individual students over time. And in order to compare approaches, we need linked data sets. McKinsey & Company and other grantees of our Education Program are working with policymakers in Sacramento to design and implement a high-quality data collection and aggregation system to fix this problem and provide educators and policymakers with the tools they need to improve student learning.

Sometimes the sheer volume of data makes it difficult for citizens and policymakers to assimilate and interpret them. In the face of such information overload, distillation is

essential. The Hewlett Foundation was a charter supporter of The State of the USA (SUSA), which is developing a free Web site of indicators about American society, its economy, and its environment. SUSA works with The National Academies to present data on issues including international trade, education, the labor supply, our national resources, and the state of health care that can be compared across cities, states, and regions. Much thought is going into making the site an easy-to-use tool for evidence-based decisionmaking. Rather than engaging in abstract discussions of the rising problem of obesity, Americans will be able to compare obesity levels in California with those in Nebraska and to see how the problem breaks down by age and gender.

At the international level, our grantees have begun to make budget revenue and expenditure data more accessible. In the planning process for the Foundation's Global Development Program, improving transparency and accountability was among the long-term strategies deemed most likely to improve the income, health, and overall well-being of people living on less than \$2 a day. With our support, the Mexican Institute for Competitiveness (IMCO) is building a data platform for information about state revenues and expenditures that will include two Web-based calculators. One will show how variation in Mexican oil prices affects state revenues. The other will let municipal mayors determine, based on federally mandated budgeting, exactly what proportion of oil revenues their state governments should be providing them.

In other development work, the Hewlett and Gates Foundations are collaborating to promote transparency by international donors. A new project encourages multilateral, national, and philanthropic donors—from the World Bank to China and the United States to the Hewlett Foundation—to post data about their grants to an online database. Though much of this information is published, each donor has its own format—sometimes fairly obscure—and it is difficult to track the resource flows into a particular country. The new system will organize the information to provide policymakers and nonprofit organizations timely and comprehensive access. Eventually, we hope this platform will encourage better use of data in donor decisions and recipient requests—which will enable countries to take greater ownership of their economic development processes. Only with knowledge of the money they have and sound predictions of the money they will receive in the future can governments design and implement long-term strategies for growth.

From RSS feeds and XML tagging to wikis and social networks, technology has dramatically improved the availability and presentation of data. Computer-based GIS (geographic information systems) are increasingly used to overlay social and environmental information on maps of neighborhoods, cities, and regions. For example, Healthy City, a project of our grantee, Advancement Project, provides an information-mapping platform that combines Los Angeles County demographic data with community resource information, showing how citizens' needs do and don't match up to the distribution of preschools, violence prevention centers, and other social service agencies throughout the city. Los Angeles policymakers planning to allocate \$100 million in funding for new preschools were concerned that placing them in certain low-income neighborhoods would be impractical and expensive. Healthy City provided data on the costs of different sites and the populations that would benefit. Residents of one low-income community were able to use this information to convince policymakers that siting the preschools in their neighborhood would be both feasible and cost-effective. Healthy City aspires to create a model of how an interactive data platform can contribute to decisions like this nationwide, to strengthen regional social service sectors and facilitate data-driven city planning.

Healthy City, the California Cultural Data Project, IMCO, and SUSA are just a few examples of how Web 2.0 has affected the ways that people interact with data and, indeed, with the world itself. These are matters of particular interest to the

Foundation's Philanthropy Program<sup>3</sup> and Open Educational Resources Initiative, and are of such breadth to be left to another day.

## V. DATA ANALYSIS

Sometimes data just speak for themselves. Often, though, further analysis is necessary to translate data into knowledge useful for formulating policy. The Hewlett Foundation supports such analyses in every area of its grantmaking—for example, trying to understand the effect of arts education on students' outcomes and future well-being, or the relationships among population size, poverty, economic growth, and global warming. The Foundation's Global Development Program is funding two initiatives focused on analysis. One is designed to encourage program, or "impact," analysis—analysis of the effects of particular social programs; the other seeks to create the infrastructure for policy analysis in developing countries.

Especially in comparison to the evaluation of domestic programs (which has a tradition going back at least to the Ford Foundation's creation of Manpower Development Resource Corporation), the evaluation of social programs in developing countries has been sporadic and weak. As a founding member of a consortium of donors, we have helped create the International Initiative for Impact Evaluation (3IE) to promote and coordinate evaluation efforts globally. In the words of William Savedoff and Ruth Levine of the Center for Global Development:

*For decades, development agencies have disbursed billions of dollars for programs aimed at improving living conditions and reducing poverty; developing countries themselves have spent hundreds of billions more. Yet the shocking fact is that we have relatively little knowledge about the net impact of most of these programs. In the absence of good evidence about what works, political influences dominate, and decisions about the level and type of spending are hard to challenge.*

3IE will support high-quality analyses of what interventions do and do not work in international development. It will also act as a research hub, helping policymakers access the most recent and relevant research.

The work of another Hewlett grantee, the Abdul Latif Jameel Poverty Action Lab (J-PAL) at MIT, exemplifies the sort of impact evaluation that the 3IE initiative will support. J-PAL recently conducted a randomized controlled study of Pratham's one-on-one tutoring initiative in India. The results indicated that the program significantly increases students' math and reading test scores, that it works best for those having the most trouble learning, and that it is almost seven times more cost-effective than hiring additional teachers. The initiative is now being implemented in twenty cities and will likely be expanded with the support of philanthropy and donor governments.

Of course, impact evaluations can also show that development initiatives did not achieve their intended outcomes. For example, J-PAL studied a program that subsidized rural women's groups in Kenya with the dual goals of improving women's leadership skills and expanding the services the groups provided to their communities. Contrary to expectation, the study indicated that the increased funding pushed the original women out of the group by attracting younger, more educated, wealthier, and more male participants, with no apparent benefit for the populations funders hoped to help.

Impact evaluations like these have the potential to benefit people beyond those directly served by the evaluated programs. When a program succeeds, funders and

governments can often replicate it, with reason to believe that it will work under similar circumstances in other places.<sup>4</sup> By the same token, when a program has no impact, funders can reallocate their resources to more promising interventions.

A majority of the universities, think tanks, and specialized organizations that conduct impact and other policy-relevant analysis are located in the United States and Europe. This only reinforces the tendency of international donors to identify research priorities that are not necessarily relevant to a given country context, or are not realistic in light of local political realities and timelines. In general, policies work best when they are designed and implemented by local actors rather than advisors who live in different societies thousands of miles away.

With the aim of establishing strong local research institutions that are able to develop effective working relationships with decisionmakers, the media, and other civil society organizations, the Hewlett Foundation is collaborating with the Canadian International Development Research Centre and other international donors to support a Think Tank Initiative that will provide long-term institutional support to policy research organizations in Africa, South Asia, and Latin America. Through both financial and technical assistance, the Initiative will help equip local think tanks with the resources they need to perform country-specific policy analysis and to provide civil society actors with the unbiased information necessary to participate constructively in policy debates.

## VI. DATA USE

All the efforts to improve the collection and presentation of data are for naught—or of purely academic interest—unless they affect the practices and decisions of governments, businesses, and citizens. And absent a norm of using data, there is little incentive to improve its supply.

Yet much policymaking—and not just in developing countries—is more responsive to ideology than to data. For example, the U.S. government has allocated over a billion dollars to abstinence-only sex education programs in the face of robust studies, including a large, federally funded evaluation, showing that such programs do not reduce teenage pregnancy or sexually transmitted diseases. Only recently, as a result of the intensive work of family-planning organizations to educate the legislators, has Congress shown signs of moving toward evidence-based decisionmaking in this highly controversial realm.

Ideological commitments also have diminished the use of data in the environmental regulatory process. Perhaps in reaction to federal regulatory agencies' one-sided use of cost-benefit analysis to deny protections, some environmental advocates have all but abandoned this essential policymaking tool. New York University Law School's new Institute for the Study of Regulation aims to educate organizations and government officials alike about the value and techniques of balanced cost-benefit analysis in designing effective health, safety, and environmental controls.<sup>5</sup>

Here is an example of how basing government policies in sound data analysis can have real impact. At the request of the Chinese government, The Energy Foundation and the International Center on Energy and Transportation helped analyze the costs and benefits of automobile fuel economy standards and develop regulations based on this analysis. The China Automotive Technology and Research Center estimates that in the four years since the regulations were adopted, Chinese drivers, and the world, have saved 1.18 million tons of gasoline.

## VII. THE ROLE OF FOUNDATIONS

Reliable, comprehensive data systems are essential to a society's progress. Yet even though the long-term consequences can be enormous, it often takes years or decades for high-quality data systems to translate into better policymaking and improved welfare—and even then the effects may be uncertain and difficult to trace. As with many public goods, market and governance failures leave gaps in the data landscape. Individuals tend to under-invest because they find it difficult to capture the benefits of their investments. Politicians often under-invest because immediately pressing problems get their attention and election cycles don't match the long time horizons necessary for data analysis to pay off.

Foundations have the comparative advantages of seeking social rather than financial returns on their philanthropic investments, of having long time frames, and of not being politically accountable to electorates. Although philanthropy cannot and should not take the place of government in maintaining essential data systems, it can help jump-start them through demonstration projects, advocacy, and other means. Foundations can collaborate with government officials and policymakers and, where appropriate, goad them into action. The Hewlett Foundation's support for data systems has increased in recent years as we have come to appreciate their importance in every area of our concern. If our assessment is correct, these investments will produce significant social returns over time.

---

\* I am grateful for Emily Warren's assistance in writing this essay.

<sup>1</sup> Rachel Nugent and Danielle Kuczynski. "Improving Access and Use of Demographic and Health Data: A roadmap of activities aimed at strengthening the collection, access and use of demographic and related development data, with a focus on Africa." Center for Global Development. 03/2008.

<sup>2</sup> Frederick Mosteller was the founding chairman of the Harvard statistics department and popularized the discipline's use in medicine, sports, and politics. See "Milestones." Time. July 30, 2006.  
<<http://www.time.com/time/magazine/article/0,9171,1220541,00.html?promoid=googlep>> Accessed 08/04/2008.

<sup>3</sup> Cite to last president's statement.

<sup>4</sup> A well-designed impact evaluation study can determine a program's impact in the particular circumstances. Replication to other situations depends on the generalizability or "external validity" of the study, determining which is as much craft as science.

<sup>5</sup> The Program is directed by Dean Richard L. Revesz, co-author of *Retaking Rationality: How Cost-Benefit Analysis Can Better Protect the Environment and Our Health* (Oxford University Press 2008).