New Technology for the 2020 Census

Census Task Force Meeting

Amman, Jordan

November 2015

Topics

- Background
- Sourcing Data
- Detecting Change
- Continually Improving Data Quality

Purpose

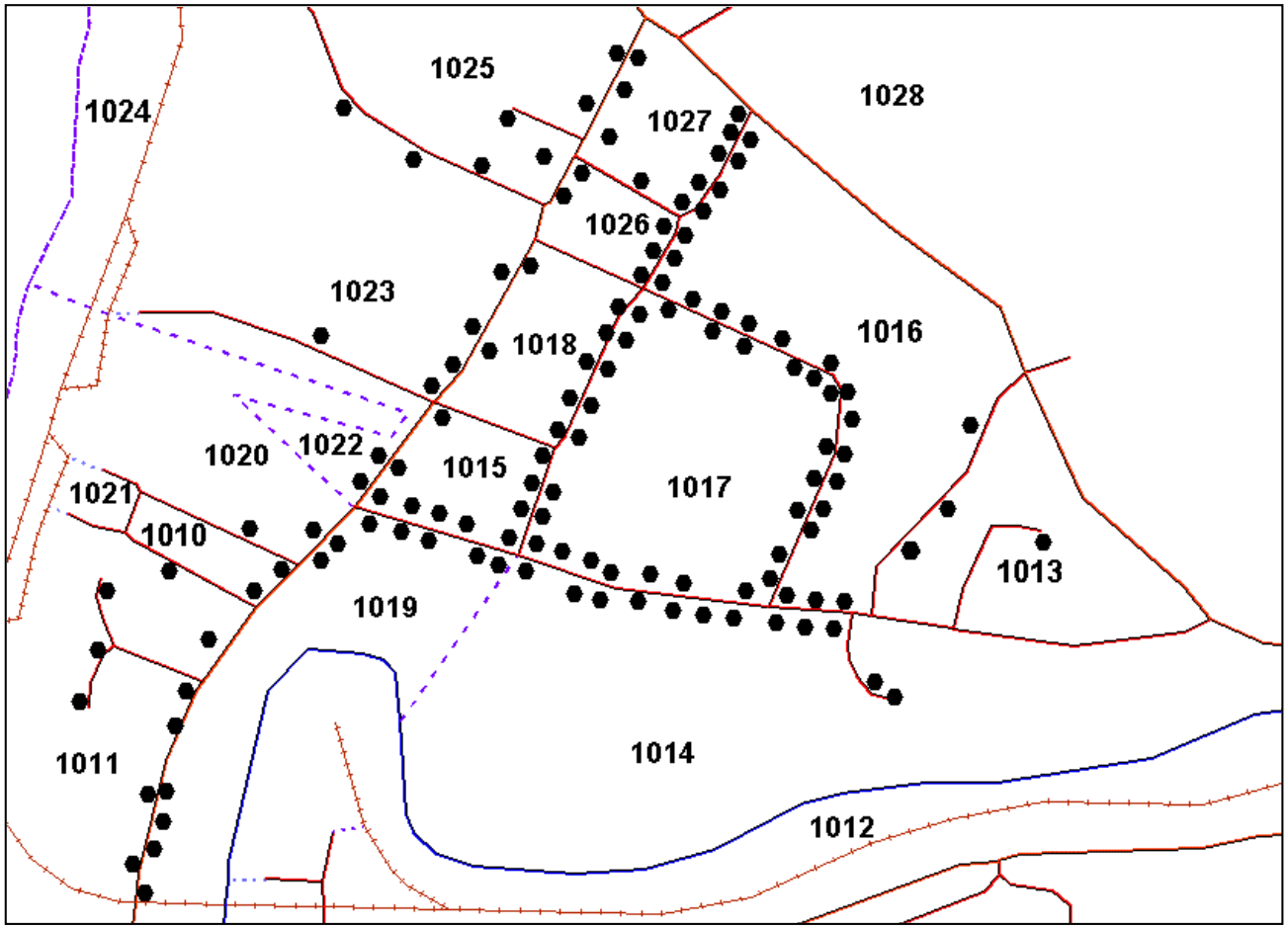
- As we approach the 2020 census round, the use of new technologies will **save time** and **improve the quality** of census geography.
- However, national statistical offices **must carefully evaluate new technology** before implementation.
- This presentation will show some examples of new technology adaptation from the U.S.
  - E.g.: databases, satellite imagery, GPS.

U.S. Statistical Geography

U.S. Statistical Geography

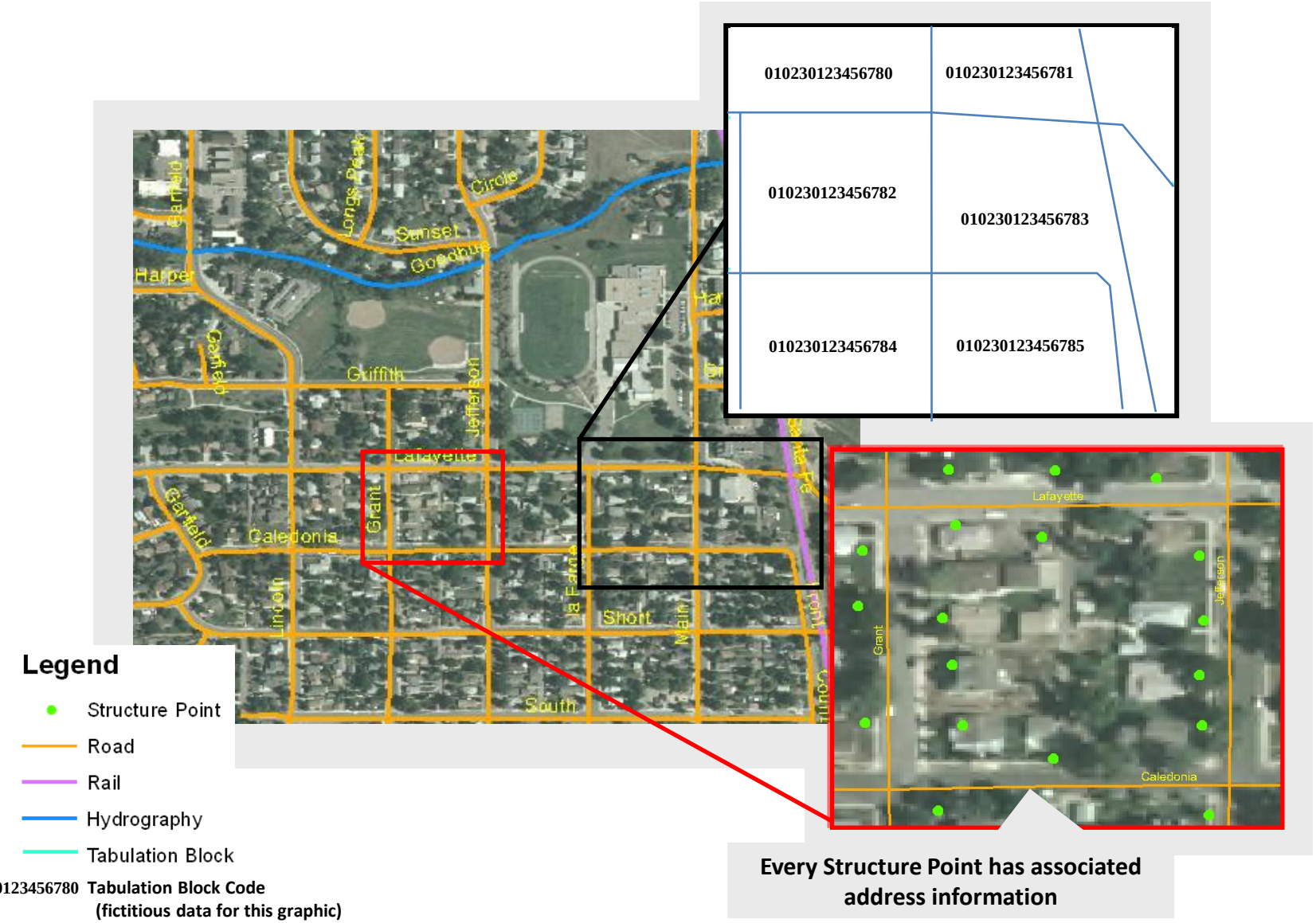
- Historically:
  - Addresses: **Master Address File (MAF)**
  - Geography: **Topologically Integrated Geographic Encoding and Referencing (TIGER)** System
- For 2010 Census, integrated into one system: **MAF/TIGER Database (MTDB)**.
  - Greater availability of data and computer processing power.
  - More powerful and user friendly Commercial Off the Shelf (COTS) GIS software.
  - Improved accuracy of Global Positioning System (GPS) devices.
  - High resolution satellite imagery.
  - Demand for geocoded housing units.

MAF/TIGER Database (MTDB)



MAF/TIGER Attributes

- Postal Code (ZIP Code)
  - Mail
  - Location
- Mailing Address
- Location Address
- Location Description
- Coordinates
  - GPS or Manual
  - Preferred
  - Position
- Type
  - Housing Units (HU)
  - Group Quarters (GQ)
  - Transitory Locations (TL)
  - Nonresidential
- Vacant, Seasonal, and Rental Status
- Lowest Level of Government
- Road/Street Edge Linkage
- Address Range Linkage
- Polygon (Face) Geocode
- Block Geocode
  - Decennial Census Tabulation Block
  - Current Tabulation Block
  - Decennial Census Collection Block
- Subdivision/Development/Apartment Complex/Mobile Home Park/Building name
- Telephone Area Code
- Year Built
- Source and History Data



Allows integration of multiple datasets

Current MAF/TIGER System

- Oracle Database
  - 10g → 11g → 12c
- Environment
  - IBM Blade Servers → Exadata Servers
- Oracle Spatial
  - MDSYS Schema
  - Spatial Indexing
- Oracle Topology Data Model
- Persistent topology
- Batch processing environment
- Benchmarks – Copied twice annually for products; full suite of edits
- COTS-based software, but **heavily customized** to Census Bureau's business needs

MAF/TIGER Database (MTDB)

- MTDB relationship tables (number of records included)
  - EDGE$ 73,187,478
  - FACE$ 22,597,539
  - NODE$ 160,123,508
  - RELATION$ 3,054,850,670

- Number of tables (layers) = 99

These Integrated Components of MAF/TIGER

are needed to support Census and Survey operations

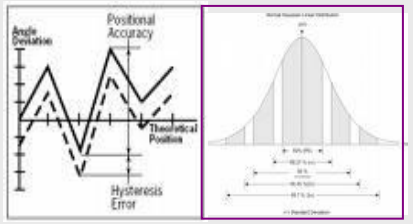
Address and HU Updates

123 Testdata Road
Anytown, CA 94939

Lat 37 degrees, 9.6 minutes N
Lon 119 degrees, 45.1 minutes W

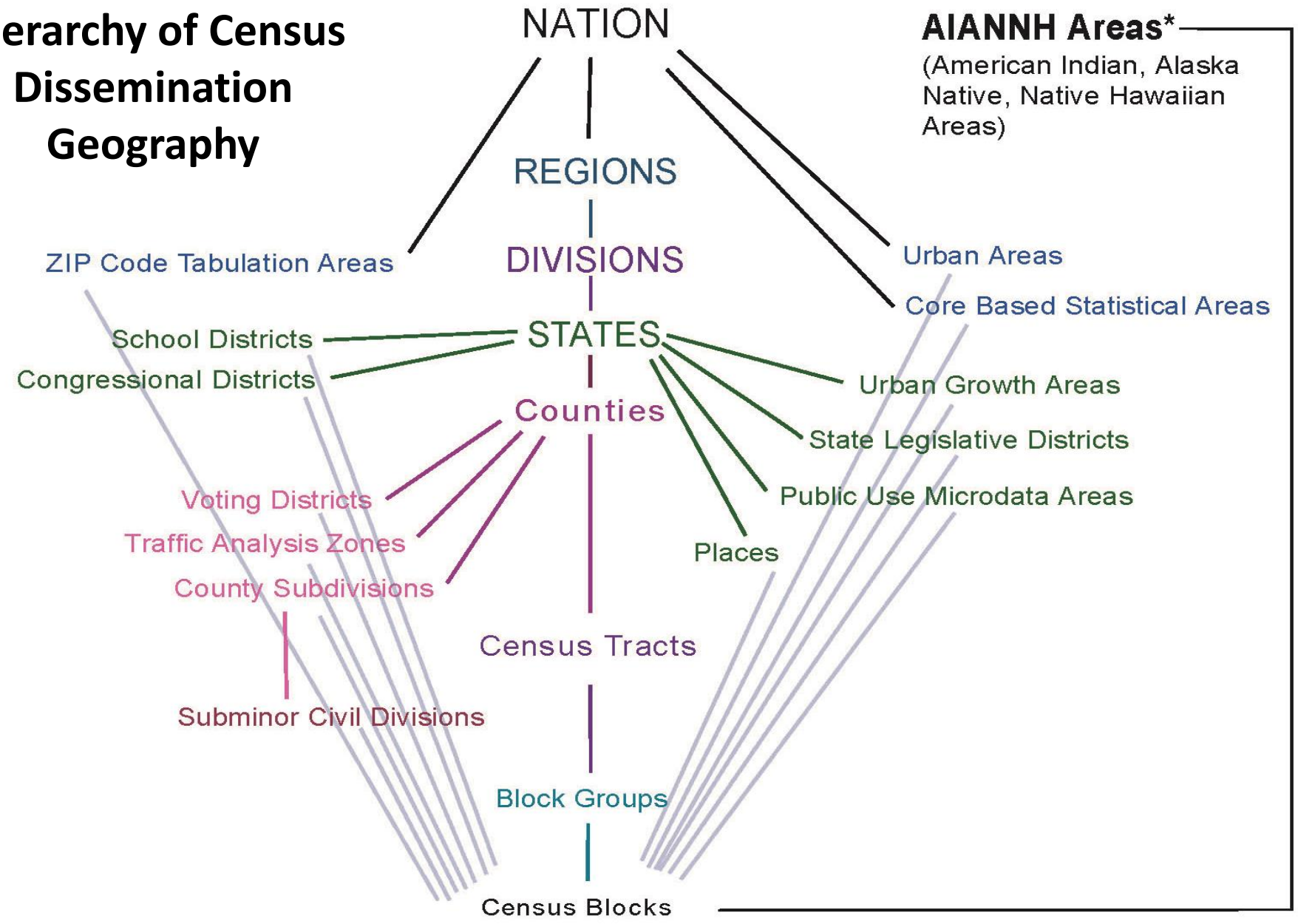
Street/Feature/Boundary Updates & Imagery

Quality Measurement, Quality Assurance & Change Detection



…and to link to statistical data

Hierarchy of Census Dissemination Geography



Collection Geographic Area Hierarchy

Local Census Office

↓

Field Operations Supervisor District

↓

Crew Leader District

↓

Assignment Area

↓

Collection Block

2010 Delineation Process

- Start with MTDB features (roads, rivers, etc)
- Create collection blocks in automated process
- Geocode housing to collection blocks
- Determine Type of Enumeration Area (TEA)
- Create Assignment Areas (AAs) in automated process
- Review AAs interactively
- Load AAs on handheld device
- Assign work by AA

2010 Collection Geography

Criteria:

1) **Bounded by:**
   - ✓ roads, shorelines, county boundaries, American Indian reservation & trust land boundaries, military installation boundaries, and/or minor civil divisions in some states.

2) **Contiguous**

Guidelines:

1) **Compactness**
2) **Size of Area**
   - ✓ Remain consistent through all census operations.
   - ✓ No housing unit count requirement for collection blocks.

Collection Blocks

2010 Collection Geography

Assignment Areas (AAs)

Criteria:

1) **Built from:**
   - ✓ one or more 2010 collection blocks
2) **Cannot cross:**
   - ✓ county boundaries, American Indian reservation & trust land boundaries, military installation boundaries, and/or minor civil divisions in some states.
3) **Contiguous**

Guidelines:

1) **Compactness**
2) **Land Area**
   - ✓ A small geographic area consisting of a collection block or group of blocks established by the Census Bureau as a basic unit for data collection by a single enumerator for a single operation.

2010 Collection Geography

- Assignment Areas different by field operation

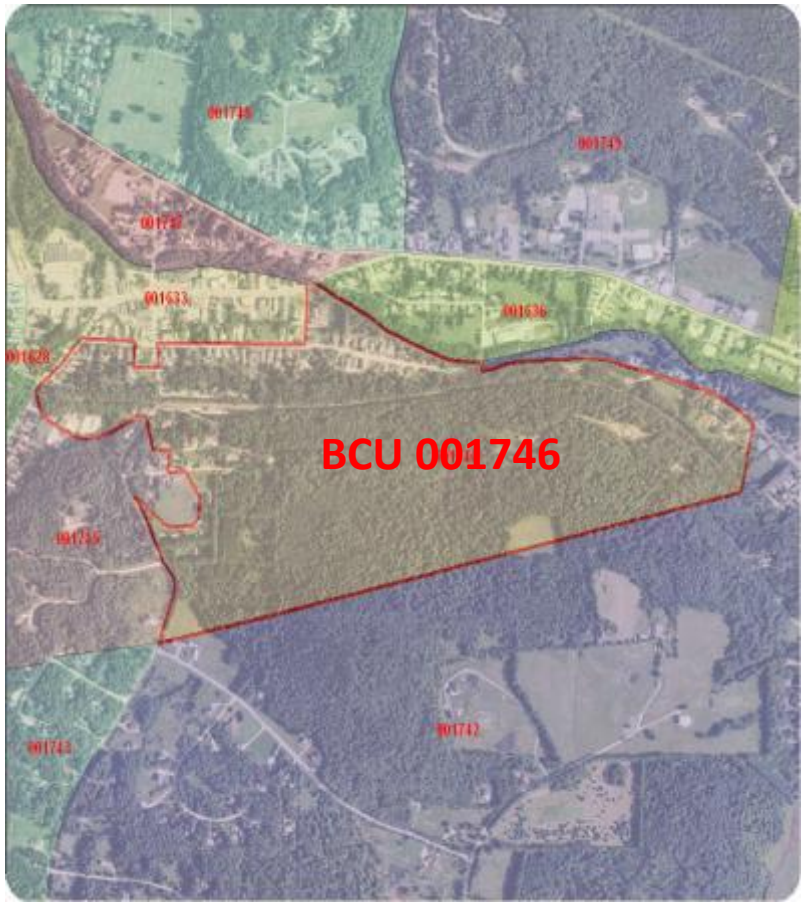
Address Canvassing

Census Coverage Measurement

The Basic Collection Unit (BCU)

- A Proposal for the 2020 Census
- Replaces the collection block and assignment area concepts
- Uses weighted BCU Concept.
  - Incorporates new homogeneity requirements (i.e. housing types and address types)
  - Updated throughout the decade
  - Designed to meet the needs of different field operations (navigability and manageable size)

More accurate estimates of work rates



Work rates are based on 2010 collection data for LCO /ELCO 2146 (ELCO 2146 for Address Canvassing).

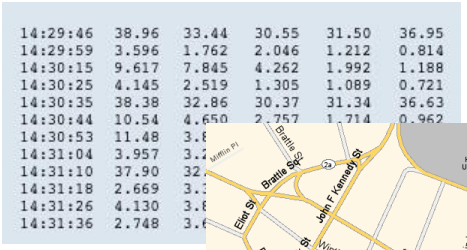
BCU Information (001746)	Statistics
Road Length (BCU 001746)	3.57 miles
Average BCU Road Length (LCO 2146)	24.02 miles
Area (BCU 001746)	0.27
Average BCU Area (LCO 2146)	5.00
Ratio of Road Length to Area (BCU 001746)	12.99
Ratio of Road Length to Area (LCO 2146)	18.08
Total Housing Units (BCU 001746)	151 HUs
Average BCU Housing Units (LCO 2146)	182 HUs

Workload Factors	BCU Estimate
ADDCAN Work Rate	12.88 cases/hr.
U/L Work Rate	5.49 cases/hr.
U/E Prod Enum. Work Rate	1.72 cases/hr.
U/E Prod RI Work Rate	0.68 cases/hr.
NRFU Prod Enum Work Rate	1.20 cases/hr.
NRFU RI Work Rate	1.18 cases/hr.

Data Sources

Source Data for MAF/TIGER

- **Geospatial data**
  - Local GIS files
  - MAF/TIGER Partnership Software (MTPS)
  - Digitized maps
  - Imagery
  - Field updates



14:29:46	38.96	33.44	30.55	31.50	36.95
14:29:59	3.596	1.762	2.046	1.212	0.814
14:30:15	9.617	7.845	4.262	1.992	1.188
14:30:25	4.145	2.519	1.305	1.089	0.721
14:30:35	38.38	32.86	30.37	31.34	36.63
14:30:44	10.54	4.650	2.757	1.714	0.962
14:30:53	11.48	3.8			
14:31:04	3.957	3.2			
14:31:10	37.90	32.			
14:31:18	2.669	3.1			
14:31:26	4.130	3.8			
14:31:36	2.748	3.6			

- **Address Data**
  - GPS coordinates from handheld devices
  - Delivery Sequence Files (DSF) from US Postal Service (USPS)
  - Existing Master Address List

Data Sources - Partnerships

- **Federal government agencies (examples)**
  - U.S. Postal Service's (USPS) Delivery Sequence File & other files
  - U.S. Geological Survey's (USGS) Geographic Names & Codes (GNIS), National Hydrologic Database (NHD),
  - U.S. Department of Agriculture - Farm Service Agency's (USDA - FSA) Imagery
  - U.S. National Oceanic and Atmospheric Administration's (NOAA) Continuously Operating Reference Stations (CORS) program – Geodetic Reference Frame (survey control, datum, and GPS correction)

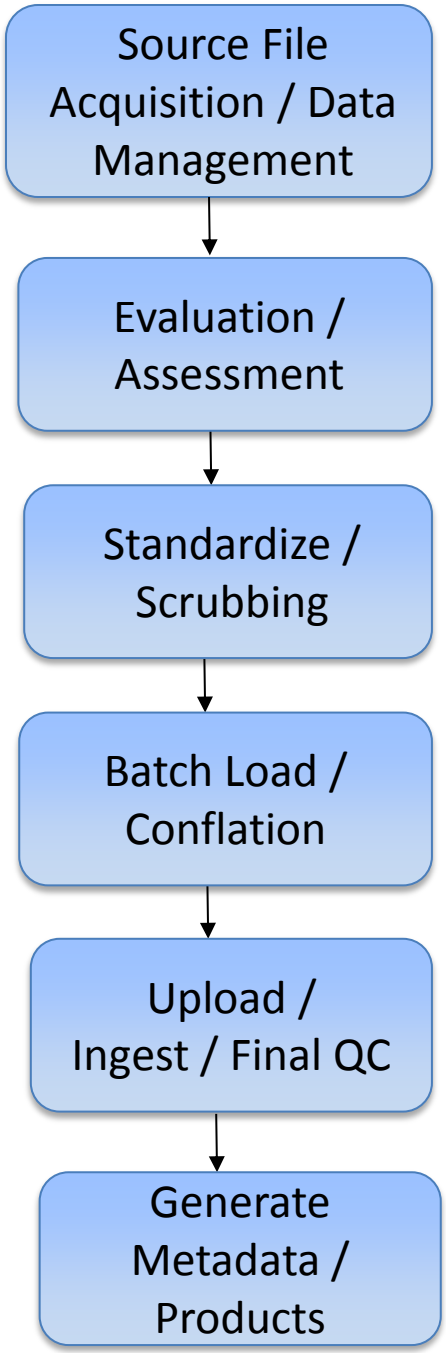
- **Tribal, State, and local governments**
  - Create, maintain, and provide the greatest volume and quality of spatial data

- **Private sector**
  - Data acquisition, software development, database design and management, data integration

- **Academic institutions**
  - Research and collaboration

U.S. Census Bureau's Geospatial Data Review Process

- Focus on ongoing maintenance and update of the MAF/TIGER Database (MTDB) spatial and address data.
- Add and improve spatial content, accuracy, and address geocoding in support of a targeted Address Canvassing.
- Incorporates five broad activities including several automated and interactive tasks that control tracking, assessment, and processing of GIS files and resulting updates



Mobile Devices

2010

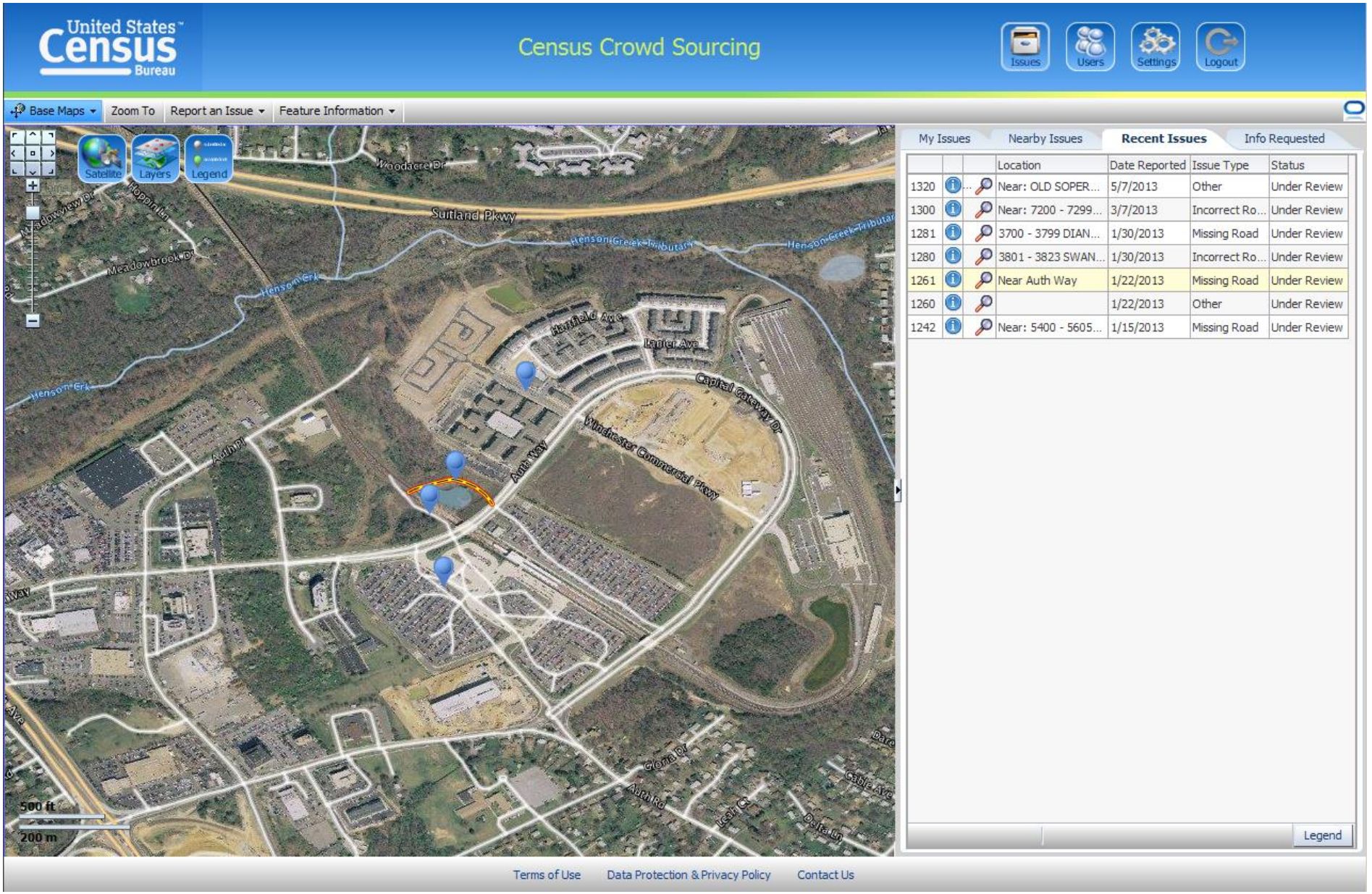
- Unique designed handheld computers were used for Address Canvassing operation to update 2010 MAF
- MTDB data were used for address and spatial data
- Included collection geography, address list for assigned area, and any geographic information to get enumerator/lister to their assigned area
- GPS capability to capture location
- Manual location of housing units (HUs) also required
- Visualization/mapping of data done using Esri's ArcPad

2020

- Platform agnostic – run on any mobile device that can support software and minimal GPS accuracy/precision requirement
- MTDB data are used for address and spatial data
- Ability to use outside data sources for navigation and routing
- Manual location of housing units (HUs) also required
- Imagery included
- Ability to complete forms / interviews

Crowd Sourcing Application

- Allows a user to indicate areas or specific features where data need to be updated or corrected.
- **Web based**, can run on any device as long as there is browser support.
- Shows TIGER features and attributes.
- Simple yet powerful way to detect change(s).
- Needs to be incorporated into Geography Division's workflow.


United States Census Bureau
Census Crowd Sourcing
Issues
Users
Settings
Logout
Base Maps
Zoom To
Report an Issue
Feature Information
Satellite
Layers
Legend
Woodacre Dr
Suitland Pkwy
Meadowview Dr
Hoppin Ln
Meadowbrook Dr
Henson Creek Tributary
Henson Crk
Hatfield Ave
Lapier Ave
Capital Gateway Dr
Winchester Commercial Pkwy
Auth Way
Auth Pl
Auth Rd
Gloria Dr
Leah Ct
Delta Ln
Cable Ave
500 ft
200 m
My Issues
Nearby Issues
Recent Issues
Info Requested
Location
Date Reported
Issue Type
Status
1320 Near: OLD SOPER... 5/7/2013 Other Under Review
1300 Near: 7200 - 7299... 3/7/2013 Incorrect Ro... Under Review
1281 3700 - 3799 DIAN... 1/30/2013 Missing Road Under Review
1280 3801 - 3823 SWAN... 1/30/2013 Incorrect Ro... Under Review
1261 Near Auth Way 1/22/2013 Missing Road Under Review
1260 1/22/2013 Other Under Review
1242 Near: 5400 - 5605... 1/15/2013 Missing Road Under Review
Legend
Terms of Use
Data Protection & Privacy Policy
Contact Us

Detecting Change

Imagery data detects rapid landscape change

- Census 2000:
  708 housing units
  - 621 occupied
  - 87 vacant
- 2010 Census:
  30 housing units
  - 10 occupied
  - 20 vacant

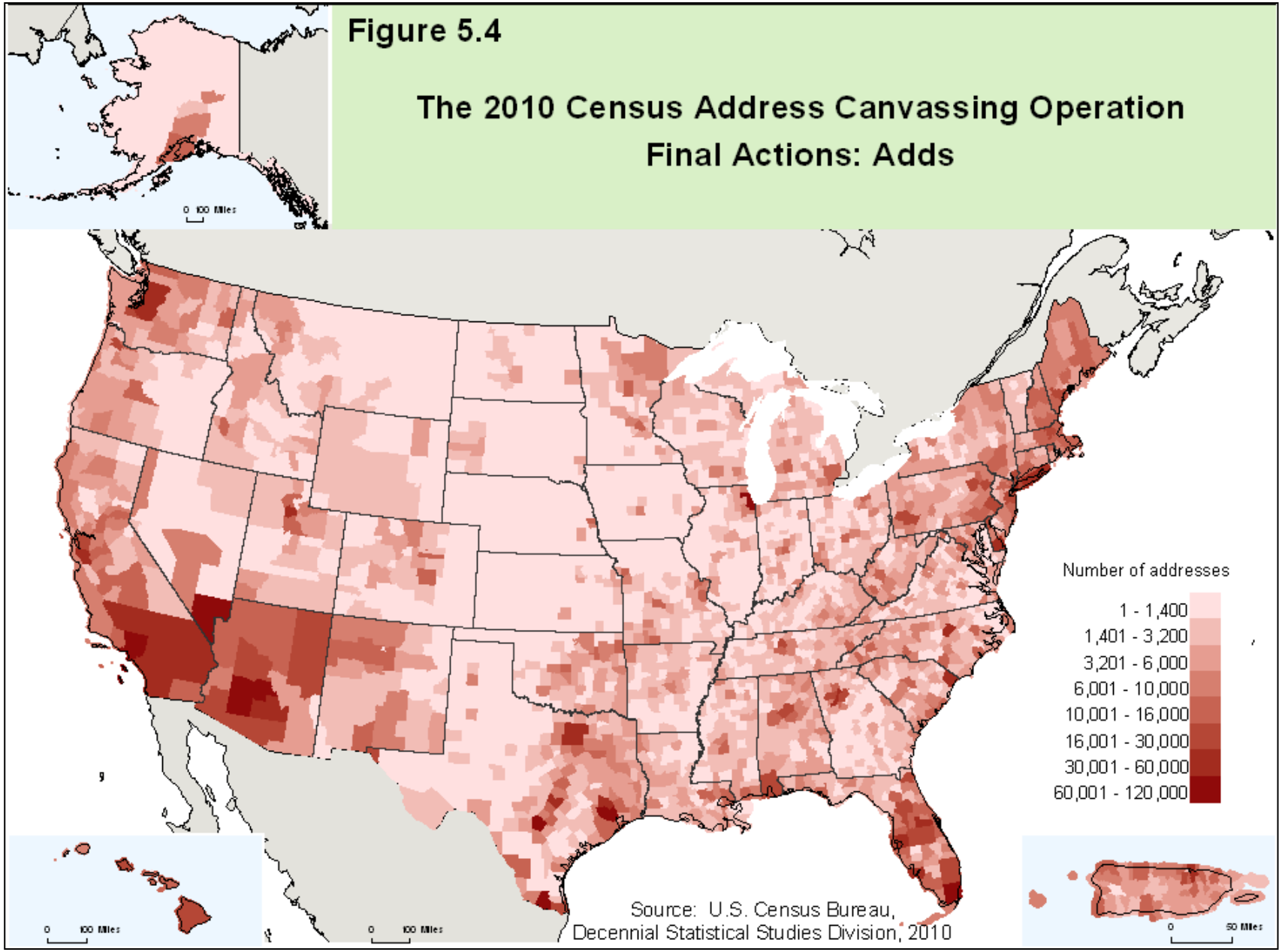


Figure 5.4
The 2010 Census Address Canvassing Operation
Final Actions: Adds
Number of addresses
1 - 1,400
1,401 - 3,200
3,201 - 6,000
6,001 - 10,000
10,001 - 16,000
16,001 - 30,000
30,001 - 60,000
60,001 - 120,000
Source: U.S. Census Bureau,
Decennial Statistical Studies Division, 2010

Number of Housing Unit additions within a Collection Block during the 2010 Address Canvassing Operation

Number of Units Added per Collection Block on the MTdb	Number of Blocks	Percent of total
Total Blocks	4,774,329	100.00
0	2,819,511	59.06
1	671,212	14.06
2-9	934,053	19.56
10-19	145,131	3.04
20-59	113,791	2.38
60-99	25,748	0.54
100-499	41,057	0.86
500-999	8,051	0.17
1000+	15,775	0.33

Source: GQV Extract Files, as defined by the matched MAFSRC and ACTION operation variables and COLBLKST, COLBLKCOU, and COLBLK extract variables.

MAF/TIGER: Updates for the 2020 Census

- An integrated program that utilizes partnerships for:
  - Improved address coverage
  - Continual address and spatial feature updates
  - Enhanced quality assessment and measurement

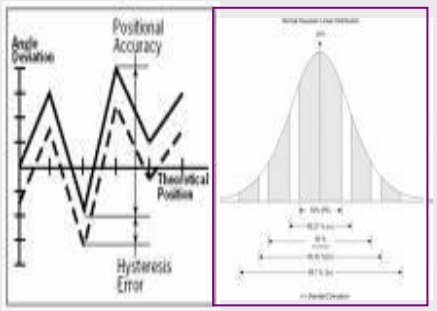
Address Updates

123 Testdata Road
Anytown, CA 94939

Lat 37 degrees, 9.6 minutes N
Lon 119 degrees, 45.1 minutes W

Street/Feature Updates

Quality Measurement



Partial Block Canvassing:
Testing A New Canvassing Methodology for the 2020 Census

- **Partial Block Canvassing (PBC)** focuses fieldwork on a specified location or area containing a cluster of updates within a census block rather than canvassing the entire block.

- **In-office imagery review** identifies the area to be updated. In-office review identifies inconsistencies between numbers of addresses in the Master Address File (MAF) and housing units visible in imagery. These are areas that will be updated through full or partial block canvassing.

- **Initial findings from the in-office review indicate that the vast majority of blocks-- over 80% -- are not exhibiting change,** and are consistent when comparing the imagery against the MAF, and therefore would not need to be canvassed through either a full or partial block canvassing operation.




Comparison of Full Canvass vs Partial Canvass


Continually Improving Data Accuracy

Using Satellite Imagery and Advanced Methodologies

MAF/TIGER Accuracy Improvement Program (MTAIP)

- Obtained GIS files from local partners
  - Generally county, state, tribal or regional planning officials
- Files were evaluated for completeness and spatial accuracy
  - Evaluated against TIGER and imagery for completeness
  - Spatial accuracy tested independently
- Obtained the National Hydrography Dataset (NHD) from USGS
  - High resolution 1:24,000 used in most places

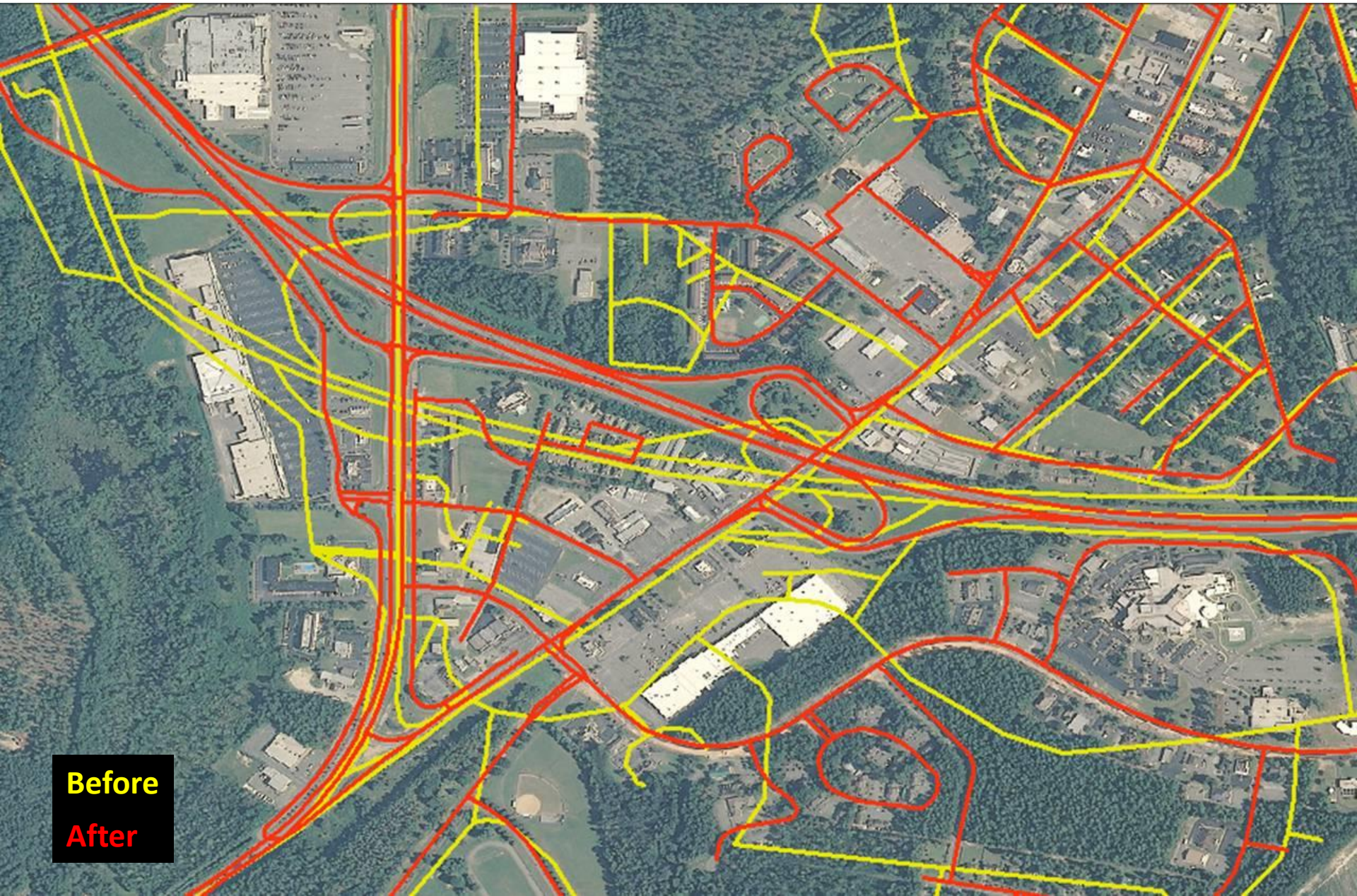
Implementing MTAIP

- If local sources did not meet spatial accuracy and no other source was available
  - Spatial enhancement of local file using imagery
- If local sources were not available
  - Extracted features from imagery
    - Limitation was that no attribute information was collected
  - Contractor drove every road in the county collecting a GPS trail
    - Attributes collected

MTAIP Highlights

- 3047 counties processed through Harris for MTAIP
- A total of 99,302,399 segments were returned by the contractor:
  - 73% of the segments represented existing features
  - 27% of the segments represented new features:
    - 7,794,002 were road features;
    - 115,602 were rail features
    - 11,901 were miscellaneous transportation features
    - 188 were other physical features
    - 77 were non visible features
    - 18,853,830 were hydrographic features
    - 16,005 were walkways.

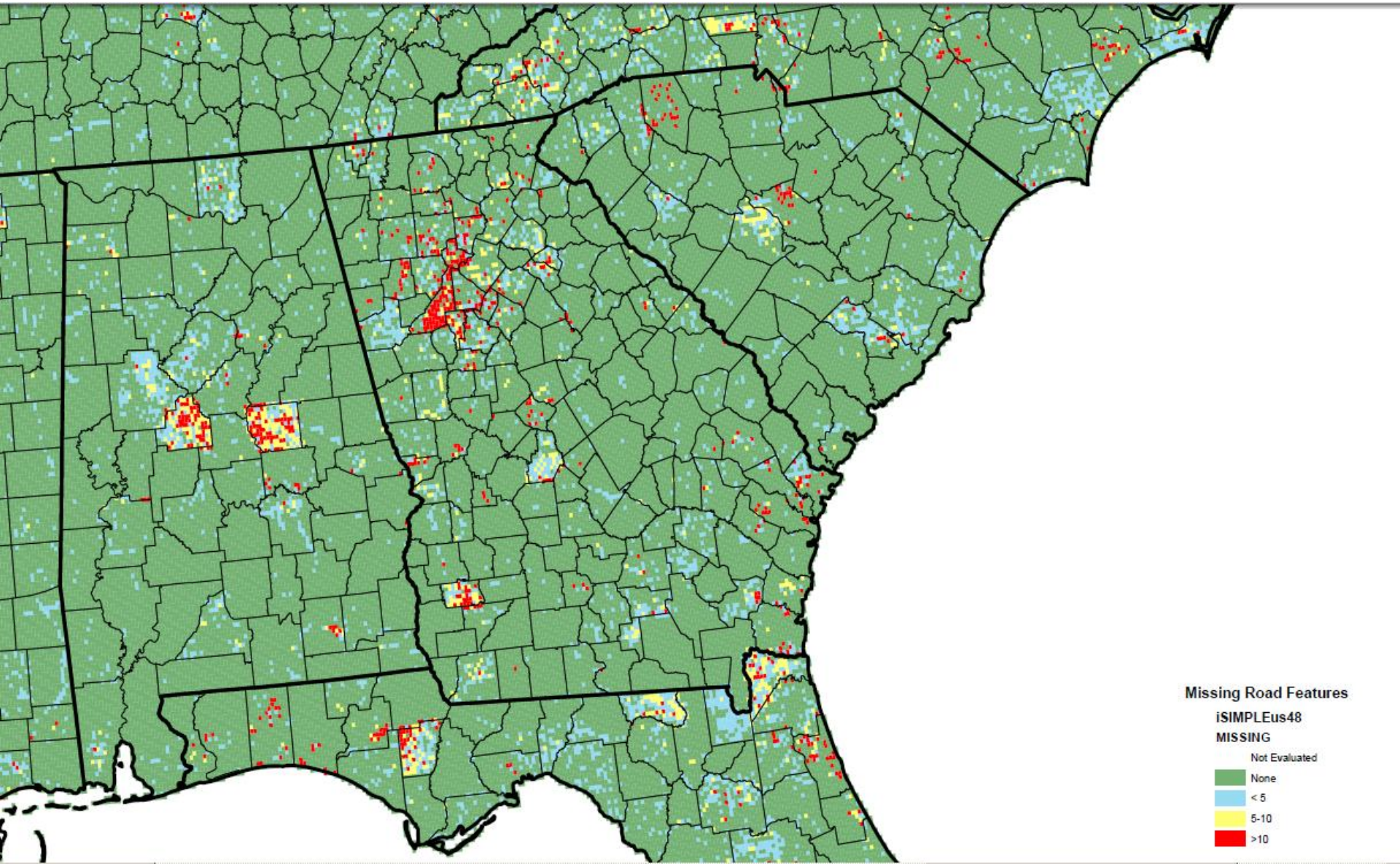
Before & After



Systematic Analysis of MAF/TIGER

- Evaluation of road features in TIGER
- Is TIGER consistent with imagery?
- 852,090 grid cells reviewed
  - 94% had NO missing features
  - 5% had 4 or less missing features
  - 70% had NO misaligned features
  - 26% had 4 or less misaligned features
- First web service based review
- Research will assist with targeting efforts

Missing Road Features



Summary

- Clearly specify your needs before implementing a census geography program.
- Continually update and maintain your statistical geography during the intercensal period.
- Use existing data sources wherever possible to avoid redundant effort and save time and money.
- Use new technologies to save resources and improve data quality.
- Thoroughly evaluate new technology before rushing to implement.