
United States Census 2020

2018 End-to-End Census Test Non-ID Processing Operational Assessment Study Plan

Non-ID Processing Integrated Project Team
(IPT)

Draft Pending Final Census Bureau Executive Review and Clearance.

Table of Contents

- I. Introduction.....1
- II. Background.....1
- III. Assumptions.....2
- IV. Scope of Assessment Content and Questions To Be Answered.....3
- V. Methodology.....5
- VI. Table Headers.....6
- VII. Risks/Limitations.....7
- VIII. Measures of Success.....7
- IX. Division Responsibilities.....8
- X. Review/Approval Table.....9
- XI. Document Revision and Version Control History.....9
- XII. Glossary of Acronyms.....9
- XIII. References.....10

I. Introduction

The 2018 End-to-End Census Test is an important opportunity for the Census Bureau to ensure an accurate count of the nation's increasingly diverse and rapidly growing population. It is the first opportunity to apply much of what has been learned from census tests conducted throughout the decade in preparation for the nation's once-a-decade population census. The 2018 End-to-End Census Test will be held in Providence, Rhode Island.

The 2018 End-to-End Census Test will be a dress rehearsal for most of the 2020 Census operations, procedures, systems, and field infrastructure to ensure there is proper integration and conformance with functional and non-functional requirements. The test also will produce prototypes of geographic and data products. Note that the 2018 End-to-End Census Test results cannot be generalized to the entire United States. Non-ID Processing is one of the 2020 Census operations that will be a part of the 2018 End-to-End Census Test.

Non-ID Processing utilizes real-time automated processing to match each non-ID respondent-provided address to an existing Master Address File (MAF) record during internet self response and Census Questionnaire Assistance interviews, or assign census geographic codes (i.e., state, county and basic collection unit "geocodes") in the event of a nonmatch. This initial step is referred to as Real-Time Non-ID Processing (RTNP). However if a MAF match is not the outcome, the non-ID address passes to Asynchronous Processing. Here elements of the response data (respondent name, date of birth, telephone number, and address) are compared with administrative records (AR) to enhance the respondent-provided address. If AR enhancement corrects errors or supplies missing/omitted address data, another attempt is made to match to an existing MAF record. If this additional attempt does not result in a match to a MAF record, or the respondent address was not enhanced, it is sent to Clerical Processing. During Manual Matching and Geocoding, the first component of Clerical Processing, responses are interactively reviewed and if errors are corrected or missing/omitted data is obtained, a final attempt is made to match to an existing MAF record. If this final attempt to match is unsuccessful, there is an attempt to assign a geocode to the response. During Office-Based Address Verification (OBAV), the other component of Clerical Processing, an attempt is made to verify geocodes acquired during RTNP, Asynchronous Processing, or Manual Matching and Geocoding. The intent of this operation is to significantly reduce the number of cases that require fieldwork to verify the existence and location of nonmatched addresses from non-ID responses.

This study plan documents how Non-ID Processing will be assessed, as guided by questions to be answered.

II. Background

During Census 2000 and the 2010 Census, non-ID cases made up less than 1 percent of self-response cases. However, given the goal of optimizing self-response for the 2020 Census, specifically by allowing respondents to submit a questionnaire without a unique identification code, the Census Bureau expects the number of non-ID cases for the 2020

Census, thus non-ID workload, will increase substantially over 2010 Census levels. Accordingly, the Non-ID Processing team has continued researching ways to improve RTNP to increase the percentage of non-ID addresses that match to an existing MAF record. Also given the goal of utilizing AR and third-party data for the 2020 Census, the Non-ID Processing team has continued researching ways to improve Asynchronous Processing, specifically AR enhancement. Every match to an address record in the census universe equates to a reduction in cost, such as follow-up mailings to nonresponding households and efforts to manually match or verify the existence and location of nonmatching addresses. Further, every match to a MAF record with a confirmed geocode¹ means that no further work is required for the non-ID response to be considered complete. In other words, if the source of the geocode for the matching record meets certain business rules, it can be accepted without any further work to confirm it, such as a clerical operation or fieldwork to confirm both the existence and location of the living quarters.

III. Assumptions

The Census Bureau put in place the following overarching assumptions to guide all research and testing. These assumptions can be revised to prepare for 2020 Census design decisions:

- Self-Response in 2020
 - We will promote internet response as the contact strategy for self-response.
 - We will achieve internet response for 47 percent of households².
 - We will achieve a significant reduction in paper data capture operations and infrastructure.
- Preferred records³ for all MAF units in the MAF will be used by DITD in attempts to match.
- The results from Non-ID Processing are made available through the Universal Tracking System (UTS) by the data repository staff.
- Delivering the unmatched non-ID cases to the Production Environment for Administrative Records Staging, Integration, and Storage (PEARSIS) for processing will result in addresses that are more likely to yield a match when returned to DITD.

The number of non-ID responses can be impacted by many of these overarching assumptions, in a variety of ways. Some of the impacts might be the situations listed below.

- The strategies related to increasing self-response, including advertising campaigns promoting participation online and paper reminders left on doors by Nonresponse

¹ A “Confirmed geocode” in this document is defined as matched to a “In Census” Housing Unit, Transitory Unit, or Group Quarters

² 2020 Census Operational Plan

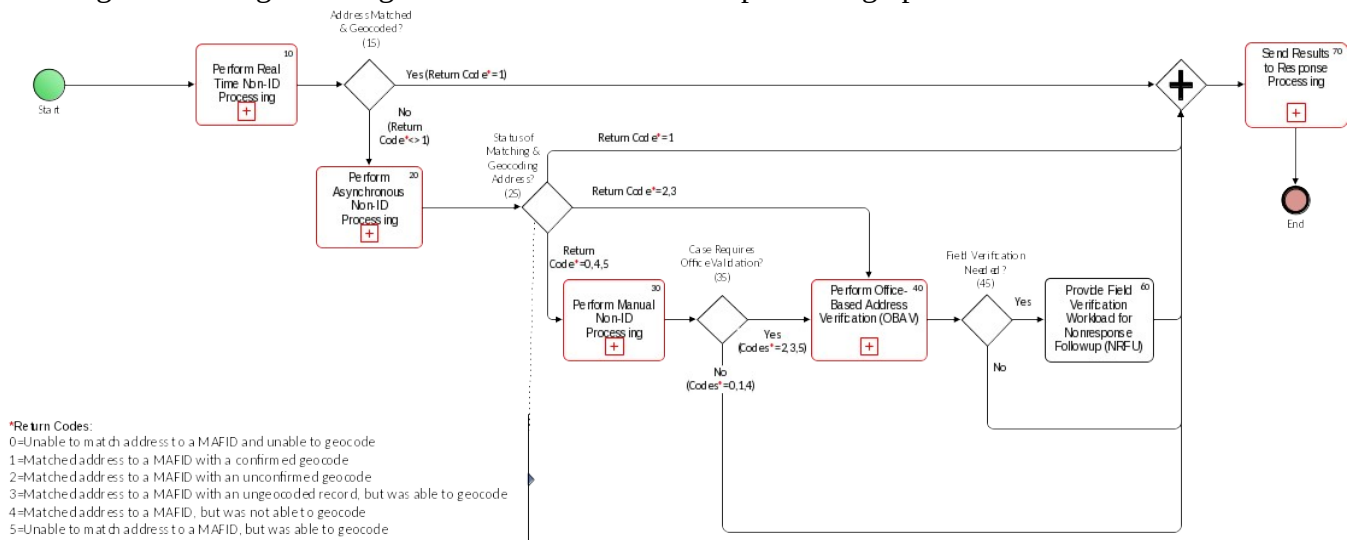
³ The selection of the preferred address(es) is based on current available information, scoring hierarchy of MAF source codes (ranking of address sources), and ongoing research conducted by the Geography Division. Multiple addresses can exist for a single MAF unit. To indicate which record is preferred for location and mailing purposes, MAF address record flags are used. A preferred record would have this flag set and a nonpreferred record would not.

Follow Up (NRFU) enumerators, could result in an increase in people going to the internet site to complete their questionnaires, whether or not they have or can locate their IDs. The absence of a paper questionnaire can also increase internet activity.

- Cases incorrectly identified as vacant in AR and removed from the NRFU workload could potentially result in late non-ID responses if the people remember that they have not yet responded and respond late.
- Various situations exist where a respondent may not respond with a Census-ID. For example, language barriers present in NRFU could lead to a respondent using the internet site to complete the questionnaire in their preferred language, or a respondent responding via cell phone without having their Census-ID readily available.

IV. Scope of Assessment Content and Questions To Be Answered

The diagram below gives a high level view of the non-ID processing operations workflow.



The Non-ID Processing 2018 E2E CT Assessment will evaluate the following aspects of the operation:

Production Rates

1. What was the resolution rate for the cases matched and resolved through RTNP?
2. How did the AR enhancement increase, if any, the matching and resolution rate?
3. How many cases were sent to Manual Matching and Geocoding?
4. For Manual Matching and Geocoding what was the production and matching rate for:
 - a. Matched and geocoded?
 - b. Matched and geocoded but requires manual processing through OBAV?
 - c. Unmatched and geocoded but requires manual processing through OBAV?
 - d. Unable to match or geocode an address in the MAF?
5. What was the production rate of Manual Matching and Geocoding cases worked per hour by the National Processing Center (NPC) staff?
6. Of the cases in or out of sample (excluding unmatched records which can not be determined to be in or out of sample), approx..how many were sent to OBAV from:

- a. RTNP?
- b. Asynchronous Processing?
- c. Manual Matching and Geocoding?
- d. All components of Non-ID?
7. For OBAV what was the rate of:
 - a. Address and geocode verified?
 - b. Address verified but geocode needs correction?
 - c. Linked to an existing MAF record?
 - d. Address deleted?
 - e. Requires Field Verification (FV)
8. What was the average number of OBAV cases worked per hour by NPC staff?

Budget

1. What was the overall cost of:
 - a. Manual Matching and Geocoding?
 - b. OBAV?
 - c. Clerical Processing?
 - d. Other?
2. What was the average cost per record (average labor cost per hour / average number of records completed per hour) for:
 - a. Manual Matching and Geocoding?
 - b. OBAV?
 - c. Clerical Processing?

Training

1. How was training provided (in person, online, group setting, one-on-one, etc) for:
 - a. Manual Matching and Geocoding?
 - b. OBAV?
2. How many Headquarters (HQ) staff were utilized to conduct for training for:
 - a. Manual Matching and Geocoding?
 - b. OBAV?
3. What was the ratio of HQ staff to NPC staff for training for:
 - a. Manual Matching and Geocoding?
 - b. OBAV?

Staffing

1. How many NPC staff were hired and how many NPC staff were made available for:
 - a. Manual Matching and Geocoding?
 - b. OBAV?
 - c. Clerical Processing?
2. Was the staffing level adequate to perform the work in the allotted time frame for:
 - a. Manual Matching and Geocoding?
 - b. OBAV?
 - c. Clerical Processing?
3. How much space was provided at NPC for:
 - a. Manual Matching and Geocoding?

- b. OBAV?
- c. Clerical Processing?
- 4. Was HQ staff required to support the NPC staff for:
 - a. Manual Matching and Geocoding?
 - b. OBAV?
 - c. Clerical Processing?
- 5. What was the ratio of HQ staff to NPC staff for:
 - a. Manual Matching and Geocoding?
 - b. OBAV?
 - c. Clerical Processing?

Schedule

- 1. Did Non-ID Processing complete on schedule?

V. Methodology

To answer the questions identified in Section IV the NID 2018 E2E CT Assessment will utilize the following sources:

- 1. 2018 E2E CT NID Analysis Dataset
- 2. 2018 E2E CT Performance Measurement Reports
- 3. 2018 E2E CT Budget Financial Management Reports
- 4. Matching and Coding Software (MaCS) Reports
- 5. 2018 E2E CT Integrated Master Schedule
- 6. 2018 E2E CT Non-ID Processing Lessons Learned Documentation

A brief outline of the approach for some of the questions is provided in the table below.

Question (from section IV)	Methodology	Table(s) – see section VI
Production Rates: 1-4, 6, 7	We will use the 2018 Analysis Dataset, built from the datasets sent to DSSD from CDL. We can determine the rate at which cases through RTNP were matched or resolved by taking only the RTNP address evolutions of all records, and getting the frequency of the match status. We can perform similar steps in order to acquire the rates for AR matching, OBAV, and Manual Processing. A match status of 1 indicates a record is resolved, while a match status of 1-4 indicates a record is matched. Lastly, we can match to the sampling universe to determine whether the matched records were in sample or out of sample.	1,2,3,4,5
Production Rates: 5,8 Training 1-3 Staffing 1-6	This information will be available in the 2018 E2E CT Performance Measurement Reports and the 2018 E2E CT Budget Financial Management Reports.	Information in report but not within a table.
Budget	This information will be available in the 2018 E2E CT Performance Measurement Reports and the 2018 E2E CT Budget Financial Management Reports.	6,7

VI. Table Headers

In order to best represent the analysis and respond to the research questions, tables and charts will be provided as part of the analysis report. Some of the table headers are presented below, along with example rows.

Table 1: RTNP Processing

Type of Match	Number	Percent	Cumulative Percent
Matched, Confirmed Geocode (Resolved)	N	%	%
Matched, Unconfirmed Geocode	N	%	%
etc..			

Table 2: Matching following Address Enhancement

Type of Match	Number	Percent	Cumulative Percent
Matched, Confirmed Geocode	N	%	%
Matched, Unconfirmed Geocode	N	%	%
etc..			

Table 3: Manual Matching and Geocoding Outcome

Type of Geocode	Number	Percent	Cumulative Percent
Matched and Geocoded	N	%	%
Matched and Geocoded but requires OBAV	N	%	%
etc..			

Table 4: Sources of OBAV records (In Sample Records)

OBAV Sources	Number	Percent	Cumulative Percent
RTNP	N	%	%
Asynchronous Processing	N	%	%
etc..			

Table 5: OBAV outcome

OBAV Sources	Number	Percent	Cumulative Percent
Address and geocode verified	N	%	%
Address verified but geocode needs correction	N	%	%
etc..			

Table 6: Budget Costs

OBAV Sources	Cost	Percent of Overall Budget	Cumulative Percent
Manual Matching and Geocoding	N	%	%
OBAV	N	%	%
etc..			

Table 7: Average Cost per Case

Process	Average Cost
Manual Matching and Geocoding	N
OBAV	N
etc..	

VII. Risks/Limitations

If a consistent framework for process management in terms of data analysis is not established, then Non-ID Processing will continue to struggle to access necessary data to conduct solid analysis.

VIII. Measures of Success

Successful completion of 2018 E2E CT Non-ID Processing will include the following criteria:

- Non-ID data from ISR and CQA flows smoothly through the non-ID processes, or with minimal issues.
- Analysis data sent to DSSD contains all requested data to perform analysis.
- The NPC staffing levels were sufficient to complete the clerical processing work by the deadline, without requiring overtime or other mitigation.
- Asynchronous Processing functions as intended, or with minimal issues.
- Clerical Processing (both Manual Matching and Geocoding and OBAV) is completed on or under budget.
- Non-ID Processing (both Automated and Clerical Processing) is completed on schedule.

IX. Division Responsibilities

DSSD is responsible for the following:

- Receiving the 2018 E2E CT Non-ID analysis data
- Conducting the analysis for the 2018 E2E CT Non-ID Processing Operational Assessment
- Drafting the 2018 E2E CT Non-ID Processing Operational Assessment

DITD is responsible for the following:

- Attempting to match the Internet responses in the 2018 E2E CT in real-time to the MAF.
- Calling PEARSIS to enhance non-ID responses with AR data.
- Attempting to match the AR enhanced non-ID responses to the MAF.
- Providing tallies daily of the processed non-ID data, as defined above.
- Providing a view of the PEARSIS-processed non-ID data for the Geography Division to view.
- Making the processed non-ID data available to Enterprise Censuses and Surveys Enabling platform (ECaSE) Operational Control System (OCS).
- Providing tallies of the non-ID data that DITD made available, as defined above.

DCMD is responsible for the following:

- Project management of the 2018 E2E CT Non-ID Processing operation
- Drafting the 2018 E2E CT Non-ID Processing Operational Assessment

GEO is responsible for the following:

- Providing subject-matter expertise
- Assisting with the analysis of the 2018 E2E CT Non-ID Processing results

X. Review/Approval Table

Role	Electronic Signature	Date
Fact Checker or independent verifier		
Author's Division Chief (or designee)		
DCMD ADC		
DROM DCMD co-executive sponsor (or designee)		
DROM DSSD co-executive sponsor (or designee)		
Associate Director for R&M (or designee)		
Associate Director for Decennial Census Programs (or designee) and 2020 PMGB		

XI. Document Revision and Version Control History

VERSION/ EDITOR	DATE	REVISION DESCRIPTION	EAE IPT CHAIR APPROVAL
v. 0.1/Dan Donello	7/28/2017	First draft	
v. 0.2/Dan Donello	7/31/2017	Post-Non-ID Branch Chief review updates	
v. 0.3/Dan Donello	8/2/2017	Post-ADC review updates	
v. 0.4/Dan Donello	8/2/2018	Final Draft	

XII. Glossary of Acronyms

Acronym	Definition
ADC	Assistant Division Chief
DCCO	Decennial Census Communications Office
DPMO	Decennial Program Management Office
DROM	Decennial Research Objectives and Methods Working Group
DSSD	Decennial Statistical Studies Division
EXC	Evaluations & Experiments Coordination Branch
IPT	Integrated Project Team
PMGB	Portfolio Management Governance Board
R&M	Research & Methodology Directorate

XIII. References

Niosi, Michael (2012), "2010 Census Non-ID Processing Assessment Report," 2010 Census Planning Memorandum No. 201 (Reissue), U.S. Census Bureau, July 13, 2012.

Jackson, Geoff, et. al. (2012), "2010 Census Be Counted and Questionnaire Assistance Center Assessment Report," U.S. Census Bureau, May 8, 2012

U.S. Census Bureau (November 2016), "2020 Census Operational Plan, Version 2," October 31, 2016