

AUGUST 15, 2024

Center for the REstoration of Economic Data: The Past Is Indeed Prologue

From Rasters to Rows: New Methods and
Applications in Automated Data Extraction
Conference
Philadelphia

Patrick T. Harker

President and Chief Executive Officer
Federal Reserve Bank of Philadelphia

The views expressed
today are my own and not
necessarily those of the
Federal Reserve System or
the Federal Open Market
Committee (FOMC).

Center for the REstoration of Economic Data: The Past Is Indeed Prologue

From Rasters to Rows: New Methods and Applications in Automated Data Extraction
Conference

Philadelphia

August 15, 2024

Patrick T. Harker
President and CEO
Federal Reserve Bank of Philadelphia

Good afternoon, everyone, and welcome once again to this first-ever conference sponsored by CREED, the Center for the REstoration of Economic Data here at the Philly Fed.

I must begin by giving my thanks to the Philadelphia Fed's very own Dan Moulton for the introduction, and to the conference organizers for their efforts in bringing us together and leadership in bringing the concept of CREED to fruition. CREED is the result of a collaboration between the Consumer Finance Institute and the Research Department, and as such is an excellent example of our Bank coming together across departments and disciplines to bring something new and exciting to life.

Teams like CREED don't come together without the support of departmental leadership. I thank Bob Hunt and Julia Cheney of the Consumer Finance Institute and Research Director Roc Armenter for their ongoing support of this tremendous and time-intensive effort.

And at the outset I must recognize the Philadelphia Department of Records and City Archives as well as the Athenaeum of Philadelphia whose partnership in this long-term endeavor has not only been significant but also vital.

Now, if you've attended a conference here at the Philadelphia Fed, you know there's one important piece of business we need to take care of before we get to the heart of any matter — the Fed disclaimer! The views I express today are my own and do not necessarily reflect those of my colleagues in the Federal Reserve System or on the Federal Open Market Committee (FOMC).

So, with that out of the way, let's get into it.

Ours is a world in which we are figuratively swimming in data and, if we still lived in a paper society, we could quite possibly literally swim in it, too!

Moreover, we are continually creating *new* data. In this, we're not merely adding new numbers or interpretation to long-established data sets. Like a baseball general manager using so-called "sabermetrics," in some cases, we're giving researchers the data to create entirely new means of statistical analyses.

But none of that devalues existing data.

In fact, in this data-driven present, I would argue that historical data — whether from decades or centuries ago — is more valuable to us today than ever. As we saw recently with the [introduction of Negro League statistics](#)¹ to the official Major League Baseball record books, historical data that might have been written off as lost or buried in the pages of time can be brought back to life to better inform how we regard our past. Much of this data was culled through the painstaking, archival work of poring through old newspaper articles, box scores, and scoresheets.

And because of this work, when someone searches for career batting average leaders, for example, they will find Negro League great Josh Gibson atop that list and no longer Ty Cobb.

This effort, which has produced a fuller and clearer picture of the history of our national pastime, set the bar for what the effort being undertaken by our CREED team can be in other fields. It is all about finding the data that had largely been left behind.

Yes, it is easy to say that we are in "unprecedented times." And, true, this economy has certainly invented some pitches some of us had never before seen cross the plate. But, almost invariably, someone has faced a similar pitch before. And when we allow ourselves to go back to when those pitches were invented, we can learn how best to take our swings today.

In this sense, our past can be seen very much as prologue to both our current conditions and our future prospects.

And there's no sense in gatekeeping this data here at the Philly Fed. What we aim to do through CREED is to make this history available to everyone. To achieve this goal, we're doing the hands-on work of locating and scanning archival records and using state-of-the-art tools and techniques to digitize and structure it so that others — from economists to social scientists, from journalists to students, from historians to the history-curious — can have it at their fingertips.

We're taking unstructured, analog data and converting it into useful, accessible, and high-quality data.

¹Castrovince, Anthony, "What to Know About Negro Leagues Stats Entering MLB Record," MLB.com, June 6, 2024, <https://www.mlb.com/news/faq-negro-leagues-stats-major-league-record>.

Now, I don't need to preach to you of all audiences about the importance of data quality, particularly regarding a publicly offered dataset. While many data digitization tools are on the leading edge of artificial intelligence and machine learning, the advanced quality of these tools *do not* guarantee data quality. The second goal that CREED hopes to achieve is to improve on these tools and methods, and again, make this knowledge available to the public.

As Dan and others will lay out in their presentations, many of today's tools are still works in progress in some respects. But as we continue to apply them to more varied projects, we can help refine the tools themselves and also help researchers understand how best to deploy them.

Simply put, the promise of CREED is using the technology of today to unlock the past — an oftentimes forgotten past. And, as I'll describe in a moment, it is critical that we unlock the past to help us answer the question, "How did we get here?"

In answering this question, there is one subject area, in particular, where the CREED initiative is already proving itself as an invaluable research tool: housing.

In his own groundbreaking research, Larry Santucci and his research team combed through and digitized [thousands of Philadelphia housing deeds](#), creating a map tracing the growth and spread of racial covenants that, a century ago, created two separate Philadelphias within the city's singular footprint. The racial covenants map on its own is a tremendous historical tool — but it is also an invaluable tool for policymakers today.

Also here at the Philly Fed, Jeffrey Lin, Jeff Brinkman, and Kyle Mangum are currently working on research that is finding that *proposed* highway segments have a permanent negative impact on central city neighborhoods. Let me say that again — *proposed*. Some were never built yet have had long-term detrimental effects. I do look forward to their final paper.

Allow me to pause here to make two short declarative statements. First, small decisions can have big impacts. Second, these impacts can persist for a long time, even if the initial choices are made irrelevant. To quote William Faulkner, "The past is never dead. It's not even past."

The research I have just noted suggests evidence for these two statements.

Philadelphia today remains the poorest large city in America as well as among the nation's most racially segregated. These rankings didn't happen by mere chance. Long ago, policymakers, developers, realtors, and homeowners made conscious decisions about where and whether certain individuals could live in certain parts of this city. Looking at the racial covenants map, which shows us Philadelphia's myriad housing and social

challenges over time, you can't help but wonder how much these decisions shaped the city we live in today.

But more broadly, the work being undertaken by CREED can help provide historical context that current data sets simply cannot. Again, we can look at the housing market to see the promise of bringing old data to light.

There is no doubt that home prices and affordability — especially within urban areas — is a distinct challenge. I hear it directly from my contacts throughout the Third District. Young families are increasingly priced out of the neighborhoods in which they wish to live and raise their families. More growing families, perhaps already in their starter home, are incapable of affording their next home, even with the increase in their current home's market value. Investors purchasing properties can increase rents faster than current residents can absorb.

We can easily look back — or even think back — across the past 20 years to how prices have risen and fallen and risen again. But that's just a snapshot of a housing market ebbing and flowing.

Allison Shertzer, now at the Philly Fed, and her colleagues utilized digitized newspaper real estate sections from 30 major markets from 2006, and going in some cases all the way back to 1890, to [trace housing prices through time and across multiple generations](#). The picture that this newly constructed, historical data paints is one in which, yes, current home prices and rents are at all-time highs, but these data are filling in previously blank spots in economic history.

I know Allison will be presenting at this conference, so I don't want to steal her thunder. But suffice it to say her work in bringing these data back from the great beyond is going to arm researchers — not just her and her colleagues, but those across the country — with new and invaluable inputs to better understand housing markets. And it is going to help put current challenges into a clearer perspective.

When Major League Baseball announced the inclusion of Negro League stats into the official record book — rewriting the history of our national pastime to be more inclusive and more complete — MLB's historian, John Thorne, said, "[I believe that the past is a living, breathing thing that informs every present moment](#)."²

And it is in this sentiment where I believe the CREED effort will likewise shine.

² Dorsey, Russell, "Negro Leagues Statistics to Be Officially Integrated Into MLB Historical Record," Yahoo, May 28, 2024, <https://sports.yahoo.com/negro-leagues-statistics-to-be-officially-integrated-into-mlb-historical-record-212207260.html>.

Now, I'm not going to say that CREED is the solution to the problems facing our modern economy. But I will say that the data that CREED is producing for the public sphere is certainly data that policymakers and others can use to better inform our present moment.

And, it is my hope that, in some cases, it will be the basis upon which broad-based solutions to those problems may be crafted.

With that all said, I thank you for spending your day here at the Philadelphia Fed for this inaugural conference. And I look forward to hearing and reading about all the ways that the data being unlocked today will benefit our common tomorrow.