

# **COVID-19 AND DISPLACEMENT OF VULNERABLE COMMUNITIES**

UTILIZING RADICAL DATASET ANALYSIS TO IMPROVE  
HOUSING POLICIES AND PREVENT EVICTIONS

SCITAC

# 1. PROJECT OBJECTIVES

---

SCI-Arc proposes to investigate an unprecedented use of machine learning to integrate the unconventional usage of wide-ranging data streams (real estate tools, job listings, social media and others) with the datasets typically used by nonprofit organizations, scholars, and advocates to inform housing policy (The Household Pulse Surveys, LA County Assessor Parcels Data, The Bureau of Labor Statistics). This radical data set analysis addresses the need for more granular geo-spatial data in real time to fully understand contemporary conditions at the neighborhood level.

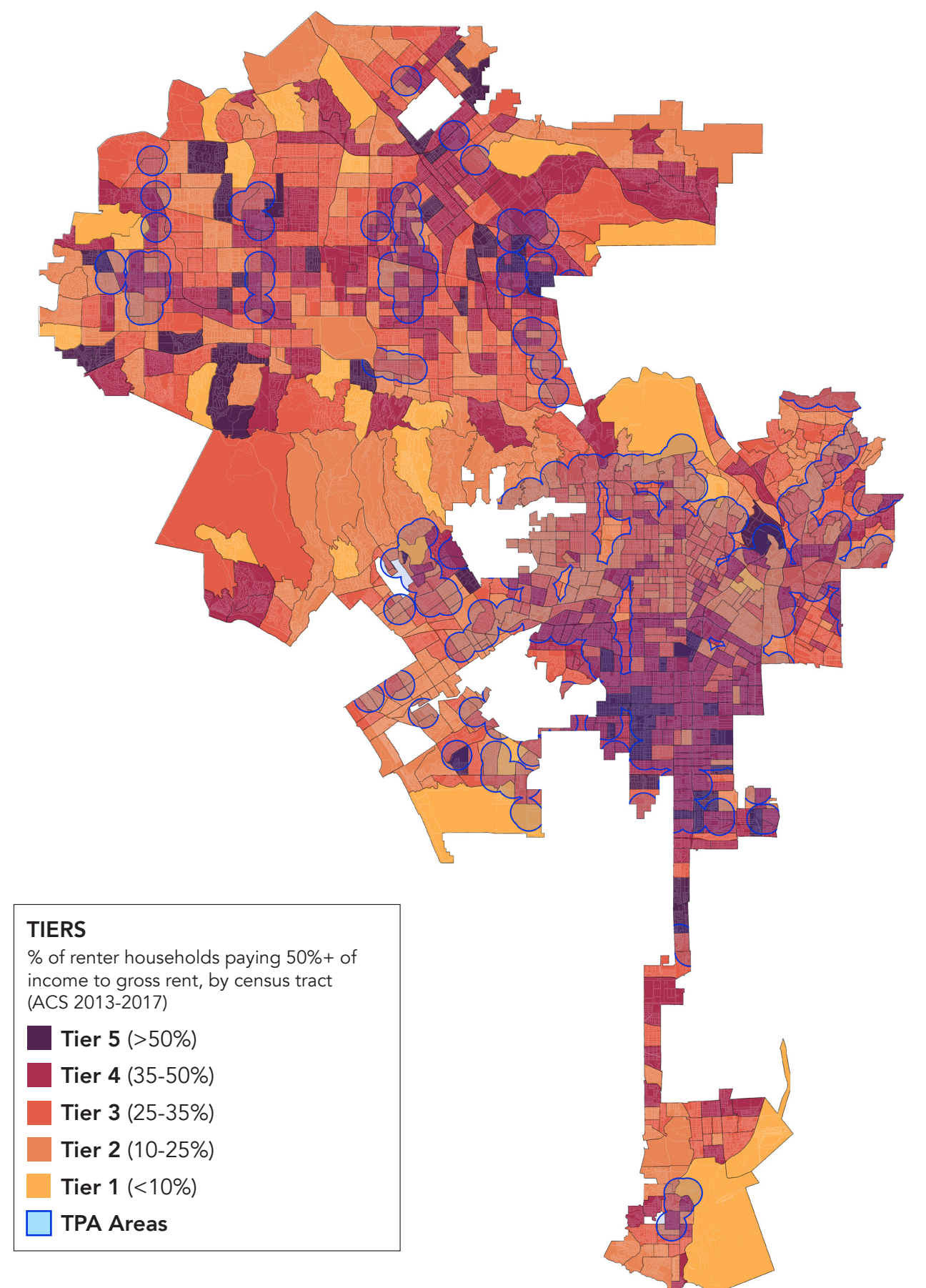
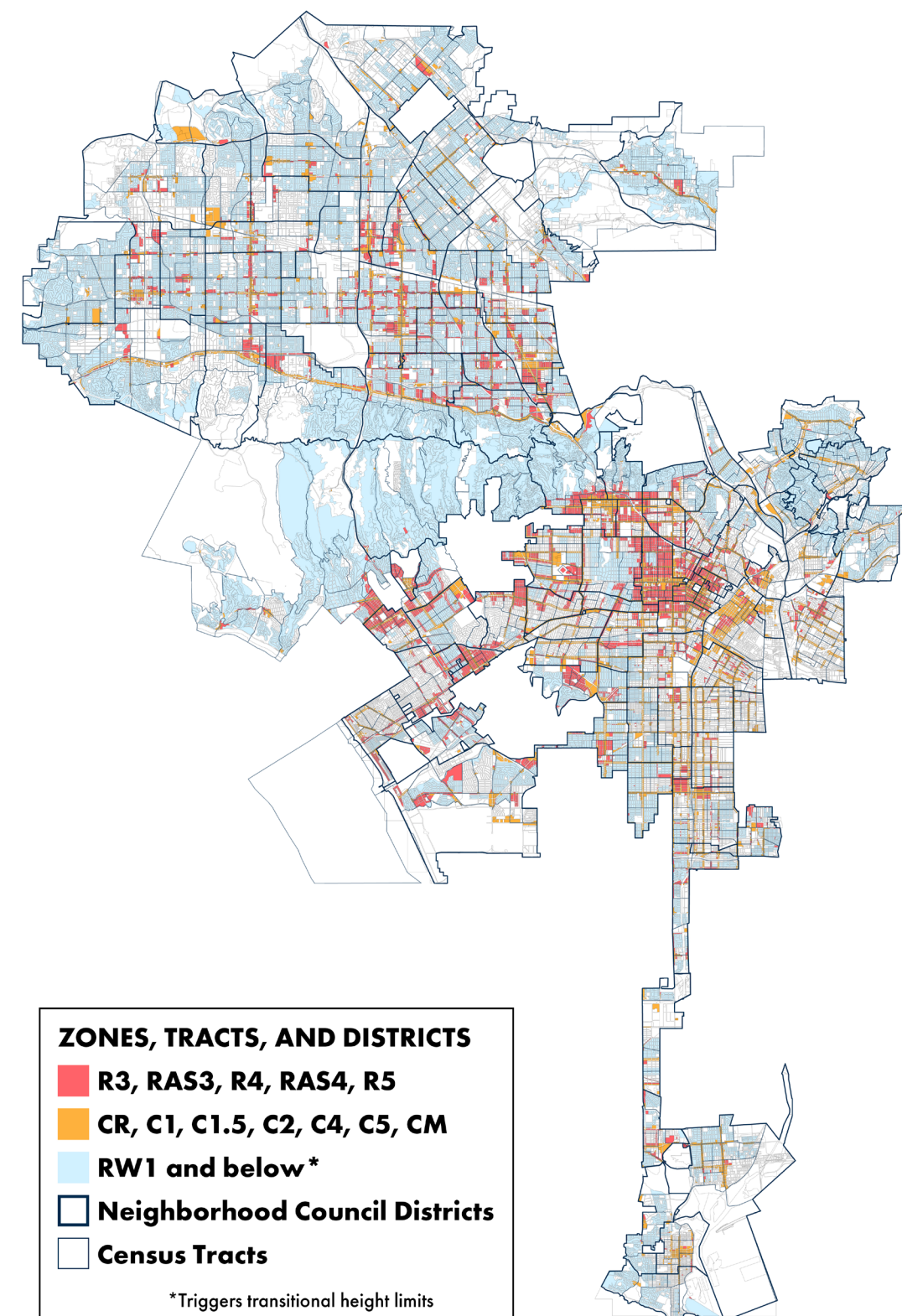
This project has dual objectives: 1) to develop a variety of customized toolkits based on the specific circumstances of distinct constituencies among Los Angeles's most vulnerable communities, allowing experts and policymakers to more clearly see the effects of Covid-19 and subsequent policies; and 2) to demonstrate the efficacy of this novel methodology to vastly improve results for advocates and scholars seeking to shape affordable housing strategies and policies.

## 2. PRELIMINARY RESEARCH

In Fall 2019, in partnership with the Los Angeles Chamber of Commerce, SCI-Arc performed an initial analysis of policy elements to calibrate an inclusionary zoning mechanism as part of the Neighborhood Protection and Affordability (NPA) Ballot Initiative, which aims to produce, preserve and protect a diverse housing stock, while creating mixed-income communities. This Initiative targets the growth in the most cost-burdened census tracts and requires affordable units to be built in the same location. Allocating incentives at the level of the census tract allows for fine-grained responsiveness to the variety of dynamics within the city.

In developing the methodology for measuring the level of the cost burden, we recognized the limitations of U.S. Census datasets that do not capture real estate market conditions. To perform a geographical analysis of housing precarity, we partnered with one of the largest GIS software suppliers, ESRI. Even these tools (CityEngine and ArcGis Urban) proved insufficient for relating zoning requirements, median rents, median incomes, existing regulations (Density Bonus, TOD Incentives) and other datasets. Findings demonstrated the need for a new approach: radical dataset analysis.

The methodology of the proposed project, describing geopolitical, demographic, economic and spatial conditions, capitalizes on SCI-Arc's unique capacity to bring together emerging technologies, speculative design strategies and sophisticated visualization.



# 3. RESEARCH FIELD: COVID-19 AND DISPLACEMENT OF VULNERABLE COMMUNITIES

---

## SIGNIFICANCE

Academic research institutions in Los Angeles (USC Lusk Center) and California (Turner Center for Housing Innovation) have analyzed renter distress caused by COVID-19. Their findings, which align with results of similar research performed across the country, underscore the significance of the proposed project's tool kit development.

1. Low-income households are 5.5 times more likely to experience issues paying rent.
2. Black and Latinx households are more likely to have non-payment problems than whites and Asian Americans in Los Angeles.
3. Most of the households that have lost income since March 2020 are paying rent on time and in full relying on savings, debt, and family support.
4. These solutions significantly increase their financial insecurity and will have long-lasting impacts.
5. 7% of households did not pay rent at least one month.
6. 2% are three months behind on rent, which amounts to 40,000 households that are in danger of eviction starting from February 1st, 2021.

## RESEARCH GAPS

Most of the existing research focuses on estimating the number of renter households facing income losses due to the economic impacts of COVID-19, which is only one angle for approaching the issue of looming evictions and displacement. Moreover, we may even lose Census Data as a reliable source, since this year's count has been shut down. In other words, the need for alternative data sets is only increasing every day.

1. A reliance on one type of traditional datasets which disregards the complexities of the problems at hand.
2. A lack of granularity in the geographic analysis specifically at the neighborhood level.
3. No predictive analysis based on geo-spatial data.

## PROPOSED RESEARCH

This project uses radical dataset analysis to develop a trio of customized toolkits based on the specific circumstances of distinct constituencies within LA's most vulnerable communities. (For example, one of the three toolkits may focus on elderly occupants of rent stabilized buildings; Another will target tenants of small apartment buildings in predominantly Black and Latinx neighborhoods, specifically, households that are several months behind on their rent.) The project's goal is to help policymakers and housing practitioners keep renters housed through and after the pandemic, preventing evictions and displacement throughout LA County. Our approach addresses the existing research gaps by:

1. Identifying key factors associated with evictions and finding new unconventional data streams that can help recognize households in danger of eviction.
2. Conducting analysis based on correlations of new data streams (real estate market tools, geotagged social media posts, job listings) and traditional data sets, in conjunction with and informed by expert interviews.
3. Recognizing patterns that are specific to a building type (single-family home, duplex, fourplex, building with 5-20 units), type and length of tenancy, neighborhood and its demographics.

# 4. DATA COLLECTION

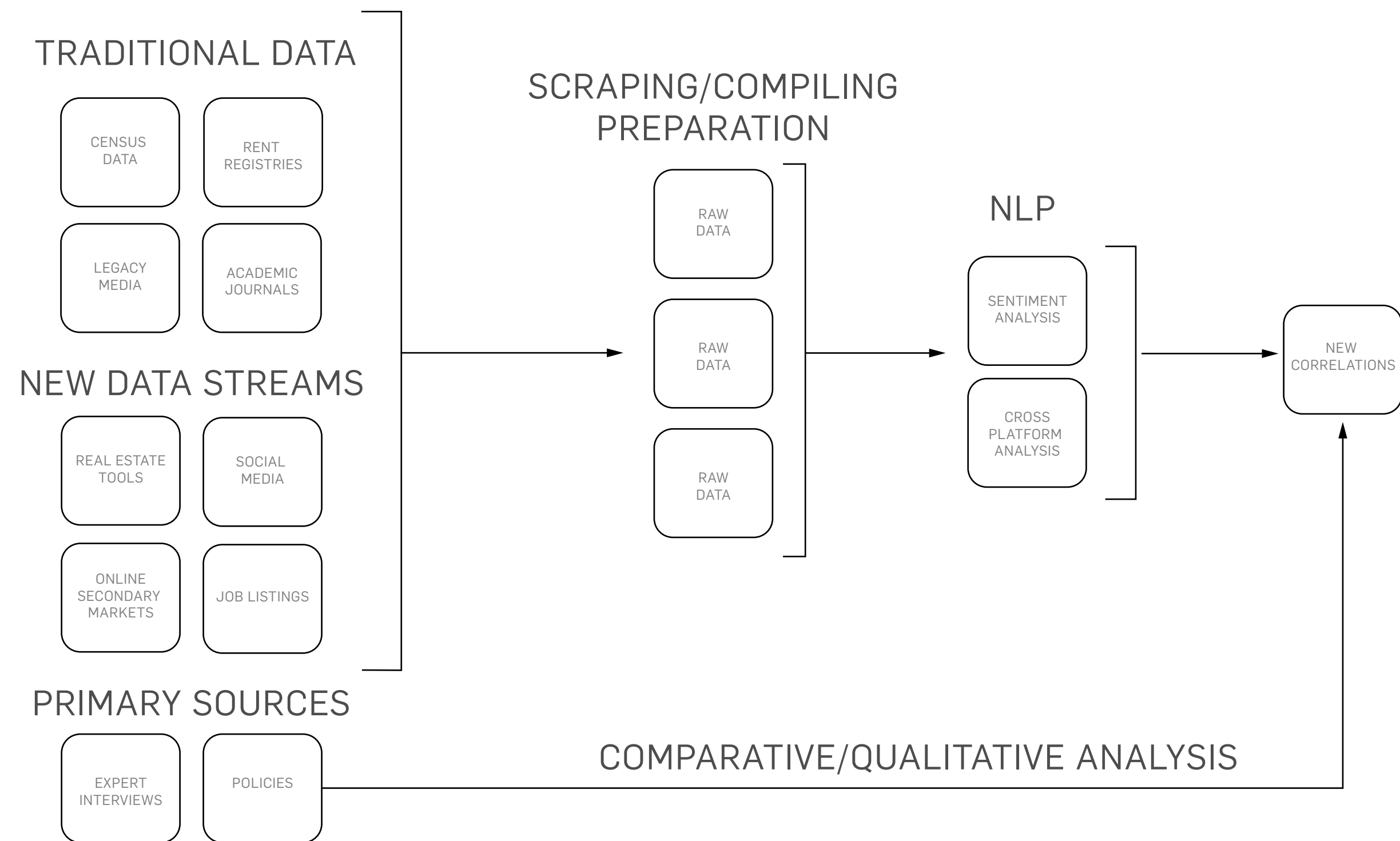
It is well established that a solid application of a Machine Learning (ML) component requires good data sets. This project aims to correlate three types of data to build and analyze these robust sets.

The goal is to combine traditional sources of data, such as census and other government sources, with more market-driven data and aggregated data from multiple platforms. This includes looking at more informal sources such as shared room listings, aggregated social media comments, and real estate tools used by market professionals.

The intention is to automate this collection and cleaning process of raw data through the use of API (Application Programming Interface) scrapers.

Cleaned raw data is then fed into a series of Natural Language Processing ML components to look for new correlations.

All of this will be continually cross-checked against primary sources such as expert interviews, community partners, and policy language.



# 5. PROJECT WORKFLOW

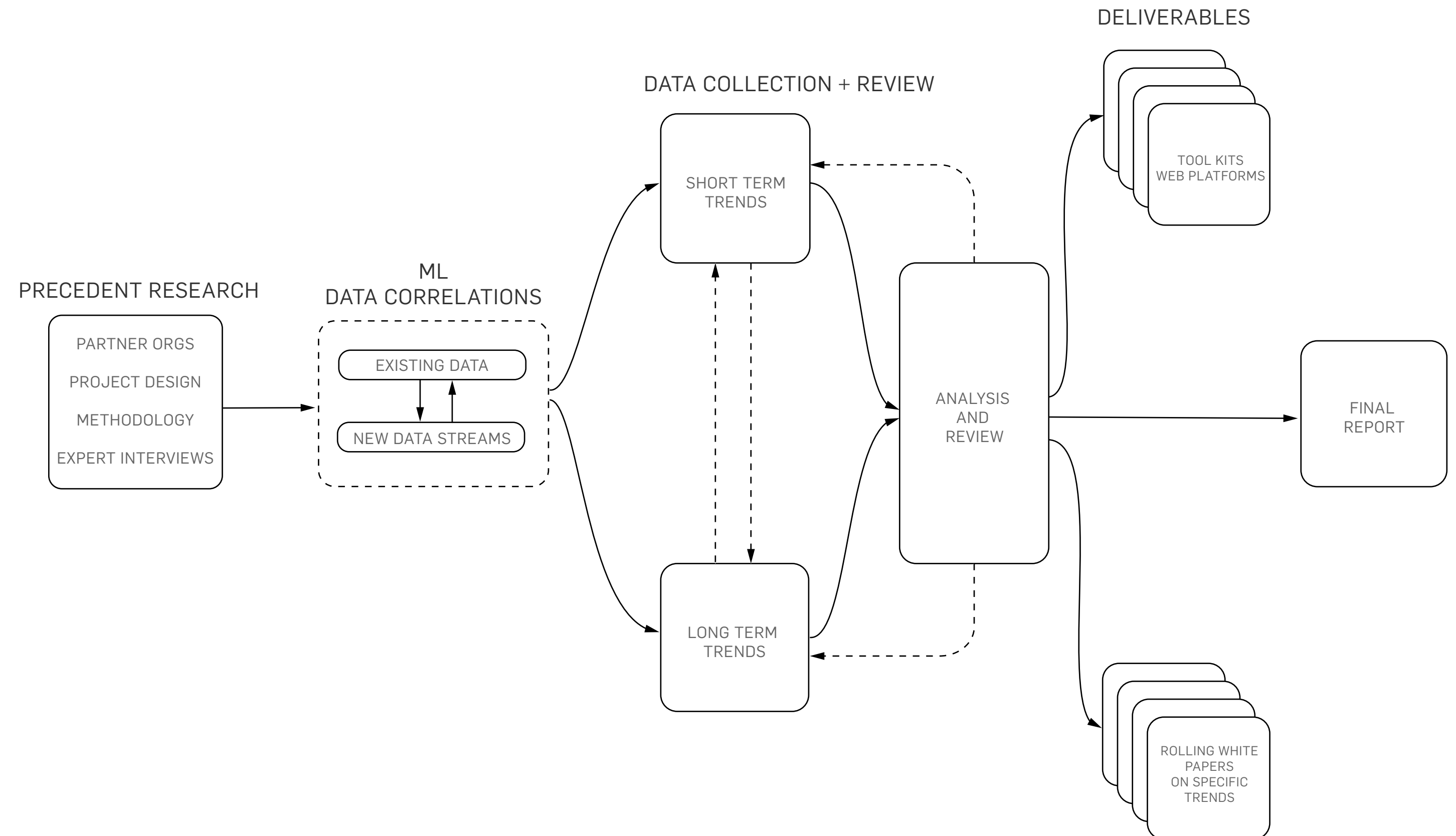
The project workflow consists of 3 core phases:

1. Precedent Research - This phase consists of building the base knowledge set by looking at traditional data sets, identifying key factors associated with potential evictions, creating institutional partnerships, and further refining the methodology for the data collection through beta testing. **PHASE 1**

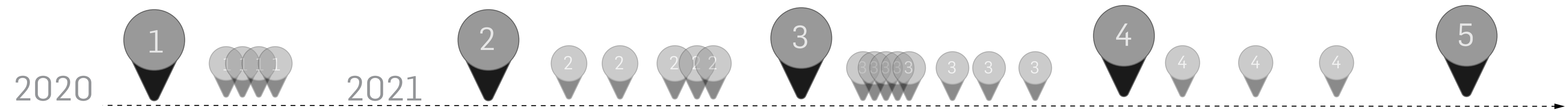
2. Data Collection + Review - Implementing a novel machine learning process combined with primary source research should result in new data streams. The project aims to segregate data into two categories, long and short term trends. Our current objective is to track multiple machinations of policy and events through various platforms. This requires a level of flexibility in one pipeline [short term analysis] that we may not want to implement on the other [long term analysis]. **PHASE 2-3**

3. Deliverables - The project intends to output multiple deliverables, with a focus on actionable real-time data tools to address the distinct needs of a specific constituency. **PHASE 4-5**

## PROCESS OVERVIEW DIAGRAM



# 6. PROJECT TIMELINE



## PHASE 1

Preliminary data collection:  
October – December 2020

Identifying policy, technical and institutional partners, making a survey of existing research and methodologies. Recognizing key factors associated with potential evictions.

## PHASE 2

Dataset Creation:  
January – June 2021

Creating a collection of traditional data sets, new data streams and primary sources. We will use American Community Service (ACS) datasets, Census Bureau's Household Pulse Surveys, Rent Registry of Los Angeles City. Utilizing Machine Learning algorithms will allow us to relate all types of data to come up with new correlations.

## PHASE 3

Analysis of datasets:  
January – October 2021

Analysis will accompany dataset creation. Starting from July 2021, the research team will work on identifying key tenants facing potential evictions, landlords in financial distress and building types where this crisis takes place. The results will be visualized on the neighborhood scale and published as a white paper.

## PHASE 4

Toolkit creation and launch:  
November - December 2021

Based on data analysis we will create toolkits that target specific groups in financial distress, like elderly occupants in rent-stabilized units, landlords of small apartment buildings in predominantly Black and Latin neighborhoods, and households that are several months behind on their rent. Online toolkits will be accessible and actionable for both policymakers and housing justice advocates (researchers, tenant organizers, non-profits).

## PHASE 5

Research report:  
December 2021

The research report will include data from the initial collection, information from the white paper, and analysis of the toolkit usage by some of our policy and community partners.

# 7. DELIVERABLES

---

## TOOLKITS

Toolkits will support housing practitioners and community organizations in producing targeted policy responses. Their focus will be determined after the analysis of data sets, interviews with various experts and continuous review of research performed by partner institutions. Each tool kit will be customized for its intended constituency.

## REPORT

The research report will include new data streams and some traditional data sets, white paper information, developed methodology, and policy recommendations.

## WHITE PAPER

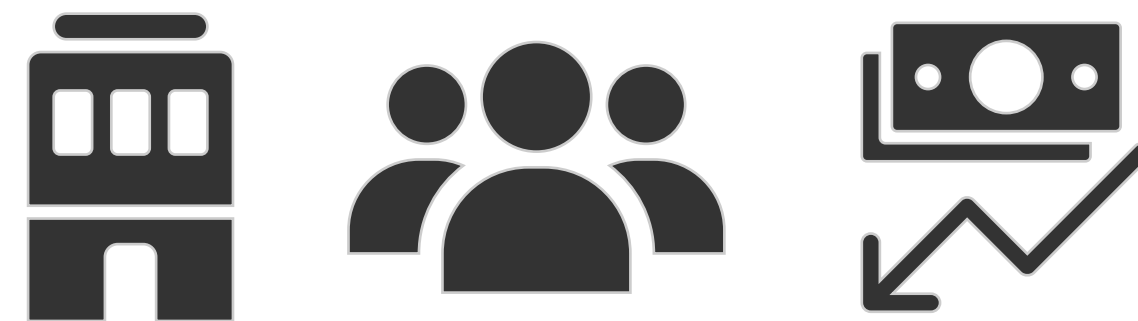
White papers will help disseminate results during the research project. They will address specific findings of economic and spatial impacts of COVID-19 on the neighborhoods with the most vulnerable households. Also, they will include visualizations of new data streams on the neighborhood level.



## 8. SAMPLE TOOLKITS

---

1. One Toolkit will be dedicated to the housing precarity of the population of renters of single-family homes. Most renters in California impacted by a recent job loss do not live in large multi-family homes (42% live in a single-family home, 16% in a building with 2-4 units, 22% in a building with 5 to 19 apartments). Single-family homes in Los Angeles are excluded from the Rent Stabilization Ordinance.



2. Second Toolkit will address tenants of small apartment buildings (under 20 units) in predominantly Black and Latinx neighborhoods. According to NAHREP (National Association of Hispanic Real Estate Professionals), 80% of their members reported a decline in their rental income and 25% have already borrowed funds.

3. Third Toolkit will illuminate locations and current level of the precarity of households with very long-term (i.e. elderly) tenants occupying rent-stabilized units in Los Angeles County. The median age of tenants in rent-stabilized units in the city is 43, but for those living in the same unit for 30 or more years, it is 68. This is a high-risk group for COVID-19.

