COUNTERPOINT TO "BIG DATA FOR PUBLIC POLICY: THE QUADRUPLE HELIX"

Ron S. Jarmin and Amy B. O'Hara

Lane argues for a "quadruple helix" approach to building the data and analytical infrastructure to support public policy in the 21st century. We agree—the augmented federal statistical system and proposed administrative records clearing house described in our Point essay is not sufficient. That said, we argue next that the federal statistical system is well positioned to provide comprehensive benchmark data. Lane's other three strands are critical to furnish missing pieces where the federal agencies are lacking. We review Lane's quadruple helix, summarizing our agreement and concerns, then discuss ways the four strands—and a fifth—can be brought together to yield the biggest payoff for public policy analysis and research.

LANE'S QUADRUPLE HELIX

The first strand in Lane's quadruple helix is city and state government agencies. We agree that state and local government agencies need to expand their role as both sources of data and centers for policy analytics. She mentions a number of excellent initiatives, mostly in large cities, but we are concerned that many state and local governments will lack resources or the urgency to improve data and analytics in a way that permits detailed and comparable policy analyses.

Lane's second strand can assist with this, especially as state and local governments partner with universities. Indeed, this is her primary call to action and has been a feature of several initiatives already underway that leverage data and develop analytical talent at the local level (e.g., University of Chicago's Urban Center for Computation and Data, Center for Government Excellence at Johns Hopkins University). However, it is unlikely that smaller cities, rural areas, and larger cities lacking a local university will devote their scarce resources to data and analytics. We are also concerned that diffuse efforts at the local level may lack the commitment to privacy and confidentiality protections that exist in more mature federal and state activities.

Lane's third strand is private sector data. Companies undoubtedly possess rich data that can support economic and social measurement as well as public policy analysis. Statistical agencies such as the Bureau of Economic Analysis, Bureau of Labor Statistics, and Census Bureau are exploring a number of ways to build more timely and detailed statistical products by integrating private sector "Big data" (Bostic, Jarmin, & Moyer, 2016). Academics have flooded many high-profile firms with data requests to support research. Companies must rethink data access as they reconcile privacy and resource issues with a desire to support research.

Finally, the fourth strand of Lane's quadruple helix, the federal statistical system, generates official statistical estimates and the survey and administrative data that underlies them. We agree with her assessment of our strand's strengths (e.g., rigorous principles and practices, privacy, and confidentiality) and weaknesses (overreliance on surveys and lack of nimbleness). Like other strands, the federal system finds itself lacking in computing and human capital resources needed to take full

Journal of Policy Analysis and Management, Vol. 00, No. 0, 1–3 (2016)

Published by Wiley Periodicals, Inc. View this article online at wileyonlinelibrary.com/journal/pam
DOI:10.1002/pam.21926

advantage of the data. As mentioned in our Point essay, we are enhancing the existing Census Bureau linkage infrastructure to improve data access, providing a core of well-curated, standardized, and comparable data for all U.S. communities.

HOW THE STRANDS OF THE QUADRUPLE HELIX SHOULD WORK TOGETHER

Both Point essays make the case that we stand on the cusp of transformative change in social and economic measurement and related policy analysis. We find ourselves in a data-rich world with challenges involving secure data access and the development of reliable statistical information and evidence for policymakers using a multitude of new data sources. Tackling these challenges requires all four strands of Lane's quadruple helix, discussed next.

Human Capital Development

Policy analysis requires the best possible understanding of how the world actually works. Successful analyses bring the right data, models, and subject matter expertise to bear. While the data scientist's tool kit is essential, statisticians, social scientists, and policy analysts are needed. The nation's universities must train the next generation of social scientists, government statisticians, and policy analysts in data curation and analytics. This training needs to be steeped in the problems faced by the four strands. The class designed recently for the statistical agencies (Jarmin et al., 2014) exemplifies cross-disciplinary training addressing real-world (not solely commercial or computer science) problems.

Infrastructure Interface

We strongly support pushing expert policy analysis to where policy is implemented and its impacts felt. The university–city partnerships Lane describes are an excellent place to start. But we argue that these efforts will provide greater evidence and foster broader analytic innovations when plugged into the federal infrastructure, specifically the administrative records clearing house at the Census Bureau. Linking to the broader statistical infrastructure will allow local efforts to leverage federal investments in data, standards, tools, and policies, permitting analyses that are comparable across jurisdictions and over time. To the extent that there are economies of scale in data curation and provisioning, state and local agencies may find it cost-effective to participate directly in the clearing house to provide analysts and evaluators access to their own data, and for negotiated access to other localities' data. These federal, state, city/local, and university solutions must also interface with private sector data sources enabled by standards (e.g., spatial data infrastructure; Hendriks, Dessers, & van Hootegem, 2012). Incentives must be identified to encourage data access, and technical solutions, such as secure multiparty computation, are needed.

Reliability and Benchmarking

There are many open questions regarding utilizing private sector data for federal statistics and estimates. Access issues predominate with questions regarding cost, reliability of access over time, data quality, and data provenance. Fitness for use is a major concern, with well-known examples where initially promising uses of private data for public measurement activities were later shown to have serious flaws, as in the case of Google Flu (Lazer et al., 2014). Still, private data will generate new insights, especially in urban environments where dense networks of sensors

and the Internet of Things provide the raw data to describe behavior in rich ways (Townsend, 2013). The quality and consistency of such data need principled analysis and documentation.

The Fifth Strand

Finally, we close with the observation that Lane's quadruple helix omits an important fifth player in conducting and publicly disseminating high-quality policy analysis in the 21st century: the media. The explosion of data and the tools to analyze them enable the media (both traditional and new outlets) to churn out increasingly complex analyses and graphics using both official statistical and new big data sources. Even in cases where policy analyses are performed by academic or government analysts, the results are broadcast through the Tweets, Blogs, and stories of data journalists on organizations such as fivethirtyeight.com and The Upshot. Thus, they represent a fifth strand to a quintuple helix.

RON S. JARMIN is the Assistant Director for Research and Methodology at the U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233 (e-mail: ron.s.jarmin@census.gov).

AMY B. O'HARA is Chief of the Center for Administrative Records Research and Applications at the U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233 (e-mail: amy.b.ohara@census.gov).

REFERENCES

Bostic, W. G., Jarmin, R. S., & Moyer, B. (2016). Modernizing federal economic statistics. American Economic Review, 106(5).

Hendriks, P. H. J., Dessers, E., & van Hootegem, G. (2012). Reconsidering the definition of a spatial data infrastructure. International Journal of Geographical Information Science, 26, 1479–1494.

Jarmin, R., Lane, J., Marco, A., & Foster, I. (2014, November). Using the classroom to bring big data to statistical agencies. Amstat News. Retrieved April 28, 2016, from http://magazine.amstat.org/blog/2014/11/01/classroom/.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). Big data. The parable of Google Flu: Traps in big data analysis. Science, 343, 1203–1205.

Townsend, A. M. (2013). Smart cities: Big data, civic hackers, and the quest for a new Utopia. New York: W.W. Norton & Company.