

Address Matching at OED

- Goals
 - Develop Batch programs for geo-coding large historical time series databases of addresses and attributes
 - Have highest precision as possible by using local level data where possible followed by lower resolution data
 - Use multiple geo-coding methods (engines) and multiple data sets for optimizing results

Address Scrubbing

Address Data File

Format Data to USPS Standard
 Upload data to IBM Mainframe
 Run Finallist (TSO)

- City/Zip match
- Address Standardization
- Street Matching
- Add Zip+4 to addresses

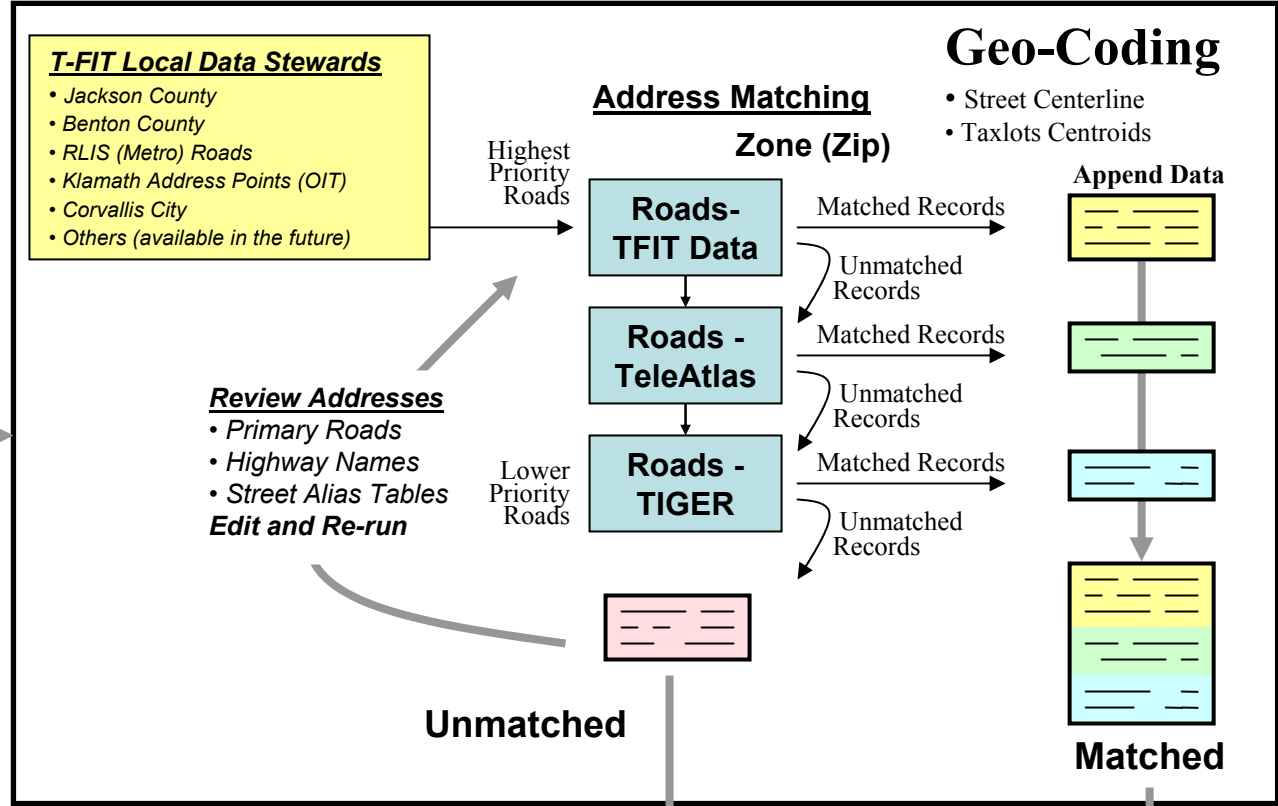
Run Easytrieve+ formatting program

Download and cross-tabulate results

Merge additional addresses from Longitudinal Data Base (LDB)

Format data for GIS input

GIS Address Matching



Manual Digitizing (Hardwired)

Data looked up off from maps, etc

Centroid Methods

Unmatched

Zip+4 Centroid

Zip-ZCTA Centroid

USPS - TIGER

P.O. Boxes & Special Zips

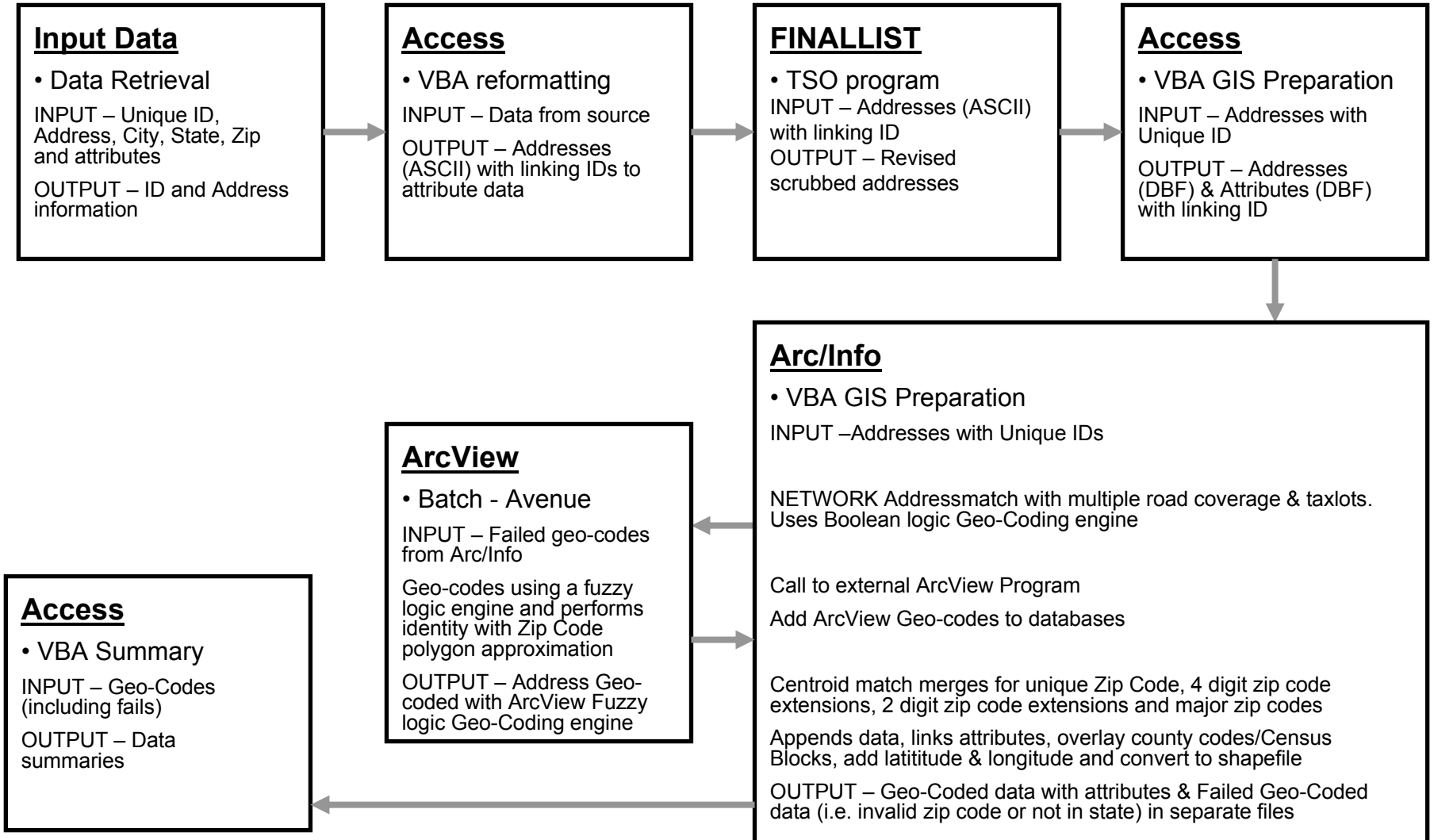
Inclusion Zip Codes

Append Data

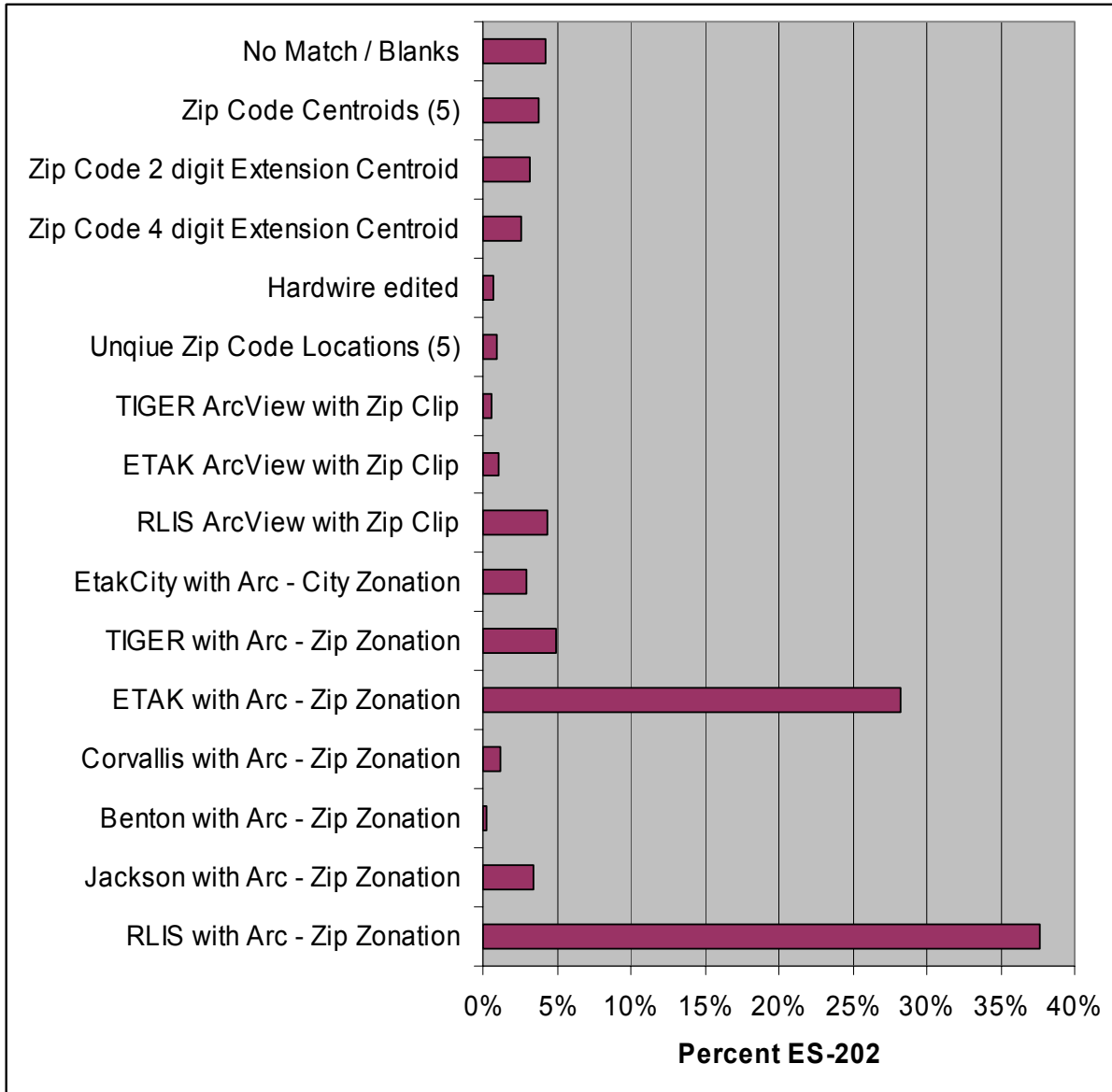
Matched



Geo-Coding Program Flow



Example: 116,000 Employers in Oregon



Lowest



Precision



Highest

Other Recently Parsed Data

- Oregon Job Applicants and Benefit Claimants (from Oregon Employment Department) – 650,000 unique SSN and addresses
- Oregon Business Licenses (from Secretary of State, Corporate Division) – 750,000 unique historical licenses and addresses
- Worker Compensation (from Department of Consumer and Business Services) – 450,000 historical unique business locations and addresses
- Oregon Employers Wages and Employment – 120,000 unique firm and addresses

The current programs can process roughly 250,000 unique addresses per hour of processing time