



9-28-2017

Urban Institute Builds Platform for Researchers to Access Linked Administrative Housing Data

Monica King
University of Pennsylvania

Follow this and additional works at: https://repository.upenn.edu/admindata_interviews

Recommended Citation

King, Monica, "Urban Institute Builds Platform for Researchers to Access Linked Administrative Housing Data" (2017). *ADRF Network Newsletter Interviews*. 2.
https://repository.upenn.edu/admindata_interviews/2

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/admindata_interviews/2
For more information, please contact repository@pobox.upenn.edu.

Urban Institute Builds Platform for Researchers to Access Linked Administrative Housing Data

Creative Commons License



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Urban Institute Builds Platform for Researchers to Access Linked Administrative Housing Data

September 28, 2017

By Monica King

Last fall, Urban Institute received funding from the [Alfred P. Sloan Foundation](#) to link administrative data on housing with Census Bureau data and make them [publicly available](#) to the broader research community. [Alanna McCargo](#), co-director of the Housing Finance Policy Center and principal investigator on the grant said, “our goal for this project was to expand the access, functionality, and availability of research data.”

Eliminating Repetition

At the core of the project is the Home Mortgage Disclosure Act (HMDA) data. For more than 40 years, the government has required financial institutions to publicly release all mortgage data, including details about each loan, property, and borrower. In 2015, almost 7,000 institutions released 14.3 million records, making HMDA an invaluable administrative dataset on housing and home ownership.

Despite containing enormous troves of data, HMDA is even more powerful when used in conjunction with the Census Bureau’s American Community Survey (ACS), which includes key social, demographic, and economic variables on American people and communities. However, researchers wanting to leverage HMDA data would often spend a lot of time on processing the data and linking it to other data sources. If their efforts are not documented properly, other researchers would have to repeat the same process again in the future.

“We saw this as an amazing opportunity for us to figure out ways to simplify not only how the data is organized but also the time it takes to manipulate it,” McCargo said.

Standardizing Geographic Variables

A major challenge that the team faced in linking HMDA and ACS data is harmonizing the different geographic variables from the two datasets. “There are a lot of administrative data that have geographical boundary definitions that are inconsistent, especially as you look at historical data. One of the key things that we really focused on was how we can link data across geographical areas and across time,” McCargo said.

Ultimately, the team created linked HMDA and ACS datasets aggregated at five geographic levels ranging from the census tract level all the way up to the state level. Along with the data, they also published a comprehensive data dictionary. The final product from this project will enable researchers to skip over the arduous process of creating crosswalks to relate the data across geographies and jump directly into analyzing the refined data.

Expanding Research Uses

Researchers at Urban Institute and around the country are already using the linked dataset to answer important research questions on housing and communities. For example, McCargo and her team wanted to understand why black homeownership rates have declined significantly in recent decades. Using the dataset, the team was able to quickly hone in on trends by age, location, and other key variables. “Now that these two data sets have been merged and made easily accessible, we have been able to quickly pull data on a variety of issues in response to questions from reporters, consumer advocates, the public and other researchers.” McCargo said. “Notably, the Sloan ADRF database allows us to quickly provide detailed data on homeownership by race and bring precision to recent conversations about black homeownership.”

More recently, Urban Institute’s Housing Finance Policy Center cited the project in a report that visualizes the [impact of Hurricane Harvey](#) on Houston’s neighborhoods.

Building Cloud Architecture

Urban Institute also used the Sloan funding to build a prototype for researchers to easily access the datasets. The result is [Spark for Social Science](#): an open-source, cloud-based platform supporting analysis in R and Python programming languages. The tool is undergoing final stages of user testing and feedback from other researchers.

Although this project centered on linking administrative data in housing research, McCargo believes that the approach and lessons learned from this project can be applied to making other sources of administrative data, across disciplines, more accessible as well.

Urban Institute’s Administrative Data Research Facility Project can be accessed at adrf.urban.org