

MPC News

WWW.POP.UMN.EDU

News and events from the
Minnesota Population Center



New Frontiers in Big Data

EXTEND THE SCOPE OF DEMOGRAPHIC RESEARCH

A major agreement between the MPC, FamilyFinder, and Ancestry.com will significantly increase the amount of census data available to social science researchers, enabling research that was previously impossible.

By Gina Rumore

By 2020, MPC will make freely available to researchers worldwide 100% count U.S. Census microdata through 1940. This dataset will include over 650 million individual-level (1850-1940) and 7.5 million household-level records (1790-1840). The microdata represents the fruition of longstanding collaborations between MPC and the nation's two largest genealogical organizations—Ancestry.com and FamilySearch—to leverage genealogical data for scientific purposes.

“The importance of this massive donation

of census data would be difficult to overstate,” says MPC Director Steve Ruggles. “This is one of the largest-scale data-entry efforts ever undertaken.”

Together, FamilySearch and Ancestry.com devoted approximately 22 million hours to transcribing individual family records, the equivalent of over 10,000 person-years of effort, in one of the largest-scale data-entry efforts ever undertaken. The data-entry cost to replicate the collection in the U.S. would be about \$420 million.

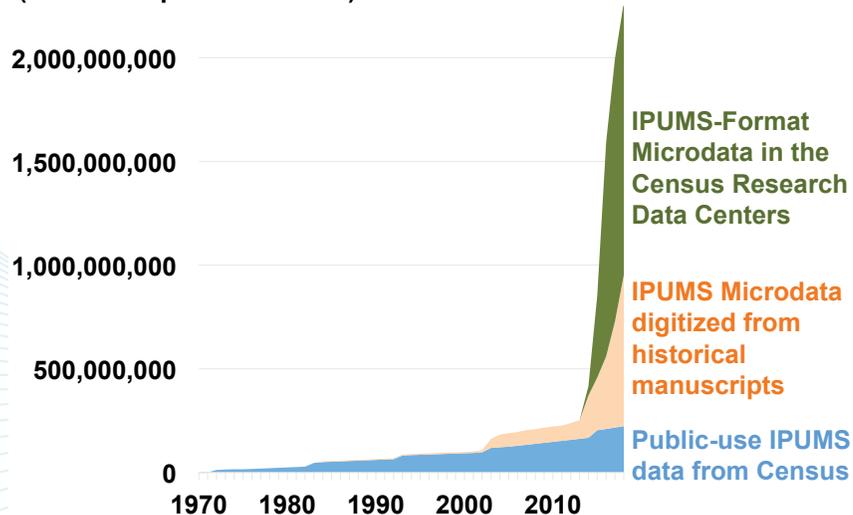
While Ancestry and FamilySearch focused

their efforts on census questions that are of greatest use for genealogical research, such as name, age, sex, marital status, birthplace, and parental birthplace, their transcription omitted crucial demographic and economic characteristics such as occupation, employment status, literacy, school attendance, and home ownership.

“Thanks to a new collaborative agreement with Ancestry.com,” says Ruggles, “we are now adding the missing variables to fill the last remaining gaps in the data series.” The additions are substantial. The MPC has

continued on page 6

Integrated microdata available for research, 1970-2020 (number of person records)



Inside

- 1 | BIG DATA
- 2 | LETTER FROM THE DIRECTOR
- 3 | DIVERSITY AT MPC
- 4 | THE 2015 MPC DIVERSITY FELLOWS
- 7 | MPC SEMINAR SERIES SCHEDULE

ABOUT THE MPC



50 Willey Hall
225 – 19th Avenue South
Minneapolis, MN 55455

Phone: 612-624-5818
Fax: 612-626-8375
E-mail: mpc@umn.edu

DIRECTOR

Steven Ruggles, *Regents Professor of History*

ASSOCIATE DIRECTOR

Catherine Fitch

TRAINING DIRECTOR

Rob Warren, *Sociology*

ADVISORY BOARD 2015-2016

Benjamin Capistrant, *Epidemiology and Community Health*

Jack DeWaard, *Sociology*

Ezra Golberstein, *Health Policy and Management*

J. David Hacker, *History*

Janna Johnson, *Humphrey School of Public Affairs*

Carolyn Liebler, *Sociology*

Lara Cleveland, *MPC Research Associate*

Jonathan Schroeder, *MPC Research Associate*

Sula Sarkar, *MPC Research Associate*

Johanna Fajardo-Gonzales, *Graduate Student*

Sarah Garcia, *Graduate Student*

Marjory Silisyene, *Graduate Student*

MPC News Editorial Staff

Editor/Writer: Pamela Vold

Co-Editor and Contributing Writer: Gina Rumore

Contributing Writer: Melissa Kelley

Graphic Designer: Shawn Welch

WWW.POP.UMN.EDU



Follow @minnpop



Like Us Minnesota Population Center

Learn more about our latest projects, researchers, and events, photos, and opinions.

FROM THE DIRECTOR

Busy Days



Steven Ruggles
Director, Minnesota
Population Center

This fall is the beginning of a very busy period for MPC. Our work here includes a renewal proposal for our Center Grant through the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD). In the process of writing this proposal, I've had the opportunity to review the work and accomplishments of the MPC Faculty, Research Staff, and Graduate Student Members over the past five years. It is an amazing honor to be the Director of such a vibrant and productive research center.

Our success is documented in many ways. Our faculty, research associates, and students produced about 900 population-related publications over the past three fiscal years, including more articles in *Demography* than any other university. MPC has more external sponsored research support than any other research center at the University of Minnesota. Among all universities, Minnesota has the largest portfolio of research grants administered by the Population Dynamics Branch of NICHD. MPC disseminates more than a Terabyte of integrated demographic and health data each week to a global audience of over 70,000 demographic researchers.

Of equal importance, we continue to improve ourselves internally. This year we welcomed our first cohort of summer Diversity Fellows to MPC. We learned a great deal from these six fine young scholars, and I hope you will take the time to read more about our fellows and the program in these pages. We also continue to expand our data projects and to fine-tune and enhance our data delivery systems to meet the needs and expectations of our users worldwide. We take great pride in the work we do at MPC, and it takes a wonderful group of employees and members to make it possible. I thank them all for continuing their work in making the MPC a truly unique and remarkable place.

Steven Ruggles
Director, Minnesota Population Center

MPC by the Numbers 97401...

1

Terabyte of integrated demographic and health data disseminated each week by the MPC

70,000

Demographic researchers worldwide using MPC data

900

Population-related publications over the past three fiscal years by MPC staff and members

Diversified Assets

EXPANDING THE REACH OF THE MPC

How the MPC created a diversity program to expose undergraduate and graduate students to the Center's work and learned a few things of its own.

By Pamela Vold

When Human Resources Associate Mia Riza and Research Associate David Haynes took on the task of creating a diversity program for the MPC in 2015, they looked to their pasts. "I thought of experiences that would have helped me as a student," says Haynes.

As an undergraduate student, Riza had participated in a diversity program with the Minnesota Historical Society run by Chris Taylor. That program provided a

program explains Riza, "We had the idea that the undergraduates would work with the graduate students on the projects so that they could develop another relationship. There is a mentor, a grad student, and an undergraduate working together on the project. The mentor works with them both, but there is also an expectation that the graduate students mentor the undergraduate students. We had the graduates start two weeks earlier to get up to speed on the projects, and then we asked them to teach while they were learning.

That's a big task, but they all seemed up for it. We had conversations with them about how they can best work together and learn from each other. We were pleased to hear during one-on-one conversations that this was happening."

Riza goes on to explain, "Part of the challenge of the program is finding the experience that is appropriate for the levels of the students: preparing undergraduate students to be able to take on graduate researcher roles and preparing graduate students to take on their specific MPC projects."

"We really worked with the project managers to figure out how they could break down their projects into tasks that each student would understand," Haynes elaborates. "The undergraduates might have more basic questions, but then we talk about it and they get exposure to what we do. This helps open up their eyes to other research paths."



David Haynes (left) and Mia Riza (right), co-directors of the MPC Diversity Program

"There are qualified students of color that are ready to contribute."

—David Haynes

model for fostering professional growth within an academic environment. Haynes and Riza then designed a program that suited the needs of the projects at the MPC. "We thought that a summer opportunity would work best for the Center and also matched students' needs as they looked for summer internships," explains Riza, "We want to develop a talent pipeline so that when we have openings later, we have qualified diverse applicants to fill those roles."

The two consulted with MPC project managers to determine which skills the students could gain and build upon over a summer term, and how those skills could be put to use to benefit the Center. Working with the project managers, Riza and Haynes created projects and goals for a group of six students, three graduate students and three undergraduates.

Recruiting students at different academic levels was an important part of the

Two professional development opportunities put on by MPC for the Diversity Fellows also provided guidance. In one event, Noro Andriamanalina from the University of Minnesota Office for Equity led a workshop on crafting CVs and resumes. In another, MPC convened a mentoring panel of faculty and professional staff from underrepresented backgrounds working in STEM fields to discuss their career paths with the fellows. "It was an intimate and informal gathering," says Riza. "The students got a lot out of it and the panelists were excited about it." Haynes adds, "One of our students made a real connection networking with a speaker after the program."

"David and I sat in on the resume workshop" says Riza, "and we came away with valuable insights as well."

Lessons Learned

After a great first experience, Haynes and Riza are starting to think about the pro-

continued on page 8

MPC's 2015 Diversity Fellows

The MPC's first cohort of Diversity Fellows include three graduate students and three undergraduate students.



Besufekad Alemu is a Ph.D. student in the Applied Economics Department at the University of Minnesota. While at the MPC, he worked with the IPUMS Historical Project to link immigrants from Norway and Sweden to the 1920 and 1930 censuses and to observe the outcomes of these groups within the United States. His current research interests focus on the intergenerational mobility of immigrants, labor economics, and development economics.



Baishali Bakshi is a Ph.D. student in the Natural Resources Science and Management graduate program researching the impact of climate-induced vegetation change on ecosystem services such as recreation and wildlife habitat. She joined

the MPC as a Diversity Fellow working on a research project linking climate change and mortality in Africa. She earned a Ph.D. in Economics from University of California-Irvine specializing on the linkages between international trade, corruption, and environmental resources. Baishali's research has appeared in *Topics in Economic Analysis and Policy* and *Energy Studies Review*.



Shoumith Jeyakumar is a graduate from the Master of Geographic Information Science at the University of Minnesota. He completed his bachelor of engineering degree in Geoinformatics from the College of Engineering Guindy - Chennai, India. Shoumith is a self-proclaimed data enthusiast and map fanatic who will forever be intrigued with new technology. At the MPC Shoumith worked on creating a web interface that would help users visualize the GIS and demographic data for the TerraPop project.



Wen Jing is an undergraduate student studying actuarial science and computer science at the University of Minnesota. During her time at the MPC, she has worked with the TerraPop IT group on developing a dashboard to help achieve data visualization. This group aims to make modifying data more user-friendly, adding graph and table capabilities. This is Wen's second year in the United States.

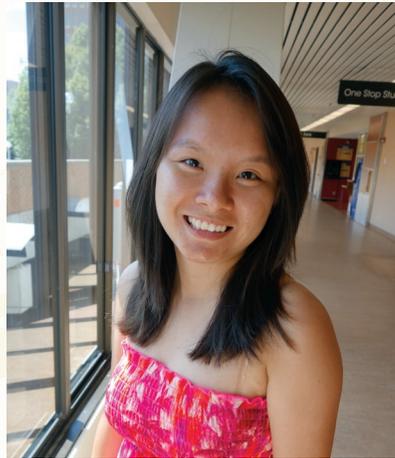


Luisa Silva Lelis is an undergraduate from Brazil studying Forest Engineering at the Federal University of Technology - Parana. As a member of the Brazilian Scientific Mobility Program, she has traveled to both the University of Wisconsin - River Falls and the University

JUST FOR FUN!

of Minnesota for her studies. Her work currently focuses on processing spatial data for a project analyzing climate change and mortality in Africa.

Giathu “Thu” Vien is an undergraduate at the University of Minnesota studying Mathematics with Actuarial Specialization. Along with another diversity fellow, Besu, she collaborates with Evan Roberts on the Historical Projects at MPC. Her work specifically focuses on how immigrants from Scandinavian countries adjusted to American life in the early twentieth century and specifically why certain individuals settled where they did. Thu is originally from Vietnam but currently resides in Prior Lake, Minnesota.



What is your favorite vacation spot?

Besufekad Alemu
Addis Ababa, Ethiopia

Baishali Bakshi
San Diego, California

Shoumith Jeyakumar
New Zealand

Wen Jing
Qingdao, China

Luisa Silva Leelis
My Grandmother's House

Giathu “Thu” Vien
Bahamas

JOIN US FOR AN MPC-SPONSORED PANEL:

An Assessment of the 1965 Immigration Act and the Future of Immigration Policy in the U.S.



Friday, October 23, 2015

10:45 a.m.-12:00 p.m.

Chair: Ryan Allen, University of Minnesota

Discussant: Jack DeWaard, University of Minnesota

Participants:

Doug Hartmann, University of Minnesota

Charles Hirschman, University of Washington

Guillermina Jasso, New York University

Douglas Massey, Princeton University

Part of “Immigrant America: New Immigration Histories from 1965 to 2015”

A conference hosted by the Immigration History Research Center and Archives at the University of Minnesota, marking the 50th Anniversary of the 1965 Immigration Act

Find more information at cla.umn.edu/ihr/research/immigrant-america-conference

Big Data *continued from page 1*

already filled in the missing information for 1850 and 1940. Now, with the support of two new grants from the National Institutes of Health (NIH), MPC will more than double the number of variables available for the period 1900 through 1930, adding an average of 22 new variables to the data for each census year.

The full count historical data will allow scientists from across the disciplines—from health policy researchers and economists to epidemiologists and environmental scientists—to analyze demographic processes and test population models in ways never before possible.

“The most exciting aspect of the full count data for me,” says MPC Director of Data Integration Matt Sobek, “is the potential to study social mobility and inequality in a way never before possible. This is the holy grail of American social history. There aren’t many bigger themes than America as the land of opportunity: of immigrants climbing the ladder of success, of settlers moving west for a better life, of African Americans moving north to escape discrimination and get better jobs. But we’ve so far had only piecemeal evidence to address whether this picture of America is true, and to what degree, and for whom. The historical census data project should finally give us some definitive answers to these fundamental questions and give us a more realistic perception of our past and maybe what we need to work on in the future.”

Data’s missing link

According to MPC Assistant Professor Evan Roberts, the large database will also

greatly enhance neighborhood research—the full count data will allow researchers to see characteristics of neighbors as well as of households—studies of small area impacts of policy and environmental differences. Roberts further points out the exciting possibilities for researchers who study aging in the United States. For example, in collaboration with Wendy Rahn, a professor of Political Science at Minnesota, Roberts is using the restricted 1940 data set [*see page inset*] to find participants from the Iowa Women’s Health Study in the 1940 Census, giving them access to details on the early life conditions of the survey participants, including where they lived (address), education and employment, migration, and family economic status.

The Roberts and Rahn project foreshadows what many feel will be the most exciting use of the full count data: linking. Not only can individuals and families be linked across censuses, many projects are now underway to link individuals in the 1940 Census to other federally-funded health and aging surveys.

Professor of Sociology and MPC Training Director Rob Warren has a new project to link five major aging surveys—HRS, PSID, WLS, NSHAP, and NHATS—to the 1940 census. While these linked data sets will have to be used in the projects’ secure data enclaves, they will be freely available to all researchers who meet the current criteria to use the survey data. An NIH reviewer for Warren’s proposal noted that “[t]his project will substantially improve basic infrastructure for demographic, social scientific, and public health research on aging, retirement, health, mortality,

economic circumstances, and well-being in later adulthood.” The full count historical census data will make countless similar research projects possible as population scientists link the census data backward and forward across time and space.

Data challenges

While the opportunities created by the massive size of this dataset makes it exciting to population scientists, it also presents some new and unique challenges to the researchers who produce IPUMS data. Processes that have worked in converting the 5% samples into IPUMS data just don’t scale up efficiently: for example, Roberts explains, “One of the hallmarks of the historical samples is that we have manually looked up a lot of the errors or potential errors we have found. With data of this scale, we’re not able to do the manual error correction we have in the past.” Sobek elaborates, “There are inevitably some contradictions in the data, either between a person’s responses—an eight-year old physician—or between the records of people within a household—a child older than a parent. Again, the data are too large to manually examine all the potential errors, and we must devise editing routines that make logical and probabilistic recodes based on close examination of specific problematic scenarios.”

Despite the many challenges, both those anticipated and those that continue to arise during the data processing process, MPC is on pace to complete the historical data project by 2020. The full count 1940 dataset was released in late 2014 via IPUMS-USA (popdata.org). And, according to MPC Associate Director Catherine Fitch, “Preliminary data releases will continue on a rolling basis, working backward from 1930, which we anticipate releasing in early 2016.” The full count household-level data from 1790-1840 will be available by the end of the year.

“We believe,” says Ruggles, “that this data will revolutionize the way research is done across the social, behavioral, economic, and health sciences.”

The MPC’s Restricted Data

The MPC also has restricted versions of all years of the complete count data, which include names. The version with names cannot be released publicly due to MPC’s agreement with Ancestry.com. For more information on obtaining the restricted data, please email ipumsres@umn.edu.

Fall 2015 Schedule

Unless otherwise noted, all MPC seminars are held from 12:15 to 1:15 PM on Mondays in the MPC Seminar Room (50 Willey Hall).

Monday, September 14

Fertility Intentions Across the Reproductive Years: What We Have Learned from the NLSY79

PHILIP MORGAN, *Carolina Population Center, University of North Carolina at Chapel Hill*

Monday, September 21

Welfare State Myths and Measurement

IRWIN GARFINKEL, *School of Social Work, Columbia University*

Monday, September 28

Environmental Impacts of Feeding 9-11 Billion People

DAVE TILMAN, *Ecology, University of Minnesota*

Monday, October 5

Integrating, Preserving, and Disseminating Linked CPS Data

SARAH FLOOD, *Minnesota Population Center, University of Minnesota*

Monday, October 12

Was There a Marriage Premium in Late Nineteenth-Century Manufacturing? Evidence from Sweden

MARIA STANFORS, *Center for Economic Demography, Lund University*

Monday, October 19

Government and the Crises of the Great Depression and Second World War

PRICE FISHBACK, *Economics, University of Arizona*

Monday, October 26

Family Caregiving for Older Adults and the Caregiver's Health

BEN CAPISTRANT, *School of Public Health, University of Minnesota*

Monday, November 2

Love, Money and HIV: Revisiting Education, Wealth, and Fundamental Cause Theory

SANYU MOJOLA, *Sociology, University of Colorado, Boulder*

Monday, November 9

Excluded Generation: The Growing Challenge of Labor Market Insertion for Egyptian Youth

RAGUI ASSAAD, *Humphrey School of Public Affairs, University of Minnesota*

Monday, November 16

Machineries of Data Power: Manual versus Mechanical Census Compilation in Nineteenth-Century Europe

CHRISTINE VON OERTZEN, *Max Planck Institute for the History of Science*

Monday, November 23

Disasters and Residential Change in the U.S.

ELIZABETH FUSSELL, *Population Studies, Brown University*

Monday, November 30

Changes in Mortality After Massachusetts Health Care Reform

SHARON LONG, *Health Policy Center, Urban Institute*

Monday, December 7

Couples Shared Time across Decades, Countries, and the Life Course

KATIE GENADEK, SARAH FLOOD AND JOAN GARCIA, *Minnesota Population Center, University of Minnesota*



Minnesota Population Center

50 Willey Hall
225 – 19th Avenue South
Minneapolis, MN 55455

WWW.POP.UMN.EDU

Diversity at MPC *continued from page 3*

gram moving forward next year, reflecting on lessons learned, and considering how the program could be improved for the next cohort of students. One of those items is the recruiting of the students themselves. Haynes and Riza had success working with Noro Andriamanalina to recruit the first cohort of students. Her connection with the Community of Scholars program provided them with a valuable tool in reaching graduate students of color. “She helped get the word out,” explains Haynes. “Noro also knew students that she could tap on the shoulder and say, hey, this might be a good opportunity for you.”

The room for growth for the MPC team is in recruiting undergraduates. Haynes explains, “There are more and better ways to get the word out about what we have to offer. Grads are really looking for opportunities like this and undergrads don’t really understand how it would be useful for them.”

Unexpected Benefits

While the primary goal for creating the program may have been to increase MPC’s future job candidate pool, the program provided some unexpected benefits for the Center. For example, never before

had there been any coordination with the Department of Actuarial Science. Through the program, the MPC received an introduction to that department and found a new audience for MPC data.

“There are people right here on campus that want, and can use, our data,” Haynes explains. “There are qualified students of color in these other fields that are ready to contribute. Actuarial science—we had around 20 students apply from that department. Now we have new departments we can talk to about ways our data can be used for research. We are building stronger connections with other groups like the School of Public Health. We are diversifying not just in terms of color or race, but in terms of the research that students are able to pursue.”

High Praise

Haynes and Riza have been diligent in soliciting feedback from everyone involved in the new program, and the two have been pleased by the responses. Several MPC staff members noted to Haynes and Riza that they were anticipating a highly

competent group of students to take on challenging projects, and they were not disappointed. Riza says, “They are pleased with the quality of the applicants. The fellows’ contributions to the projects have really exceeded their mentors’ expectations.”

Of course, the best feedback about the program has come from the students themselves. Riza explains, “They are very excited about MPC and want to stay forever. I’ve heard some of them say, ‘I want to come back here when I graduate. This has changed what I want to do.’”



Diversity Fellows Thu Vien (left) and Besu Alemu (right) presenting their work from the summer program to MPC staff.