Distribution of Quality Control to the Point of Data Collection in the Field: Impacts on Cost, User Experience, and Security

Andrea Johnson (U.S. Census Bureau)
Jonathan Krentel (Gunnison Consulting Group, Inc.)

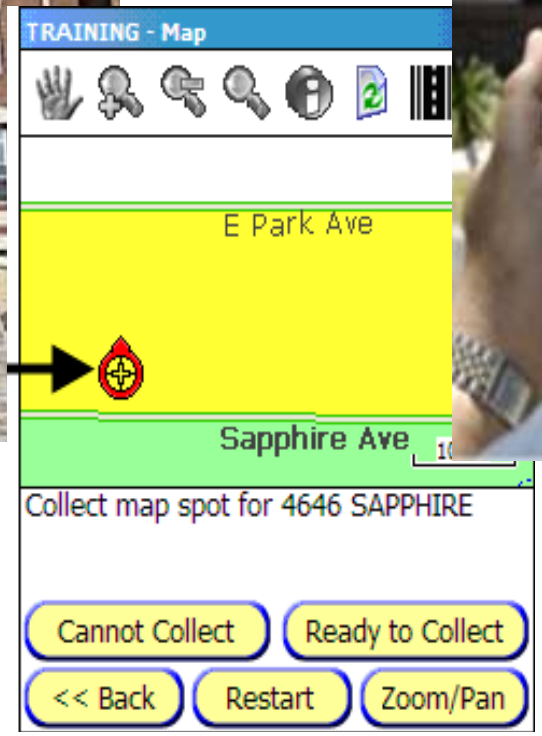
United States Census Bureau
U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU

Why Collect Housing Unit Locations?

- Accurate housing unit location matters

2010 Address Canvassing Map Spot Collection

1) Locate Housing Unit



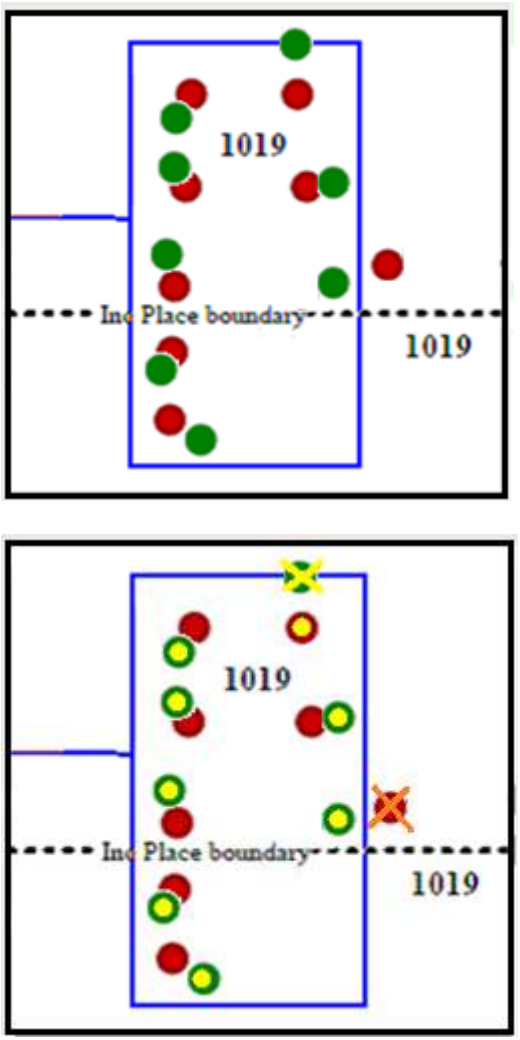
2) Locate GPS "you are here" indicator on map

3) Tap on map

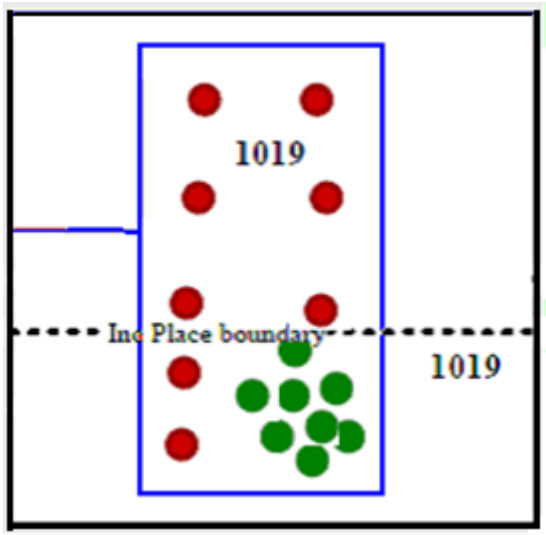
4) Confirm location of map spot (from tap)

Map Spot Collection Results

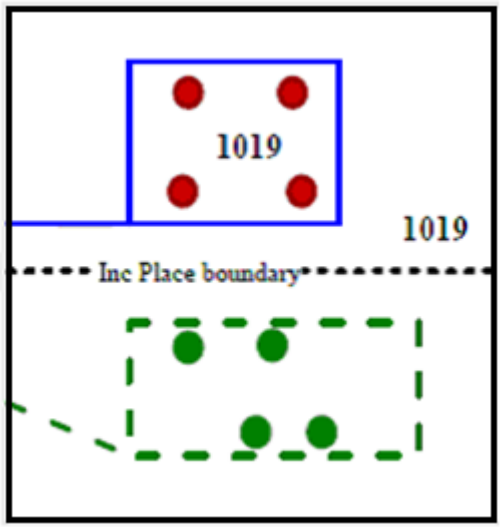
- 106,000,000 Map Spots collected
- For every map spot collected, two coordinate pairs were returned to the Census:
  - Manual
  - GPS
  - Only one point is “preferred”
- Three primary outcomes:
  - GPS = Manual
  - GPS location correct but manual location wrong
  - Manual location correct, GPS is wrong



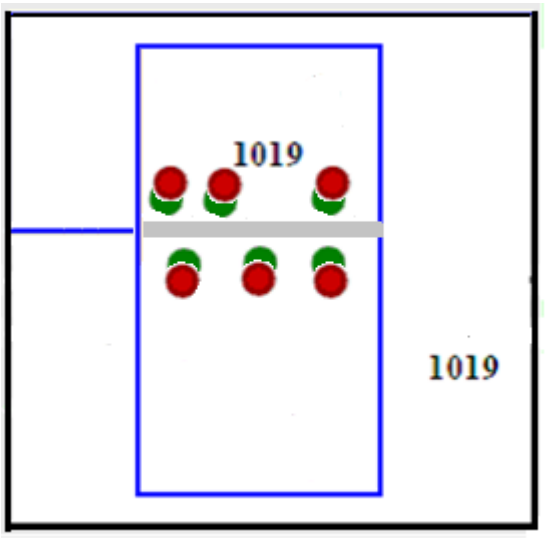
Map Spot Collection Issues that Required Resolution



Curbstone Clusters
~ Workload: 210,000

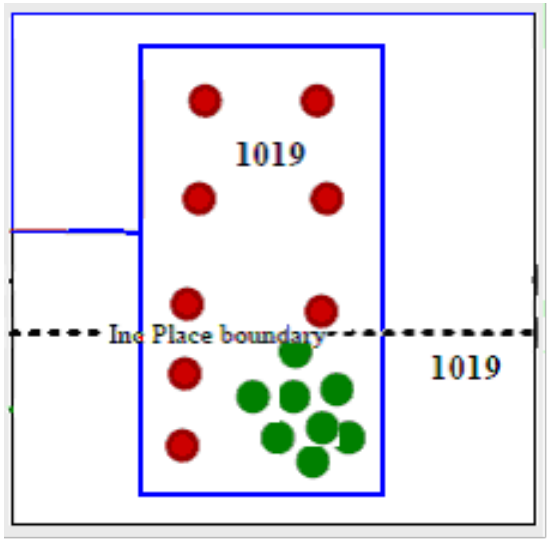


Map Spot Reconciliation
1,500,000

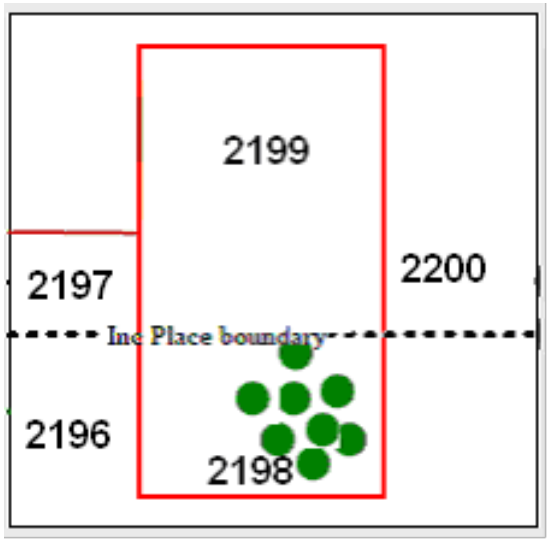


Feature Name Reconciliation
1,150,000

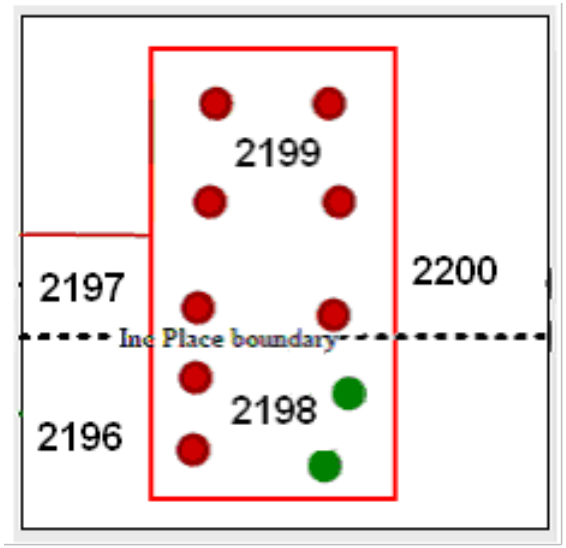
Curbstone Clusters Resolution



A) After Collection



B) If GPS Accepted and No Manual Review

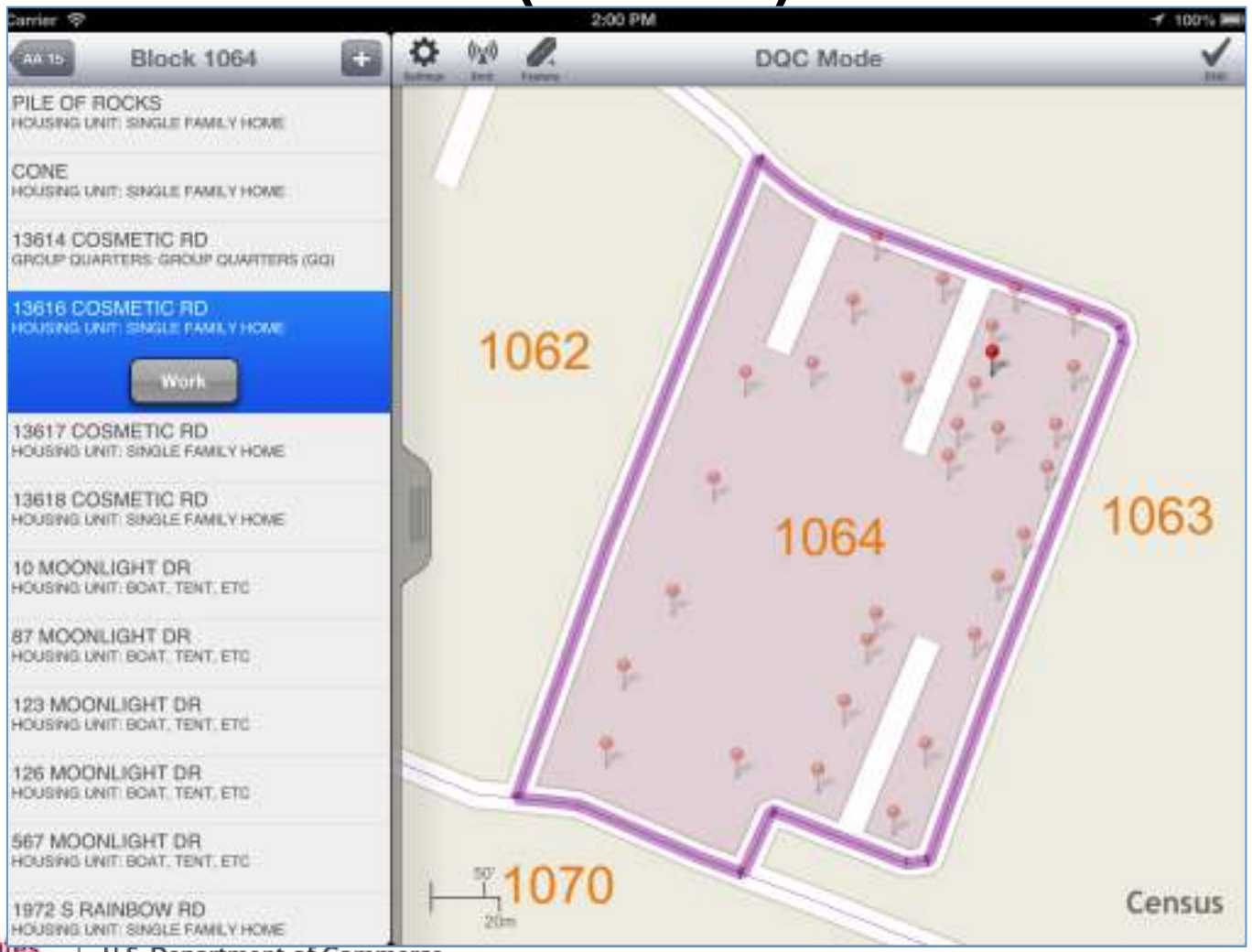


A) After Manual Review

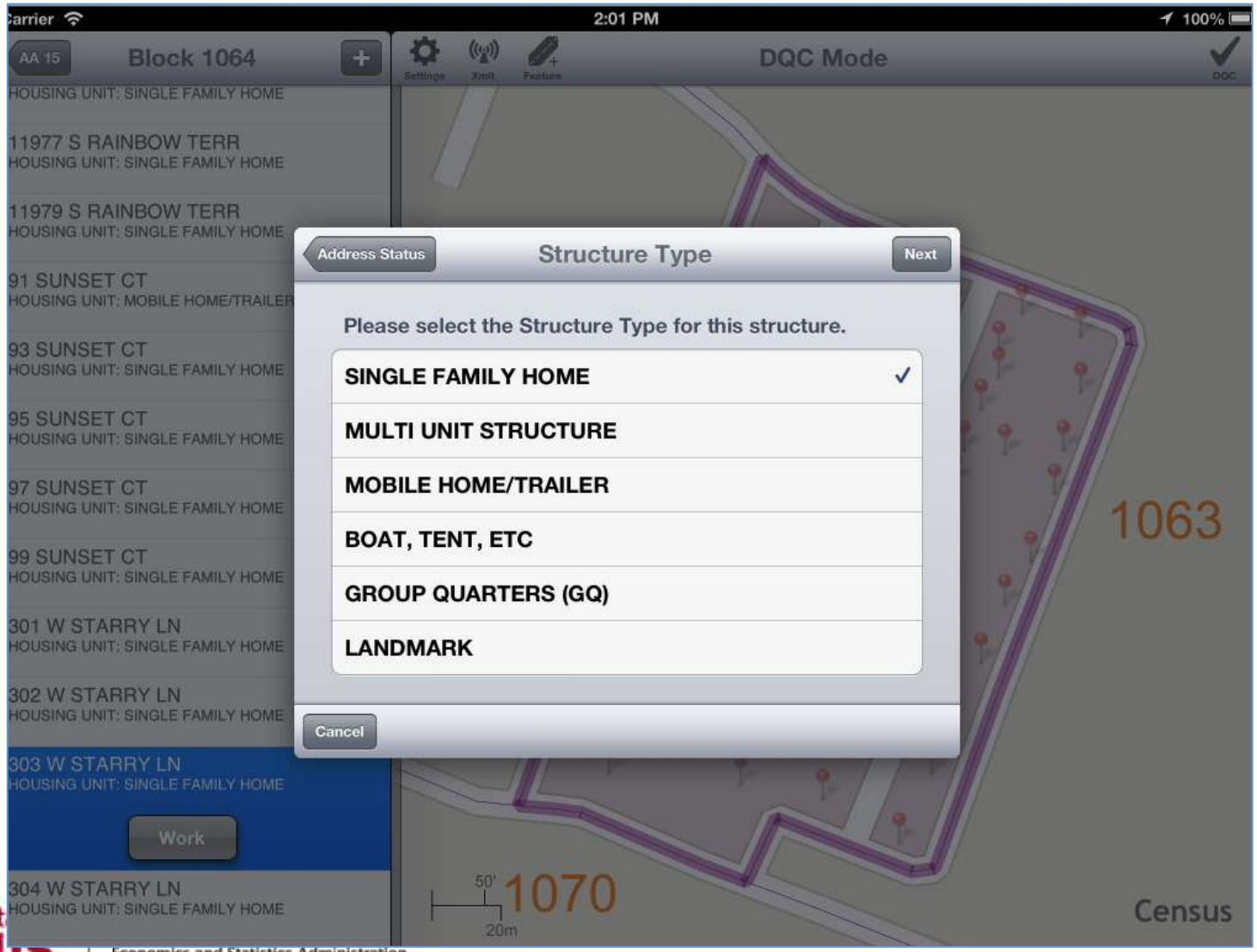
High Cost of Manual Resolution

2009 Curbstone Cluster Project Staff Resource			
Task	**People**	**Job Code**	**Total Hours**
Develop Procedures	1	Geographer	120
Conduct Training	**4**	Geographers	128
Monitor Analysis	4	Geographers	128
Resolve Referrals	3	Geographer	384
Training	27	Clerical	864
Peform Analysis	27	Clerical	8208
Conduct Quality Control	4	Supervisory Clerica	384
Misc IT (Upload, Systems Dev	3	IT Specialists	180
Unique People	**~34**		**10,396**

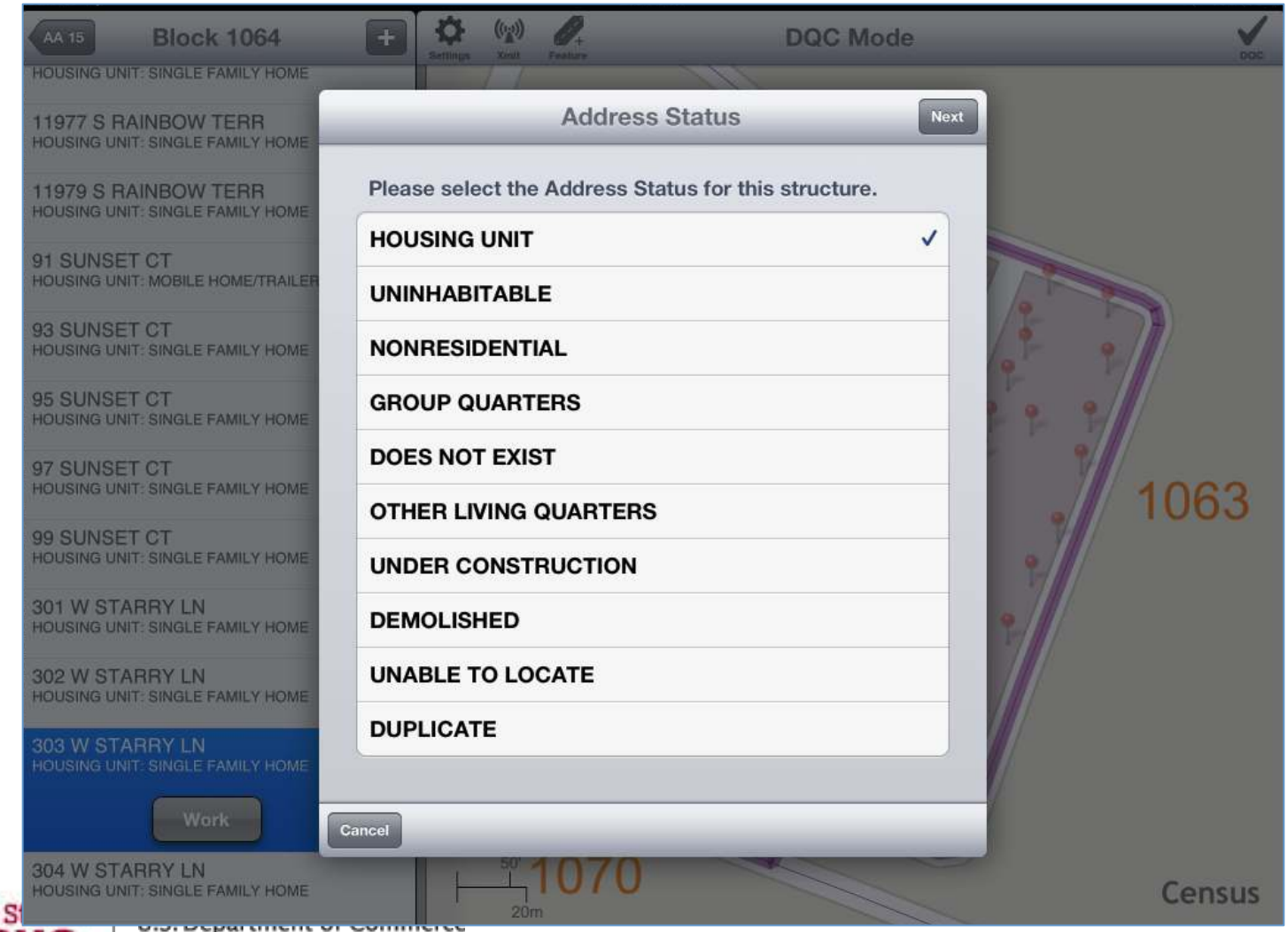
Listing & Mapping Application (LiMA)



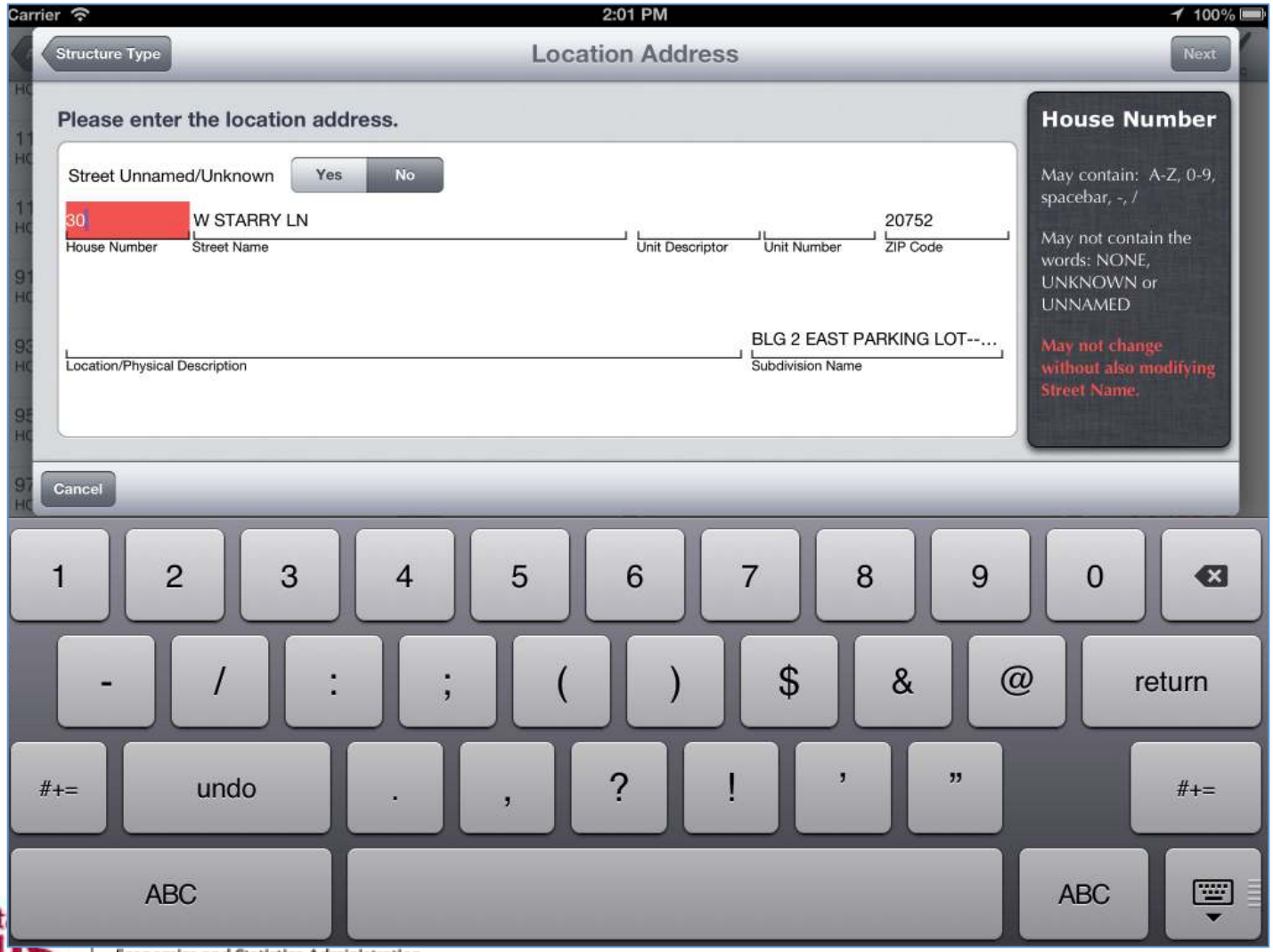
LiMA: Structure Type



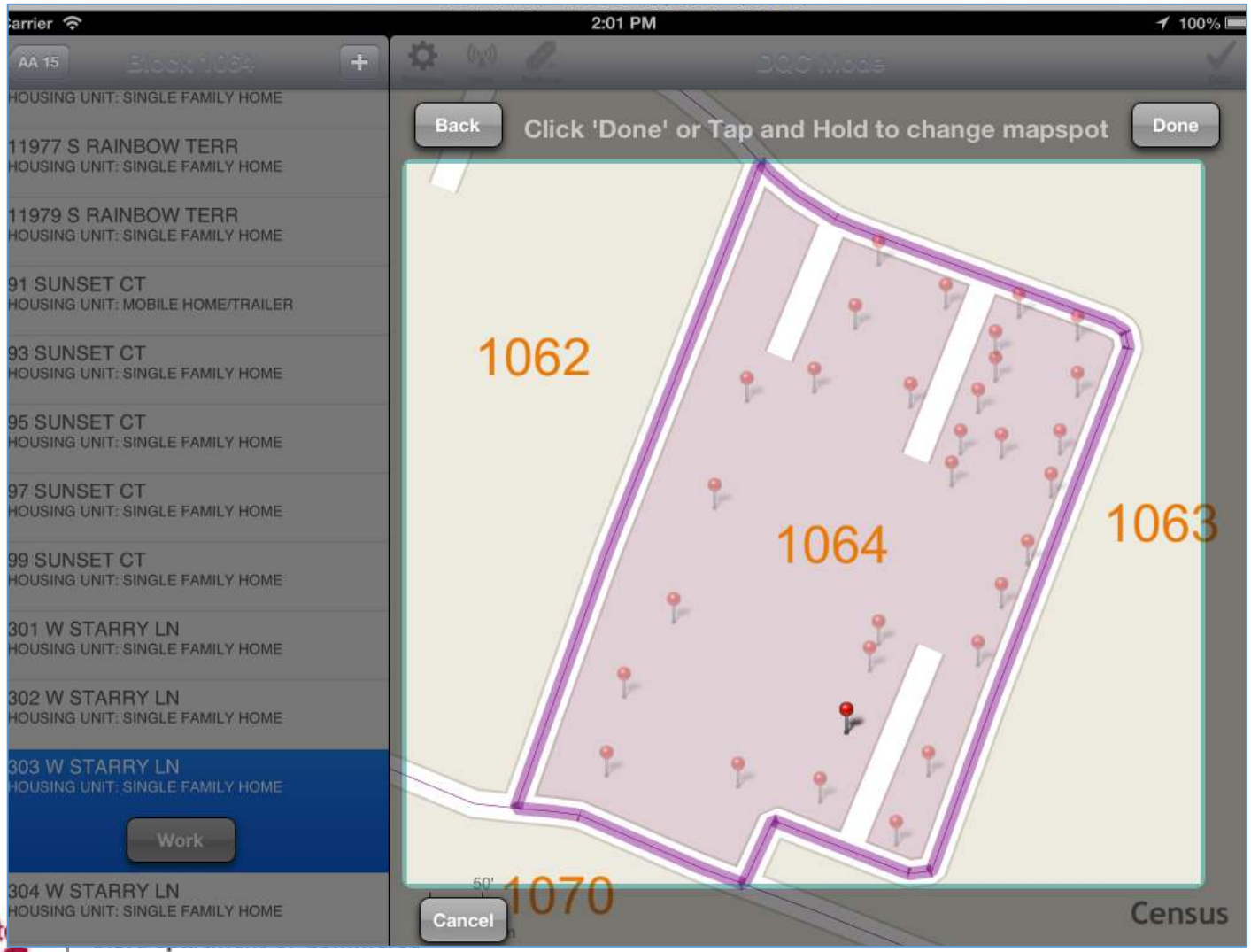
LiMA: Address Status



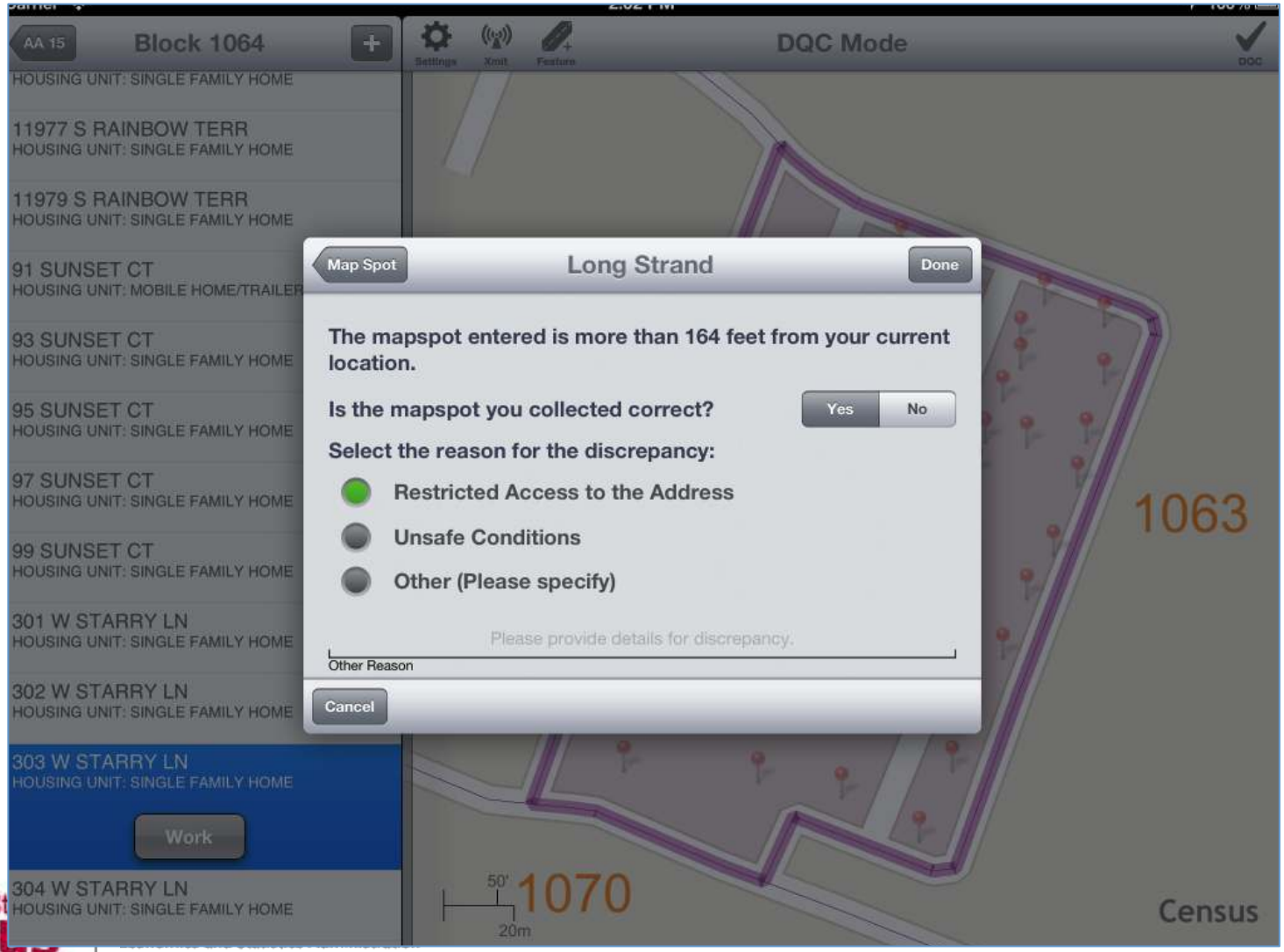
LiMA: Location Address



LiMA: Map-spotting

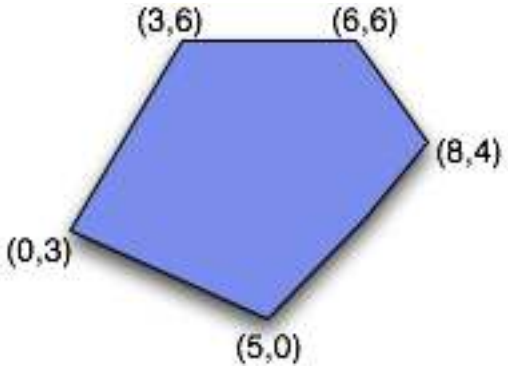


LiMA: Long Strand

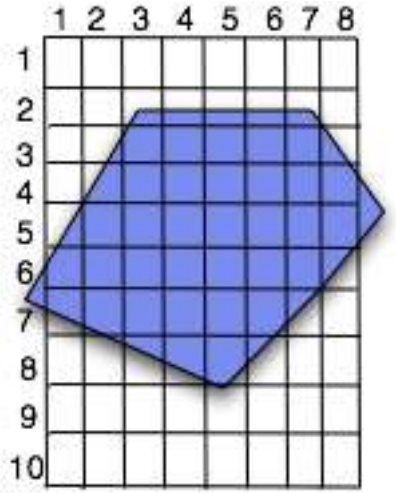


Mapping Considerations

Locally rendered vector-based maps vs. tile-based maps

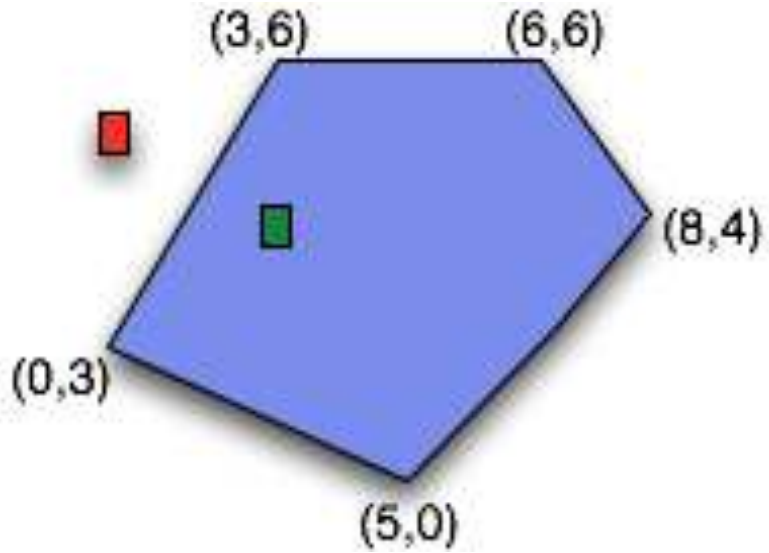


vs.



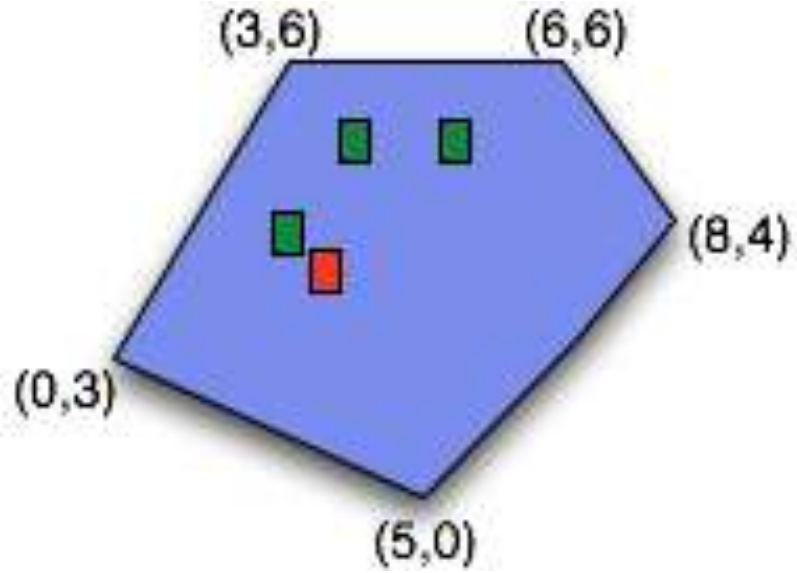
Benefits of vector-based maps (1)

Quality = Spatial-awareness



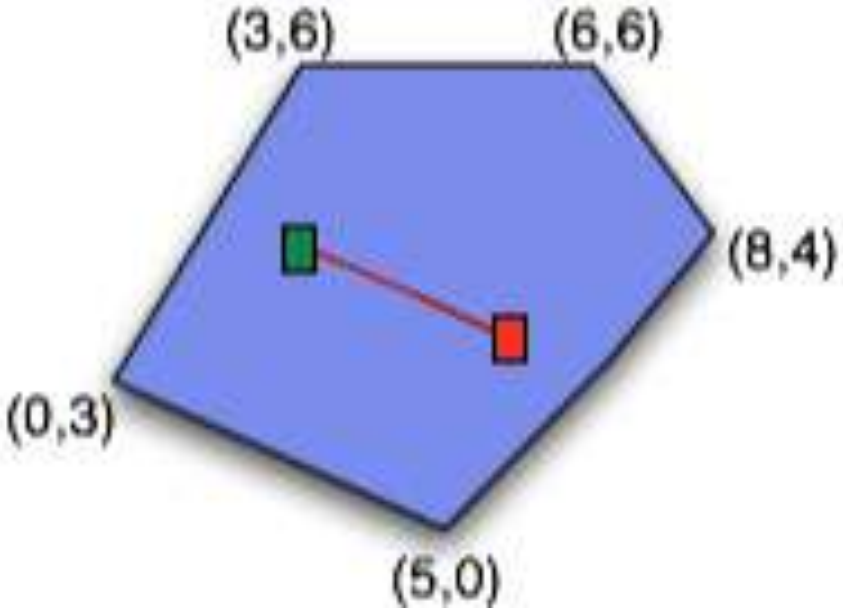
Benefits of vector-based maps (2)

Quality = Real-time detection of clustering



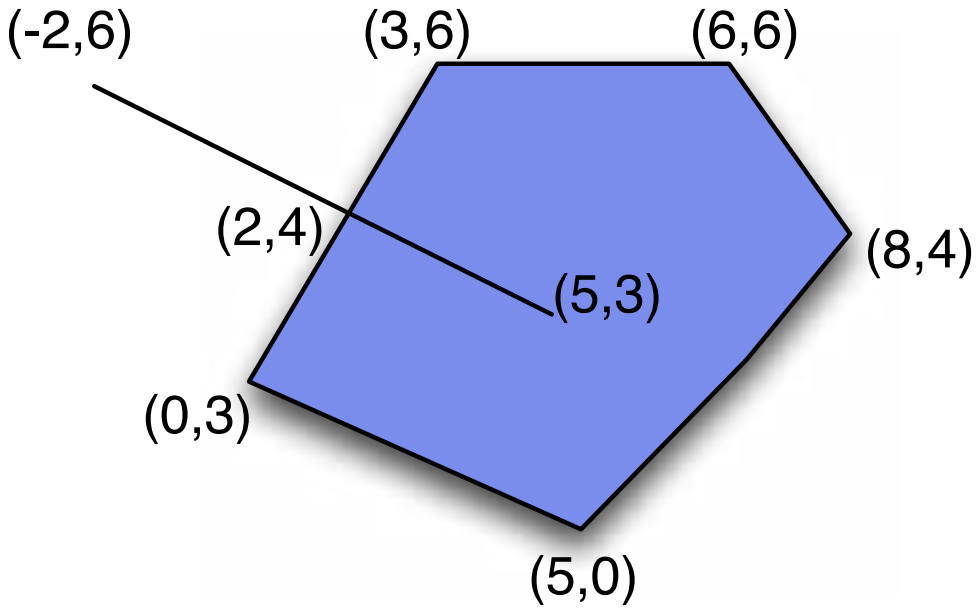
Benefits of vector-based maps (3)

Quality = Real-time long strand detection



Benefits of vector-based mapping (4)

Quality = Maintain topological integrity when adding or updating features



“Thick Client” Technical Approach

- Benefits:
  - Rich data on the device
    - All objects on the map are programmatically addressable
    - Allows for locally executed Quality Control
    - Potentially eliminates the need for expensive post-processing operations
    - Allows for QC to take place by the person in the best position to make corrections
  - Enables native user experience
  - Mitigates connectivity risk
- Risks:
  - Multiplication of code bases = BYOD risk?
  - Data on the device = Security risk?
  - Heavy processing load on the device = Performance risk?