The 2020 U.S. Census: A Time for Change

Tim Trainor

U.S. Census Bureau

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

Trends

- Adaptive design
- Mobile technologies and increased automation in the field
- Big data / paradata
- Focus on addresses for survey frames

Background

Planning for the 2020 U.S. Census

- Contain costs
  - Design and conduct a census that costs less per housing unit than the 2010 Census while maintaining high quality
  - Identify cost drivers and implement innovative enumeration methods aimed at reducing these costs
- Plan based on research and testing
  - Focus early research and testing program on major innovations to the design of the census oriented around major cost drivers of the 2010 Census

Census 2020 Objectives

- Contain costs
- Increased use of addresses
  - A redesigned address canvassing operation
- Optimize self-response program
  - Increase self-response options
  - Make use of electronic contact strategies and methods
- Maximize internet response
  - Increase awareness of the internet option
  - Encourage respondents to respond via the internet
- Continue small area geographies for data users

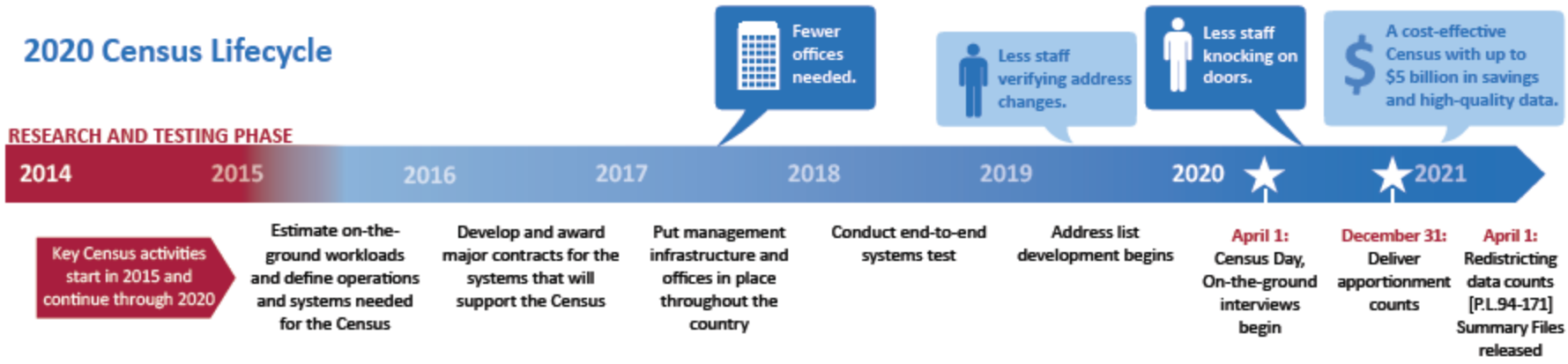
Decennial Census Cost Drivers

- Need for nationwide updating of address list prior to Census
- Diversity of the population
- Demand for improved count accuracy
- Declining response rates
- Management of major acquisitions, schedule, and budget
- Field Infrastructure

Decennial Census Research Relative to Cost-Drivers

- Redesigned Address Canvassing Operation
- Administrative and Commercial Records
- Use of Mobile Technologies
- Streamlining and Automating Field Management and Operations
- Optimizing Self Response

Key Milestones
Steps Towards 2020 Census



Adaptive design

Adaptive Design

- A data collection is adaptive to the extent that it:
  - Plans fieldwork to achieve cost and quality goals
  - Monitors process data and cost and quality indicators
  - Uses auxiliary frame data to tailor contact approaches (or impute or adjust)
  - Uses auxiliary data, paradata and response data to change contact approaches rapidly
  - Strikes data-based cost/quality tradeoffs

Adaptation is NOT New

- Sub-sampling non-respondents
- Increasing contacts
- Timing contacts
- Increasing incentives
- Tailoring survey invitations
- Tailoring refusal letters
- Switching modes

Some Adaptations ARE New

- More centralized, less ad hoc, more timely efforts, e.g.
- Using auxiliary data to tailor contacts
- Using auxiliary data, paradata and response data to alter contacts
- Switching modes based on auxiliary data, paradata and response data
- Motivated by a plan and enabled by new systems

Optimizing Self-Response

- Internet data collection
- Adaptive contact strategies
- New contact modes
  - Telephone
  - E-mail

Mobile Technologies and Increased Automation in the Field

Major Changes for Field Operations

- Using automation to support processes
  - Optimized daily enumerator assignments of respondent contact attempts
  - Near real time operations information for decision making
  - Enhanced operational control system
  - Automated training for enumerators and managers
- New field structure, including field staff roles and staffing ratios

Mobile Technologies

- Routing
- Navigation
- Data Collection

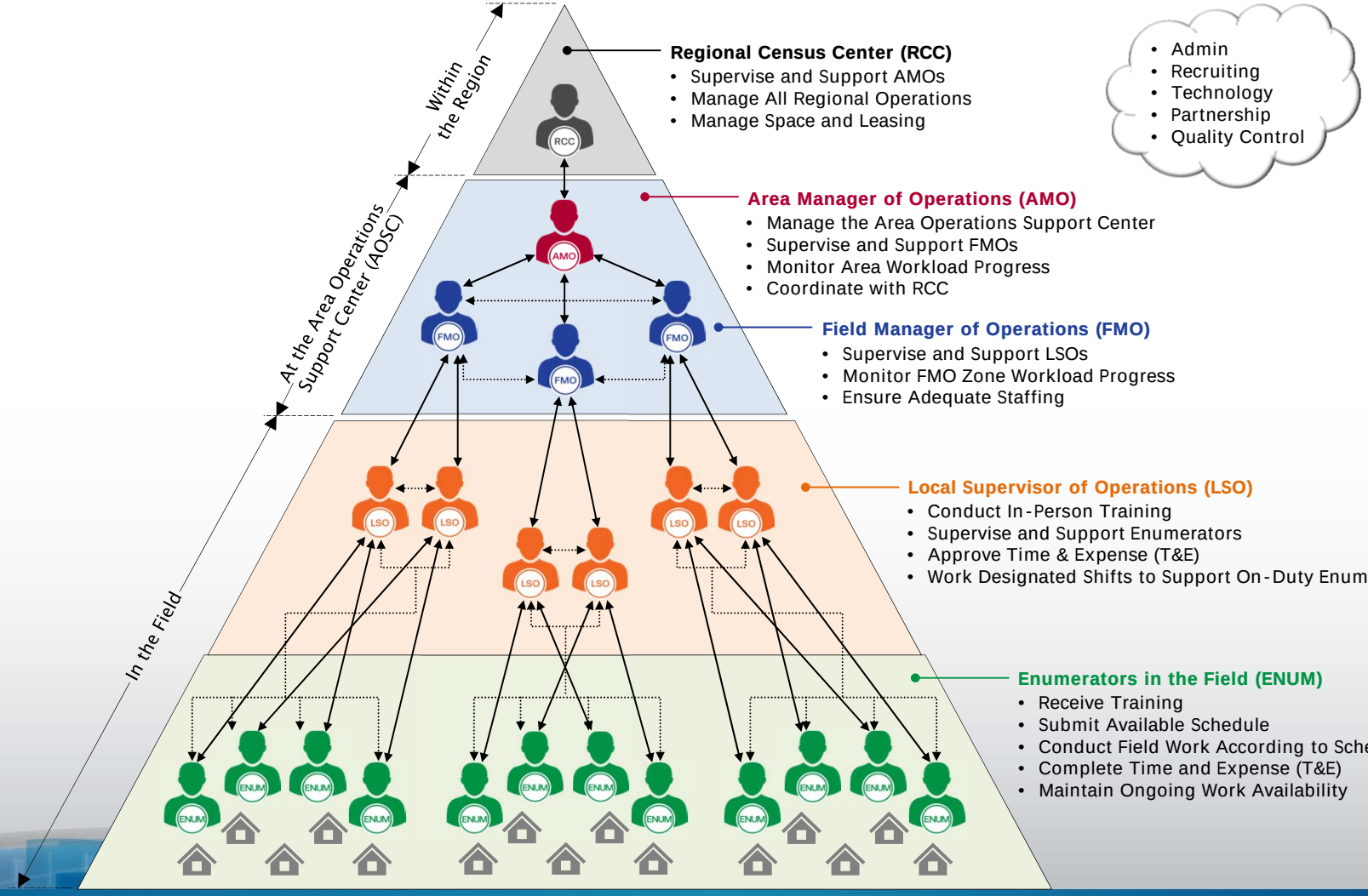
Field Reengineering and Nonresponse Followup (NRFU) using Administrative Records and Adaptive Design

- Reengineer the roles, responsibilities, and infrastructure for the field
- Evaluate the feasibility of fully utilizing the advantages of technology, automation, and real-time data to transform the efficiency and effectiveness of data collection operations
  - Move to automated training for enumerators and managers
  - Test and implement routing and/or navigation
  - Reengineer the approach to case management

Field Reengineering and NRFU using Administrative Records and Adaptive Design (cont.)

- Reduce NRFU workload and increase NRFU productivity with:
  - Administrative Records
  - Reduce cases that need to be resolved in NRFU by varying type of cases removed and timing of case removal from the workload
  - Reduce the number of contact attempts to cases resolved in NRFU
- Field Reengineering and Adaptive Design
  - Reduce the number of contact attempts
  - Leverage dynamic case management with route planning and other methodologies to improve enumerator productivity through automation
- Planned for an April 1 Census Day

Field Organizational Structure



Regional Census Center (RCC)
- Supervise and Support AMOs
- Manage All Regional Operations
- Manage Space and Leasing

- Admin
- Recruiting
- Technology
- Partnership
- Quality Control

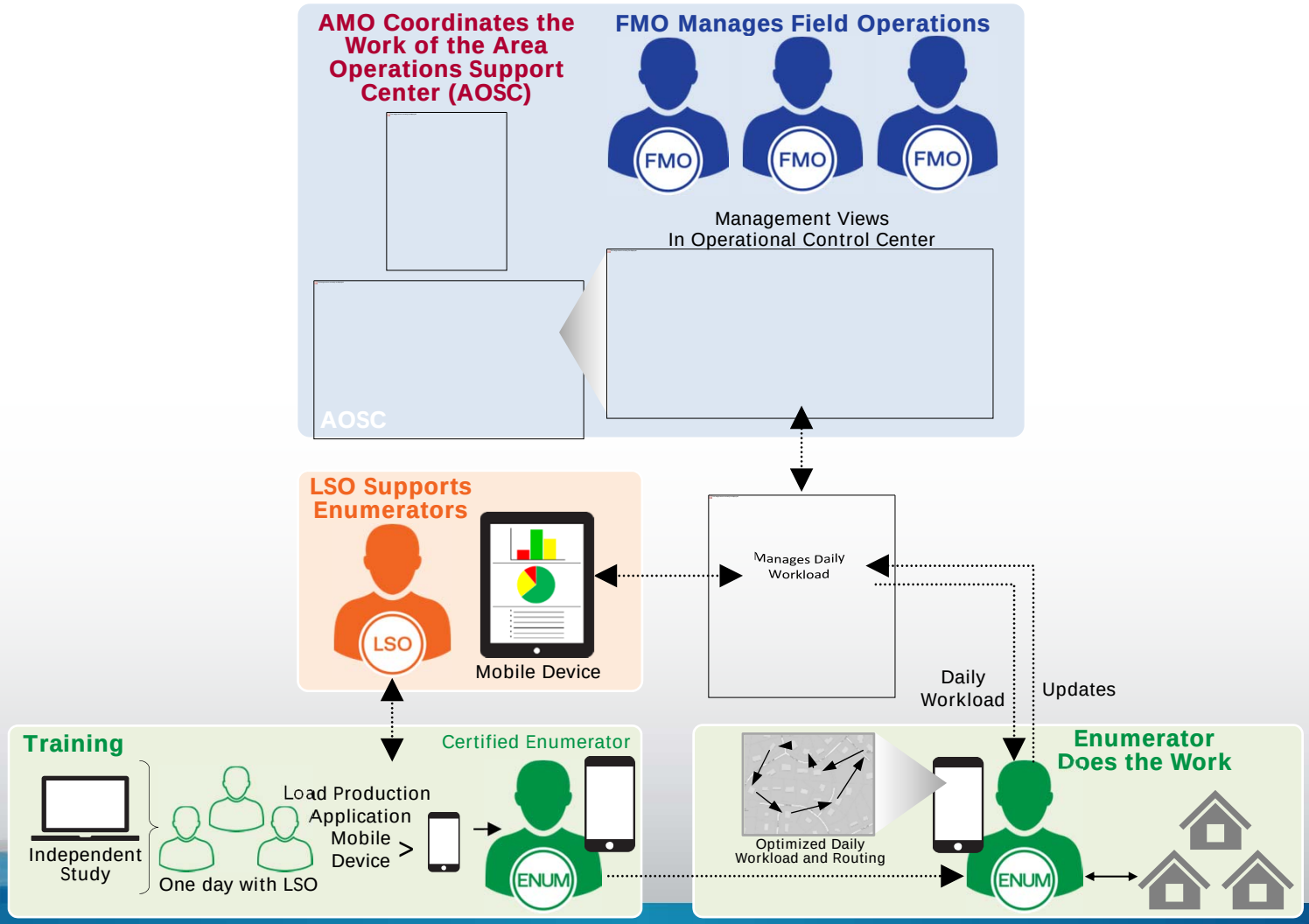
Area Manager of Operations (AMO)
- Manage the Area Operations Support Center
- Supervise and Support FMOs
- Monitor Area Workload Progress
- Coordinate with RCC

Field Manager of Operations (FMO)
- Supervise and Support LSOs
- Monitor FMO Zone Workload Progress
- Ensure Adequate Staffing

Local Supervisor of Operations (LSO)
- Conduct In-Person Training
- Supervise and Support Enumerators
- Approve Time & Expense (T&E)
- Work Designated Shifts to Support On-Duty Enumerators

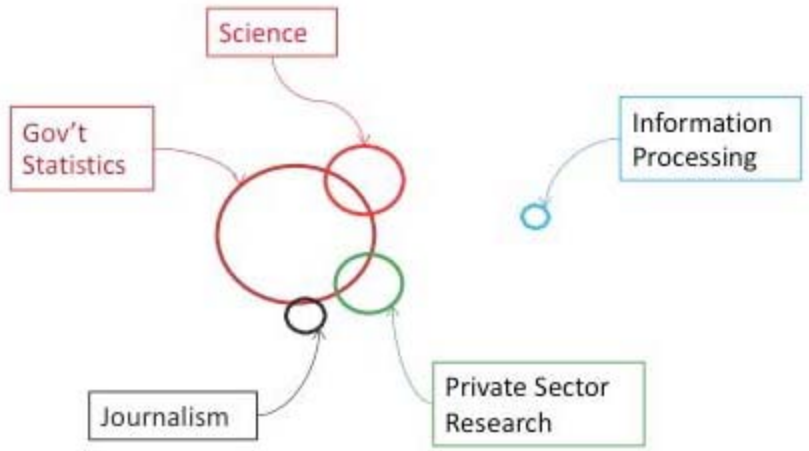
Enumerators in the Field (ENUM)
- Receive Training
- Submit Available Schedule
- Conduct Field Work According to Schedule
- Complete Time and Expense (T&E)
- Maintain Ongoing Work Availability

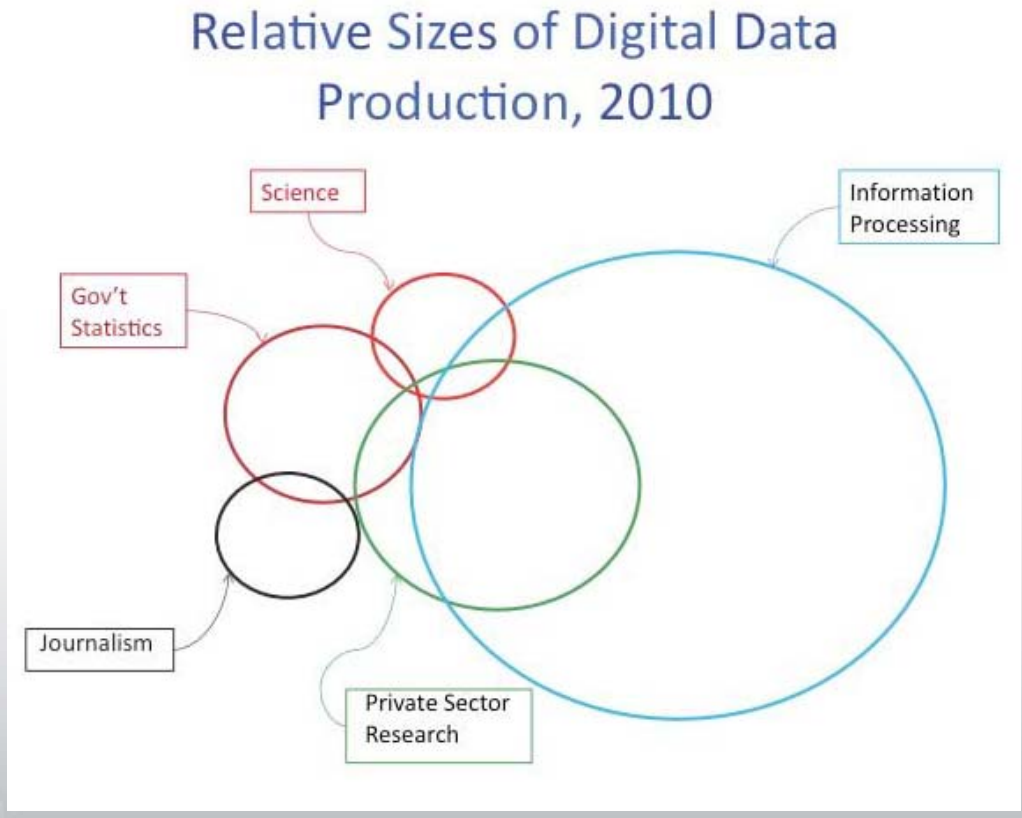
Concept of Operations



Big Data

Big Data


Relative Sizes of Digital Data Production, c.1960
Science
Gov't Statistics
Information Processing
Journalism
Private Sector Research


Relative Sizes of Digital Data Production, 2010
Science
Information Processing
Gov't Statistics
Journalism
Private Sector Research

Big Data Research

- Administrative records to improve cost and increase timeliness and accuracy
  - Quality control
  - Coverage improvement
  - Substitute for in-person visits to households that do not self respond
- Processing techniques to allow real time decision making
  - Adaptive design
  - Self response options
- Data dissemination via API's to allow creation of apps and products that combine our data with other external data sets
  - Census explorer data visualization
  - Other apps from our web site
  - More work required in this area to stimulate interest

Big Data: Concerns

- There are no currently acceptable processes or procedures for using Big Data to produce Official Statistics
  - Don’t even have a common definition of Big Data

Focus on Addresses for Survey Frames

The GSS Initiative (GSS-I)

- An integrated program of improved address coverage, continual spatial feature updates, and enhanced quality assessment and measurement
  - All activities contribute to MAF/TIGER Database improvement
  - Builds on the accomplishments of last decade's MAF/TIGER Enhancement Program (MTEP)
- Supports the goal of a **redesigned address canvassing** for the 2020 Census
- Continual updates throughout the decade support current surveys

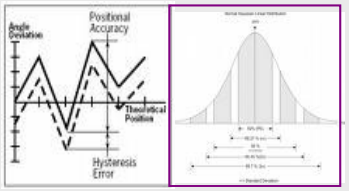
Address Updates

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



Redesigned Address Canvassing

General Questions:

- Is a traditional, on-the-ground canvassing operation necessary to ensure a complete and accurate address list for the decennial census?
- Are there areas of the country in which the address list and locational information can be kept current without canvassing?
- What characteristics identify an area that should be included in a traditional canvassing?

Research Goals

- Develop statistical models to identify geographic areas to be canvassed or not canvassed
  - Predict adds and deletes with estimated coverage error
- Interactive Review - Identify and classify areas
  - In which the number of addresses/housing units is stable and unlikely to change
  - With unique housing/addressing/mail delivery situations that may require canvassing
  - Land use/land cover is entirely non-residential
  - Where the address list can be updated and assured through administrative or operational methods

Address Canvassing Research, Model, and Area Classification

- 2009 Statistical Model
- 2013 Statistical Model
- Interactive Review
  - 27 test counties

MAF Error Model Objective

- The objective of the MEM project is to provide statistical models for the MAF that will produce estimates of coverage error at levels of geography down to the block level
  - These models could potentially inform Address Canvassing decisions

What is the MAF Error Model?

- Two predictive models developed at the block level, collectively known as the "MAF Error Model"
  - One model for the number of **adds** and one model for the number of **deletes** as functions of identified predictors
- Zero-inflated (ZI) regression models
  - Zero-inflated models can provide a model-based approach to obtaining coverage estimates
    - Provides more granularity at lower levels of geography over other common modeling approaches (e.g., logistic regression)

Address Canvassing: Master Address File (MAF) Model Validation Test and Focused Field Address Resolution Approach

Model Based Approaches

- Test our ability to use statistical modeling to measure error in the MAF and to identify areas experiencing significant change
- Inform the performance of the models used to define the Address Canvassing workloads

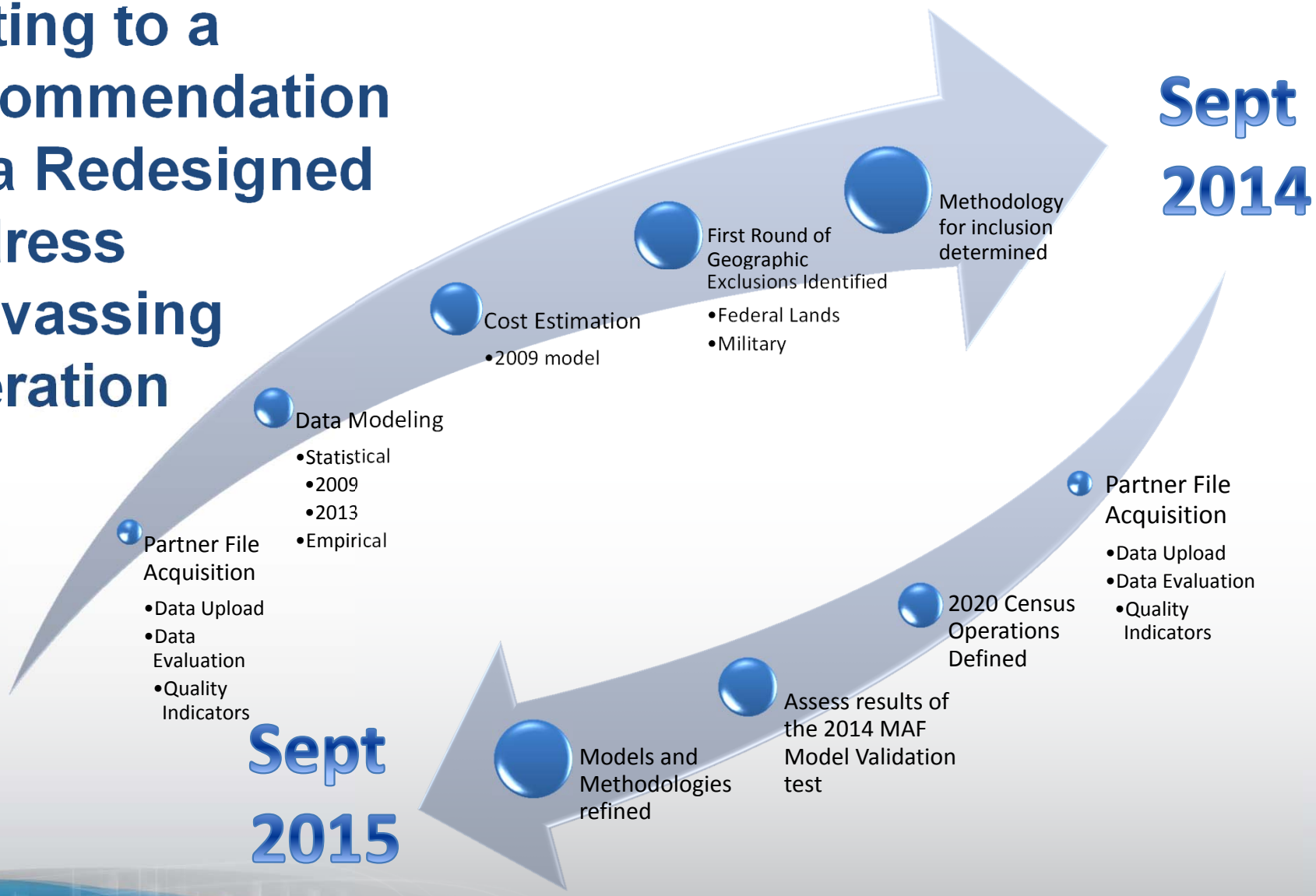
Focused Field Address Resolution ("micro-targeting") Approach

- Incorporate imagery reviews to detect changes and discrepancies
- Include field updating of addresses for portions of blocks

MAF Model Validation Test Objectives

- The purpose of the MAF Model Validation Test (MMVT) is to collect data to inform components of the Address Canvassing decision-points
  - MAF Error Model
  - Address Canvassing, Research, Model, and Classification team
  - Models for Zero Living Quarters blocks
- Test the concept of Micro-Targeting and uses of imagery

Getting to a Recommendation for a Redesigned Address Canvassing Operation



Address Canvassing Methodology Plan

Preliminary Federal Land Use and similar types of blocks

2009 TEA* Operational Overlay - Remove non-MO/MB areas (UL, UE…)

2009 Statistical Models (2020 and GSS) - Use only data available in 2009

- Preliminary Cost Estimation
- Jan 2014 - March 2014

Federal Land Use and
2009 TEA Operational Overlay

Preliminary Interactive Review 4/14
Use Aerial Imagery to add/remove blocks

2013 Statistical Models 4/14
Use only data available in 2013

- Cost Estimation
- Quality Metrics (MMVT)
- **LCAT**

Process definition occurs here and will be repeated

GEO "go/no go" Recommendation 9/14
Field Infra Decision Point 1/15

***LCAT** will examine costs on later operations and provide feedback to modify models*

MAF Model Validation Test 9/14-12/14
Data available on January 2015

- Observe and measure the performance of the models
- Update the models with more current field data (5 yr. field update)

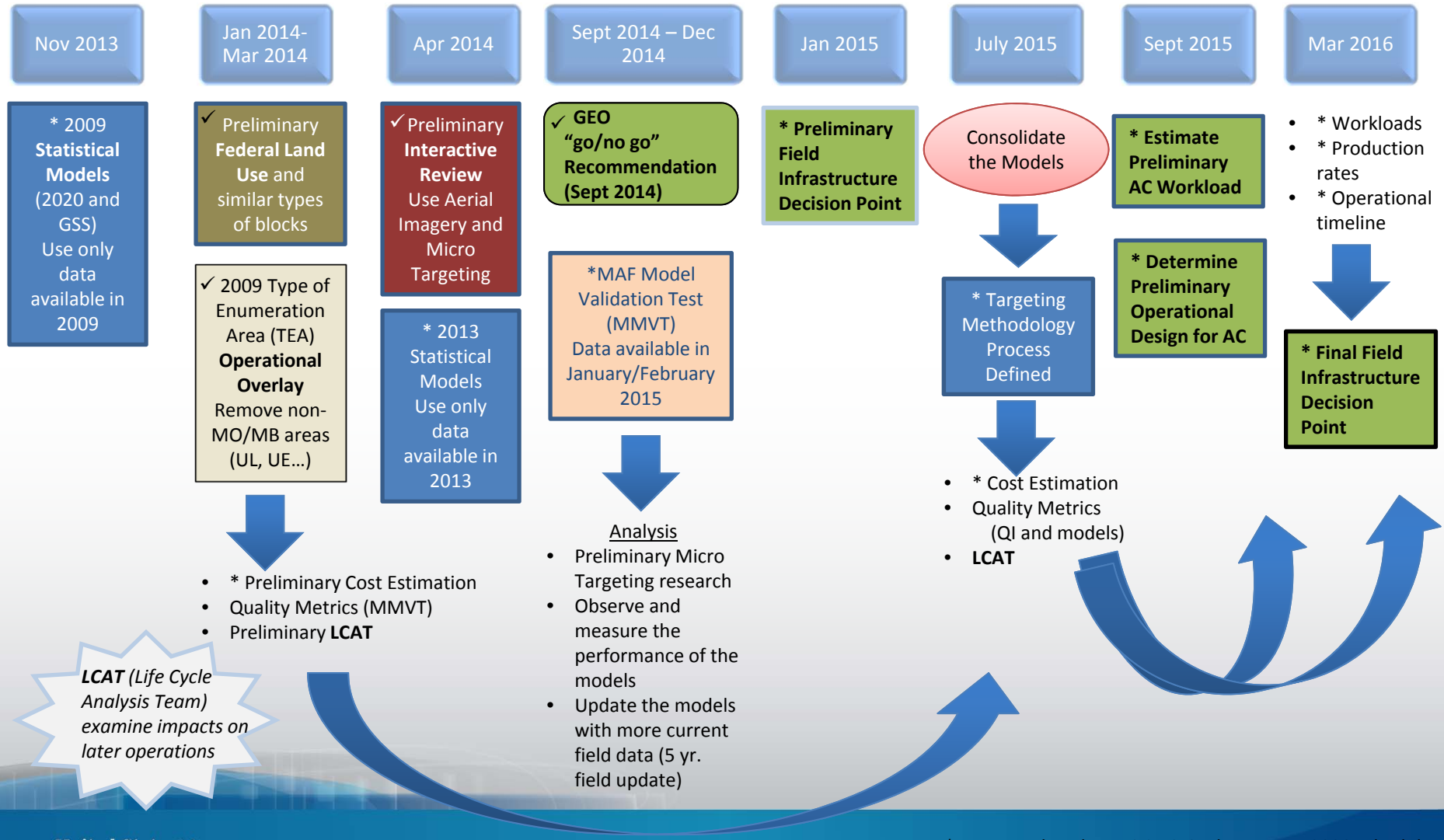
Consolidate the Models

2015 Methodology 3/15
Process Defined

- Cost Estimation
- Quality Metrics (QI and models)
- **LCAT**

- Recommendation for Integration 9/15
- Field Infrastructure Decision Point 1/16

Frame Schedule



* Denotes that the activity is in the current 2020 schedule
✓ Denotes GEO/GSS activity

Summary

- A redesigned census
- Traditional approaches are challenged
- Adds risk
- Modernization is critical
- All comes down to cost

Questions?