



This document was prepared by and for Census Bureau staff to aid in future research and planning, but the Census Bureau is making the document publicly available in order to share the information with as wide an audience as possible. Questions about the document should be directed to Kevin Deardorff at (301) 763-6033 or [kevin.e.deardorff@census.gov](mailto:kevin.e.deardorff@census.gov)

September 5, 2012

## **2010 CENSUS PLANNING MEMORANDA SERIES**

**No. 234**

MEMORANDUM FOR      The Distribution List

From:                      Burton Reist *[signed]*  
                                 Acting Chief, Decennial Management Division

Subject:                    2010 Census Universe Control and Management/Response  
                                 Processing Systems Assessment Report

Attached is the 2010 Census Universe Control and Management/Response Processing Systems Assessment Report. The Quality Process for the 2010 Census Test Evaluations, Experiments, and Assessments was applied to the methodology development and review process. The report is sound and appropriate for completeness and accuracy.

If you have any questions about this document, please contact David Spindel at 301-763-8725.

Attachment

# 2010 Census Universe Control and Management/ Response Processing Systems Assessment Report

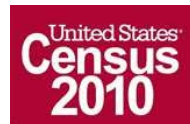
U.S. Census Bureau standards and quality process procedures were applied throughout the creation of this report.

FINAL REPORT

---

David Spindel

Decennial Statistical Studies Division



This page is intentionally left blank.

This document contains no Title 13 data or  
Personally Identifiable Information (PII).

# Table of Contents

Executive Summary .....	vii
1. Introduction .....	1
1.1 PURPOSE OF ASSESSMENT .....	1
1.2 INTENDED AUDIENCE .....	1
2. Background .....	1
2.1 UCM HIGH LEVEL OVERVIEW .....	1
2.2 RPS HIGH LEVEL OVERVIEW .....	2
2.3 ASSUMPTIONS .....	3
3. Methodology .....	3
3.1 OBJECTIVE.....	3
3.2 RESEARCH QUESTIONS AND METHODOLOGY .....	4
3.3 DATA FILES FOR THE STUDY .....	5
4. Limitations .....	6
5. Results .....	6
5.1 HOW MANY ADDRESSES WERE ADDED AND HOW MANY WERE DELETED AT EACH STAGE OF THE CENSUS, BY TYPE OF ADDRESS? .....	6
5.2 HOW MANY ADDRESSES HAD A FINAL STATUS OF "OCCUPIED"? .....	15
5.3 HOW MANY ADDRESSES HAD A FINAL STATUS OF "VACANT"? .....	15
5.4 WHAT IS THE DISTRIBUTION OF HOUSEHOLD SIZES FOR ALL RESPONSES? WHAT IS THE DISTRIBUTION AMONG VALID RESPONSES? .....	16
5.5 WHAT IS THE DISTRIBUTION OF HOUSEHOLD SIZES AT THE END OF THE PRIMARY SELECTION ALGORITHM PROCESS? .....	17
5.6 WHAT IS THE DISTRIBUTION OF THE HOUSEHOLD SIZES FOR ALL RESPONSES AT THE END OF THE CENSUS?.....	18
6. Related Assessments .....	18
7. Key Lessons Learned, Conclusions, and/or Recommendations .....	19
8. Acknowledgments .....	22
9. References .....	22
Appendix A: UCM/RPS Assessment Acronyms List.....	23

## List of Tables

Table 5.1	Distribution of Added Addresses by Census Universe.....	7
Table 5.2	Distribution of Addresses Deleted or Undeliverable .....	8
Table 5.3	Distribution of Vacant Addresses by Operation .....	8
Table 5.4	Distribution of Addresses Added by Address Type .....	9
Table 5.5	Addresses Added in the Initial File by Address Type .....	10
Table 5.6	Addresses Added to the Enumeration Universe by Address Type .....	11
Table 5.7	Addresses Added in the Supplemental NRFU Extract by Address Type .....	12
Table 5.8	Addresses Added in Coverage Followup/Field Verification by Address Type .....	13
Table 5.9	Addresses Added to the Final Collection Universe by Address Type .....	14
Table 5.10	Stateside and Puerto Rico Distribution of Addresses Added as a Percentage of the Universe Totals.....	15
Table 5.11	Occupied and Vacant Housing Units .....	15
Table 5.12	Distribution of Total and PSA Eligible Responses by Household Size .....	16
Table 5.13	Household Sizes for Primary Selection Algorithm (PSA) Selected Households .....	17
Table 5.14	Household Size for Occupied Housing Units .....	18

## Executive Summary

The Universe Control and Management system was developed to control and track the enumeration activities of the 2010 Census. The core of the Universe Control and Management is a database containing characteristics and operational status indicators for each of the addresses in the 2010 Census universe. The primary benefit of the Universe Control and Management system is that it provides a centralized source and reference point for multiple systems that are external to decennial processing systems. The system is primarily divided into two subsystems: The Universe Control and Management Operations system which is designed to track all actions related to a specific address, and the Universe Control and Management Address system which contains address specific information.

The purpose of the Response Processing System is to collect the decennial census response data and prepare it for Congressional Apportionment and decennial census data product usage. The Response Processing System is composed of nine distinct modules that capture the collected 2010 Census responses and the results of systems, including imputing missing data elements, selecting a return amongst all available returns and/or persons to represent the unit, appending tabulation geography information and adding recode variables. The combination of the operational data from the Universe Control and Management system and the response data from the Response Processing System, through the use of a unique Master Address File ID number for each record, permits a complete picture of the 2010 Census universe and response processing to emerge from this assessment.

The Universe Control and Management /Response Processing Systems assume that they completely and accurately reflect the specifications of the 2010 Census sponsors. During the 2010 Census, there were two committees formed to develop, test, and provide feedback on the quality of the system processing. The first group, internal to the Decennial Systems and Contracts Management Office, was tasked with developing the system based upon user supplied specifications. The second group, external to Decennial System and Contracts Management Office but inside the Census Bureau, was comprised of Decennial Statistical Studies Division and Population Division management and technical staff. This second group met regularly to oversee the implementation of the system specifications, and to assess the accuracy of the output based on those specifications.

In the Lessons Learned section of this assessment, detailed feedback is provided from the system developers and sponsors. Some highlights of the Lessons Learned are:

- All data exchanges and mappings need to be thoroughly tested with both subject matter experts and independent parties, including analysts.
- More rigorous “user testing” of the residual coding Quality Control component is needed to ensure that the system meets stakeholder expectations.
- The elaboration of requirements, the careful effort of the developers, and the testing protocols implemented resulted in a minimization of errors in processing for the Decennial Response File, the Census Unedited File, the Census Edited File, and the Hundred Percent Detail File.

Since the Universe Control and Management /Response Processing System contains the official 2010 Census data along with the operational variables, it is the primary source for how addresses and individuals were added and deleted at various operational stages of the 2010 Census.

Some of the significant results from the 2010 Census are:

- An analysis of the data from the Decennial Response File on the variables representing the validity of the response was performed for all responses and then for only those that were valid responses (those eligible for the Primary Selection Algorithm). The Primary Selection Algorithm consists of a set of decision processes to determine which form is to be used when multiple forms exist for an address. There were 161,712,091 total household responses received. When this group is refined into those household responses that were classified as valid responses, the number of households is reduced to 147,048,720.
- Since the Primary Selection Algorithm eligible number of housing units includes multiple returns per Master Address File Identifier, the Primary Selection Algorithm was then applied to yield the best single return for each Master Address File Identifier. Subsequently, imputation procedures were applied for housing units with unreturned forms as part of the processing to create the Census Unedited File. Therefore, following application of the Primary Selection Algorithm and the creation of the Census Unedited File, there were 133,341,676 housing units in the final 2010 Census (131,704,730 from stateside addresses and 1,636,946 from Puerto Rico). Of this number, 118,092,823, or 88.56 percent of all housing units, had a final status of being an occupied housing unit. Over one-tenth (11.44 percent) of all housing units, 15,248,853 units, in the 2010 Census had a final status as being vacant.

Finally, in an effort to improve the 2020 Census the following recommendations are offered:

- Enhance development testing of all systems by developers.
- Testing by subject matter experts and independent parties, including analysts, should occur from start to finish, with specific effort expended on testing alternative or unexpected data paths.
- Access to the data files should be made to appropriate subject matter experts as early as possible for the purpose of system testing.



# 1. Introduction

## 1.1 Purpose of Assessment

The objective of the 2010 Census Universe Control and Management (UCM)/Response Processing Systems (RPS) Assessment is to document the processes and to assess the quality of the data for the U.S. and Puerto Rico contained in the 2010 Census UCM/RPS Systems, as well as to aid in planning for future UCM/RPS type processing operations.

## 1.2 Intended Audience

This document assumes that the reader has a basic understanding of the 2010 Census operations. It will serve as input for discussion by the research, planning, and development teams when planning for the 2020 Census. If the reader does not have a basic understanding of census operations, a detailed description can be found in the document “2010 Census Operational Plan” (Project Management Branch, 2010).

# 2. Background

A similar file and report were created in Census 2000 from information on the Master Address File (MAF), the Decennial Master Address File, the Hundred Percent Census Unedited File, and the Hundred Percent Census Edited File. This information greatly aided research efforts and planning for the 2010 Census. This assessment will help planning and research efforts for the 2020 Census.

The combination of the operational data from UCM and the response data from RPS permits a complete picture of the decennial census data to emerge from this assessment. For example, it allows a comparison of the distribution of household size before and after implementing the Primary Selection Algorithm (PSA), and after creation of the final census universe as represented on the Census Unedited File (CUF). In this way we gain a clearer view of the interactions between the UCM and the RPS Systems.

## 2.1 UCM High Level Overview

The UCM was developed and is maintained by the Decennial System and Contracts Management Office (DSCMO) to control and track the enumeration activities of the 2010 Census. The core of the UCM is a database containing characteristics and operational status indicators for each of the addresses in the 2010 Census universe. While the core is the database, several layers of software surround it (Clark, 2010). These layers are designed to:

- Load, update, and maintain the addresses to be enumerated, based on MAF updates and the results of the 2010 Census Address Canvassing (AC) operation and enumeration activities.

- Select and flag records for application of various treatments as part of the design, including but not limited to Mailout/Mailback (MO/MB) and Nonresponse Followup (NRFU). Also, receive the results of these and other operations to set housing status flags in the database.

The primary benefit of the UCM is it provides a centralized source and reference point for multiple systems, external to Decennial Processing Systems. UCM operates as the central hub between Geography Division, Decennial Response Integration Systems (DRIS), Field Division, DSCMO Paper Based Operations (PBO), Population Division (POP), and the analytical areas of Decennial Statistical Studies Division (DSSD).

In 2010, the DSCMO PBO also authored control systems for field work previously performed via the Field Data Collection Automation contractor. The 2010 UCM system is built based on sponsor specifications from subject matter experts in DSSD and POP.

The UCM reflects the collected address data in the decennial census universe. It does not determine the characteristics of those address data, but provides a reference to support census activities.

## **2.2 RPS High Level Overview**

The purpose of the RPS is to prepare the decennial census response data for Congressional Apportionment and census data product usage. It is composed of nine distinct modules (McCoy, 2010).

- Interface and Acknowledgment Processing System (IAP) - The IAP System is a data integration system used to verify, validate, and ingest incoming response data from census return processing.
- Coverage Followup (CFU) Selection System - The CFU Selection System performs a variety of data calculations to identify the universe of responses for Coverage Followup. The actual process involves an RPS assessment of Data Capture Audit and Resolution variables which are set by the DRIS to determine CFU eligible cases. Then eligible cases are sent to DSSD based on specified filtering and selection criteria.
- Race and Hispanic Origin Coding System (Coding) - The Coding system consists of two components, autocoding and residual coding. The autocoding component provides the response coding of race and Hispanic origin write-in responses using the unified race and unified Hispanic origin response dictionaries of write-in responses that have previously been assigned a code value. The residual coding component provides an interactive interface system with response coding assistance files to aid the coding staff in the National Processing Center in assigning codes and implementing quality control.

- Post Capture Data Integration System (PCDI) - The PCDI system is a data integration system that provides final processing on response data that results in the production of the Decennial Response File (DRF1).<sup>1</sup>
- PSA - The PSA system applies various functionality to find the “best” representation amongst available returns and/or persons to represent the unit. The data product that results after the PSA is applied to the DRF1 is known as the final Decennial Response File (DRF2).<sup>2</sup>
- CUF - The CUF represents the static universe of addresses and return data utilized for Congressional Apportionment. The CUF is the primary data source to impute a household size for those addresses lacking a population count.

The final three modules, though actually part of the RPS, were beyond the scope of this assessment, and hence, not specifically addressed in this report.

- Census Edited File System (CEF) - The CEF system processes response data to ensure reported data fields from each record in the CUF are filled with valid values.
- Edit Review System (ERS) - The ERS is intended as a tool for subject matter analysts to review and approve the outcome of edit and allocation processing. This is done by reviewing the CUF and CEF jointly, as well by providing reports (such as matrix counts and tally counts) during edit processing.
- Hundred Percent Detail File System (HDF) - The HDF system appends tabulation geography information, adds recode variables, and applies the data disclosure avoidance algorithm to the CEF.

## 2.3 Assumptions

The UCM/RPS assumes that it completely and accurately reflects the specifications of the 2010 Census sponsors.

## 3. Methodology

### 3.1 Objective

The objective of this assessment is to provide complete documentation of the 2010 Census processing systems. Some of these tallies are provided in other reports, but all tallies are

---

<sup>1</sup> Note there was no actual delivered output file for the DRF1. The first actual delivered output file, containing response data, is the DRF2.

<sup>2</sup> Note that the distinction between the DRF1 and the DRF2 is artificial in that there is no delivered output file for the DRF1. The actual output file is the DRF2.

presented within this report in order to present a complete picture of the 2010 Census. These tallies can be used as control counts for other assessments, evaluations, and experiments.

### 3.2 Research Questions and Methodology

How many addresses were on the census address list during the formulation of the:

- a) Initial UCM Universe - the universe of Housing Units (HU) for the initial UCM build. The purpose of this universe was to facilitate the printing of Mailout/Mailback and Update/Leave (U/L) questionnaires.
- b) Enumeration Universe - this universe added Group Quarters (GQs), units that were listed as HUs during Group Quarters Validation, and new, valid HUs from the Fall 2009 Delivery Sequence File (DSF) to the initial UCM universe.
- c) Supplemental NRFU Universe and the Field Verification (FV) Universe - the Supplemental NRFU Universe list was created from a variety of sources that added HUs to the universe after the delivery of the Enumeration Universe and in time to be enumerated in the Vacant/Delete Check (VDC). These sources include Local Update of Census Addresses (LUCA) Appeals, added addresses from U/L, and new, valid HUs from the Fall 2009 DSF, among others. For the FV Universe, additions came from Non-ID sources such as Be Counted forms.
- d) Final Collection Universe - all units that were ever included in the 2010 Census universe and their associated collection geography.
- e) Final Tabulation Universe - all units in the final 2010 Census and their associated tabulation geography.

Specifically, the assessment examined:

1. How many addresses were added and how many were deleted at each stage of the 2010 Census, by type of address? (Data source: UCM tables)
2. How many addresses had a final status of "occupied"? (Data source: the CUF)
3. How many addresses had a final status of "vacant"? (Data source: the CUF)
4. What is the distribution of household sizes for all responses? What is the distribution among valid responses?<sup>3</sup> (Data source: the DRF)

---

<sup>3</sup> A return is considered valid if it is considered an eligible return for selection by the Primary Selection Algorithm. An example of eligibility criteria would be the selection of one enumeration when multiple enumeration records exist. Another example would be that a continuation form (for enumeration of large households) is not eligible if it is linked to a parent enumeration, but is eligible if it is unlinked to a parent enumeration.

5. What is the distribution of household sizes for all responses at the end of the Primary Selection Algorithm process? (Data source: the DRF)
6. What is the distribution of the household sizes for all responses at the end of the 2010 Census? (Data source: the CUF)
7. What lessons were learned as a result of the UCM/RPS processing of the decennial census files? What worked well? What could be improved? (Data source: We held meetings with the membership of the UCM/RPS teams to collect lessons learned about the processing stages. In addition, we incorporated findings of the DSSD-chaired review committee (Pennington, 2010).)

### **3.3 Data Files for the Study**

The majority of the data for this assessment were derived from extracts from the DSCMO UCM Operations and the UCM Address tables. A subset of selected variables in the UCM Operations tables were combined with all of the variables in the UCM Address table to create an analysis file for this assessment. In addition, extracts were made from the DRF and CUF (subsets of the UCM/RPS Systems).

#### *Types of Addresses Profiled in the Assessment Results*

DSSD evaluated 2010 UCM/RPS addresses by type of address information (i.e., city-style, rural route, etc.) contained on the listing pages. The addresses were defined primarily to categorize their potential to be located by Listers. Because the location house number and street name fields on the MAF were used first when populating the listing pages (while location ZIP Code was not included) these fields were used in the criteria for determining a complete city-style address.<sup>4</sup>

DSSD classified addresses into six categories, based on the highest criteria met:

- Complete city-style-stateside: Included all stateside units that had complete city-style addresses (house number and street name).
- Complete city-style-Puerto Rico: Included all Puerto Rico units that had complete city-style addresses (house number and street name or urbanization name, or apartment complex name and structure ID).
- Complete rural route: Included units that did not have a complete city-style address but did have a complete rural route address, such as “Rural Route 2, Box 3”.

---

<sup>4</sup> ZIP Code was included to identify those address records as valid city-style addresses for mailing, as mailing information was used in filling the Field Verification Address Listing Pages if location information was not present. In the absence of a block, ZIP code was required for matching and was important for accurate geocoding.

- Complete Post Office (P.O.) Box: Included units without a complete city-style or a complete rural route address, but did have a complete P.O. Box address, such as “P.O. Box 515”.
- Incomplete address: Included units that had some address information but without a complete address of any type.
- No address information: Included units that were missing house number, street name, Rural Route, and P.O. Box information.

Addresses were further delineated by whether or not they had had a physical/location description provided during a decennial census field operation (e.g., brick house, blue shutters).

## 4. Limitations

Due to the scope of this assessment, the data used to produce the study results are based on tabulations of population data derived from UCM and RPS tables which contain the 2010 Census operations files and response information. Therefore, the subsequent assessment results are as good, and with the same limitations, as the data contained in those systems. For specific variables, as identified in Table 5.3 and Table 5.4 (in section 5), for “NRFU vacants and deletes that were ultimately enumerated”, only an overlap category was available. The specific variable limitations for this situation are discussed when the data are presented in that section.

## 5. Results

### 5.1 How many addresses were added and how many were deleted at each stage of the census, by type of address?

Addresses enter the 2010 Census based on a variety of census operations. The initial census universe, which includes addresses available for the initial mailout, was developed based on the results of the 2010 Census AC operation.

By joining the ID history variable in the UCM Operations table with address information from the UCM Address table, it is possible to obtain the status of addresses in the 2010 Census for each census operation. The ID history variable for a record contains an activity indicator in the UCM Operations table for each delivery of census universes and shows what happened to that record with respect to that universe (i.e., not delivered, added, updated, or deleted). Based on this variable, there were 132,402,816 addresses available for printing labels for questionnaires in the initial universe. This initial universe is comprised of addresses for MO/MB, U/L, Alaska/Remote Update Enumerate and Update Enumerate operations. Of this initial universe, 130,827,141 were Stateside addresses, and 1,575,675 were addresses from Puerto Rico. Table 5.1 shows how many addresses were added in conjunction with each universe delivery for the 2010 Census.

**Table 5.1 Distribution of Added Addresses by Census Universe<sup>5</sup>**

	Stateside		Puerto Rico		Total	
	Number	Percent	Number	Percent	Number	Percent
<b>Enumeration</b>	2,000,038	32.49	8850	5.71	2,008,888	31.83
<b>Supplemental NRFU</b>	2,996,316	48.67	81,567	52.63	3,077,883	48.77
<b>CFU/FV</b>	325,090	5.28	4,435	2.86	329,525	5.22
<b>Final Collection</b>	835,052	13.56	60,138	38.80	895,190	14.18
<b>Total</b>	<b>6,156,496</b>	<b>100.00</b>	<b>154,990</b>	<b>100.00</b>	<b>6,311,486</b>	<b>100.00</b>

*Data Source: the UCM Operations file*

The largest number of added addresses occurred during Supplemental NRFU extract, with 3,077,883 addresses which included the results of LUCA Appeals, the Fall 2009 DSF, and U/L, as well as some additional sources of added addresses. Prior to creating the enumeration universe, 2,008,888 addresses were added. While the percentage of addresses added during Supplemental NRFU extract was approximately the same for both Stateside and Puerto Rico addresses (48.67 percent and 52.63 percent, respectively), almost one-third (32.49 percent) of the added addresses for Stateside addresses occurred in connection with creating the enumeration universe and 13.56 percent were added in the final collection universe.<sup>6</sup> In contrast to this, Puerto Rico had only 5.71 percent of its addresses added in connection with forming the enumeration universe, and almost 40 percent (38.80 percent) when the final collection file was formed.

With respect to the discussion of adds and deletes, it is important to note that the census universes, identified by the ID history variable in UCM processing, are actually scheduled “stopping points in time” to perform extracts and develop files that are necessary for other census operations.<sup>7</sup> In fact, during the census, multiple addresses were being added or being identified for deletion by multiple operations. Only during the creation of the CUF was a final determination made as to which of the addresses were to be included or excluded, to develop the final population counts. Therefore, the only operation that actually produced deletions in the UCM Operations table was the creation of the CUF. Using the CUF final status variable, almost 5 million (4,970,318) addresses were deleted from the 2010 Census.

While deletions were classified in the CUF final status, specific operations contributed to processing these deletions. Operations that resulted in deletions are presented in Table 5.2. Note, these are not equivalent to CUF deletions, and are being shown to understand the scope of where deletions occurred.

<sup>5</sup> See section 3.2 for description of these universes.

<sup>6</sup> Addresses added in forming the final collection universe are addresses that were added to the UCM system due to the results of prior census operations, such as Non-ID cases, that had not been previously added, but had complete forms and geography.

<sup>7</sup> In the UCM Operations file, the ID history variable “deletes” value was developed in order to handle address deletions that were provided to DSCMO during GEO processing. Based on the DSCMO processing in the UCM system, it was decided that this value would not be utilized for the ID history variable.

**Table 5.2 Distribution of Deleted or Undeliverable Addresses**

	<b>Stateside</b>	<b>Puerto Rico</b>	<b>Total</b>
<b>NRFU Operation</b>	3,379,036	70,984	3,450,020
<b>VDC Operation</b>	1,001,013	25,933	1,026,946
<b>Update/Leave Operation</b>	562,591	75,437	638,028

*Data Source: the UCM Operations file*

*Note: The NRFU Operation deleted addresses do not include addresses that were coded as “vacants or deletes that were ultimately VDC enumerated”.*

The use of the data in the above table is limited, due to the fact that the final determination of deletes occurs subsequently during the production of the CUF using a combination of operation outcomes. Final deletes in the CUF are based on a “double delete” rule, whereby the deletion was indicated in two different operations. The data presented in Table 5.2 are based on a “delete” value for that specific operation in the UCM, even if it was otherwise categorized in a subsequent operation. Therefore, the data in this table are strictly intended to be viewed in the context of whether that particular operation considered the address as a delete, and not the ultimate disposition of the address (i.e. a “NRFU delete, VDC vacant” value would be considered, in this context, as a NRFU delete for Table 5.2). Address records with a delete value are included in follow-up operations according to their status outcomes. In addition, the NRFU deletes in Table 5.2 and vacants in Table 5.3 are somewhat underestimated since an overlap category exists which contains “NRFU vacants or deletes that were found to be occupied by VDC, and enumerated” in the variable that determined these tables.

Likewise, in order to further understand which operations identified vacant households, a similar analysis was performed. Table 5.3 shows vacant addresses by the operation in which they were identified.

**Table 5.3 Distribution of Vacant Addresses by Operation**

	<b>Stateside</b>	<b>Puerto Rico</b>	<b>Total</b>
<b>NRFU Operation</b>	12,276,696	200,120	12,476,816
<b>VDC Operation</b>	3,321,414	145,471	3,466,885

*Data Source: the UCM Operations file*

*Note: The NRFU Operation vacant addresses do not include addresses that were coded as “vacants or deletes that were ultimately VDC enumerated”.*

The data presented in Table 5.3 are based on a “vacant” value for the specific operation, even if it was otherwise categorized in a subsequent operation. As was the case in the delete section, the data in this table are strictly intended to be viewed in the context of whether that particular operation considered the address as being vacant, and not the ultimate disposition of the address (i.e. a “NRFU vacant, VDC delete” value, for this table would be considered a NRFU vacant for



Table 5.3). In addition, as with the NRFU deletes, the NRFU “vacants” are subject to the previously cited overlap category, and hence, will be somewhat understated. The actual final number of vacancies must be derived using the final status variable from the CUF.

Table 5.4 shows the total number of addresses added to the UCM system during the 2010 Census. Addresses were considered by UCM operations to be added due to the fact that they either comprised the formation of the initial universe or because they were added in the formation of one of the subsequent census universes (Enumeration, Supplemental NRFU, CFU/FV, or Final Collection). UCM Operations ID history considers that the census universe in which an address enters the UCM system as an “add” for that universe. Almost 98 percent (97.79 percent) of the addresses were Complete City-Style Addresses (with and without location description). Of these, Complete City-Style Addresses, almost nine-tenths (88.94 percent) of all added addresses were categorized as “without location description”.

**Table 5.4 Distribution of Addresses Added by Address Type**

ADDRESS TYPE	LOCATION DESCRIPTION	Stateside		Puerto Rico		Total	
		Number	Percent	Number	Percent	Number	Percent
Complete City-Style Address	without location description	122,609,493	89.51	769,408	44.46	123,378,901	88.94
	with location description	11,727,509	8.56	542,372	31.34	12,269,881	8.85
Incomplete Address	without location description	21,947	0.02	15,229	0.88	37,176	0.03
	with location description	1,940,682	1.42	207,096	11.97	2,147,778	1.55
No Address Information	without location description	86,928	0.06	5,032	0.29	91,960	0.07
	with location description	169,309	0.12	44,442	2.57	213,751	0.15
Complete P.O. Box Address	without location description	1,140	0.00	163	0.01	1,303	0.00
	with location description	132,191	0.10	35,430	2.05	167,621	0.12
Complete Rural Route	without location description	3,945	0.00	2,286	0.13	6,231	0.00
	with location description	290,493	0.21	109,207	6.31	399,700	0.29
<b>Total</b>		<b>136,983,637</b>	<b>100.00</b>	<b>1,730,665</b>	<b>100.00</b>	<b>138,714,302</b>	<b>100.00</b>
	without location description	122,723,453	89.59	792,118	45.77	123,515,571	89.04
	with location description	14,260,184	10.41	938,547	54.23	15,198,731	10.96

*Data Source: the UCM Operations file and the UCM Address file*

*Note: Since this table includes all of the initial addresses to facilitate printing of the Mailout/Mailback and U/L questionnaires, it consists of all addresses that were added, based on being added for the initial UCM build in the UCM Operations file*

The following tables (Table 5.5 through Table 5.9) will present similar analyses for the Initial, Enumeration, Supplemental NRFU, Coverage Followup/Field Verification and Final Collection Universes.

Table 5.5, which includes the results of the 2010 Census AC, shows that 132,402,816 addresses were added to the UCM in the initial file creation. Of these added addresses, 98.81 percent (130,827,141) came from Stateside locations, and 1.19 percent (1,575,675) came from Puerto Rico. Almost all of the addresses (97.96 percent) in the initial operation were complete city-style addresses (89.19 percent and 8.77 percent, without and with location description, respectively.)

**Table 5.5 Addresses in the Initial File by Address Type**

		Stateside		Puerto Rico		Total	
		Number	Percent	Number	Percent	Number	Percent
<b>Complete City-Style Address</b>	<b>without location description</b>	117,359,387	89.71	725,737	46.06	118,085,124	89.19
	<b>with location description</b>	11,088,362	8.48	517,601	32.85	11,605,963	8.77
<b>Incomplete Address</b>	<b>without location description</b>	862	0.00	116	0.01	978	0.00
	<b>with location description</b>	1,843,628	1.41	168,477	10.69	2,012,105	1.52
<b>No Address Information</b>	<b>without location description</b>	0	0.00	0	0.00	0	0.00
	<b>with location description</b>	128,009	0.10	27,342	1.74	155,351	0.12
<b>Complete P.O. Box Address</b>	<b>without location description</b>	90	0.00	2	0.00	92	0.00
	<b>with location description</b>	128,821	0.10	34,038	2.16	162,859	0.12
<b>Complete Rural Route</b>	<b>without location description</b>	110	0.00	1,320	0.08	1,430	0.00
	<b>with location description</b>	277,872	0.21	101,042	6.41	378,914	0.29
<b>Total</b>		<b>130,827,141</b>	<b>100.00</b>	<b>1,575,675</b>	<b>100.00</b>	<b>132,402,816</b>	<b>100.00</b>
	<b>without location description</b>	117,360,449	89.71	727,175	46.14	118,087,624	89.19
	<b>with location description</b>	13,466,692	10.29	848,599	53.86	14,315,192	10.81

*Data Source: the UCM Operations file and the UCM Address file*

For the enumeration universe, Table 5.6 shows that approximately 2 million (2,008,888) addresses were added to the 2010 Census. While the percentage of complete city-style addresses is roughly the same as the initial universe (98.03 percent), the percentage of those addresses without a location description decreased and those with a location description increased (82.14 percent and 15.89 percent for the enumeration universe as compared to 89.19 percent and 8.77 percent for the initial universe, respectively.)

**Table 5.6 Addresses Added to the Enumeration Universe by Address Type**

		Stateside		Puerto Rico		Total	
		Number	Percent	Number	Percent	Number	Percent
<b>Complete City-Style Address</b>	<b>without location description</b>	1,645,442	82.27	4,742	53.58	1,650,184	82.14
	<b>with location description</b>	316,567	15.83	2,699	30.50	319,266	15.89
<b>Incomplete Address</b>	<b>without location description</b>	759	0.04	81	0.92	840	0.04
	<b>with location description</b>	13,273	0.66	1,038	11.73	14,311	0.71
<b>No Address Information</b>	<b>without location description</b>	136	0.01	4	0.05	140	0.01
	<b>with location description</b>	22,851	1.14	68	0.77	22,919	1.14
<b>Complete P.O. Box Address</b>	<b>without location description</b>	11	0.00	8	0.09	19	0.00
	<b>with location description</b>	498	0.02	115	1.30	613	0.03
<b>Complete Rural Route</b>	<b>without location description</b>	2	0.0	0	0.00	2	0.00
	<b>with location description</b>	499	0.02	95	1.07	594	0.03
<b>Total</b>		<b>2,000,038</b>	<b>100.00</b>	<b>8,850</b>	<b>100.00</b>	<b>2,008,888</b>	<b>100.00</b>
	<b>without location description</b>	1,646,350	82.32	4,835	54.63	1,651,185	82.19
	<b>with location description</b>	353,688	17.68	4,015	45.37	357,703	17.81

*Data Source: the UCM Operations file and the UCM Address file*

The Supplemental NRFU extract produced the largest number of addresses added to the 2010 Census following the creation of the initial universe. In the extract that formed this universe, just over three million (3,077,883) addresses were added to the census. Of the added addresses, 95.23 percent were Complete City-Style Addresses (88.92 percent without location description and 6.31 percent with location description).

**Table 5.7 Addresses Added in the Supplemental NRFU Extract by Address Type**

		Stateside		Puerto Rico		Total	
		Number	Percent	Number	Percent	Number	Percent
<b>Complete City-Style Address</b>	<b>without location description</b>	2,716,731	90.67	20,095	24.64	2,736,826	88.92
	<b>with location description</b>	180,016	6.01	14,314	17.55	194,330	6.31
<b>Incomplete Address</b>	<b>without location description</b>	6,026	0.20	7,278	8.92	13,304	0.43
	<b>with location description</b>	68,697	2.29	24,920	30.55	93,617	3.04
<b>No Address Information</b>	<b>without location description</b>	606	0.02	1,245	1.53	1,851	0.06
	<b>with location description</b>	7,876	0.26	5,627	6.90	13,503	0.44
<b>Complete P.O. Box Address</b>	<b>without location description</b>	412	0.01	132	0.16	544	0.02
	<b>with location description</b>	2,780	0.09	956	1.17	3,736	0.12
<b>Complete Rural Route</b>	<b>without location description</b>	1,586	0.05	758	0.93	2,344	0.08
	<b>with location description</b>	11,586	0.39	6,242	7.65	17,828	0.58
<b>Total</b>		<b>2,996,316</b>	<b>100.00</b>	<b>81,567</b>	<b>100.00</b>	<b>3,077,883</b>	<b>100.00</b>
	<b>without location description</b>	2,725,361	90.96	29,508	36.18	2,754,869	89.50
	<b>with location description</b>	270,955	9.04	52,059	63.82	323,014	10.49

*Data Source: the UCM Operations file and the UCM Address file*

Table 5.8 shows that 329,525 addresses were added in the Coverage Followup/Field Verification universe. The largest group of additions during this operation came from addresses with a “Complete City-Style Address without location description” (93.17 percent).

**Table 5.8 Addresses Added in Coverage Followup/Field Verification by Address Type**

		Stateside		Puerto Rico		Total	
		Number	Percent	Number	Percent	Number	Percent
<b>Complete City-Style Address</b>	<b>without location description</b>	305,689	94.03	1,339	30.19	307,028	93.17
	<b>with location description</b>	12,804	3.94	823	18.56	13,627	4.14
<b>Incomplete Address</b>	<b>without location description</b>	4,439	1.37	436	9.83	4,875	1.48
	<b>with location description</b>	267	0.08	1,659	37.41	1,926	0.58
<b>No Address Information</b>	<b>without location description</b>	1,177	0.36	34	0.77	1,211	0.37
	<b>with location description</b>	382	0.12	62	1.40	444	0.13
<b>Complete P.O. Box Address</b>	<b>without location description</b>	280	0.09	1	0.02	281	0.09
	<b>with location description</b>	7	0.00	25	0.56	32	0.01
<b>Complete Rural Route</b>	<b>without location description</b>	35	0.01	5	0.11	40	0.01
	<b>with location description</b>	10	0.00	51	1.15	61	0.02
<b>Total</b>		<b>325,090</b>	<b>100.00</b>	<b>4,435</b>	<b>100.00</b>	<b>329,525</b>	<b>100.00</b>
	<b>without location description</b>	311,620	95.86	1,815	40.92	313,435	95.12
	<b>with location description</b>	13,470	4.14	2,620	59.08	16,090	4.88

*Data Source: the UCM Operations file and the UCM Address file*

Table 5.9 shows that 895,190 addresses were added during the Final Collection operation. In contrast to Coverage Followup/Field Verification, where over nine of ten (93.17 percent) addresses came from “Complete City-Style Address without location description”, “Complete City-Style Address without location description” accounted for only two-thirds (67.00 percent) of the added addresses in the final collection operation. This was followed by the “Complete City-Style Address with location description” with 15.27 percent of the added addresses. The third largest group of addresses added was housing units with “No Address Information without location description”, accounting for almost 10 percent (9.91 percent) in the Final Collection operation.

**Table 5.9 Addresses Added to the Final Collection Universe by Address Type**

		Stateside		Puerto Rico		Total	
		Number	Percent	Number	Percent	Number	Percent
<b>Complete City-Style Address</b>	<b>without location description</b>	582,244	69.73	17,495	29.09	599,739	67.00
	<b>with location description</b>	129,760	15.54	6,935	11.53	136,695	15.27
<b>Incomplete Address</b>	<b>without location description</b>	9,861	1.18	7,318	12.17	17,179	1.92
	<b>with location description</b>	14,817	1.77	11,002	18.29	25,819	2.88
<b>No Address Information</b>	<b>without location description</b>	85,009	10.18	3,749	6.23	88,758	9.91
	<b>with location description</b>	10,191	1.22	11,343	18.86	21,534	2.41
<b>Complete P.O. Box Address</b>	<b>without location description</b>	347	0.04	20	0.03	367	0.04
	<b>with location description</b>	85	0.01	296	0.49	381	0.04
<b>Complete Rural Route</b>	<b>without location description</b>	2,212	0.26	203	0.34	2,415	0.27
	<b>with location description</b>	526	0.06	1,777	2.95	2,303	0.26
<b>Total</b>		<b>835,052</b>	<b>100.00</b>	<b>60,138</b>	<b>100.00</b>	<b>895,190</b>	<b>100.00</b>
	<b>without location description</b>	679,673	81.39	28,785	47.86	708,458	79.14
	<b>with location description</b>	155,379	18.61	31,353	52.14	186,732	20.86

*Data Source: the UCM Operations file and the UCM Address file.*

*Note: Addresses added in the final collection operation are addresses that were added to the UCM system due to the results of prior census operations, such as non-ID cases, that had not been previously added, but had complete forms and geography.*

The highest percentage of Stateside added addresses occurred in forming the Enumeration universe (99.56 percent). The lowest percentage of additions for Stateside addresses was 93.28 percent, in the Final Collection operation. The highest percentage of added addresses for Puerto Rico was 6.72 percent, occurring in developing the Final Collection universe.

**Table 5.10 Stateside and Puerto Rico Distribution of Addresses Added as a Percentage of the Universe Totals**

Operation	Stateside		Puerto Rico		Total	
	Number	Percent	Number	Percent	Number	Percent
<b>Enumeration</b>	2,000,038	99.56	8,850	0.44	2,008,888	100.00
<b>Supplemental NRFU</b>	2,996,316	97.35	81,567	2.65	3,077,883	100.00
<b>CFU/FV</b>	325,090	98.65	4,435	1.35	329,525	100.00
<b>Final Collection</b>	835,052	93.28	60,138	6.72	895,190	100.00
<b>Total</b>	<b>6,156,496</b>	<b>97.54</b>	<b>154,990</b>	<b>2.46</b>	<b>6,311,486</b>	<b>100.00</b>

*Data Source: the UCM Operations file*

## 5.2 How many addresses had a final status of "occupied"?

There were 133,341,676 housing units in the final 2010 Census. Occupied housing units accounted for 118,092,823, or 88.56 percent of all housing units. It can be seen that the percentage of occupied units is almost 5 percent lower in Puerto Rico than for the Stateside (84.09 percent and 88.62 percent, respectively).

**Table 5.11 Occupied and Vacant Housing Units**

Housing Unit Status	Stateside		Puerto Rico		Total	
	Number	Percent	Number	Percent	Number	Percent
<b>Occupied</b>	116,716,292	88.62	1,376,531	84.09	118,092,823	88.56
<b>Vacant</b>	14,988,438	11.38	260,415	15.91	15,248,853	11.44
<b>Total</b>	<b>131,704,730</b>	<b>100.00</b>	<b>1,636,946</b>	<b>100.00</b>	<b>133,341,676</b>	<b>100.00</b>

*Data Source: the 2010 CUF*

## 5.3 How many addresses had a final status of "vacant"?

There were 15,248,853 vacant housing units in the 2010 Census (Table 5.11). Over one-tenth (11.44 percent) of all housing units were vacant. The percentage of vacant units is significantly higher in Puerto Rico than for the Stateside (15.91 percent and 11.38 percent, respectively).

#### 5.4 What is the distribution of household sizes for all responses? What is the distribution among valid responses?

There were 161,712,091 total household responses in the 2010 Census. When this group is refined into valid household responses, based on PSA eligibility (see footnote 2 on page 12), the number of households is reduced to 147,048,720.

We examined the number of data defined people<sup>8</sup> in the total housing unit response universe and by those that were eligible for PSA. Almost half (46.68 percent) of the households in the total response group are inhabited by 1 or 2 people. If we look at only the PSA eligible responses this percentage increases to 49.09 percent.

**Table 5.12 Distribution of Total and PSA Eligible Responses by Household Size**

Household Size	Total Responses		PSA Eligible		Percent of Responses that are PSA Eligible by Household Size
	Number of Households	Percent	Number of PSA Eligible Households	Percent	
<b>Vacant</b>	31,230,049	19.31	22,503,662	15.30	72.06
<b>1</b>	34,071,633	21.07	32,283,307	21.95	94.75
<b>2</b>	41,409,832	25.61	39,903,943	27.14	96.36
<b>3</b>	21,094,615	13.04	20,140,678	13.70	95.48
<b>4</b>	17,840,689	11.03	17,058,326	11.60	95.61
<b>5</b>	8,884,349	5.49	8,384,092	5.70	94.37
<b>6</b>	3,590,472	2.22	3,415,174	2.32	95.12
<b>7</b>	1,804,271	1.12	1,695,792	1.15	93.99
<b>8</b>	845,563	0.52	796,679	0.54	94.22
<b>9</b>	413,725	0.26	391,066	0.27	94.52
<b>10</b>	250,243	0.15	226,052	0.15	90.33
<b>11 to 12</b>	231,941	0.14	208,066	0.14	89.71
<b>13 to 49</b>	44,664	0.03	41,842	0.03	93.68
<b>50 to 97</b>	45	0.00	41	0.00	91.11
<b>TOTAL</b>	<b>161,712,091</b>	<b>100.00</b>	<b>147,048,720</b>	<b>100.00</b>	<b>90.93</b>

Data Source: the 2010 DRF

<sup>8</sup> A data defined person is determined by observing two demographic characteristics corresponding to the reference for that person on the corresponding census form.



For household sizes from one occupant through nine occupants the percentage of responses that are PSA eligible is approximately 95 percent for each household size. However, for household sizes 10 and over this drops to around 91 percent. Over 70 percent (72.06 percent) of vacant households were deemed to be PSA eligible. Note that complex PSA procedures will be applied to the vacant, and occupied, households to determine the best return of the available returns, but ultimately CUF creation determines the final 2010 Census status.

### 5.5 What is the distribution of household sizes at the end of the Primary Selection Algorithm process?

There were 133,243,950 housing units at the end of the PSA process. The post-PSA number of households with one or two residents (based on the number of data defined people in the household) accounts for more than half of the households (51.19 percent).

**Table 5.13 Household Sizes for Primary Selection Algorithm (PSA) Selected Households**

Household Size	Stateside		Puerto Rico		Total	
	Number	Percent	Number	Percent	Number	Percent
<b>Vacant</b>	17,316,958	13.16	263,728	16.35	17,580,686	13.19
<b>1</b>	30,173,770	22.92	320,828	19.89	30,494,598	22.89
<b>2</b>	37,331,110	28.36	380,665	23.60	37,711,775	28.30
<b>3</b>	18,352,687	13.94	274,270	17.00	18,626,957	13.98
<b>4</b>	15,515,289	11.79	224,350	13.91	15,739,639	11.81
<b>5</b>	7,534,123	5.72	98,646	6.12	7,632,769	5.73
<b>6</b>	3,046,877	2.31	31,594	1.96	3,078,471	2.31
<b>7</b>	1,196,573	0.91	10,658	0.66	1,207,231	0.91
<b>8</b>	571,530	0.43	4,561	0.28	576,091	0.43
<b>9</b>	263,861	0.2	1,718	0.11	265,579	0.20
<b>10</b>	154,447	0.12	1,055	0.07	155,502	0.12
<b>11 to 12</b>	137,149	0.10	779	0.05	137,928	0.10
<b>13 to 49</b>	36,560	0.03	124	0.01	36,684	0.03
<b>50 to 97</b>	40	0.00	0	0.00	40	0.00
<b>TOTAL</b>	<b>131,630,974</b>	<b>100.00</b>	<b>1,612,976</b>	<b>100.00</b>	<b>133,243,950<sup>9</sup></b>	<b>100.00</b>

Data Source: the 2010 DRF

<sup>9</sup> The total number of housing units cited here (133,243,950) is slightly lower than the final total cited in Table 5.11 (133,341,676) due to the fact that in some cases PSA was not able to select a form since not all housing units returned their forms or were enumerated during one of the enumeration operations. Imputation procedures were applied to units with no returns and account for the difference of 97,726 housing units between the final 2010 Census total and the PSA selected total.

## 5.6 What is the distribution of the household sizes for all responses at the end of the census?

Table 5.14 presents the distribution of occupied housing units by household size. Over one-quarter (26.70 percent) of all occupied households have only one occupant. When households with two occupants are added to this analysis, almost three-fifths (59.42 percent) of all occupied households at the end of the 2010 Census have two or fewer occupants. In Puerto Rico, the percentage of households with up to two occupants is lower than for the Stateside (52.24 percent and 59.51 percent, respectively). When households with up to four occupants are viewed, the difference nearly disappears (88.96 percent and 89.20 percent, respectively).

**Table 5.14 Household Size for Occupied Housing Units**

Household Size	Stateside		Puerto Rico		Total	
	Number	Percent	Number	Percent	Number	Percent
1	31,204,909	26.74	327,560	23.80	31,532,469	26.70
2	38,242,628	32.77	391,452	28.44	38,634,080	32.72
3	18,757,985	16.07	280,818	20.40	19,038,803	16.12
4	15,625,246	13.39	227,988	16.56	15,853,234	13.42
5	7,538,631	6.46	99,560	7.23	7,638,191	6.47
6	3,074,699	2.63	31,434	2.28	3,106,133	2.63
7	1,180,961	1.01	10,692	0.78	1,191,653	1.01
8	544,251	0.47	4,090	0.30	548,341	0.46
9	257,147	0.22	1,638	0.12	258,785	0.22
10	137,885	0.12	724	0.05	138,609	0.12
11 to 12	122,608	0.11	487	0.04	123,095	0.10
13 to 49	29,308	0.03	88	0.01	29,396	0.02
50 to 97	34	0.00	0	0.00	34	0.00
<b>Total</b>	<b>116,716,292</b>	<b>100.00</b>	<b>1,376,531</b>	<b>100.00</b>	<b>118,092,823</b>	<b>100.00</b>

*Data Source: the 2010 CUF*

## 6. Related Assessments

2010 Decennial Response Integration System (DRIS) Paper Questionnaire Data Capture Assessment

## 7. Key Lessons Learned, Conclusions, and/or Recommendations

Originator: (DSSD)

Lessons Learned:

- Early access to the DRF was extremely helpful for facilitating review of the CUF, but some time was still spent on understanding the variables during the critical CUF review period.
- While errors were prevented with the testing that occurred during the 2010 Census, this both shows the benefit of such testing and demonstrates the need for testing that is as thorough as possible.
- While the PSA information is statistically sensitive, the 2010 implementation required broad operational and data understanding from a select number of people. This put a particular strain on staff during the exhaustive testing process, but there were also some data elements that could have been understood more completely. Responses to operational changes during the census would have benefited from additional scrutiny.

Recommendations:

- Continue to work to ensure that affected parties understand the file variables and variable interactions of interest for their work as early as possible.
- All data exchanges and mappings need to be thoroughly tested, with both subject matter experts and independent parties, including analysts. Testing should occur from start to finish, with specific effort expended on testing alternative or unexpected data paths.
- Include more staff members from more areas in the PSA determination process.

Originator: (POP)

Lessons Learned:

- Update GQ Usual Home Elsewhere MAFIDs (UHE\_MAFID) when the original UHE\_MAFID value is updated through the Surviving MAFID process.
- Implement more thorough testing to ensure that all operations/forms are feeding in expected data. This became a problem with the Shipboard Census Report forms originally not providing data for geocoding because the screener flags were mismapped by DRIS. We might have caught this earlier; however we caught this in production and implemented a fix.

- There should be more testing earlier of systems such as the Edit Review System, which was designed to support POP/Housing and Household Economic Studies review of decennial data. Also, there was a problem with lack of flexibility in requirements management -- needed changes could not be implemented due to timing problems.
- POP was the sponsor of the DSCMO Count Review system, the design of which did not support functionalities the program later determined necessary. As a result, the DSCMO system had to be supplemented by an additional contractor-built system, which conformed much more closely to final program requirements.
- The elaboration of requirements, the careful effort of the developers, and the testing protocols implemented resulted in few (if any) errors in CEF data.
- Some design constraints on the Universal Response database Schema (namely, the use of the paper questionnaire fields and field lengths) as standards for file layouts containing data from multiple modes proved too limiting, particularly for long and complicated strings that may have been collected for race and Hispanic origin. However, the interactive design process, incorporating both DRIS and Headquarters stakeholders in common meetings to review and comment, was an effective approach.

Recommendations:

- Employ greater flexibility in early design and development processes.
- Implement more thorough testing to ensure that all operations/forms are feeding in expected data.

Originator: (DSCMO)

Lessons Learned:

More rigorous 'user testing' of the residual coding Quality Control (QC) component is needed to ensure system meets stakeholder expectations. Specifically, POP questioned the accuracy of the QC program. Following wave 2 of residual coding, POP went in and fixed items. Due to constraints in the programming schedule due to the developer leaving, testing may have been conducted too briskly.

Recommendations: Make user testing of the QC program part of the schedule.

Originator: (DSCMO)

Lessons Learned:

The Graphical User Interface (GUI) did not meet requirements at first demonstration on the Race and Hispanic Origin Coding System. The code

development proceeded for several months without review to ensure that the interface was meeting specifications and usability standards. Once a demonstration was given (near the end of the development period) it was clear that the system was not ready and nearly unusable. There was a great deal of development and rework on the code even during production.

Recommendations:

There should be periodic reviews of the GUI interface to ensure that requirements are being met.

Originator: (DSCMO)

Lessons Learned:

Future deliveries of label files (or similar) to outside service providers such as printing vendors with whom we have no Interface Control Document or Product Services interface, must continue to go through a designated delivery coordinator. Specifically, the original plan for delivery of label files to print vendors for the 2010 Census was for electronic transfer. A variety of security issue caused changes to the original plan, eventually leading to use of removable optical disk storage media for file transfer, rather than electronic means. The variability in the potential number of print vendors and file delivery methodologies led to a flow design whereby DACMO contracted with the testing site staff (BETA) to act as a provider to; a) receive the label files from Head Quarters Processing (HQP), b) prepare appropriate transfer media with applied security measures, and c) coordinate with DACMO and the printing vendors for successful transfer and receipt of the label files. HQP's technical responsibilities were to deliver quality files to BETA.

Analysis: While there was increased risk due to additional parties being involved between file production by HQP and file receipt by the printing vendor, the risk was acceptable in consideration of the value added by having a qualified third party (such as BETA) serve as an intermediary.

Recommendations:

Because HQP does not support secure removable media, or other non-Product Services secure data delivery mechanisms appropriate for external customers, including but not limited to printing vendors, it is appropriate and necessary that a qualified staff (such as BETA) continue to be tasked as an intermediary for such deliveries.

## 8. Acknowledgments

I would like to thank the following individuals for their valuable contributions to this assessment:

### DSSD

Jerry Imel  
Robin Pennington  
Ryan King  
Keith Wechter

### POP

Mike Cardella

### DSCMO

Charles Kahn  
Michael Clark  
Gerard Moore  
Thomas McCoy  
James Cope

## 9. References

Clark, Michael (2010) - 2010 Universe Control & Management High Level Design - DOCUMENT NO.: PSS-DEV-UCMHLD-2010SD - Revision 6.0, US Census Bureau

Cope, James (2010) - 2010 Universe Control & Management (UCM) System Requirements Specification - DOCUMENT NO.: PSS-MGMT-UCMSRS-2010SD - Revision 8.0, US Census Bureau

Jonas, Kim (2003) - Census Unedited File Creation - Final Report - Census 2000 Evaluation L.4, US Census Bureau

McCoy, Tom (2010) - RPS High Level Overview, US Census Bureau, September 20, 2010

Pennington, Robin A. (2010) - Clarification for Analysts using the Decennial Response File, Census Unedited File, and Census Edited File, US Census Bureau

Project Management Branch, Decennial Management Division - 2010 Census Operational Plan – 2010 Census Information Memoranda Series #2, US Census Bureau

## Appendix A: UCM/RPS Assessment Acronyms List

AC	Address Canvassing
CFU	Coverage Followup
CEF	Census Edited File
CUF	Census Unedited File
DRF	Decennial Response File
DRIS	Decennial Response Integration System
DSCMO	Decennial Systems and Contract Management Office
DSF	Delivery Sequence File
DSSD	Decennial Statistical Studies Division
ERS	Edit Review System
FV	Field Verification
GQ	Group Quarters
GUI	Graphical User Interface
HDF	Hundred Percent Detail File
HQP	Head Quarters Processing
HU	Housing Unit
IAP	Interface and Acknowledgment Processing
MAF	Master Address File
MAFID	Master Address File Identifier
MAFX	MAF Extract
MO/MB	Mailout/Mailback
NRFU	Nonresponse Followup
PBO	Paper Based Operations
PCDI	Post Capture Data Integration System
POP	Population Division
PSA	Primary Selection Algorithm
QC	Quality Control
RPS	Response Processing System
U/L	Update/Leave
UCM	Universe Control and Management
VDC	Vacant/Delete Check