

ISSUES

IN SCIENCE AND TECHNOLOGY

SPRING 2018

NATIONAL ACADEMIES OF
SCIENCES, ENGINEERING,
AND MEDICINE

THE UNIVERSITY
OF TEXAS AT DALLAS

ARIZONA STATE
UNIVERSITY

Preparing for

THE NEXT FLOOD

A Response to Reactionary Populism

Straight Talk on Workforce Skills

New Rules for Self-Driving Cars

Infrastructure in the Anthropocene

The Downton Abbey University System

Using Data to Make Better Decisions

USA \$12.00 / CAN \$13.00

8 1 >



0 56698 57433 0

Helping Communities Use Data to Make Better Decisions

Communities, especially small and rural ones, need to take advantage of new techniques for collecting and analyzing data to better serve their residents. Here's a plan to help them succeed.

Across the United States, cities and local communities are responsible for addressing the everyday societal needs of their residents. Their list of responsibilities includes protecting public safety, providing critical health services, and fostering economic development, among others. The good news is that communities can increasingly participate in the emerging big data revolution, which is providing new opportunities for them to build insights and capacity to meet current needs and prepare for their futures.

Capitalizing on this opportunity, however, will require communities of all sizes to implement new approaches for making decisions that utilize the full array of available data. Many communities—especially small and mid-sized ones, as well as rural communities—will need assistance in building their capacity to access data and leverage statistical, computational, and social science expertise necessary to manage, analyze, and interpret data. Here too, there is good news. Several states have developed innovative programs that feature collaboration between communities and nearby universities to aid in data-enriched decision-making, and their efforts are expected to be quickly adopted elsewhere.

Our experiences may help to tell this tale. In joint efforts at Virginia Tech and Iowa State University, we have created a model for data-driven community

engagement and community-based research, which we call Community Learning through Data Driven Discovery (CLD3). The key innovation in CLD3 is, as its name suggests, community-based research where the community participates in asking and answering the questions that drive information gathering and provide insights relevant to program or policy decisions. We have successfully tested the model in several communities in our states.

Based on this work, we propose a strategy to bring the CLD3 process to local governments of all sizes across the nation by leveraging the expertise of land-grant universities, which have been effectively disseminating information to communities and rural areas since the mid-nineteenth century. The plan also capitalizes on the expertise of several groups, notably the Cooperative Extension System (CES) and Regional Rural Development Centers (RRDCs), that are associated with the land-grant system and have their own demonstrated track record on transferring up-to-date research information to various communities.

Building data bridges to good decisions

Achieving data-enriched local governance, where information is shared among community leaders, officials, and residents to inform evidence-based policy-making, is crucial for an efficient and effective

democracy. CLD3 seeks to build capacity and resources for data-driven governance at local levels where relevant expertise may be lacking. The model essentially brings together data science and community empowerment. It incorporates mechanisms for merging local, state, and federal information on housing, transportation, health care, recreation, and public services, including social services, police, fire, and emergency medical services, libraries, and utilities. Then university researchers mining that mountain of data will provide community officials with the modern decision-making capacity that can enable better-informed, evidence-based policies. Importantly, the process extends through a continuous, sustainable, and controlled feedback loop that allows for updating and modification as needed. CLD3 thus enables communities of any size to sustainably manage social and economic challenges.

Community learning will provide government leaders with the capability to tackle the issues confronting their communities and to learn what works, for whom, and in what context. The complete community learning cycle includes:

- Exposing local leaders to data-driven learning through a forum across government departments and offices. The forum is structured to bring local program and agency leaders together through a dialogue process that is designed to discuss issues and data sources that span their boundaries.
- Working with local leaders to develop CLD3 governance structure, such as data access and sharing agreements, and defining their critical questions and issues facing their community by asking them “What keeps you up at night?”
- Identifying and wrangling the data sources that cut across programs to provide the knowledge and insight for building an agile and scalable statewide CLD3 data infrastructure.
- Using statistical and geospatial learning along with the communities’ collective knowledge to develop hypotheses to inform policy decisions relevant to issues.
- Developing and deploying intervention strategies and education programs to address identified issues, evaluating hypotheses using rigorous experimental design principles, and adapting strategies based on continuous evaluation.

- Training the next generation of local government civil servants to be data savvy.
- Engaging entrepreneurs to develop products and tools using local data to support economic development and local, state, and federal governments.

Land-grant universities, which are at the heart of CLD3, operate in every state and most US territories. They trace their roots to the Morrill Act of 1862, passed by Congress and signed into law during the Civil War by President Abraham Lincoln. The law gave states that remained in the union public lands to be sold or used for profit to establish at least one college that would teach agriculture, mechanical arts (engineering), and military tactics, in addition to traditional classical subjects, to all people, including those of average means. Later iterations of the Morrill Act extended the program to the former secessionist states and to a number of Native American tribal groups. Over time, most of the land-grant colleges transformed into full-fledged universities charged with teaching, research, and transferring knowledge beyond their campuses.

The Cooperative Extension System, another pillar of the CLD3 initiative, was created under the Smith-Lever Act of 1914. It now has units embedded in over 3,000 counties and cities across the United States, focused primarily on translating and disseminating to the public the evidence-based research findings produced by land-grant universities (or other public universities) and their agricultural experiment stations to address issues in agriculture and natural resources, human sciences, and community and economic development. CES is a cooperative activity involving the federal government (funded through the US Department of Agriculture’s National Institute of Food and Agriculture); the states, through the land-grant institutions; and local governments, through the CES network of cooperative extension professionals. Cooperative extension professionals work with university researchers to translate their science-based research results into language and decision tools appropriate for targeted audiences. Cooperative extension professionals work with local government, residents, and interest groups to solve problems, evaluate the effectiveness of learning tools, and collect grassroots input to prioritize future research. By living and working in communities, cooperative extension professionals rely on

We envision an enhanced role for the Cooperative Extension System that builds upon its current roles and experience gained through the Regional Rural Development Centers to bring data in service of the public good through deepening the partnership between communities and land-grant universities.

existing relationships to respond to local needs, build trust, and engage effectively with communities.

The Regional Rural Development Centers, also central to CLD3, were created by the Rural Development Act of 1972. There are four of them across the country: the North Central Regional Center, hosted at Michigan State University; the Northeast Regional Center, hosted at Penn State; the Southern Rural Development Center, hosted at Mississippi State University; and the Western Rural Development Center, hosted at Utah State University. The centers serve as sources of economic and community development data, decision tools, education, and guidance in rural communities. They connect rural areas to the nationwide network of land-grant university researchers, educators, and practitioners to provide useful information and hands-on, community-level training. Each center uses its regional network of land-grant universities to conduct research and develop education and outreach programs, often carried out by cooperative extension professionals, to teach rural communities how to make science-based decisions about their community and economic development investments.

CLD3 innovates in offering a new approach for cooperative extension professionals and professionals at the RRDCs to enhance their effectiveness by working with local community leaders and officials in a format that incorporates the knowledge gained through research and education into local decision-making to create positive change for the community. The key medium for these interactions is data, especially local data integrated with other data sources.

Taking community snapshots

Providing communities with the ability to harness the data revolution and gain data-driven insights into problem solutions they identify will require

cooperative extension professionals to form new relationships with data science specialists at land-grant universities and to partner with local government officials in new ways. What we have done at our universities using CLD3 may illustrate the process. Both universities have on-campus big data initiatives, experienced data scientists practicing engaged scholarship, and a commitment from the university leadership to steward collaborative processes.

The Virginia Tech and Iowa State senior academic, research, and cooperative extension leadership have taken the first step of embracing the CLD3 vision through allocating some resources to implement the process. They recognize that in repositioning CES beyond the current approaches and program area boundaries, CLD3 provides a more holistic and data-driven engagement with their communities and a more direct engagement with local governments. For both institutions, this is an outgrowth of the current data science revolution.

Using Virginia and Iowa as initial CLD3 deployment exemplars was an intentional choice. They are complementary in their style of local governance. Virginia is primarily county-based, governed by boards. Iowa communities are primarily city-based, governed by mayors and councils. Communities in both states have indicated potential issues that could be addressed in early roll outs. Common issues include anticipating and preparing for meeting the day-to-day needs of vulnerable populations, providing access to food and healthy life style options, adapting to rapidly changing demographics, improving education, and expanding access to jobs, while understanding how to better serve everyone in their communities. Looking deeper, consider what we found in three particular communities:

Pulaski County, Virginia. This is a rural community in the southwestern part of the state,

with a population of about 34,000 people in 2016. During the CES-led first meeting with its local government leaders, we learned that employment in the manufacturing sector has been steadily declining, with the county being hit especially hard in 2008-09. At the same time, drug use increased, and other problems followed, including high teen pregnancy and incarceration of parents. County officials told us that they have attracted businesses to Pulaski, but the current residents have a hard time getting jobs at these businesses because they cannot pass the drug test.

Working together, we were able to provide a data profile of the county as a first step toward identifying the vulnerable populations, and we also produced data comparisons with other surrounding communities. These data-driven insights can be used to develop and discuss interventions by the local government as well as new CES programming to better target community needs. Some early insights include noting that a smaller percentage of Pulaski residents had jobs within the county in 2015 (40%) compared

populations and large increases in enrollment of students in English language learning classes.

Most of these new arrivals work in the manufacturing plants in Marshalltown and the county. During the data discovery forum with university researchers, cooperative extension professionals, and local government leaders, the city administrator said she would like to use data to better inform decisions about how to integrate this growing population into the community. Among particular needs, she wants better access to data that would help improve government officials' understanding of how public transportation is meeting the needs of this demographically changing population, how current fee structures governing access to and use of parks and recreation programs might be changed to increase use by diverse populations, and what strategies might help in maintaining the quality of the neighborhoods through improvement projects.

One initial set of data they have found particularly insightful is the substantial shifts in

Community learning will provide government leaders with the capability to tackle the issues confronting their communities and to learn what works, for whom, and in what context.

with 2002 (53%), resulting in many county residents facing longer commutes, which may exacerbate family problems. Also, though local stakeholders were aware of issues around teen pregnancy, the data insights demonstrated that the issue is not consistent across the region, a finding that may help officials better target intervention programs. Finally, the local economic vulnerability indicators that we developed may help geo-locate potential vulnerable populations as well as changes in these populations over time.

Marshalltown, Iowa. This is a rural city of about 28,000 people, located within Marshall County. The community has a growing immigrant population. Hispanics comprised less than 1% of the total population in 1994, 13% in 2000, and now over 27%. The overall population growth during this time has been 8%. In addition, people from other countries are coming to the community, and there are now over 50 languages spoken in the public schools, and ethnic and racial minority populations are expected to exceed 50% of Marshalltown's population in the near future. This demographic shift has resulted in large pockets of linguistically isolated

commuting patterns of their residents between 2002 and 2015. These data from the US Census Bureau's OnTheMap, a web-based mapping and reporting application, show that more workers are commuting longer distances to work, and they are more likely to head southwest (toward Des Moines) and less likely to commute to the northeast of Marshalltown (toward Waterloo and Cedar Rapids).

Fairfax County, Virginia. This is an urban county with a population of almost 1.1 million people in 2016. County officials told us that among particular issues, the Department of Health and Human Services is interested in improving its capacity for long-term program planning through identifying patterns of activity throughout the county. The officials' initial interest centered on reducing obesity in young people, and they want to understand how integrating federal data at the county level with sub-county local levels of data could provide needed new data-driven insights.

The geographic boundaries of most interest to county officials for program planning around healthy lifestyles are school and political districts administered by elected supervisors. We found that

these geographies do not align with traditionally used demographic boundaries of census block groups or tracts, so metrics such as poverty, participation in subsidy programs, and transportation availability need to be calculated over new geographic regions within the county. By using statistical methods, we were able to align estimates based on the Census Bureau's American Community Survey data to new geographic boundaries of interest to county officials. These methods preserve multivariate relationships in the data and are able to coherently integrate margins of errors, providing confidence intervals for the new estimates.

We also looked at data on the locations of rental and owned housing units, derived by examining local tax assessments and rental records, to map distances from home to areas of interest, such as farmers markets (healthy food), fast food restaurants (unhealthy food), and recreational activities (opportunities for exercise). The results may help county officials tailor health and development programs to meet the specific needs of particular neighborhoods.

Partners eager to work

When we entered into our project, we found rich ground for tilling. Within the Association for Public and Land-Grant Universities, the Extension Committee on Organization and Policy expressed great enthusiasm about deploying CLD3 through the Cooperative Extension System. As an initial step, the committee conducted a survey during June and July of 2017, asking cooperative extension professionals at land-grant universities to describe the sources of data and analyses they utilize in their community-based work, as well as their access to data expertise on campus. The results showed that they use data in new ways to support program development. Of the professionals from the 22 universities that responded to the survey, about three-fifths use data from federal statistics, compile data from multiple sources, and generate new tools to guide decision-making. Many of them do not access or use local (municipal) data, such as land use and natural resource data. About one-third of the institutions have data analysis expertise among their leadership teams, and approximately half of them have access to similar expertise on campus that could be leveraged into data-driven engaged scholarship. Similarly, the Regional Rural Development Centers are also ready and prepared to participate, having gained considerable experience in working with communities to frame economic development plans.

In our initial plans worked out with the Extension Committee on Organization and Policy, Iowa State

University and Virginia Tech, working in concert with their respective RRDCs, would provide leadership in seeding projects in each state that would test CLD3 as a model of community engagement that could be adopted nationally. We would introduce communities in Virginia and Iowa to CLD3 over a three-year time frame. Working with the RRDCs, lessons learned and tools developed will be shared such that CLD3 can expand rapidly across the country. This will include working collaboratively to develop, deploy, and curate data sciences practices and processes and establish communities of practice with academic researchers, CES professionals, local government officials, and other community stakeholders. Cooperative extension professionals and specialists would convene with communities to leverage academic research in response to the identified problems of local and state governments.

In addition to fostering more efficient, informed government decisions, one additional benefit of this new role would be developing a community workforce pipeline to supply new professionals with the data skills and the interest in applying these approaches, working with local governments. The universities are responsible for developing this workforce pipeline of data scientists and encouraging interest in social and public good applications, as well as exposing them to career opportunities in the Cooperative Extension System, local governments, and other stakeholder organizations, including nonprofits, within the community. Universities will do this through community research opportunities and engaged scholarship experiences. In our model, this will be in direct partnership with cooperative extension professionals and local government leadership.

One highly successful experiential learning model we have implemented is Virginia Tech's Data Science for the Public Good program. Student fellows in this program work in teams that are integrated horizontally across the disciplines needed to address complex issues, including statistics, data science, social and behavioral sciences, and public health, as well as vertically to provide collaboration with project stakeholders at all levels, undergraduates, graduates, postdoctoral associates, research faculty, and local, state, and federal agency leadership and nonprofits. The teams research sponsor-driven questions by exercising the CLD3 framework. Sponsors and questions come from local, state, and federal agencies. The program includes intensive data science training and serves as an incubator to educate and train the next generation of government data scientists by exposing them to data science projects that integrate data from

the municipal, state, and federal levels of government.

In parallel, we will develop local, regional, and state Communities of Practice, along with an overarching national Community of Practice. The use of online forums and integrated websites will help to create and maintain these communities of practice. As they evolve, there will be focused training efforts, meetings, and workshops where our practitioners, academics, and government colleagues share CLD3 research findings and best practices.

Going national

Many of the building blocks are already in place to move CLD3 nationwide as it demonstrates its capabilities. The current authorizing language, mission, and vision for the Cooperative Extension System allows for expansion. However, this initiative could be strengthened at the federal level by seizing the opportunity to enhance the CES role

in communities, such as their economic condition, activities, and health. Thus, we would propose adding another specific function to the mandated list: to provide infrastructure for data-driven governance to inform decisions and policy-making through collaboration with local governments, cooperative extension, and the land-grant system.

Additional resources will also be needed to encourage CES and local leaders to move beyond the status quo and to implement this vision behind CLD3 into their communities. Title VII (Research, Extension, and Related Matters) in the 2018 Farm Bill would be the appropriate place for new legislation to support the rollout of community learning. Our estimates for pilot deployment of CLD3 across Virginia and Iowa is \$10 million over three years. During that period, we would work with the RRDs to lay the groundwork for adoption in other states. After that, continued annual funding to deploy

Even as Congress deliberates, progress is possible. Communities across the country collect a wide range of data, but they are experiencing difficulties in accessing their own data along with other relevant open data to gain insights to problems they are experiencing.

and to authorize funding to support and accelerate adoption of CLD3, specifically through the National Agricultural Research, Extension and Teaching Policy Act of 1977 and the 2018 Farm Bill now under consideration in Congress.

The 1977 legislation lays out eight functions that the Agriculture Department's research and educational programs should serve. One of them is to "support agricultural research and extension to promote economic opportunity in rural communities and to meet the increasing demand for information and technology transfer throughout the United States agriculture industry." As the data revolution has transformed the way that universities, businesses, and government at all levels work and interact, this purpose is still relevant and an important role for land-grant and public universities working through CES. What has changed is the data that can be used to provide evidence-based insights into community infrastructure, such as operations, resilience, and sustainability; environmental conditions, such as water quality, air quality, and noise; and the lives of the people living

pilots to other states would be needed. We expect accelerated adoption and economies of scale as learning in states conducting pilot studies expands to other states, thus reducing start-up costs.

The vision for this legislation is to expand the CES role to connect with local government officials and civil servants through the CLD3 process to build capacity and ultimately create the foundation for evidence-based governance, using new approaches to integrate and model information from disparate data sources. To do this, CES would expand its current situational analyses and programming reports through use of local administrative data and social media. CES would also increase its collaborations with university researchers through engaged scholarship and by creating and curating processes to support data discovery, sharing, access, analytics, and evaluation for data-driven decision-making. This will also require development of a workforce with the skills to undertake this data science approach by engaging students in these community-based research projects.

There has been some congressional interest already.

In a draft report, the Committee on Appropriations noted that under the section regarding appropriations for Agriculture, Rural Development, Food and Drug Administration, and Related Agencies for fiscal year 2018, the bill directed the Agriculture Department's Economic Research Service (ERS) "to use data and evidence to address local challenges." Further, it said that "As part of the bipartisan effort to improve government capacity for evidence-based policy-making, the Committee encourages ERS to explore ways to assist rural communities in using data and evidence to address local challenges. In particular, ERS should examine ways in which local governments in rural communities could access the research and data expertise of public land-grant universities to help communities address local needs and priorities." Unfortunately, this language does not appear to have remained in the final appropriations bill, but the early language is a first step toward realizing the value of using local data and evidence to address local issues.

And even as Congress deliberates, progress is possible. Communities across the country collect a wide range of data, but they are experiencing difficulties in accessing their own data along with other relevant open data to gain insights to problems they are experiencing. A number of cooperative extension programs across the country are experimenting with different models of engagement around data science approaches, but they remain disconnected and uncoordinated. CLD3 provides the opportunity to bring a cohesive approach. We envision an enhanced role for the Cooperative Extension System that builds upon its current roles and experience gained through the Regional Rural Development Centers to bring data in service of the public good through deepening the partnership between communities and land-grant universities. This will provide the processes and resources to enable local governments to become data-driven learning communities and expand their capacity for data-driven governance. And we propose that universities adapt the Virginia Tech Data Science for the Public Good program as a model for equipping new generations of scientists with skills they need to provide policy-makers and government leaders with data analysis support that can inform intelligent decision-making. This approach will also bring data science to CES directly by enhancing its professionals' skills and catalyzing their engagement with university faculty and students.

Our hope is that through these steps, the Community Learning through Data-Driven Discovery model will bring the data revolution to local governments of all sizes in urban and rural

areas. And opportunities extend beyond the United States as well. The United Nations, recognizing the urgency of the need, is embedding urban research and data into several of its initiatives. UN leaders recently noted the need to bring evidence-based scientific analyses to city executives through engagement at the local, national, and multilateral levels, highlighting the critical need for community-university partnerships and the sharing of clear examples of success.

Sallie Keller is a professor of statistics and director of the Social and Decision Analytics Laboratory within the Biocomplexity Institute of Virginia Tech; Sarah Nusser is professor of statistics and vice president for research at Iowa State University; Stephanie Shipp is a research professor and deputy director of the Social and Decision Analytics Laboratory within the Biocomplexity Institute of Virginia Tech; and Catherine E. Woteki is a professor of food science and human nutrition at Iowa State University and a visiting scholar at the Social and Decision Analytics Laboratory within the Biocomplexity Institute of Virginia Tech.

Recommended reading

- Michele Acuto, "Global Science for City Policy," *Science Magazine* 359, no. 6372 (Jan. 12, 2017).
- Agriculture, Rural Development, Food and Drug Administration, and Related Agencies Appropriation Act, H. R. 3268, 115th Congress (2017).
- Committee on Appropriations, Agriculture, Rural Development, Food and Drug Administration, and Related Agencies Appropriations Bill, 2018, available online: https://appropriations.house.gov/uploadedfiles/hrpt-115-hr_.pdf
- S. A. Keller, V. A. Lancaster, and S. S. Shipp, "Building Capacity for Data Driven Governance: Creating a New Foundation for Democracy," *Statistics and Public Policy* 4, no. 1 (2017).
- S. A. Keller, S. E. Koonin, and S. S. Shipp, "Big Data and City Living: What Can It Do for Us?" *Significance* 9, no. 4 (2012): 4-7.
- S. A. Keller, S. Shipp, G. Korkmaz, E. Molino, J. Goldstein, V. Lancaster, B. Pires, D. Higdon, D. Chen, and A. Schroeder, "Harnessing the Power of Data to Support Community-Based Research," *Wires Computational Statistics* (2018).
- Regional Rural Development Centers, information available online: http://rrdc.info/about_us.html
- National Research Council, *Colleges of Agriculture at the Land Grant Universities: A Profile* (Washington, DCL The National Academies Press, 1995).
- US Department of Agriculture, National Institute of Food and Agriculture, Cooperative Extension Service, information available online: <https://nifa.usda.gov/extension>