APPLYING OPERATIONS RESEARCH PRINCIPLES TO IMPROVE GEO-DATABASE QUALITY

Lean Principles and Value-Stream Mapping as Tools for Improving Geocoding Success

2014 ESRI International User's Conference
Data Compilation Case Studies: QA/QC
July 17, 2014
San Diego, CA

Anirudh Kannan Vinakayaram, MEME
Research Scientist
Geospatial and Population Studies
University of New Mexico

Ron Lumia, PhD
Professor, Mechanical Engineering
University of New Mexico

Ganesh Balaksharan, PhD
Assistant Professor, Electrical and Computer Engineering
Center for High Tech Materials, CHTM
University of New Mexico

Jack Baker, PhD
Sr. Research Scientist
Geospatial and Population Studies
University of New Mexico

Nathan Crouse, MS
Analyst Programmer
Geospatial and Population Studies
University of New Mexico

PRESENTATION OUTLINE

- Potential and Pitfalls in Geocoding for Small-Area Demographic Estimates: Project Goals
- Geocoding as a Manufacturing Process
- Lean Principles & Value-Stream Mapping
- Case Study & Simulation
- Results & Recommendations

Geospatial and Population Studies (GPS) at the University of New Mexico

GEOCODING AND SMALL-AREA DEMOGRAPHIC ESTIMATES

GEOSPATIAL AND POPULATION STUDIES AT THE UNIVERSITY OF NEW MEXICO

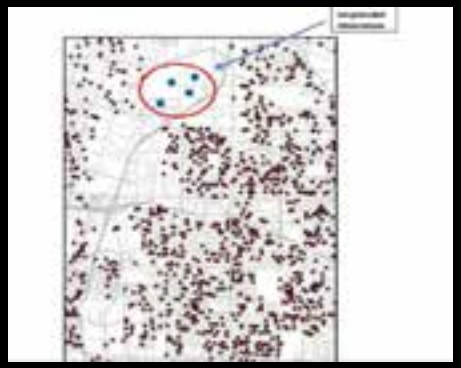
- Makes official population estimates for State of New Mexico.
- Estimates:
  - Form basis of health surveillance and reporting.
  - Inform public sector decision-making (legislation)
  - Support private sector decision-making (real estate, business development, etc.)

- Input data used in GPS estimation process:
  - Informs the US Census Bureau's Population Division in making their State, County, and Municipal-level estimates.
  - Directly fed into the updating of the US Census Bureau's Master Address File through programs including the Local Update of Census Addresses and the 2010 Count Review Program.

POTENTIAL AND PITFALLS IN GEOCODING FOR DEMOGRAPHIC ESTIMATES

Pre-Directional	House Number	Street Name	Street Type	Post Directional	City	State	Zip
	1313	Mockingbird	Ln		Anywhereville	NM	87478
N	2300	Lane	St		Anywhereville	NM	87478
	1100	Clod	Dr	SE	Whoville	NM	87478

1. Geo code address data.
2. Summarize counts within geographic units.
3. Make demographic estimates.



GEOID10	Tract Description	Total Population
35001000107	Census Tract 1.07; Bernalillo County; New Mexico	1,410
35001000108	Census Tract 1.08; Bernalillo County; New Mexico	1,386
35001000109	Census Tract 1.09; Bernalillo County; New Mexico	1,406
35001000110	Census Tract 1.10; Bernalillo County; New Mexico	1,884
35001000111	Census Tract 1.11; Bernalillo County; New Mexico	1,500
35001000112	Census Tract 1.12; Bernalillo County; New Mexico	1,081
35001000113	Census Tract 1.13; Bernalillo County; New Mexico	1,692

PROJECT GOALS

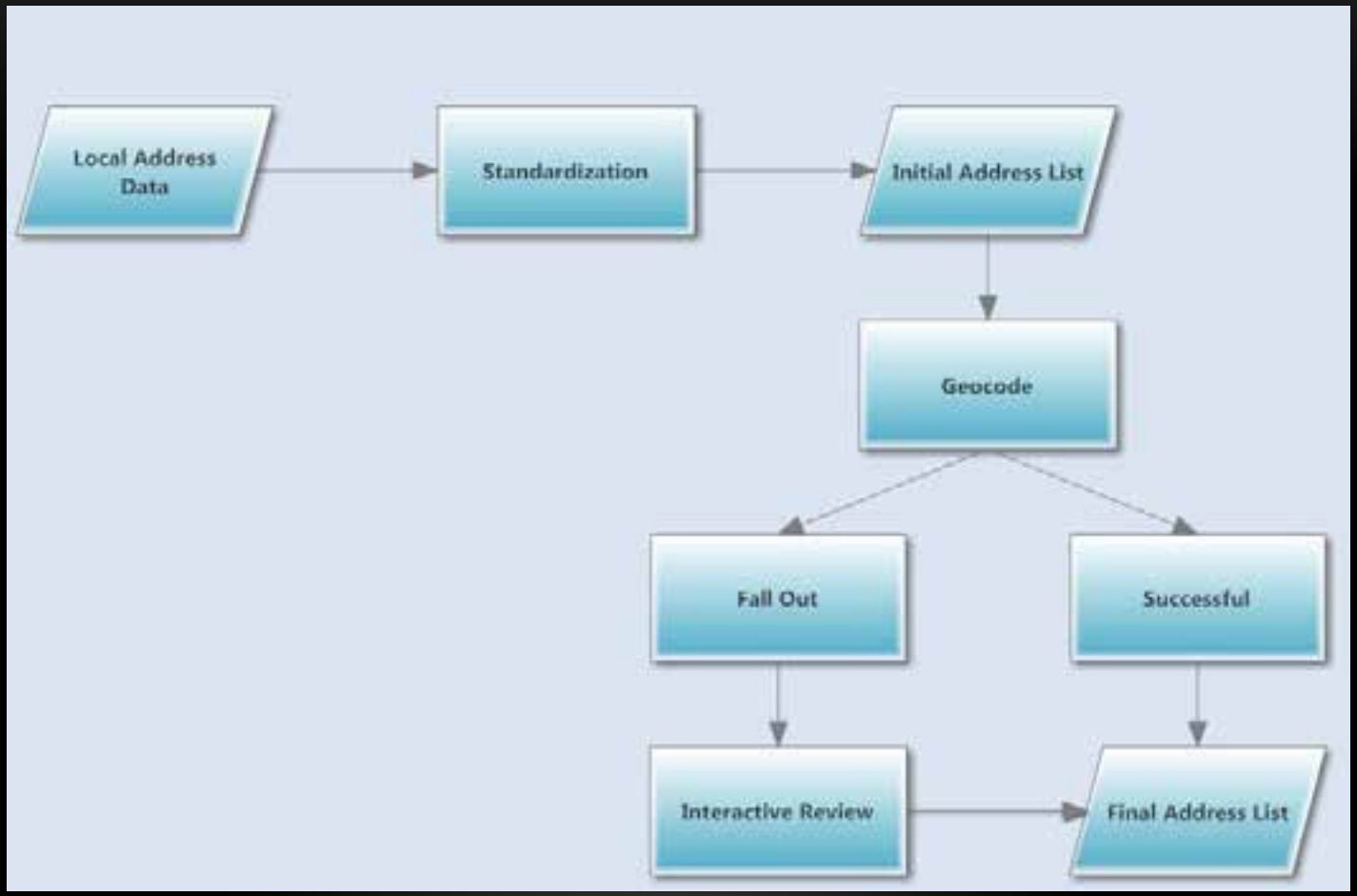
- Improvements in geocoding mean more accurate estimates.
- Geocoding success rates vary:
  - 75 to 90 % in urban areas
  - 10% to 40% in rural areas

- Goal: Improve geocoding in urban areas by 5%--moving from 90% to 95%.

- This improvement could result in a savings of approximately 1/5th of the GPS operating budget for geocoding and improve the accuracy of small-area demographic estimates by as much as 10 percentage points.

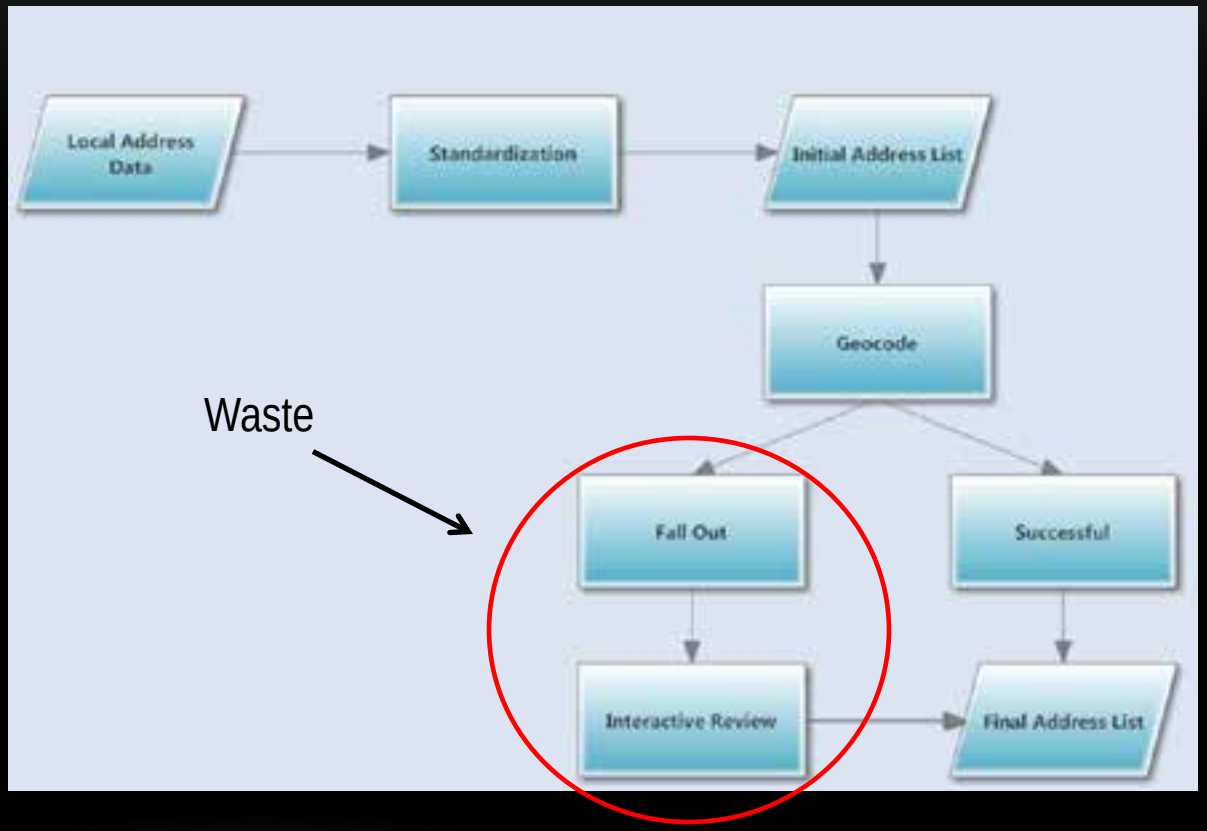
LEAN PRINCIPLES AND VALUE-STREAM MAPPING

GEOCODING AS A PRODUCTION PROCESS



LEAN PRINCIPLES IN GEOCODING

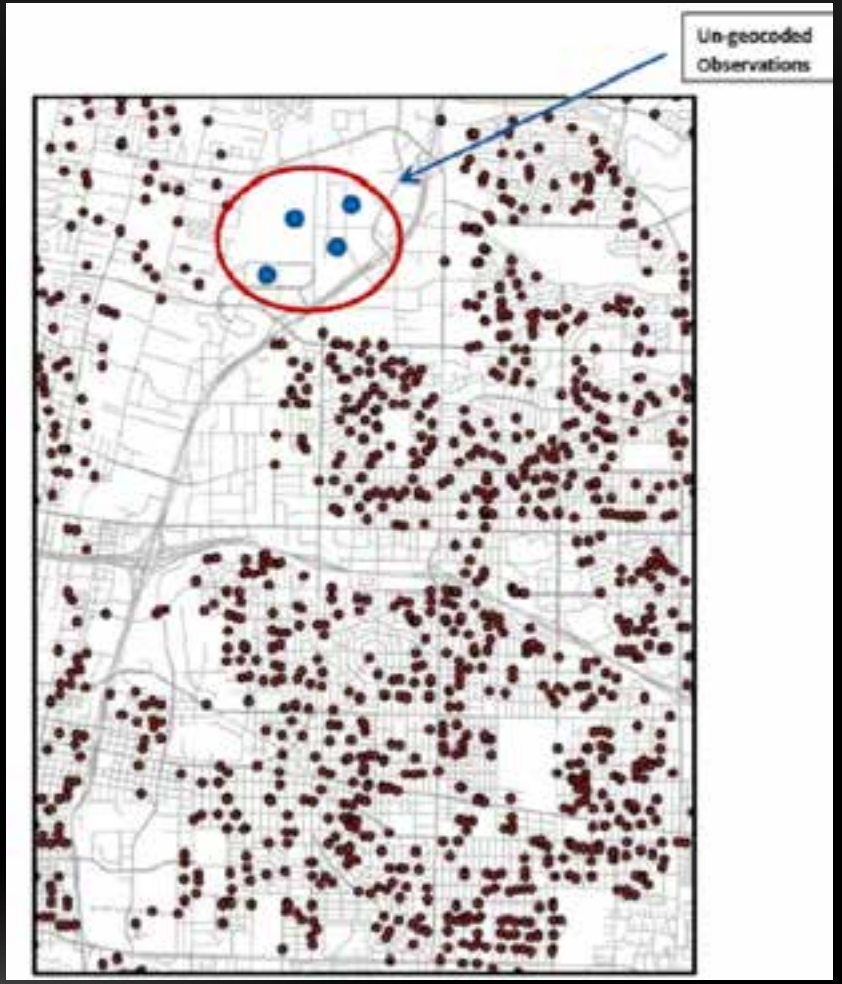
- Eliminate waste
- Waste adds cost with no value
- Define “value”
- Goal is to minimize waste to maximize value.



CONCEPTS OF WASTE (MUDA)

7 types of waste defined by Toyota

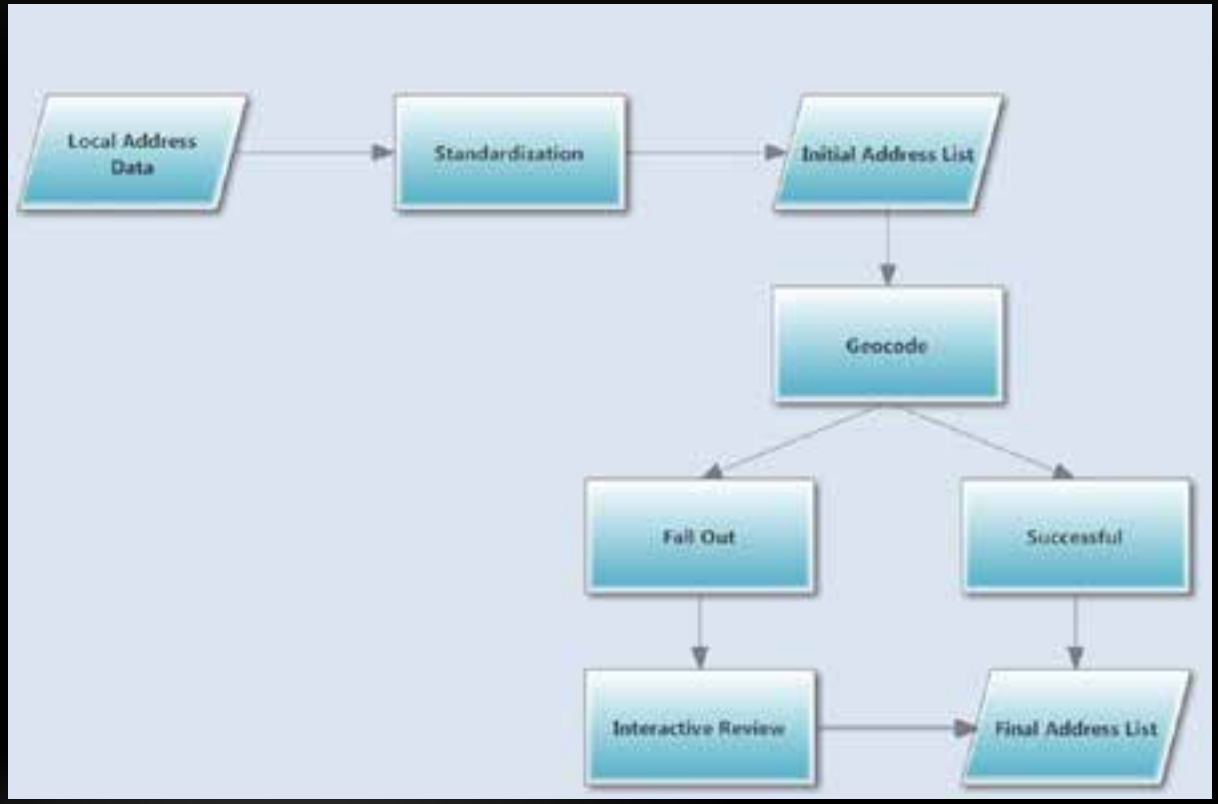
1. Overproduction
2. Waiting
3. Transporting
4. **Process Defects (Quality Control)**
5. Excess Inventory
6. Unnecessary Motion
7. Inappropriate Processing



VALUE STREAM MAPPING AS A TOOL

- Value Stream Mapping
  - Operationalizes Lean Principles
  - Represents a process graphically
  - Uses this graphic as a means of identifying waste
  - Suggests avenues for process improvement (eliminate waste)
- Graphic = “State Map”
  - Current
  - Future (lean principles applied)

Current State Map for GPS Geocoding Process



CASE STUDY AND SIMULATION

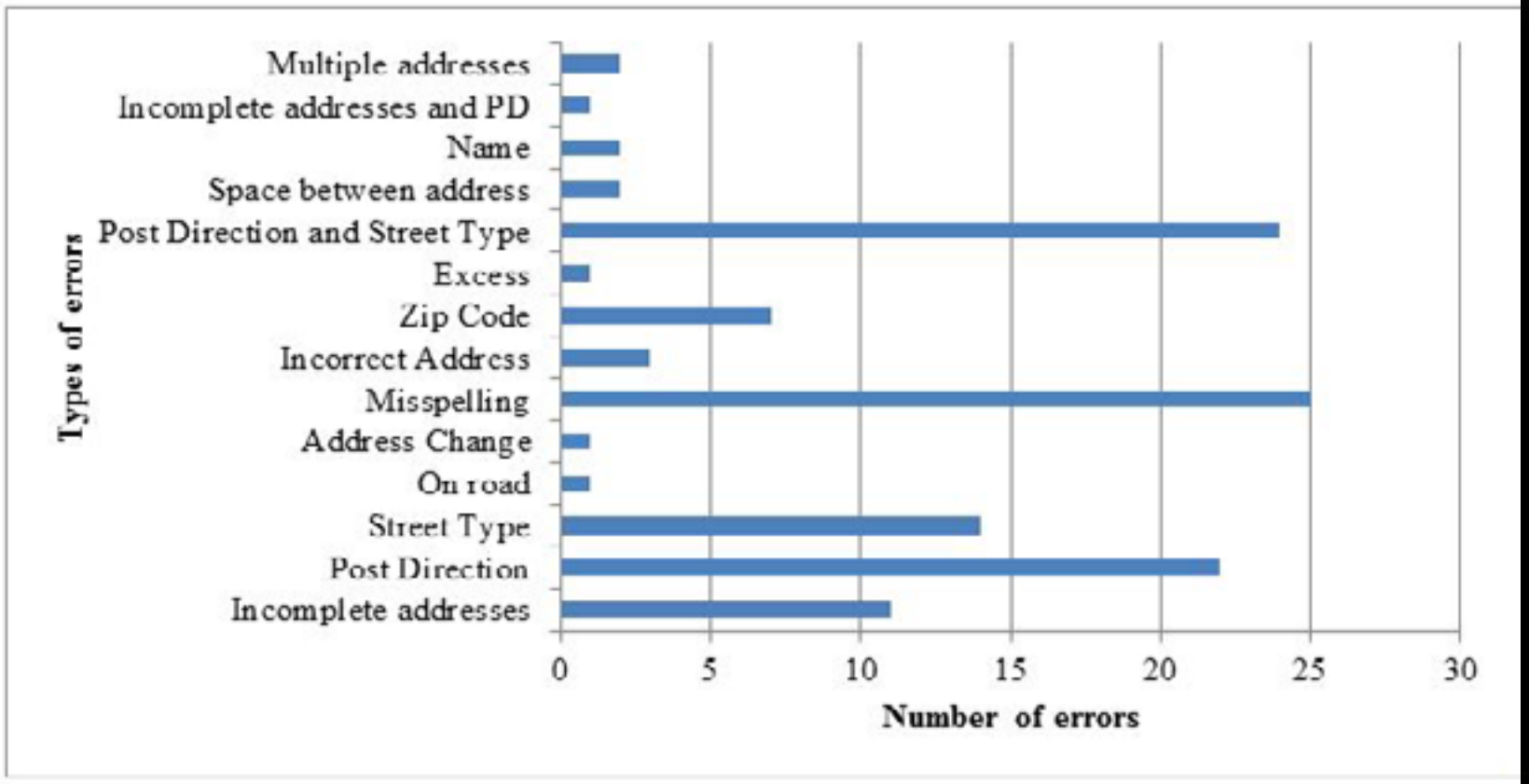
CASE STUDY

- Case study is used to identify defects in standardization process.

- Provides information on the specific types of errors and their frequency.

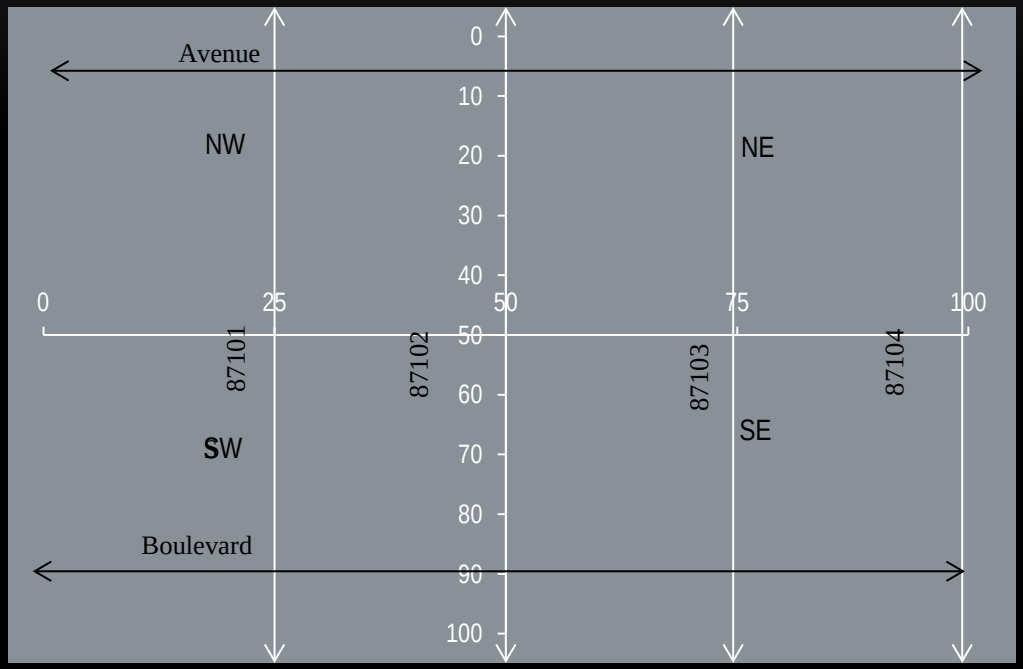
Incomplete addresses	11
Post Direction	22
Street Type	14
On road	1
Address Change	1
Misspelling	25
Incorrect Address	3
Zip Code	7
Excess	1
Post Direction and Street Type	24
Space between address	2
Name	2
Incomplete addresses and PD	1
Multiple addresses	2

RESULTS FROM CASE STUDY



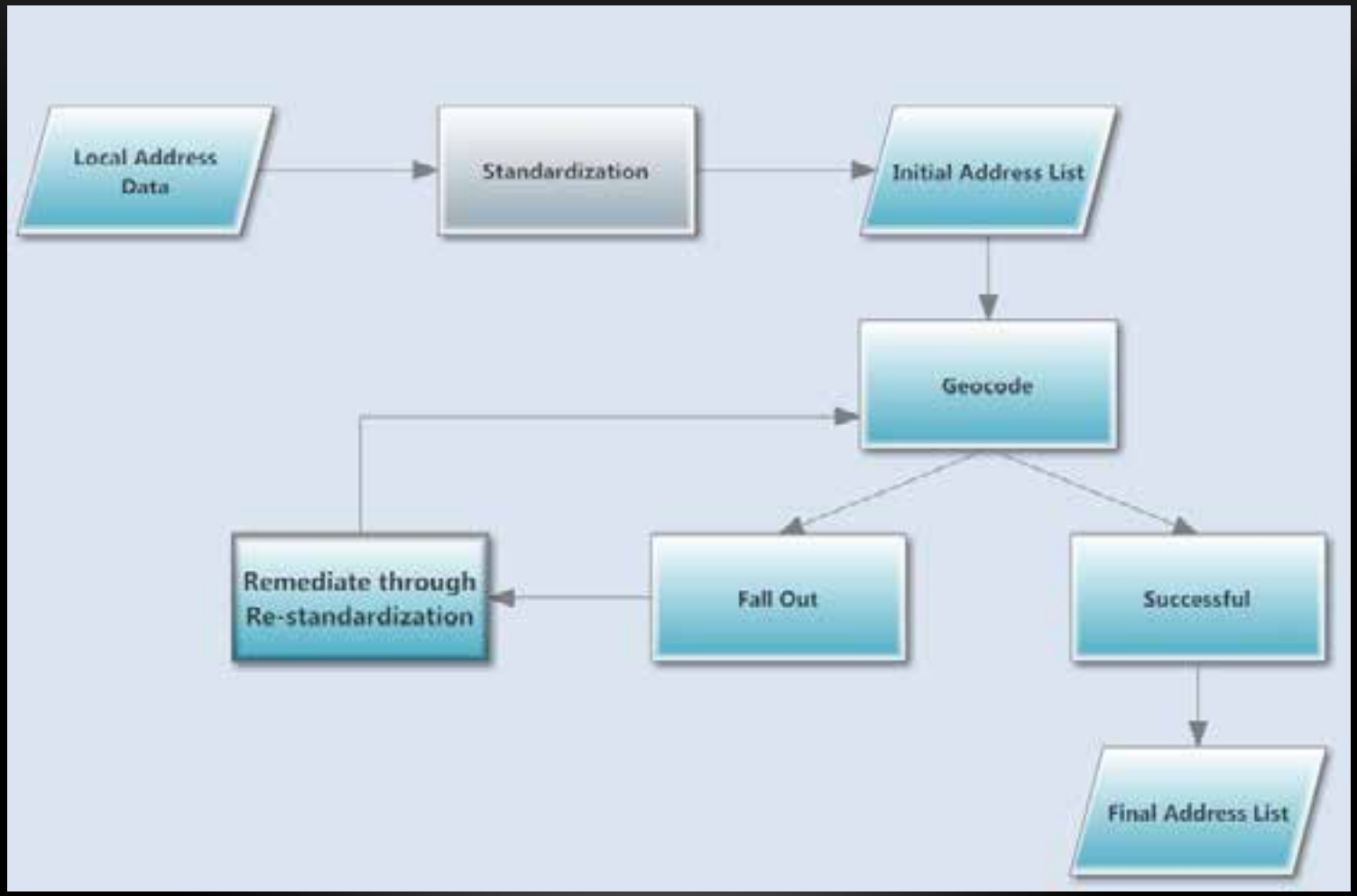
SIMULATION STUDY

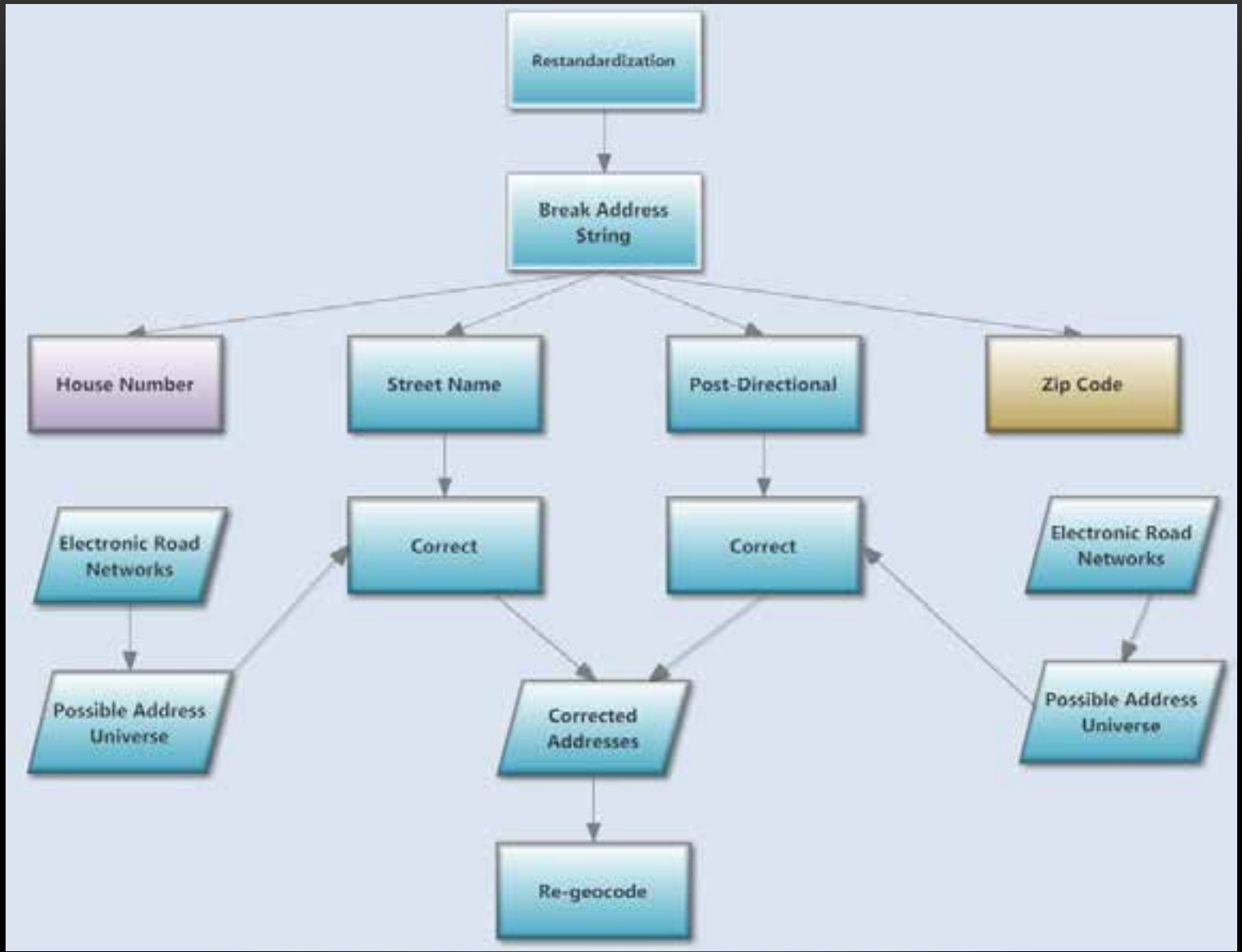
- Used to study the production system and mimic the error-generating process.
- An artificial city was built with 100 x 100 houses.
- 4 Zip Codes, 4 Post Directions and 2 Street Types were used.
- Addresses were corrupted according to frequency and type of error identified in case-study.
- **Correction algorithms were designed in light of results.**



FUTURE STATE MAP—SIMULATION TO ACTION

A NEW STATE MAP AND REMEDIATION PROCESS




Restandardization
Break Address String
House Number
Street Name
Post-Directional
Zip Code
Electronic Road Networks
Correct
Correct
Electronic Road Networks
Possible Address Universe
Corrected Addresses
Possible Address Universe
Re-geocode

RESULTS AND RECOMMENDATIONS

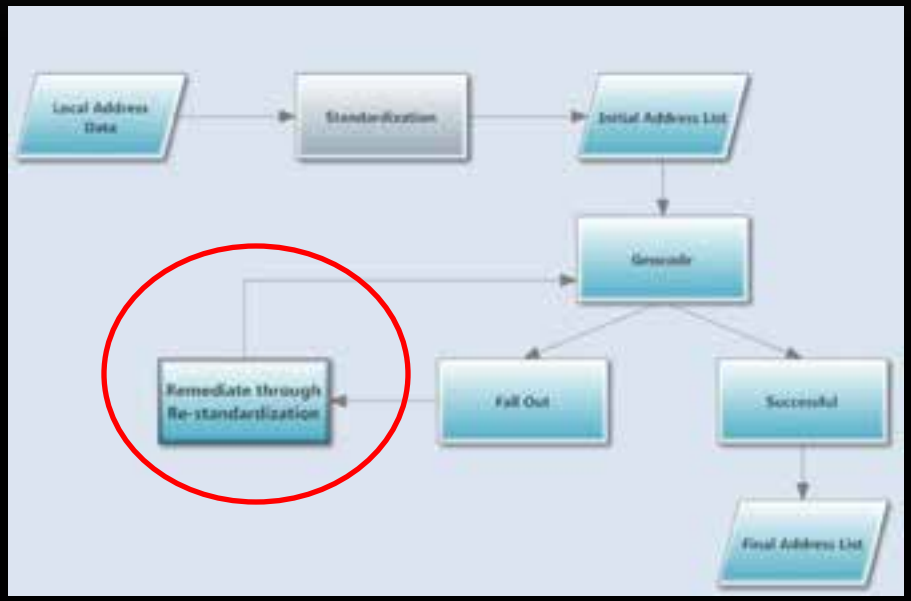
RESULTS

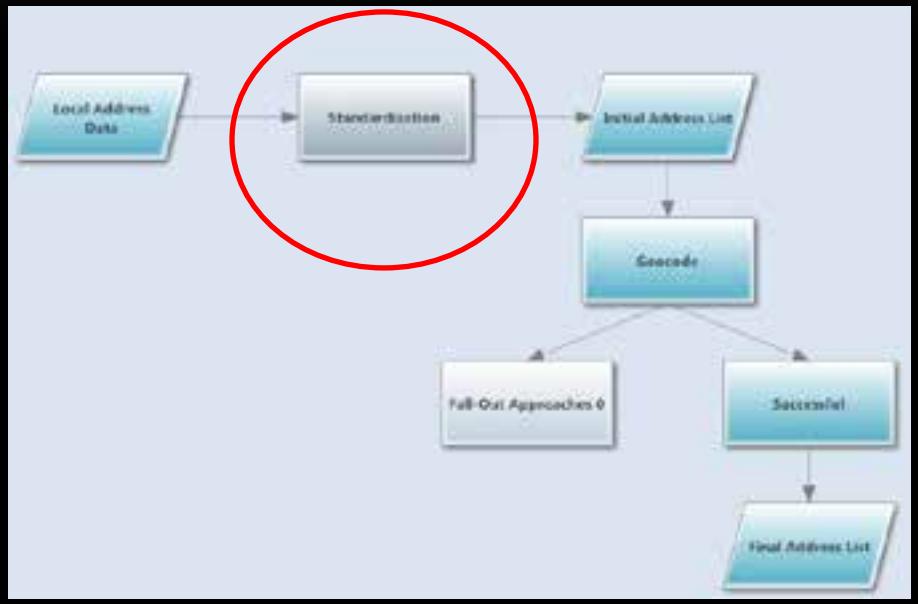
- New State Map/Process Results:
  - 95.79% of defective addresses were corrected.
    - House numbers (missing or out of range)
    - Incorrect Zip Code designations
    - PO Boxes
  - 79% of the remediated addresses were successfully re-geocoded.
  - In the final tally, geocoding increased from 78% to 95%--a 17 percentage point improvement.

COST OFFSET

Activity	Number of Records	Cost Per Record	Total Cost	Cumulative Cost	Percent of Data Capture
Geo-coding	20,379	$0.50	$10,189.50	$10,189.50	78.00
Interactive Hand-Matching	2,874	$1.00	$2,874.00	$13,063.50	88.00
Aerial Imagery Review/Google Maps	1,725	$5.00	$8,625.00	$21,688.50	94.65
Fieldwork	1,149	$15.00	$17,235.00	**$38,923.50**	100.00
Geo-coding	20,379	$0.50	$10,189.50	$10,189.50	78.00
Proposed Process	4,540	$.23	$1,023.75	$11,213.25	95.28
Left Imagery/Fieldwork	1,208	$10.00	$12,080.00	**$23,293.25**	100.00

RECOMMENDATION: FUTURE STATE MAP





FUTURE IMPROVEMENTS

- Focus on methods for remediating errors in house numbers
  - Missing (Imputation)
  - Out of Range (Machine Learning Correction Algorithms)
  - PO Boxes (Linkages)

- Zip Codes
  - Error rates in assignment
  - Missing elements (incomplete digits)
  - Zip code changes over time

ACKNOWLEDGMENTS

- Dr. Adelamar Alcantara, Director of Geospatial and Population Studies
- Dr. Michael Sonksen, Assistant Professor of Statistics, University of New Mexico

CONTACT INFO

Anirudh Kannan Vinayakaram

Research Scientist

Geospatial and Population Studies

MSC06 3510

1 University of New Mexico

Albuquerque, NM 87131

(505) 277-1704

akannan@unm.edu