



NSF-Census Research Network Newsletter

Vol. 3, Issue 4

Taking stock of NCRN's impact as an era ends

Change is in the air. There is finality in this newsletter. It is the last of a series that started in November of 2013, which in turn documented earlier activities of the NSF-Census Research Network going back to Fall of 2011, when the first awards in the network were made. You can find all past newsletters at <https://www.ncrn.info/news/newsletter>, and they are archived at <http://hdlhandle.net/1813/33762>.

Some of the NCRN research nodes are going to continue as organized entities; others will live on in the memory and research of their alumni. We've presented some of the students that earned their spurs at the nodes in these pages, and have now gone on to other positions in academia, government, and the private sector, and you can meet a few more in this issue.



Lars Vilhuber

Many of the researchers that started out as "FedStat neophytes" are now continuing to work on issues related to the federal statistical system, and sometimes as part of the federal statistical system. A document providing a more systematic overview of the outcomes of the NCRN is being prepared, and will be available at our website this fall; a presentation of that document is currently being planned for JSM 2018 as part of the SPAIG award activities (see next article) – stay tuned. At the Coordinating Office, we are preparing and supporting the wind down of network-wide activities. We will continue to collect and index materials from the nodes up until the end of 2017, such as

Continued on page 5

NCRN honored with SPAIG award from ASA

BALTIMORE, MD – The National Science Foundation-Census Research Network (NCRN), a node of which has been hosted at Cornell University for the past six years, has been awarded the American Statistical Association's Statistical Partnerships Among Academe, Industry, and Government (SPAIG) Award. The award recognizes excellence in collaboration between academe, industry, and government that results in significant contributions to statistics with

applications to real-world problems. The award was announced at the ASA's Joint Statistical Meetings in Baltimore, MD, on July 30, 2017.

The NCRN is composed of nodes at eight universities (Carnegie Mellon University, University of Colorado joint with the University of Tennessee, Cornell University, Duke University joint with the National Institute of Statistical Science, University of Michigan, University of Missouri, University of Nebraska, and Northwestern University) and the United States Census Bureau, and was established in 2011 to provide support for a set of research nodes, staffed by teams of researchers conducting interdisciplinary research and educational activities on methodological questions of interest and significance to the broader

Continued on page 3

In This Issue

NCRN Alumni Update - 2
Node News - 4
NCRN Spring 2017 Meeting - 5

Alumni recall connections, collaborations that came with their participation in NCRN program

We recently caught up with a few alumni of the NSF-Census Research Network, which has funded the research of more than 50 graduate and post-doctoral students. As we have reconnected with these alumni for an update on their careers, we have gotten a clearer picture of the program's impact on their lives. This is the third story in a series on NCRN alumni.

Nicole Dalzell recently received her doctorate in Statistical Science from Duke University, where she was at NCRN's node from 2015-2017. Now, she is preparing to work as an Assistant Teaching Professor at Wake Forest University, starting in the fall of 2017. Dalzell said she gained a great deal of experience working



with large, real-world data in the NCRN program. "Through the NCRN, I've also had the opportunity to work on an interdisciplinary project. Specifically, I'm the statistician working on a project for a team in Economics. It's been extremely rewarding to work on an application I'm excited about, and to learn from my colleague, who is an expert on the data," she said. "I have learned how to adapt to realities of imperfect data and have come to better appreciate the enormity of attempting to process and collect such data. I've learned about the process of thinking through early steps of a project; merging the data, verifying fields, and exploring idiosyncrasies of the data." Dalzell said the graduate training experience changed how she would consider the Census as a potential employer. "I have learned about the process of protecting the data and preserving privacy, and I find it very reassuring to know how carefully the data is protected. I am also intrigued by the sheer variety of data collected by the Bureau. My work with NCRN helped me better understand some of what Census does, as well as how, and has made me more interested in the Census as a potential employer," she said, adding that she has been very grateful for the opportunity to work with NCRN as part of her **dissertation work**. "I have learned a great deal, and have enjoyed the opportunity to work on a collaborative, real-world project."

Jared Murray was at the Carnegie-Mellon University, Duke University/National Institute of Statistical Sciences (NISS) nodes from 2012-17. Post-NCRN, he completed his PhD at Duke, with Jerry Reiter. Murray then joined CMU as a visiting assistant professor, working with the CMU NCRN node. He says his current



placement at CMU was a result of his NCRN experience. "Next year, I'll join the University of Texas at Austin as an assistant professor of statistics and machine learning," he said. Murray said the NCRN exposed him to problems and collaborators he would not have encountered otherwise. "My experience with the Duke node was instrumental in getting my current position at CMU," Murray said. His research agenda has been strongly influenced by his experience with the NCRN, he added. "For example, probabilistic record linkage (combining and/or deduplicating databases which lack a unique identifier) and methods for appropriately drawing inferences using the linked data constitute a significant component of my near-term research agenda. It's a problem that wouldn't have even been on my radar but for the NCRN." Murray added that the opportunity to interact with researchers at other nodes and with folks at the Census through the regular meetings, workshops and virtual seminars has been "immensely valuable ... it's been great!"

Continued on page 3

NCRN Alumni Share Experiences (Continued from page 1)



Anne-Sophie Charest, who received her doctorate in Statistics, was at the Carnegie-Mellon University node of NCRN in 2011-12. After her graduate training, she got a tenure-track position in Statistics at Université Laval in Quebec City, where she now is an Associate Professor of Statistics.

Charest said the NCRN program gave her the opportunity to meet and work with other researchers interested in confidentiality. “It was a great occasion to present my own work, and be exposed to varied research from other members of the program. I even collaborated with some people that I met there after I completed my PhD,” she said. “I was already far along my PhD when I joined the program, so the main positive aspect for me was developing connections with other researchers.”



Leah Ruppanner was at the University of Nebraska-Lincoln node, and received her doctorate in Sociology. Prior to her current appointment with University of Melbourne Sociology faculty, she was a Research Assistant Professor at the University of Nebraska-Lincoln where

she worked on the project “**Reducing Error in Computerized Survey Data Collection.**”

Ruppanner said her appointment at the University of Melbourne is somewhat a result of her NCRN experience, and that the program has changed her views of the Census as a potential employer. “It made me think that working for the Census would be really cool!” she said.

By Dory Knight-Ingram

Award (Continued from page 1)

research community and to the federal statistical system, particularly the U.S. Census Bureau. The Coordinating Office for the NCRN, created in 2012 and managed by Cornell University and Duke/NISS, organized conferences and workshops, compiled the NCRN newsletter, and served as the primary marketing mechanism for NCRN public forums and work products. The NCRN tackled “challenging issues that could not have been addressed so productively by either academic or government researchers acting on their own,” said Prof. Sarah Nusser, Vice-President for Research at Iowa State University, and Dr. Nancy Potok, Chief Statistician of the United States, noted that based on her own observation, “the collaboration has been highly successful.”

Just a small sample of key outputs and successes of the partnership include support and engagement of over 90 students pursuing research projects and dissertations directly relevant to the statistical agencies,

development of full length courses/short courses including courses on problems statistical agencies face; social and economic data; spatio-temporal statistical modeling (see current list at <https://www.ncrn.info/education>), publicly available software and datasets including “Naturally Occurring Data” public-use datafile; multiple imputation R package, and more (full list at <https://www.ncrn.info/software>), applied work products related to spatial-temporal modeling, geo-spatial statistics, disclosure avoidance, synthetic data, Big Data applications, advances in editing and imputation, spatial visualization of uncertainty and user-defined geographies to reduce margin of error; and “smart-agent” internet data collection designs. A full biography of publications and working papers can be found at <https://www.ncrn.info/documents/bibliographies>.

More information on the NCRN and its nodes can be found through the website at www.ncrn.info.

Node News

A realistic base file for testing privacy-preserving data analysis and publication algorithms

CORNELL UNIVERSITY – The [Labor Dynamics Institute](#), supported by the [Sloan Foundation](#), has now made a new data asset available.

We have built a synthetic population of the United States entirely from public-use data in the American Community Survey. Unlike some of the other synthetic data projects we have undertaken in the past, these data were constructed to provide a realistic base file for testing privacy-preserving data analysis and publication algorithms. The records in the synthetic population are all actual records from the responses to the American Community Survey (2010-2014) and the data dictionary is the Census Bureau's [data dictionary for the ACS PUMS 2010-2014](#) (also available at <http://doi.org/10.3886/E100486V1>).

For more information:

[Read more about this](#) on the NCRN website.

U.S. Census Bureau Director John Thompson steps down

John Thompson stepped down in June 2017 as director of the U.S. Census Bureau, after three decades of service in the federal government and at NORC. Thompson has been a strong supporter of the increased interaction between the Census Bureau and the academic community, including the NSF-Census Research Network. His support for NCRN has been strong since his appointment in August 2013, when he took over the reins from his predecessor Robert Groves. We wish him well in his future endeavors.



John Thompson

CED²AR 2.8.2.0

CED²AR 2.8.2.0 is available at <https://github.com/ncrncornell/ced2ar/releases/tag/2.8.2.0>, (also 10.5281/zenodo.495191) along with a list of new features and bug fixes.

ASA Differential Privacy webinar viewable online

The ASA Differential Privacy webinar, hosted by the ASA Committee on Privacy and Confidentiality, is now online! <https://www.youtube.com/watch?v=hzHBUYEkgHM> (via @ASA_SciPol) The Webinar, given by Daniel Kifer, Pennsylvania State University, and Vishesh Karwa, Harvard University, and moderated by Aleksandra Slavkovic, Pennsylvania State University, took place on Friday, February 10, 2017, 1:00 p.m. – 2:00 p.m. Eastern time.

Jerry Reiter and Lars Vilhuber speak at Ryan Murray Commission for Evidence Based Policymaking

Jerry Reiter and Lars Vilhuber (NCRN node at Cornell) spoke to the Commission on Evidence Based Policymaking on February 24, 2017. They discussed the role of statistical methodologies and technology in protecting data privacy. The Evidence-Based Policymaking Commission Act of 2016 was jointly introduced by U.S. Senator Patty Murray (D-WA) and Representative Paul Ryan (R-WI), and signed into law on March 30, 2016. The Commission is described at <https://www.cep.gov/about.html>. The announcement of the talks is at <https://www.cep.gov/meetings/2017-02-24.html>.

Continued on page 5

NCRN Spring 2017 Meeting

On April 24, 2017, the Principal Investigators and Senior Researchers from the NCRN nodes met with Census Bureau staff, and presented on a number of research topics in the Census Bureau's auditorium.

About 40 participants, including researchers from a variety of federal statistical agencies, attended the scientific presentations. The NCRN Coordinating Office's Principal Investigator Lars Vilhuber and John Eltinge, Associate Director for Research and Methodology at the U.S. Census Bureau, welcomed the participants. Jared Murray (Carnegie Mellon University), David Folch (Florida State University representing the Colorado-Tennessee node), Matthew Simpson (University of Missouri), and Carol Caldwell (US Census Bureau) presented in the morning. Jerry Reiter (Duke University) in the afternoon. Presentations were a mixture of results from past research within the network, and preliminary results from ongoing projects that had come out of the network collaborations. The NCRN welcomed the principal investigators of the new collaborative agreements the Census Bureau has signed

with Georgetown University and Purdue University. Kobbi Nissim (Georgetown) presented on "Formal Privacy Models and Title 13," and Chris Clifton (Purdue) presented on "Practical Issues in Anonymity." After the scientific presentations and meetings between NCRN members and Census staff, a lively discussion was continued at the traditional dinner at the Lebanese Taverna on Connecticut Avenue. Because some presentations were of preliminary material, not all presentations materials are available online. More details can be found at <https://www.ncrn.info/event/ncrn-meeting-spring-2017>. Currently, no further NCRN Meetings are planned. Some of the research will be presented in final form at conferences such as the Joint Statistical Meetings, and we look forward to discussing these topics and those and other forums.

Taking stock (Continued from page 1)

the NCRN bibliography (<https://www.ncrn.info/documents/bibliographies>) and the list of NCRN-funded software (<https://www.ncrn.info/software>). We will work with the nodes to ensure archiving of their respective websites, aiming to provide persistent traces of the network beyond the publications that are the most tangible outcome.

The NCRN will wind down in its formal structure, but the network of collaborations it fostered and boosted live on. If you have any questions, do not hesitate to contact the members of the Coordinating Office (<https://www.ncrn.info/about>) or the Census Bureau.

(Node News, Continued from page 4)

TCRN postdocs and students accept new positions

DUKE/NISS – Several current and former Triangle Census Research Node (TCRN) trainees are moving on to new careers in fall 2017. Graduate student Nicole Dalzell will be an assistant teaching professor of statistics at Wake Forest University. Postdoc Mauricio Sadinle will be an assistant professor of biostatistics at the University of Washington. Postdoc Maria De Yoreo will be a statistician at RAND Corporation. Jared Murray, former graduate student and current visiting professor at CMU, will be an assistant professor of statistics at the business school at the University of Texas at Austin.

The NCRN Newsletter is published quarterly by the NCRN Coordinating Office.

Newsletter staff:

Lars Vilhuber, Principal Investigator,
NCRN Coordinating Office
Dory Knight-Ingram, Managing Editor
Melissa Colbeth, Assistant Editor

You may reach us at:

275 Ives Hall,
Cornell University, Ithaca, NY 14853
(607) 255-2744
info@ncrn.info
www.ncrn.info