Introduction: History and Motivation

By
SANDRA L. HOFFERTH,
EMILIO F. MORAN,
BARBARA ENTWISLE,
J. LAWRENCE ABER,
HENRY E. BRADY,
DALTON CONLEY,
SUSAN L. CUTTER,
CATHERINE C. ECKEL,
DARRICK HAMILTON,
and
KLAUS HUBACEK

Big data, that is, data that are byproducts of our lives rather than designed for research purposes, are the newest of the information highway innovations. One of the important challenges to social and behavioral science data collection, curation, and dissemination for the foreseeable future is to link diverse forms of data in a way that is cumulative, representative, meaningful, and accessible to a broad range of researchers. It is critical to explore the new questions these data can address and to develop new methods to address them, including linking persons and information about them and their environments across different data platforms while maintaining confidentiality and privacy. Linking a broad array of information—from administrative data (local and state and regional), to social media (Twitter, Facebook), to census and other surveys, to ethnographic data, and data from experiments such as randomized controlled trials—to address how humans and their communities make decisions is challenging. This issue was addressed by papers presented at a conference on New Data Linkages convened by the Social Observatories Coordinating Network in 2016; those articles are brought together in this volume.

Keywords: big data; social science; systems science; data analytics; knowledge generation; survey research; administrative data

The nation and the world are changing rapidly, yet scholars attempt to understand such changes with tools and infrastructure that

Sandra L. Hofferth is a professor emerita, School of Public Health, and a research professor, Maryland Population Research Center, University of Maryland. Her research focuses on Americans' use of time; economic disadvantage, parental behavior, and child health and development; fathers and fathering; and immigrant youth's transition to adulthood.

Emilio F. Moran is Hannah Distinguished Professor at Michigan State University. He is the author of eleven books, fifteen edited volumes, and more than 200 journal articles and book chapters. His research addresses how humans interact with the environment. He was elected to the U.S. National Academy of Sciences in 2010.

Correspondence: hofferth@umd.edu

were developed more than 50 years ago. In the last decade, hurricanes and floods have devastated coastal areas of the United States; sudden migratory flows due to environmental, political, and economic events abroad have raised concerns about the assimilation of refugees and immigrants; and the increased financial burden of coping with unexpected health shocks has drawn attention to inequities in the health care system and exposed patterns of ecological, economic, and social vulnerabilities across the nation. Finally, a large and robust middle class—one of America's greatest achievements—has been steadily eroding, leading to anger at our social and political institutions.

Barbara Entwisle is Kenan Distinguished Professor of Sociology and a fellow of the Carolina Population Center at the University of North Carolina at Chapel Hill. Her research focuses broadly on the study of social, natural, and built environments and their consequences for a range of demographic and health outcomes.

J. Lawrence Aber is the Willner Family Professor of Psychology and Public Policy at NYU Steinhardt, and codirector of the international research center, Global TIES for Children. His basic research examines the influence of poverty and violence on the social, emotional, behavioral, cognitive, and academic development of children and youth.

Henry E. Brady is the dean of the Goldman School of Public Policy and Class of 1941 Monroe Deutsch Professor of Political Science and Public Policy, University of California, Berkeley. His research and writings address economic and political issues including electoral politics and political participation, social welfare policy, political polling, and the dynamics of public opinion.

Dalton Conley is the Henry Putnam University Professor in Sociology at Princeton University. His research focuses on how socioeconomic status and health are transmitted across generations and on the public policies affecting those processes. He studies sibling differences in socioeconomic success; racial inequalities; and how health and biology affect (and are affected by) social position.

Susan L. Cutter is a Carolina Distinguished Professor of Geography at the University of South Carolina and director of its Hazards and Vulnerability Research Institute. Her research focuses on vulnerability and resilience metrics especially as they relate to social inequalities in disaster response and recovery.

Catherine C. Eckel is the Sara & John Lindsey Professor of Economics and university distinguished professor at Texas A&M University, where she directs the Behavioral Economics and Policy Program. She uses experimental methods to investigate social and economic influences on behavior, including topics such as discrimination, charitable giving, risk-taking, and trust.

Darrick Hamilton is the director of the doctoral program in policy and an associate professor of economics and urban policy at The New School in New York. He is a stratification economist, whose work examines the causes, consequences, and remedies of racial and ethnic inequality including intersections of identity, racism, colorism, and socioeconomic outcomes.

Klaus Hubacek is an ecological economist and professor in the Department of Geographical Sciences at the University of Maryland, College Park. His research focuses on conceptualizing and modeling the interaction between human and environmental systems and modeling scenarios of future change.

Scientific explanations of these phenomena have been fragmentary and discipline-bound, so our prescriptions for reversing these trends are piecemeal and there is a high degree of uncertainty as to their effectiveness. Moving forward will require new ways of thinking and new ways of accessing, curating, and analyzing the existing but not always accessible information. Social scientists generally work with survey, administrative, observational, and experimental data, data gathered for specific purposes. Social scientists' infrastructure and tools consist of planned data collection and well-established analytic methods. Survey methods permit generalizing questionnaire responses to known populations, observational methods add depth, and experimental methods facilitate causal inference. Linked with surveys, administrative data could expand our understanding of individual behavior but have remained largely inaccessible. Examples of large-scale survey data of the federal statistical system include data held and disseminated by the National Center for Health Statistics, the U.S. Census Bureau, the Bureau of Labor Statistics, and the Economic Research Services of the Department of Agriculture. However, big data—large, diverse, and heterogeneous datasets, often by-products generated from business and Internet transactions, email, social media, health care facilities, and various sensors and instruments—have produced large archives of data that are not organized in a way that can be easily analyzed by social scientists (Foster et al. 2017). Population characteristics are often unknown and inferential statistics inappropriate. However, these new data provide information, such as emotional responses, that may complement traditional types of data and are more immediately accessible. To make this situation even more challenging, new social phenomena, particularly social media (e.g., Twitter, LinkedIn, Facebook), have arisen that cannot be studied with traditional methods, but our academic programs do not prepare the new generation of social scientists to link new media with other kinds of social data. We need to improve our tools and the information that we generate to better serve business, government, and the social and economic needs of the population.

Developments in information technology offer an unprecedented opportunity to collect diverse data at fine-grained spatial and temporal scales, and present a remarkable chance to change the way social science is conducted and to greatly expand the questions that can be addressed. Today, the proliferation of new data coming from the Internet and social media requires new ways to collaborate across social science disciplines and to link social science with genetic, linguistic, medical, environmental, biological, and earth systems science. This is an opportune time to rethink the primary ways in which data are collected, gathered, coded, curated, documented, archived, and disseminated in the United States. This volume builds on a series of workshops sponsored by the National Science Foundation (NSF) in 2005, 2007, 2008, 2009, 2010, and 2011 that led to a consensus for the creation of a network of regional data centers that could, when fully developed, represent the entire U.S. population and its diverse regions. In 2012, the NSF began to support a set of researchers and academic faculty from across the United States, known as "the Social Observatories Coordinating Network (SOCN)," to take on this challenge.[1] In our several years of discussions, the network settled on a challenge for the future—to design a national network

of regional data centers that could be coordinated through common objectives, sharing of protocols, and data sharing (Moran et al. 2014).

The network concluded that Americans' social outcomes and behavior are so situation- and place-specific that it is practically impossible to use widely dispersed national samples of populations to draw conclusions about processes in any one place. Although we can describe average educational attainment across 4th graders in the United States using data from the National Assessment of Educational Progress, the averages do not elucidate the social and cultural processes underlying educational underachievement in specific school districts. They can suggest general trends, but they cannot identify at-risk counties or school districts and the causes of underachievement. Populations tend to be spatially clustered by characteristics. This has implications for important issues of national concern. For example, the United States has always been characterized as the land of opportunity, where anyone can, through effort, succeed and attain the American dream. Yet a recent study showed that one out of four children raised in the middle class has slipped downward by their early 40s (Acs 2011); even more surprising is the finding that the chances for upward mobility and its maintenance depend specifically on where people live (Olinsky and Post 2013; Chetty et al. 2013). Similarly, a University of Washington study revealed that life expectancy for American males (and females) varies by up to 18 years, depending in which American county they reside (Institute for Health Metrics and Evaluation 2013). That is a remarkable range, and social scientists are just now mapping the social indicators to document and put it into context. Research focusing on context has the promise of pointing to pockets of concentrated disadvantage and poor health, where resources can be targeted to do the most good.

Why is there so much place-based variability in such outcomes as health and mobility? Social policy, economic conditions, race/ethnic/immigrant composition, 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 fostering opportunity and mobility processes, and the impact of their policies on individuals, families, and neighborhoods. 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 of regional data centers that we envision can address such questions and others central to our understanding of who we are as a people and a nation.

We need a new national framework or platform that is both scalable and flexible for addressing these new challenges, one that allows for rapid response to local crises such as the devastating impact of tornadoes and hurricanes or social unrest, and those provoked by prolonged regional drought or economic decline. We need a platform to address national crises such as the declining middle class and why in many schools across the country our children are not learning optimally. Over the past few years, experts have called for increasing our capability in cyber-infrastructure for the social, behavioral, and economic (SBE) sciences. We propose building this national network of social observatories to ensure that the

SBE sciences have an effective scientific infrastructure to contribute to and collaborate with other sciences in addressing questions of national importance.

What Should a National Network of Regional Data Centers Look Like?

Such a network needs to be capable of representing the people and the places where people live both in the aggregate and in the fine detail—detail that captures local and regional differences. In doing so, the network would be able to paint a picture that is representative of the nation's population and a picture of population dynamics as detailed as is currently captured by our national surveys and research infrastructure. However, it would go far beyond the national surveys and would also be a place-based sample. Unlike most existing research platforms, this place-based capability will ensure that we understand not only the urban places where the majority of the population lives, but also the important medium and low density places that represent a vast majority of the communities and land area in the nation. To do so, the observatories would identify several hundred census tracts to be systematically studied over time and space by regional data centers spread across the country. Unlike the census, taken every 10 years, such regional data centers will be able to provide a continuous stream of information for the nation and thereby better address the dynamics of change in our society, allowing researchers to be able to quickly see how national policies affect local places. This is a totally different effort from that undertaken by our national surveys. It is not meant to replace them but rather will both offer a broader national picture of the population and also deepen our understanding of the population in places across the country using fine-grained methods. By addressing the context of individual activities and decisions, social observatories would provide a more complete understanding of socioeconomic success and failure in our society and what we might do to promote the former.

The observatories would study the social, behavioral, and economic experiences of the population and its physical and environmental context in fine detail. They will do so by using complementary methods, including ethnography, experiments, surveys, observations, geographical information systems, systems science, records searches, and historical and archival methods. These observatories will work closely with local and state governments to gain access to administrative data that will provide not a sample of the population of those hundreds of tracts, but complete records on the whole of the population in those tracts, thereby ensuring a depth of understanding, and integration of knowledge, heretofore never achieved in analyses. A network of regional data centers would do all this while at the same time being less invasive—and the data less prone to declining participation and response rates than national surveys—because they would be more closely tied to the local community through agreements with local private and public institutions and not need to rely on telephone calls. No matter how

the nation changes or where its people move, the observatories would be able to describe how people and places change over time.

To date, the best neighborhood studies feature specific cases, such as Chicago or New York City, where substantial investments have been made to create systems of linked data such as we propose for these regional centers. They have been created because our larger cities have recognized that they need better data to serve their citizens while minimizing costs of such data acquisition and use. With a network of regional data centers, we can contemplate the possibility of a national sample of neighborhood contexts that can be studied at multiple levels and in multiple ways. Linked with health outcomes, it will be possible to consider the effects of, say, poverty while taking into account chemical exposures; it will also be possible to consider the effects of chemical exposures while controlling for poverty and other social characteristics of local contexts. There are virtually no studies to date that have done this.

Our Proposed Network

Our group, and earlier workshops on cyber-infrastructure for the social and behavioral sciences, has proposed the development of twenty to twenty-five regional data centers located across the United States. Each center would collect, organize, create, and disseminate data. These regional data centers or "social observatories" would follow about 400 census tracts over time and space from these twenty to twenty-five regions across the country. Working closely with local and state governments, they will access administrative data that will provide not a sample of the population of those several hundred census tracts, but complete records on all the population in those tracts (i.e., circa two million Americans). The centers would serve as data collection facilities wherein data are cleaned, linked, and made available for legitimate research purposes through a secure integrated data dissemination system. Although they would be charged with keeping data on their particular geographic region, some of these centers may have a national focus as well. These centers may conduct surveys, but they would also use data sources that until now have not been part of the toolbox of the social sciences, and they would connect these data to local context and place without losing the capacity to aggregate and serve as a national sample of people and places in the United States. Our vision is that, collectively, they would offer a nationally representative sample, one that was highly clustered so as to capture local context and variability.

A national framework provides an enormous advantage. First, it allows generalization across multiple contexts. The national framework permits comparison of variables and questions across multiple locations. It provides improved conceptual models that are not specific to place and can take into account variability. It ensures national representativeness. Second, the framework provides a rapid response capability. Over the observational period, at least some of the sites are likely to experience emergencies or crises. Because of the time dimension, the design will have an improved ability to disentangle causality. Third, having a decentralized structure permits each center to have a unique substantive focus.

Substantive foci of the regional data centers

Two major concerns about America's future are the adequacy of its physical infrastructure and the robustness of its economic structure. Many people worry that our transportation, environmental, water, sewer, educational, and even governmental infrastructure are outdated and in need of replacement. Yet efforts to move this agenda repeatedly fail to gain support in legislatures and among the public, despite the obvious benefits to business and citizens. People also worry that structural changes in the U.S. economy are making the nation less competitive in a global world and that the economy offers opportunities to some but not others. *The Economist* (2016) noted that the U.S. economy lacks vitality and competitiveness, and that it appears to be moving toward ever greater concentration of wealth, yet it is failing to benefit the larger population on which the economy depends. High profits are absorbed by ever more concentrated institutions, rather than being passed on to consumers or invested in innovation. This is a formula that reminds one of oligopolistic behavior: very high returns on capital, ever greater concentration, and control of prices in the hands of a handful of firms that cannot but lead to greater wealth concentration and inequality, according to the *Economist* article.

The major substantive foci of the data centers would initially be on questions of (a) change and adaptation and (b) opportunity and mobility, both broad questions that require data linkages and granularity in data sources. Because they would focus on place and context as well as people, researchers could identify the kinds of investments in infrastructure that provide the greatest opportunities for improvements in well-being with the fewest barriers. For example, one of the growing challenges for poor neighborhoods has been the exodus of grocery stores and therefore access to fresh food at reasonable prices. Creating opportunities for businesses to provide better access to healthy food can be investigated as a way to improve the lives of people who may be at an economic disadvantage. Regional data centers could also provide information on the organizational structures of communities that may facilitate or hinder appropriate adaptation to ongoing economic, social, and environmental change. This detailed understanding can inform public agents about how local economies might need to be reconfigured to better compete for jobs, for example. Of course, identifying needs in particular areas of the country creates the potential of having to ameliorate or reverse structural inequities and could lead to conflict among local groups. However, this does not mean that we should not move ahead to identify such needs and seek solutions.

What kinds of data would be collected?

To study local contexts such as communities and neighborhoods, we need spatially referenced administrative data, GPS-enabled cellphone data on the movements of individuals through their day, social media data, remotely sensed and observational data such as Google Street View, and survey data. These new data will be rich in detail. But while detail is important, the key to their use for

social and behavioral research is linking them across different levels. Linking individual information to administrative data or to other characteristics of communities in which individuals live, linking information on social media activity to health or other events occurring within the area, and linking medical records with housing and health data are only some such examples.

Designed properly and operated efficiently, these networked regional data centers will provide a nimble platform to incorporate changing sources of information that are being created by social media companies and on other media platforms such as cellphones. The goal is to gain access to the new forms of communication used by the nation to understand how they transform how people think and what motivates them to act in certain ways, and how they construct virtual and real social networks and communities. The task here will be to improve the granularity of data; provide in-depth context to data; and address issues of social, time, and spatial scales. New cutting edge approaches such as data trawling and web scraping will produce detailed accounts of movement, social networks, and other forms of community building that require interpretation by bringing social theory and history to inform the analysis of tweets and other data moving across cyberspace.

Having these networked regional data centers will transform how the SBE sciences go about their work; they will encourage the integration of the SBE sciences, rather than promoting the fragmentation that we have experienced since the 1960s. The latter was a necessary phase to achieve greater depth through specialization but has over the years had the effect of making it ever more difficult for SBE scientists to share methods and approaches to address issues of national importance. The regional data center network will explicitly promote what is now a broad call from the National Academy of Sciences to integrate the social and physical sciences to address issues of importance with the best tools available without regard for disciplinary origins. Regional centers across the nation, with the explicit charge of ensuring that teams of scientists are working together around questions of national interest, will help to integrate the sciences and serve the nation better by providing diagnostic and policy-relevant solutions at a variety of scales from local to state to national to international issues. Although there are critical issues of privacy to be addressed in this geocoded world, the observatories will be a place where these concerns can be addressed systematically and lead to the creation of standards for ensuring privacy of sensitive information.

Regional and local data centers are already happening

A number of communities across the United States are developing collaborative regional data gathering efforts to document the linkages between people and place that go beyond specific city or state boundaries. The National Neighborhood Indicators Partnership (Kingsley and Pettit 2011) is active in more than thirty-seven cities. This partnership collects and shares data to better serve their communities and learn from one another. In addition to this existing network, a community of scholars has been working on individual elements of an ambitious network of regional data centers to ensure that the American people have

available in a timely fashion nationally scaled and locally relevant information to make better business, health, education, and other important decisions. These include scholars at the New York Academy of Medicine and the New York City Department of Health and Mental Hygiene, the University of Pennsylvania, Chapin Hall–University of Chicago, the University of Colorado–Boulder, Portland State University, the University of Dallas, the Ohio State University, and American University, to name just a few examples. Cities such as Chicago have built impressive spatially explicit data bases that allow for quicker responses to social needs, and have provided a public portal so that citizens can engage the government to be more responsive and can be engaged with what happens in their city. We have seen a proliferation of these efforts across the nation. There are efforts to articulate some of these endeavors, but a larger and more systematic effort is needed to ensure that these efforts coalesce and provide a more complete picture of both local and national processes.

Methodological advances: Linkages across data

The important challenge to data collection for the foreseeable future is linking diverse data in a way that is cumulative, accurate, and accessible to a broad range of researchers. For SBE research, it is critical to develop new methods that can link persons and information about them and their environments across different data platforms. The proposed regional data centers would undertake the challenging task of linking a broad array of information—from administrative data (local and state and regional), to social media (Twitter, Facebook), to census and other surveys, to ethnographic data, to data from experiments such as randomized controlled trials—to address how different human communities make decisions. This is the issue that was addressed by the conference convened by the SOCN in 2016 on New Data Linkages, and that this volume addresses.

How This Volume Will Move Us Forward

The NSF SOCN sponsored a conference in the Washington, DC, area on March 24–25, 2016. The purpose of this conference was to bring together researchers involved in different regional data collection and linking efforts to (1) promote synergies across projects and (2) explore what types of issues have arisen that could be facilitated by a regional system of data centers. The call for papers was issued in spring 2015 across a number of academic disciplines, including Demography, Sociology, Economics, Psychology, Anthropology, Geography, Hazards and Environment Risk, Political Science, and Statistics. From this call, we gathered eight different research teams to discuss their work and explore potential collaborations across these projects. Besides the principal investigators of these projects, we invited members of the SOCN and guests from funding agencies such as the National Institutes of Health, NSF, the National Academy of Science, and private foundations to serve as discussants and observers.

This volume has the following structure: This article, the introduction, documents our thinking about how to develop a network of regional centers into a national platform capable of serving the SBE sciences and how several research groups are already making progress in linking data. In the following three sections, authors give concrete examples of how linked data are advancing research in (1) community characteristics and quality of life, (2) individual and community factors and health, and (3) change and adaptation and disaster planning. Although these topics are not the only issues in which such a network could be actively involved, community quality of life, health and health care, and adaptation to immigration and climate change represent important areas of concern for families and for future public policy discussion in the United States, and all are issues that have attracted a great deal of attention and on which solutions remain incomplete and so far unsatisfactory. We then offer our conclusions and suggestions for future research.

Community characteristics and quality of life

In the first article of this section, Michael Bader and colleagues report on a new initiative in the Washington, DC, area that brings physical characteristics of communities into research on behavior. His team addresses how to link data from Google Street View to better assess physical infrastructures that can facilitate or hinder the mobility of the aged. Second, Christopher Browning and colleagues, a team from Ohio, summarize their research on adolescent behavior through the use of new tools such as smartphones that map activity spaces instead of only neighborhoods, and link individual and activity space data with administrative data. Their article addresses the characteristics that affect exposure to violent locations, which could threaten mental/physical health. Finally, Amy O'Hara, Rachel Shattuck, and Robert Goerge report on both U.S. Census Bureau and Chapin Hall–University of Chicago initiatives to integrate data sources for improved research on families. They link federal surveys with federal and state administrative data to better measure families and households, obtain more extensive information on families, evaluate survey coverage and accuracy, and evaluate participation in social welfare programs.

Individual and community factors and health

The first contribution in this section—from Hongying Dai, Brian Lee, and Jianqiang Hao, a team from the Children's Hospital and the University of Missouri—uses novel linkages of Twitter data, health survey data, and socioeconomic data from the U.S. Census Bureau to predict community asthma burden. Next, a team from the Population Council extends the study of health to the use of modern contraception in African communities. Jean Digitale and colleagues describe how individual and community factors are associated with contraceptive use among young women in Malawi and their selection of a contraceptive provider. The authors link data from a survey of youth with data from a survey of

family-planning service providers. The third contribution is from a team at the University of Dallas and University of Texas Southwestern Medical Center, which examines the health of residents of low-income areas in the city of Dallas. Tammy Leonard, Amy Hughes, and Sandi Pruitt examine the coping strategies of families experiencing a health shock, by linking medical record data with community service administrative data and housing appraisal data.

Adaptation and disaster planning

Coming from the Population Studies and Training Center at Brown University where she has been studying the consequences for New Orleans of Hurricane Katrina, Elizabeth Fussell and her colleagues offer a study of how past population trends, population density, cumulative weather-related losses, and weather events intersect at the county level to influence future population change. Also in this section, Elyakim Kislev studies what will likely be one of the major challenges for the United States in the not too distant future: the integration of European immigrants whose country of origin is Africa or the Middle East. Kislev demonstrates how a variety of linked data can be used to track the mobility of immigrants in Western Europe, examines the characteristics of those who subsequently immigrated to the United States, and compares the successes of these immigrants with different origins to U.S. natives.

In the concluding article to the volume, Barbara Entwisle, Sandra Hofferth, and Emilio Moran offer thoughtful reflections on the articles in this volume and how and in what ways we might move forward to achieve the promise presented by data linkages within the context of a national network of regional centers that can enhance the SBE sciences and, in doing so, serve society.

We invite readers to join us in working toward the advancement of interdisciplinary science by taking on the challenge of linking relevant data and utilizing innovative methods to elucidate social dynamics and solve challenging problems all around us now and in the foreseeable future.

Note

1. See materials at www.socialobservatories.org.

References

Acs, Gregory. 2011. *Downward mobility from the middle class: Waking up from the American dream.* Washington, DC: Pew Charitable Trusts.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2013. *The economic impacts of tax expenditures: Evidence from spatial variation across the U.S.* Cambridge, MA and Berkeley, CA: Equality of Opportunity Project, Harvard University and the University of California, Berkeley. Available from http://www.equality-of-opportunity.org/.

Economist. 26 March/1 April 2016. Too much of a good thing: Why high profits are a problem for America, 23–28.

Foster, Ian, Rayid Ghani, Ron S. Jarmin, Frauke Kreuter, and Julia Lane. 2017. *Big data and social science: A practical guide to methods and tools*. Boca Raton, FL: CRC Press

Institute for Health Metrics and Evaluation (IHME). 2013. *The state of U.S. health: Innovations, insights, and recommendations from the Global Burden of Disease Study*. Seattle, WA: IHME.

Kingsley, G. Thomas, and Kathryn L. S. Pettit. 2011. Quality of life at a finer grain: The National Neighborhood Indicators Partnership. In *Community Quality-of-Life Indicators: Best Cases V*, eds. M. J. Sirgy, R. Phillips, and D. Rahtz, 67–96. New York, NY: Springer

Moran, Emilio F., Sandra L. Hofferth, Catherine C. Eckel, Darrick Hamilton, Barbara Entwisle, J. Lawrence Aber, Henry E. Brady, Dalton Conley, Susan L. Cutter, and Klaus Hubacek. 2014. Building a 21st century infrastructure for the social sciences. *Proceedings of the National Academy of Sciences* 111 (45): 15855–56.

Olinsky, Ben, and Sasha Post. 2013. *Middle out mobility: Regions with larger middle classes have more economic mobility*. Washington, DC: Center for American Progress.