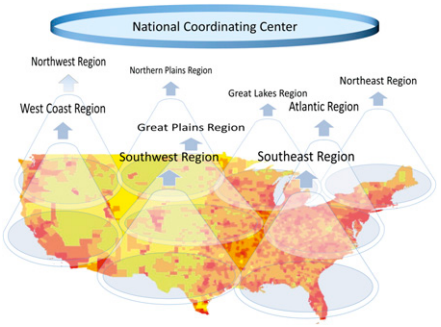
CrossMark click for updates

OPINION

Opinion: Building a 21st-century infrastructure for the social sciences

Emilio F. Moran[a], Sandra L. Hofferth[b,1], Catherine C. Eckel[c], Darrick Hamilton[d], Barbara Entwisle[e], J. Lawrence Aber[f], Henry E. Brady[g], Dalton Conley[h], Susan L. Cutter[i], Klaus Hubacek[j], and John T. Scholz[k]

[a]Department of Geography, Michigan State University, E. Lansing, MI 48823; [b]Department of Family Science and [i]Department of Geographical Sciences, University of Maryland, College Park, MD 20742; [c]Department of Economics, Texas A&M University, College Station, TX 77843; [d]Milano School of International Affairs, Management and Urban Studies, The New School for Public Engagement, New York, NY 10011; [e]Department of Sociology, University of North Carolina, Chapel Hill, NC 27599; [f]Department of Applied Psychology, New York University, New York, NY 10003; [g]Goldman School of Public Policy, University of California, Berkeley, CA 94720; [h]Department of Sociology, New York University, New York, NY 10012; [i]Department of Geography, University of South Carolina, Columbia, SC 29208; and [k]Department of Political Science, Florida State University, Tallahassee, FL 32306



The Regional Data Center Network: A vision for social science research.

The social sciences are poised for a transformation. New data from the Internet and social media, when combined with newly available administrative and transactional data, have the potential to greatly expand the questions that can be addressed, as well as the spatial and temporal scales at which they can be addressed. For example, by using data from social networking sites, business transactions, smart phones, and online experiments, we can learn about labor market and consumer behavior and assess vulnerability to weather events and the impact of local and national policies and programs in real time. Indeed, behavior and human well-being are situation and place specific, and vulnerable populations tend to be spatially clustered rather than randomly distributed. To generate analyses that incorporate this spatial variability will require entirely new ways to collect, and link, diverse data to improve population health, reduce environmental and social vulnerabilities, and better prepare people for the jobs of the future. It is time to embrace the "big data" revolution and capitalize on advances in information technology and data science as they apply to the specific needs and challenges of the social sciences (1). In short, we need a new social science data infrastructure for the 21st century.

Social science and related communities across the United States are already envisioning the components of such an infrastructure. The authors of this article are members of the Social Observatories Coordinating Network, funded by the National Science Foundation (socialobservatories.org) to work with disciplinary communities and other stakeholders to plan, develop, and coordinate a flexible national framework that can simultaneously address what is happening across the country, in local communities, and in people's daily lives. Representing anthropology, economics, geography, political science, psychology, and sociology, members have been working on the design of an ambitious national network of regional data centers to establish such a cyberinfrastructure and ensure that the American people have available, in a timely fashion, nationally scaled and locally relevant information to make better business, health, education, and governance decisions.

Our vision is for an integrated nationally representative but distributed framework or platform that is both scalable (from local to national) and flexible (having core data but also data that are site specific and locally relevant), that permits new questions to be addressed as needed, that allows for informed and rapid response and adaptation to local shocks such as extreme weather events or natural resource windfalls. It should also facilitate an understanding of local manifestations of national or global phenomena such as how economic downturns affect different regions of the country. The proposed infrastructure of regional data centers would undertake the challenging task of linking a broad array of information—administrative data (local and state and regional), media and social media (broadcast, printed, Twitter, Facebook), census and other surveys, ethnographic data, and data from experiments such as randomized controlled trials—to address how different human communities make decisions. These new data will be rich in detail. An important challenge is how to link their diverse forms in a way that is cumulative, accurate, and accessible to a broad range of researchers, policy makers, and other stakeholders.

To elucidate complex social dynamics and provide rich descriptions at the local and national level, the Social Observatories Coordinating Network recommends that these regional data centers be developed in the context of a well-designed representative national sample of people and localities. That sample could take different shapes. One possibility would be to partner with an existing survey, using its sample design as a platform for further development. Another possibility is to develop a sample from the ground up. Either way, the definition of locality would begin with the sampling units (e.g., census tracts, counties) and the people who reside in these units. Core data about individuals and their social relationships would be gleaned from specially designed national surveys, administrative sources, companies, public agencies, and social media

Author contributions: E.F.M., S.L.H., C.C.E., D.H., B.E., J.L.A., H.E.B., D.C., S.L.C., K.H., and J.T.S. wrote the paper.

[1]To whom correspondence should be addressed. Email: hofferth@umd.edu.

Any opinions, findings, conclusions, or recommendations expressed in this work are those of the authors and do not necessarily reflect the views of the National Academy of Sciences.

and Web-scraped data. Moreover, because the regional data centers will collect place and contextual data as well, it will be possible to understand not just individual variation, but also the impacts of spatial and contextual variation. This national representative sample would allow aggregation to the national level and comparisons across the nation with regard to how local and national processes affect social and economic mobility and adaptation to change.

There will be challenges. The social science data infrastructure we envision has as a starting point newly available administrative and transactional data, as well as new social media and Internet data that were not designed for research purposes. Careful attention must be given to data documentation, harmonization, curation, and integration, as well as to methods to support discoverability, use, and reuse of the data in a responsible way. Social media and the growing ease of collecting, storing, and sharing sensitive data require a thoughtful approach to privacy issues and enhanced safeguards.

The policy implications are great, as well as the potential rewards. For instance, the United States has always been characterized as the land of opportunity, where anyone can, through effort, succeed and attain the American dream. But a recent study shows that one of four children raised in the middle class has slipped downward by their early 40s (2) but perhaps even more important is the finding that the chances for upward mobility and its maintenance depend importantly on where people live (2–4). Why? Social policy, economic conditions, and the size of the middle class in a city provide a context in which individuals can thrive or stagnate, leading to important questions about the roles of government, nonprofits, and for-profit organizations in opportunity and mobility processes and the impact of their policies. When and how does neighborhood context matter? What are the consequences of economic and social conditions and change in those conditions for individual economic opportunity and mobility? The social science platform we envision can address such questions and others central to our understanding of who we are as a people and a nation. This endeavor would constitute a social science Apollo Project, worthy of the Information Age, and our future as a country may depend on our willingness to launch it.

ACKNOWLEDGMENTS. The Social Observatories Coordinating Network is funded by National Science Foundation Grant SES1237498.

1 King G (2011) Ensuring the data-rich future of the social sciences. *Science* 331(6018):719–721.
2 Acs G (2011) *Downward Mobility from the Middle Class: Waking Up From the American Dream* (Pew Charitable Trusts, Washington, DC).
3 Chetty R, Hendren N, Kline P, Saez E (2014) *Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States. NBER Working Papers 19843* (National Bureau of Economic Research, Cambridge, MA).
4 Olinsky B, Post S (2013) *Middle Out Mobility: Regions with Larger Middle Classes Have More Economic Mobility* (Center for American Progress, Washington, DC).