Paradata at the U.S. Census Bureau (and Where Blaise Fits In)

Michael K. Mangiapane, U.S. Census Bureau

1 Introduction

Paradata, or data about the survey process itself, has been collected in various forms for decades. In recent years it has become a topic of interest among statistical agencies as they look for ways to perform their survey operations more efficiently and at a lower cost. The U.S. Census Bureau is pursuing a number of different efforts to collect and analyze current and future paradata to improve the quality of the data they collect, reduce burden on their Field Representatives (FR's), and bring the costs of a survey down in a time of leaner budgets.

2 Current and Future Paradata Efforts

There have been a number of paradata-related efforts at the Census Bureau in use for several years.

- **CHI** – We have been collecting and analyzing information about interview contact attempts for some surveys using our Contact History Instrument (CHI) since 2004. In 2011 we implemented CHI for all of our demographic surveys and developed a facility-based CHI for our Ambulatory Medical Care surveys. The CHI will be discussed in more detail in this paper.
- **PANDA/Giant PANDA** – The Demographic Surveys Division (DSD) developed the Performance and Data Analysis (PANDA) tool to summarize key metrics of the quality of data collection, such as item missing rate and length of interview (from audit trail fields). This system is being expanded to handle more of our demographic surveys.
- **CARMN** – The Technologies Management Office (TMO) developed the Cost and Response Management Network (CARMN) system to report on field representatives' cost and performance, using data from the CAPI case management system and the FR's payroll system.
- **CARI** – The Census Bureau has been testing Computer-assisted Audio Recorded Interviewing (CARI) for several years now and last year conducted a test with a new CARI monitoring system with the American Community Survey (ACS). We have yet to put any CARI projects into full-time production though.

However, there are some new, more comprehensive efforts taking place now.

- **Efforts Related to Field Restructuring –** In 2011, the Census Bureau embarked on a major restructuring of our field offices and management structure. Our twelve (12) regional offices (RO's) will be reduced to six (6) by the end of 2012, and many of the survey supervisors who previously worked out of an RO will now be working out of their homes. The supervisory chain for the FR's has also changed to be a Field Supervisor (FS) instead of a survey supervisor in the RO. Due to this significant change, specific efforts have been undertaken to monitor the data quality and response data during this period of change. Some of the work done with the Data Combing Process (DCP), described in more detail in this paper, was related to providing data for these efforts.
- **UTS –** The ultimate goal for the Census Bureau's paradata efforts are the development of a Unified Tracking System (UTS) that will provide integrated survey progress, cost, and data quality metrics in a centralized repository for use in survey management and analysis. This system is discussed in more detail in this paper.

3 Objectives of This Paper

This paper will go into more detail about the various ways that the Census Bureau is using Blaise to collect paradata and how using this data will continue the Census Bureau's mission to serve as the leading source of quality data about the United States' people and economy. The topics to be covered include:

- Using CHI for gathering household, facility, or person-based contact attempt information for a case.
- Creation of a Data Combing Process (DCP) that runs in Manipula and uses a list of variables provided by a stakeholder to capture and output paradata to a Blaise Audit Trail or a separate ASCII file.
- Parsing and analysis of Blaise Audit Trails and CARI files and logs to provide paradata.
- Using the above-named tools to provide input to the new Unified Tracking System (UTS) under development at the Census Bureau.

4 Contact History Instrument (CHI)

In 2002, the Census Bureau and a group known as the Interagency Household Survey Nonresponse Group (IHSNG) held a conference to discuss a concern about the decline of response rates and a rise in refusals for household surveys. The summit primarily focused on the National Health Interview Survey (NHIS) and Consumer Expenditure Quarterly (CEQ) survey. Among the topics discussed was the idea to collect "call records" that would contain data about contact attempts for a case. They wanted to see the history of a case with information such as how many times and when a household was contacted, reasons that the respondent was reluctant to complete the survey, and the contact strategies the FR implemented in attempted to complete the interview.

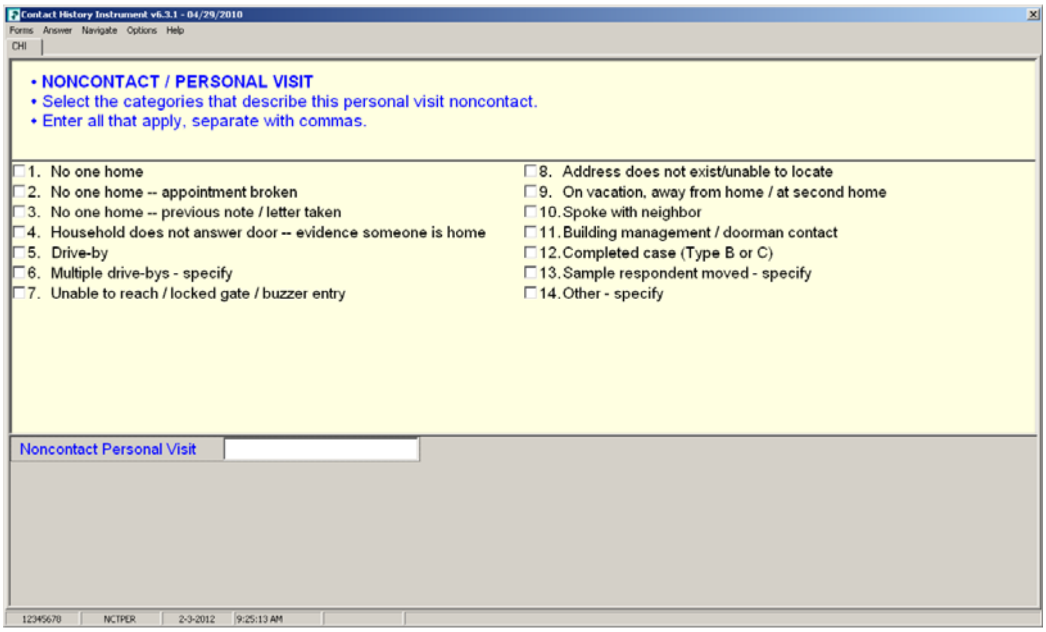
They hoped to be able to use this data to better discern why respondents tended to refuse surveys, when was the best time to attempt an interview and what strategies should be implemented to get responses. The group decided that it would be best to automate this process despite the initial costs. They also decided that the process should contain standard questions and answers across surveys so that data could be compared and evaluated the same way. The questions and answer choices could be modified if all areas agreed the modification was needed. From this effort, the Contact History Instrument (CHI) was created.

4.1 CHI Process

The CHI automatically runs after the FR exits the survey instrument so that the interviewer keys in the contact attempt while it is still fresh. It can also be called manually from Case Management for situations where the FR does not open the survey instrument during a contact attempt. After the FR exits the Blaise CHI instrument, an ASCII output file is generated and processed by Case Management so the data can be transmitted back to the Regional Office Survey Control (ROSCO) system.

There are three main sections of CHI used to collect information about a contact attempt – 1. Noncontact information, 2. Concern/Behavior/Reluctance information, and 3. Contact Strategies Attempted.

4.1.1 Noncontact Information


Contact History Instrument v6.3.1 - 04/29/2010
Forms Answer Navigate Options Help
CHI

• NONCONTACT / PERSONAL VISIT
• Select the categories that describe this personal visit noncontact.
• Enter all that apply, separate with commas.

□1. No one home
□2. No one home -- appointment broken
□3. No one home -- previous note / letter taken
□4. Household does not answer door -- evidence someone is home
□5. Drive-by
□6. Multiple drive-bys - specify
□7. Unable to reach / locked gate / buzzer entry
□8. Address does not exist/unable to locate
□9. On vacation, away from home / at second home
□10. Spoke with neighbor
□11. Building management / doorman contact
□12. Completed case (Type B or C)
□13. Sample respondent moved - specify
□14. Other - specify

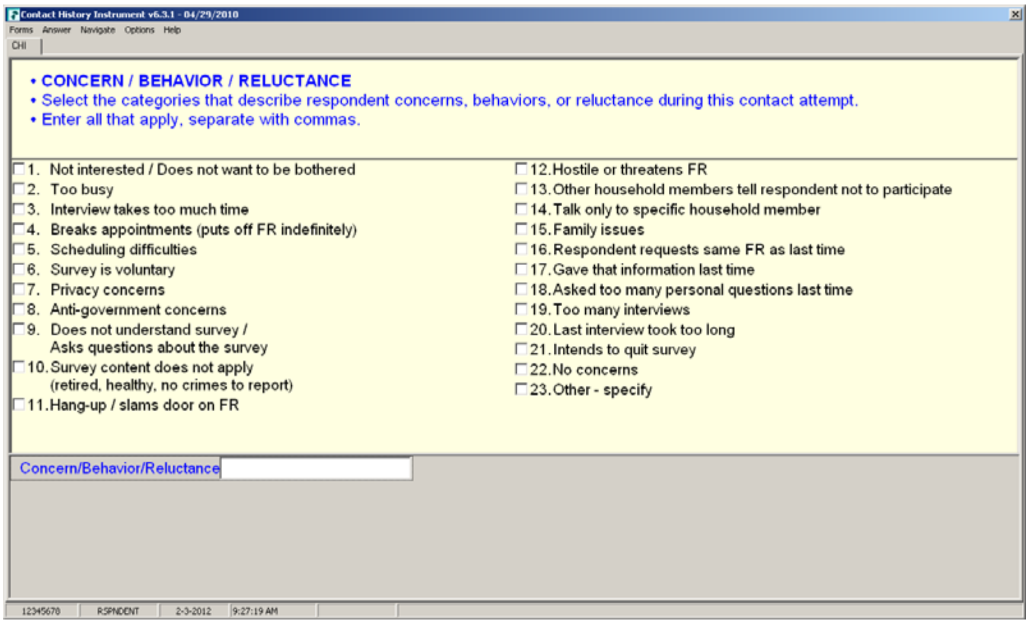
Noncontact Personal Visit

12345670 NCTPER 2-3-2012 9:25:13 AM

Fig. 1 - CHI Non-Contact screen

Inside the CHI, the FR is asked how they attempted to make contact with the respondent, by personal visit or by phone. If they did not make contact with the respondent, they are taken to this section to mark the outcome of their non-contact (e.g. No one home; locked gate; spoke with neighbor). They then enter any strategies they attempted to make contact in the "Contact Strategies Attempted" section. The CHI then exits and creates output.

4.1.2 Concern/Behavior/Reluctance Information


Contact History Instrument v6.3.1 - 04/29/2010
Forms Answer Navigate Options Help
CHI

• CONCERN / BEHAVIOR / RELUCTANCE
• Select the categories that describe respondent concerns, behaviors, or reluctance during this contact attempt.
• Enter all that apply, separate with commas.

□1. Not interested / Does not want to be bothered
□2. Too busy
□3. Interview takes too much time
□4. Breaks appointments (puts off FR indefinitely)
□5. Scheduling difficulties
□6. Survey is voluntary
□7. Privacy concerns
□8. Anti-government concerns
□9. Does not understand survey / Asks questions about the survey
□10. Survey content does not apply (retired, healthy, no crimes to report)
□11. Hang-up / slams door on FR
□12. Hostile or threatens FR
□13. Other household members tell respondent not to participate
□14. Talk only to specific household member
□15. Family issues
□16. Respondent requests same FR as last time
□17. Gave that information last time
□18. Asked too many personal questions last time
□19. Too many interviews
□20. Last interview took too long
□21. Intends to quit survey
□22. No concerns
□23. Other - specify

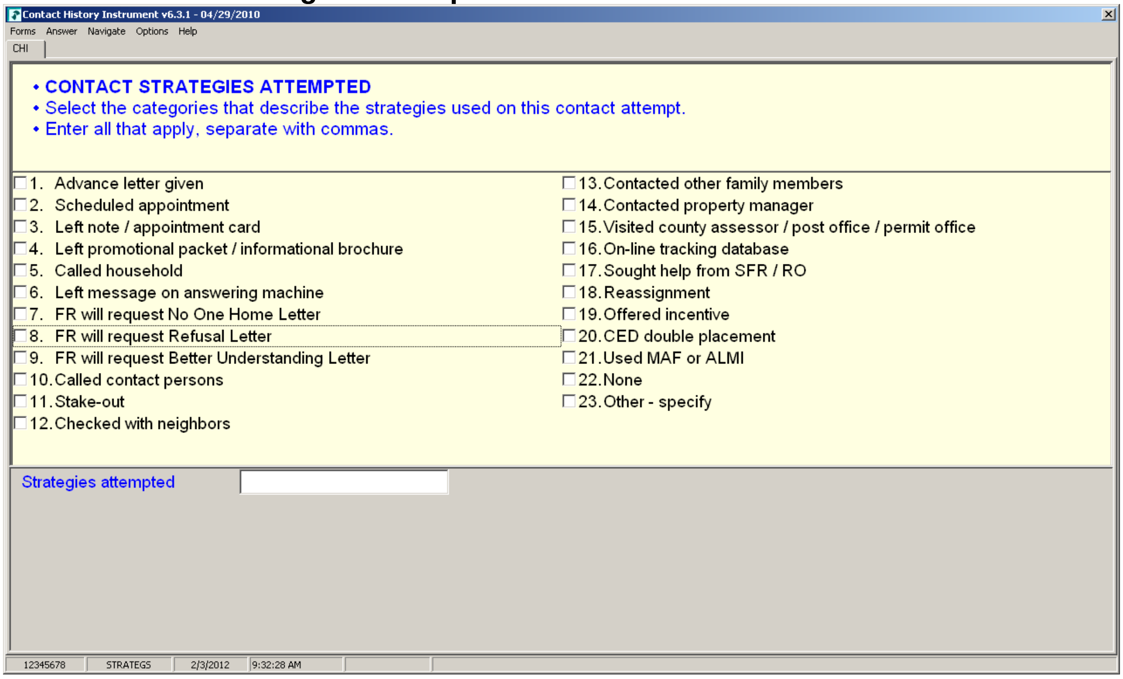
Concern/Behavior/Reluctance

12345670 RSPNDENT 2-3-2012 9:27:19 AM

Fig. 2 – CHI Concern/Behavior/Reluctance Screen

If the FR does make contact with a respondent, they are taken through this section to answer a few questions about the nature of the contact. They are asked to enter in all Concerns/Behaviors/Reluctances that apply to this contact attempt.

4.1.3 Contact Strategies Attempted



- CONTACT STRATEGIES ATTEMPTED
- Select the categories that describe the strategies used on this contact attempt.
- Enter all that apply, separate with commas.

□ 1. Advance letter given
□ 2. Scheduled appointment
□ 3. Left note / appointment card
□ 4. Left promotional packet / informational brochure
□ 5. Called household
□ 6. Left message on answering machine
□ 7. FR will request No One Home Letter
□ 8. FR will request Refusal Letter
□ 9. FR will request Better Understanding Letter
□ 10. Called contact persons
□ 11. Stake-out
□ 12. Checked with neighbors
□ 13. Contacted other family members
□ 14. Contacted property manager
□ 15. Visited county assessor / post office / permit office
□ 16. On-line tracking database
□ 17. Sought help from SFR / RO
□ 18. Reassignment
□ 19. Offered incentive
□ 20. CED double placement
□ 21. Used MAF or ALMI
□ 22. None
□ 23. Other - specify

Strategies attempted

12345678 STRATEGS 2/3/2012 9:32:28 AM

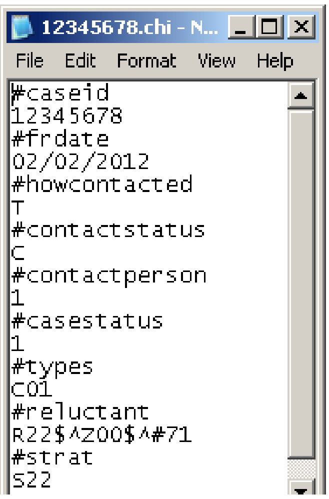
Fig. 3 – CHI Contact Strategies Attempted Screen

Regardless of the outcome of a contact attempt, the FR will need to complete the "Contact Strategies Attempted section. Here they will mark all strategies that apply to this contact attempt.

Once the FR completes and exits the CHI, output is created.

4.2 CHI Output

Upon exit, the CHI transaction process creates an ASCII file containing information from the CHI. The output is coded for processing efficiency and so that future enhancements can be easily implemented.

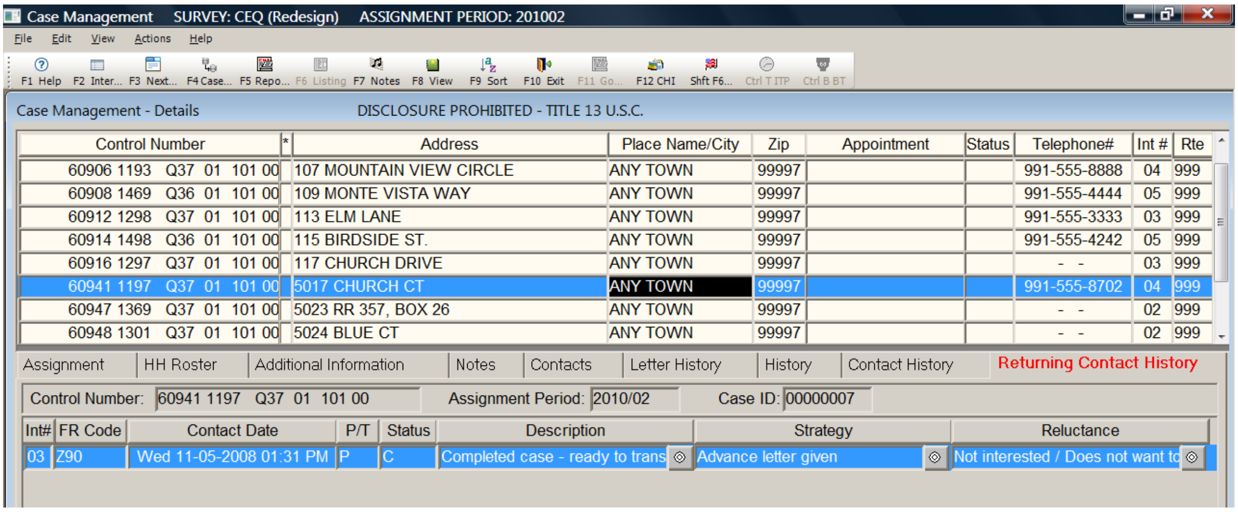


```
#caseid
12345678
#frdate
02/02/2012
#howcontacted
T
#contactstatus
C
#contactperson
1
#casestatus
1
#types
C01
#reluctant
R22$^Z00$^#71
#strat
S22
```

Fig. 4 – Example CHI Output

For this particular case, the output shows that for this contact attempt, the respondent was contacted by telephone and the attempt resulted in a completed case (How contacted, Contact Status, Types). The FR reported no concerns with respondent behavior (R22 in #reluctant), and did not employ any strategies (S22 in #strat). It is also reported that the CHI case was open for 71 seconds. This CHI data and thousands of others like it are collected and analyzed to learn about the ever-changing nature of our survey processes and respondents. The CHI data also stays with the case so when the case comes back to an FR in the following month, quarter, or year, the FR can review it before they attempt to contact the respondent.

Below is an example of returning CHI data in Case Management. The CHI data was recorded when a contact attempt was made for this case when it was previously interviewed. (November 2008).


Case Management SURVEY: CEQ (Redesign) ASSIGNMENT PERIOD: 201002
File Edit View Actions Help
F1 Help F2 Inter... F3 Next... F4 Case... F5 Repo... F6 Listing F7 Notes F8 View F9 Sort F10 Exit F11 Go... F12 CHI Shft F6... Ctrl T ITP Ctrl B BT
Assignment HH Roster Additional Information Notes Contacts Letter History History Contact History Returning Contact History

Case Management - Details DISCLOSURE PROHIBITED - TITLE 13 U.S.C.

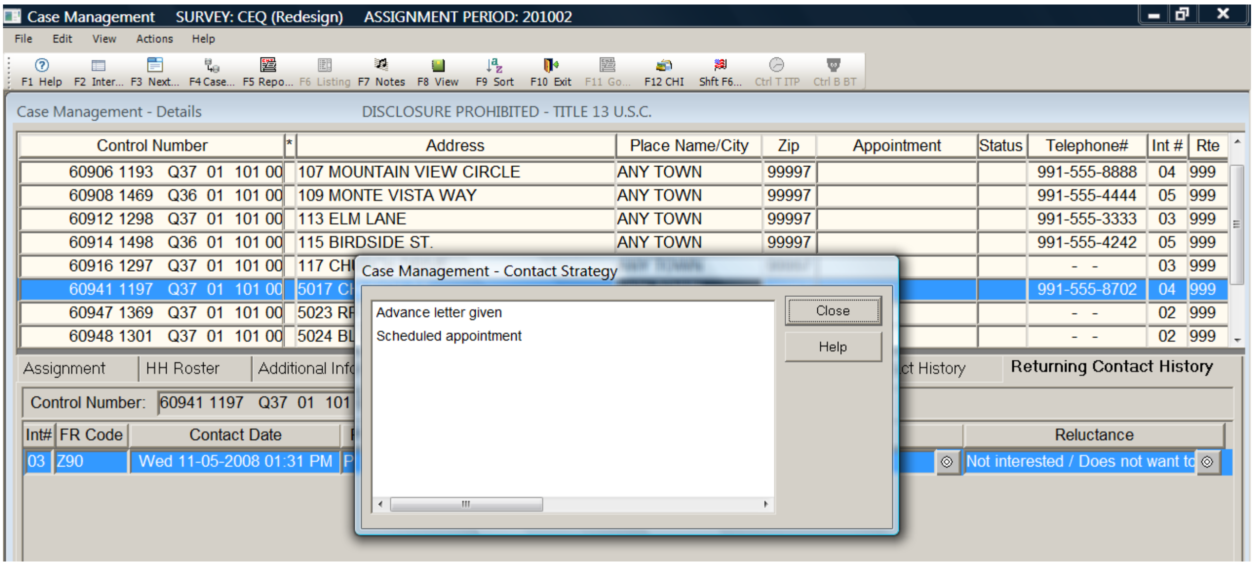
Control Number	*	Address	Place Name/City	Zip	Appointment	Status	Telephone#	Int #	Rte
60906 1193 Q37 01 101 00		107 MOUNTAIN VIEW CIRCLE	ANY TOWN	99997			991-555-8888	04	999
60908 1469 Q36 01 101 00		109 MONTE VISTA WAY	ANY TOWN	99997			991-555-4444	05	999
60912 1298 Q37 01 101 00		113 ELM LANE	ANY TOWN	99997			991-555-3333	03	999
60914 1498 Q36 01 101 00		115 BIRDSIDE ST.	ANY TOWN	99997			991-555-4242	05	999
60916 1297 Q37 01 101 00		117 CHURCH DRIVE	ANY TOWN	99997			- -	03	999
60941 1197 Q37 01 101 00		5017 CHURCH CT	ANY TOWN	99997			991-555-8702	04	999
60947 1369 Q37 01 101 00		5023 RR 357, BOX 26	ANY TOWN	99997			- -	02	999
60948 1301 Q37 01 101 00		5024 BLUE CT	ANY TOWN	99997			- -	02	999

Control Number: 60941 1197 Q37 01 101 00 Assignment Period: 2010/02 Case ID: 00000007

Int#	FR Code	Contact Date	P/T	Status	Description	Strategy	Reluctance
03	Z90	Wed 11-05-2008 01:31 PM	P	C	Completed case - ready to trans	Advance letter given	Not interested / Does not want to

Fig. 5 – Returning CHI Data in Case Management

The FR is able to see which FR last made a contact attempt, when the contact attempt was made, the type of contact and the status of the contact. They can also review any attempted strategies and behavior or reluctance concerns. The FR may click icons next to the Description, Strategy, or Reluctance fields for more information.


Case Management SURVEY: CEQ (Redesign) ASSIGNMENT PERIOD: 201002
File Edit View Actions Help
F1 Help F2 Inter... F3 Next... F4 Case... F5 Repo... F6 Listing F7 Notes F8 View F9 Sort F10 Exit F11 Go... F12 CHI Shft F6... Ctrl T ITP Ctrl B BT
Assignment HH Roster Additional Info ct History Returning Contact History
Close
Help

Case Management - Details DISCLOSURE PROHIBITED - TITLE 13 U.S.C.

Control Number	*	Address	Place Name/City	Zip	Appointment	Status	Telephone#	Int #	Rte
60906 1193 Q37 01 101 00		107 MOUNTAIN VIEW CIRCLE	ANY TOWN	99997			991-555-8888	04	999
60908 1469 Q36 01 101 00		109 MONTE VISTA WAY	ANY TOWN	99997			991-555-4444	05	999
60912 1298 Q37 01 101 00		113 ELM LANE	ANY TOWN	99997			991-555-3333	03	999
60914 1498 Q36 01 101 00		115 BIRDSIDE ST.	ANY TOWN	99997			991-555-4242	05	999
60916 1297 Q37 01 101 00		117 CHU					- -	03	999
60941 1197 Q37 01 101 00		5017 C					991-555-8702	04	999
60947 1369 Q37 01 101 00		5023 RR					- -	02	999
60948 1301 Q37 01 101 00		5024 BL					- -	02	999

Control Number: 60941 1197 Q37 01 101

Int#	FR Code	Contact Date	P	Reluctance
03	Z90	Wed 11-05-2008 01:31 PM	P	Not interested / Does not want to

Case Management - Contact Strategy

Advance letter given
Scheduled appointment

Fig. 6 – Detailed look at previous contact strategy in returning CHI data in Case Management

4.3 CHI in the Field

CHI was first put into production for NHIS in 2004. It is now in production for every one of the Census Bureau's household surveys, and we have a facility-based CHI in production with our Ambulatory Medical Care surveys. There is an undertaking to create a person-based CHI, which will be explained later in this paper. During the testing process prior to production, there were some concerns from the FR's that CHI data would be used against them (i.e. "Big Brother") but this was not what the sponsors were interested in. The FR's found that the CHI was useful as a quick way to review their contact attempts with the household, or to review contact attempts that other FR's had made with the household. Prior to CHI, they were entering notes about their contact attempts in their case-level notes for each case. CHI standardized the process of recording their contact attempts, and it meant they spent less time typing notes about their contact. On average, an new FR spends 90 seconds in CHI, and an experienced FR spends 60 seconds.

4.4 CHI Data Usage

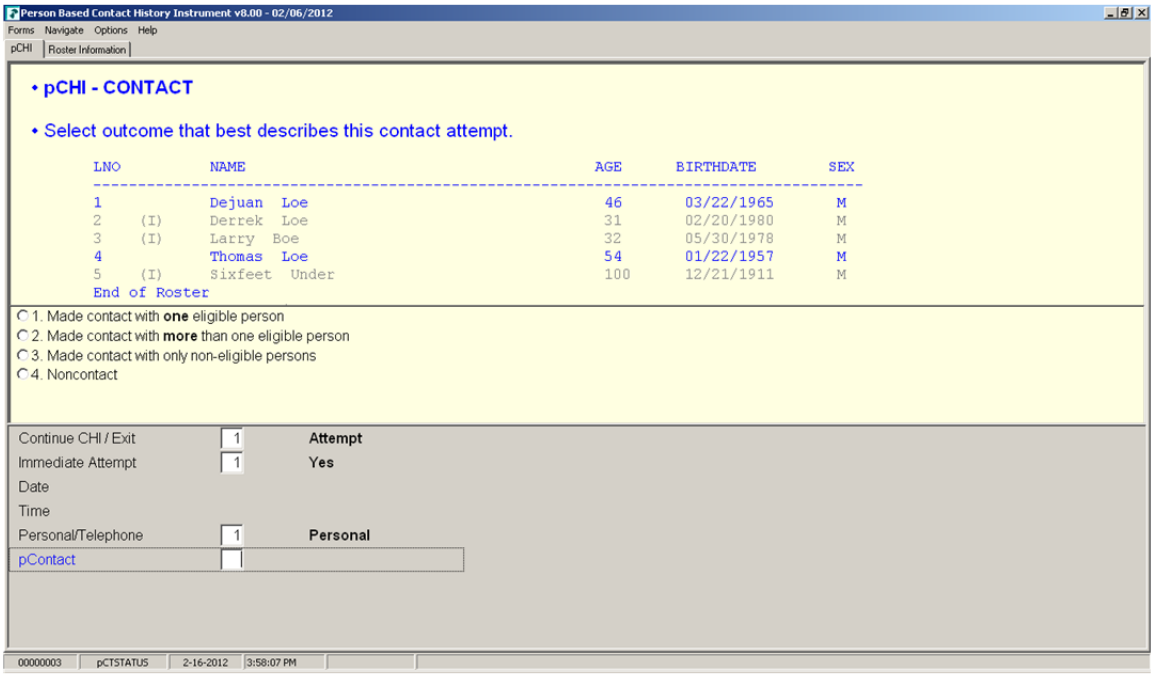
The CHI data collected has become key paradata for the Census Bureau in how they make contact with households and the responses they get. There have been a number of papers written about usage of the CHI data and what they have found from it, including which behaviors are most likely to get an interim refusal, meaning the FR will be refused before they get a survey response, and which ones will be a final refusal, meaning the FR will not get a survey response. The data also indicated that some regions of the United States are more likely to have final refusals than interim refusals, and that personal visits had a better success rate of getting survey responses than telephone contacts (Bates, Dalhamer, and Singer 2008:591). Other research has shown which times may be the best times for making the first contact attempt and the average number of contact attempts made that result in an interview or non-interview (Tan 2011:9).

CHI made it easier for the survey sponsors to review respondent behaviors and effectiveness of contact strategies because they were not reviewing case notes for thousands of cases to record responses vs. non-responses, the CHI data can be aggregated and analyzed instead.

4.5 Person-based CHI (pCHI)

When the National Crime Victimization Survey (NCVS) implemented CHI, the sponsor was concerned about collecting contact attempt data for each member of the household. NCVS interviews all persons 12 and older in the household; that is, every eligible household member has to provide their own answers to the survey questions. Since the standard household CHI assumes that you are only attempting to contact one person at the household, enhancements were requested so that contact attempt information could be collected for each respondent in the household. A new person-based CHI instrument is being developed at the Census Bureau.

The person-based CHI will read in the household roster when the instrument is started. When the FR reaches the question to ask about contact or non-contact, they will indicate if they made contact with one or more eligible persons, no eligible persons, or if it was a non-contact.


Person Based Contact History Instrument v8.00 - 02/06/2012
Forms Navigate Options Help
pCHI
Roster Information

• pCHI - CONTACT

• Select outcome that best describes this contact attempt.

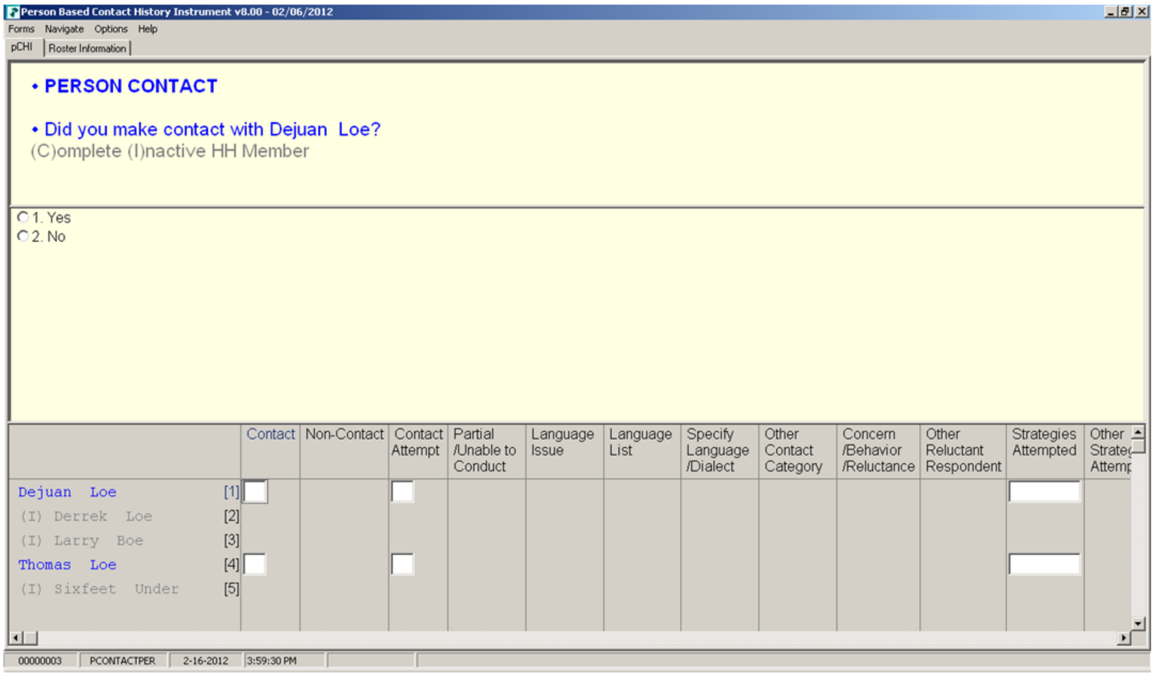
```
LNO          NAME                                   AGE       BIRTHDATE       SEX
-----------------------------------------------------------------------------------
1            Dejuan  Loe                             46        03/22/1965      M
2     (I)    Derrek  Loe                             31        02/20/1980      M
3     (I)    Larry  Boe                              32        05/30/1978      M
4            Thomas  Loe                             54        01/22/1957      M
5     (I)    Sixfeet  Under                          100       12/21/1911      M
End of Roster
```

○ 1. Made contact with **one** eligible person
○ 2. Made contact with **more** than one eligible person
○ 3. Made contact with only non-eligible persons
○ 4. Noncontact

Continue CHI / Exit	1	**Attempt**
Immediate Attempt	1	**Yes**
Date		
Time		
Personal/Telephone	1	**Personal**
pContact		

Fig. 7 – Person-Based CHI Contact Screen

If contact was made, the FR is taken to a table that displays the full roster and puts on-path the person(s) who were eligible for the survey and require CHI data to be collected. The CHI data is then collected for each person.


Person Based Contact History Instrument v8.00 - 02/06/2012
Forms Navigate Options Help
pCHI
Roster Information

• PERSON CONTACT

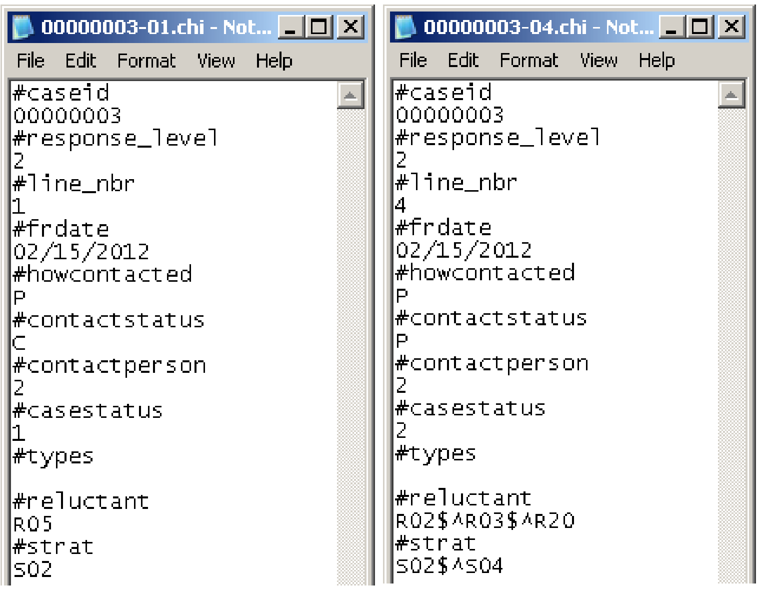
• Did you make contact with Dejuan Loe?
(C)omplete (I)nactive HH Member

○ 1. Yes
○ 2. No

		Contact	Non-Contact	Contact Attempt	Partial /Unable to Conduct	Language Issue	Language List	Specify Language /Dialect	Other Contact Category	Concern /Behavior /Reluctance	Other Reluctant Respondent	Strategies Attempted	Other Strateg Attemp
Dejuan Loe	[1]	□		□								□	
(I) Derrek Loe	[2]												
(I) Larry Boe	[3]												
Thomas Loe	[4]	□		□								□	
(I) Sixfeet Under	[5]												

Fig. 8 – CHI data collection table for eligible household members.

In the above example, only two members of the roster are eligible household members in NCVS so they are the only ones that CHI will collect data for. Just like CHI, Person-based CHI will collect the same information asked at the household-level, but for each eligible person in the roster. After the CHI is finished, an output file is produced for each line of the roster that had CHI information collected. There are a few lines that are different from CHI, but the codes used for "reluctance" and "strategies" are the same.



```
#caseid
00000003
#response_level
2
#line_nbr
1
#frdate
02/15/2012
#howcontacted
P
#contactstatus
C
#contactperson
2
#casestatus
1
#types

#reluctant
R05
#strat
S02
```

```
#caseid
00000003
#response_level
2
#line_nbr
4
#frdate
02/15/2012
#howcontacted
P
#contactstatus
P
#contactperson
2
#casestatus
2
#types

#reluctant
R02$^R03$^R20
#strat
S02$^S04
```

Fig. 9 – Example CHI output for person-based CHI for two eligible members

5 Data Combing Process (DCP)

Prior to the creation of the Unified Tracking System (UTS), the sponsors of the NCVS were interested in capturing their own paradata from Blaise. The original request was to capture the values of certain variables and include them in the audit trail file. They also wanted to know which FR completed each question if the case had been reassigned. Originally we pursued writing our own data to the audit trail file as the interview progressed using Maniplus[1]. For example, upon starting the interview, we would write a line with our own defined tag that included the FR's ID code so we could know that the questions from that point forward were completed by that FR. Our thinking then evolved to capturing the data we needed each time the FR exited the instrument and append it to the end of the audit trail file

5.1 Initial Research

In trying to implement the second approach, our first thought was to start with an existing Manipula script that our Master Control System (MCS) uses in their data processing to output data for a defined list of variables. This script pulls requested data from the Blaise database and outputs it as a text file for our Demographic Survey Methods Division (DSMD) to pick up key variables they need for the sample control system. With a few modifications to the code, this script would be placed on an FR's laptop where it would run and output the variables to the audit trail. Unfortunately, that script only read variables that were at the datamodel level and it required using Chameleon to create a separate datamodel and variable list before the script would run. Since this meant extra time and overhead for processing and changing our instruments (to pass variables to the instrument datamodel level) just to be able to produce output, the output script from MCS would not be ideal for running on the laptops. It was clear that creating a new script would have a better chance of fulfilling these requirements.

Since Blaise 4.8 had introduced many new features within Manipula, we decided to research its methods of obtaining values from a Blaise database to see if there was one that would fulfill our needs. The discovery of the GETVALUE method intrigued us since as a function it "Retrieves the value of a field of which the name is not known at prepare time" (Blaise Help File). Based on that

[1] To allow an external program to write to the audit trail, we set the "CloseFile" toggle in the .aif file to "1." This will open and close the audit trail every time a line is written to it, thus removing the Blaise generated lock file associated with that case and allowing Maniplus to write directly to the audit trail.

statement and the examples given, it meant that yes, we would be able to use any variable that existed in a Blaise database, so long as we used the fully-qualified field name.

5.2 Proof of Concept

The first DCP script used some hard-coded variable names as a proof of concept, and was designed to run on the FR laptop. It would open the audit trail and append the values at the end of the audit trail, in a line that looked something like this:

#PARADATA;value1;value2;value3;value4;...;#ENDPARADATA

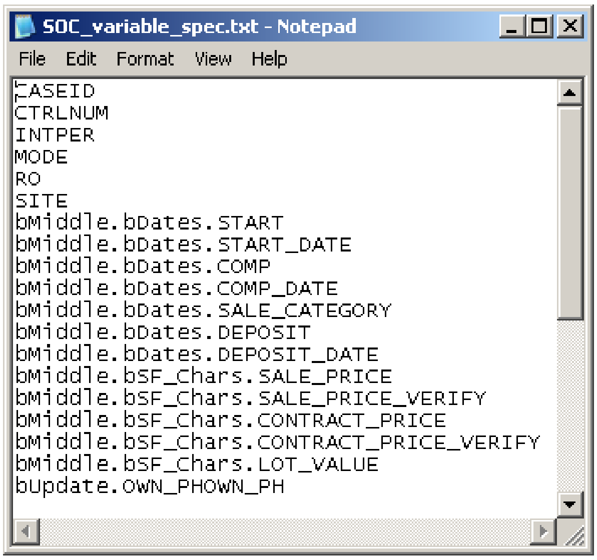
By including the #PARADATA and #ENDPARADATA, it would make it easier to parse the entire audit trail for a case and capture the included paradata.

5.3 Redesign

As this proof-of-concept was created and tested, the UTS was entering its design phase and the UTS designers wanted to capture similar data from all surveys rather than just for NCVS. At about the same time, DSMD needed to get data for numerous variables from several instruments, in order to monitor the effect of field restructuring on interview data quality. We decided to combine efforts and “comb” Blaise data for either UTS, DSMD, or a survey sponsor. With this combined effort came a redesign of the output of the script as both parties wanted to collect many variables. To reduce case processing and data transmission times on the laptops, we decided to comb the data after a case was checked back into headquarters. This also changed when the combing process would run since MCS would now handle the process rather than the FR laptops.

5.4 DCP Script Process

The DCP was built in Manipula using Blaise 4.8.2 Build 1606. It must be compiled with the datamodel of the instrument it will query, and it is set up to query a consolidated Blaise database that contains all of the case records. For our purposes we pass in the name of a Blaise database that contains all of the cases that have been checked in from the field, a list of ID’s of cases that were checked in that day, and a “qualified variable list” (QVL) of fields to be collected (provided by the sponsors in a text file). After creating or opening an ASCII file for output, the DCP steps through the list of case ID’s and begins to comb through the data to collect for output.



```
CASEID
CTRLNUM
INTPER
MODE
RO
SITE
bMiddle.bDates.START
bMiddle.bDates.START_DATE
bMiddle.bDates.COMP
bMiddle.bDates.COMP_DATE
bMiddle.bDates.SALE_CATEGORY
bMiddle.bDates.DEPOSIT
bMiddle.bDates.DEPOSIT_DATE
bMiddle.bSF_Chars.SALE_PRICE
bMiddle.bSF_Chars.SALE_PRICE_VERIFY
bMiddle.bSF_Chars.CONTRACT_PRICE
bMiddle.bSF_Chars.CONTRACT_PRICE_VERIFY
bMiddle.bSF_Chars.LOT_VALUE
bUpdate.OWN_PHOWN_PH
```

Fig. 10 - Example QVL

To collect the data, the script first checks on the status of the field. If there is a response, the script uses GETVALUE to add the value from the field to the output.

```
ConsolidatedDB.GETVALUE(InputFields.VarName, UF)
```

We also use the UF option in GETVALUE so that enumerated fields are output as a number rather than the mnemonic of the number. For non-responses, we print DK, RF, or nothing but the delimiter if the field is Empty.

After the script finishes, the output generated is one file with all the data combed out of the cases, or an individual file for each case. A typical output file will look like this (individual files will only have one line)

```
3003901N|222222222N|201010SI|0||30|DK|DK|203|20111208|20111208|121924|121945|3.28|
3003902N|333333333N|201010SI|0||30|1|1|DK|203|20111208|20111208|102202|102321|1.19|
```

We decided to use the pipe character as a delimiter because typical delimiters such as a comma or semicolon may be a part of a string field, and that would throw off the processing of the output since the values are in the order of the variable list.

A test of the script showed that it runs very fast to produce output. In a scenario of combing for 306 data values per case, the script processed 270 cases in only 32 seconds. This was much faster than expected and meant it would not have a great impact on the overhead of our processing systems.

Recently a request to enhance the script has been requested by the survey sponsors, asking if it could provide a count of the number of don't knows, refusals, and fields that were asked but left empty in an interview. While Manipula does have built-in functionality to count don't knows, refusals and responses, it does not have the functionality to count fields that were asked but left empty. We are currently exploring using ROUTESTATUS to find the empty fields, even if there is a keep on that field.

6 The Unified Tracking System (UTS)

The Census Bureau is attempting to drastically change the way it manages its survey data collection processes, including using responsive design to improve quality and/or reduce costs. There was a need for a system which would gather all of the paradata that is currently being collected by various processes such as CHI data, data extracted using the DCP, and the expansion into collecting more paradata for a survey, such as CARI data. Not only should this system be able to store data from all of these different processes, it needs to be able to present it to the users in a user-friendly format for analysis. An initiative is underway at the Census Bureau to create such a system, which we call the Unified Tracking System.

6.1 UTS Design

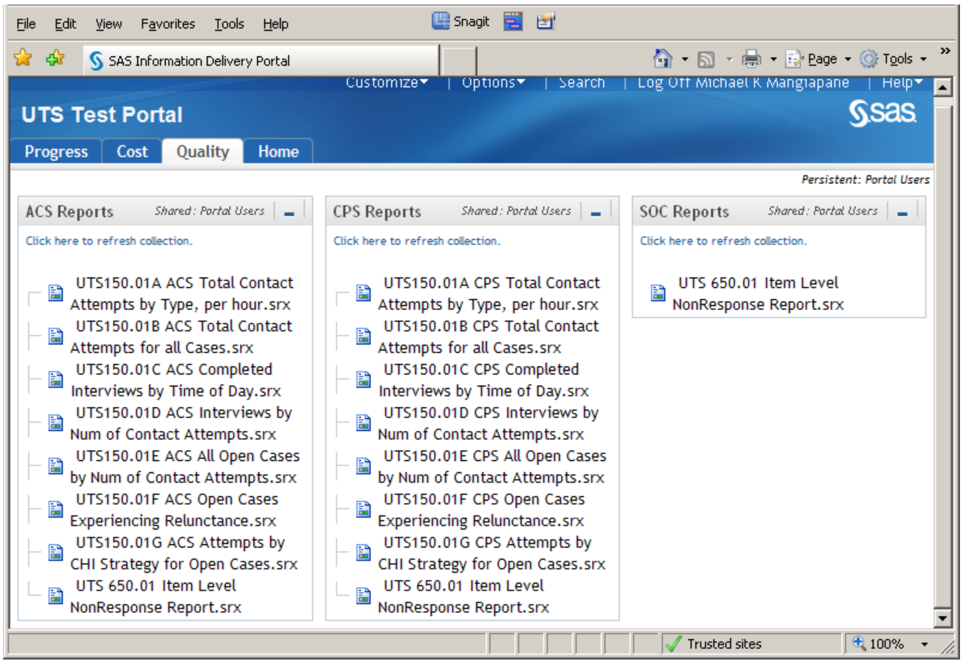
The UTS collects data from many different sources at the Census Bureau, including CHI data, DCP data, Regional Office Survey Control System (ROSCO), Cost And Response Management Network (CARMN), the Census Bureau's financial system, the Census Bureau's payroll system, Blaise audit trail files, and TMO's CATI case management system, WebCATI. Other sources may be added in the future as systems develop and data requirements change, including CARI.

The users of the UTS will be Field survey managers, internal survey sponsors, and even some of our external survey sponsors (with limited views).

All of the data presented by the UTS is inside a SAS Enterprise Business Intelligence (SAS-EBI) system in the form of "data cubes." An Oracle Data Warehouse serves as the data repository for UTS cost, progress, and quality paradata, and SAS-EBI uses this data warehouse to form the data cubes. Within a data cube, a user can create and customize views and of the data they wish to see and analyze; they have a lot of control over what they see while they are in the UTS. With a few clicks, they can go from a table of how many cases a regional office has for a survey to how many cases each FR in the region has assigned. This cuts down on waiting time to create and generate a separate report for each query the user has since SAS is doing the aggregation for them in the background. As data is added to the system, users will be able to look across all surveys rather than one at a time, and analyze data on the system as far back as five years or more, which was not possible before. The UTS allows for easy export of their data to other formats such as Microsoft Word or Excel.

6.2 Using the UTS

Upon logging in, the user is taken to their own personal portal page in the UTS, which contains links to any previously saved reports, and further access to elements of the data cube.


File Edit View Favorites Tools Help
Snagit
SAS Information Delivery Portal
Page Tools
Customize Options Search Log Off Michael K Mangiapane Help
UTS Test Portal
sas
Progress Cost Quality Home
Persistent: Portal Users
Trusted sites
100%

ACS Reports — Shared: Portal Users

Click here to refresh collection.

- UTS150.01A ACS Total Contact Attempts by Type, per hour.srx
- UTS150.01B ACS Total Contact Attempts for all Cases.srx
- UTS150.01C ACS Completed Interviews by Time of Day.srx
- UTS150.01D ACS Interviews by Num of Contact Attempts.srx
- UTS150.01E ACS All Open Cases by Num of Contact Attempts.srx
- UTS150.01F ACS Open Cases Experiencing Relunctance.srx
- UTS150.01G ACS Attempts by CHI Strategy for Open Cases.srx
- UTS 650.01 Item Level NonResponse Report.srx

CPS Reports — Shared: Portal Users

Click here to refresh collection.

- UTS150.01A CPS Total Contact Attempts by Type, per hour.srx
- UTS150.01B CPS Total Contact Attempts for all Cases.srx
- UTS150.01C CPS Completed Interviews by Time of Day.srx
- UTS150.01D CPS Interviews by Num of Contact Attempts.srx
- UTS150.01E CPS All Open Cases by Num of Contact Attempts.srx
- UTS150.01F CPS Open Cases Experiencing Relunctance.srx
- UTS150.01G CPS Attempts by CHI Strategy for Open Cases.srx
- UTS 650.01 Item Level NonResponse Report.srx

SOC Reports — Shared: Portal Users

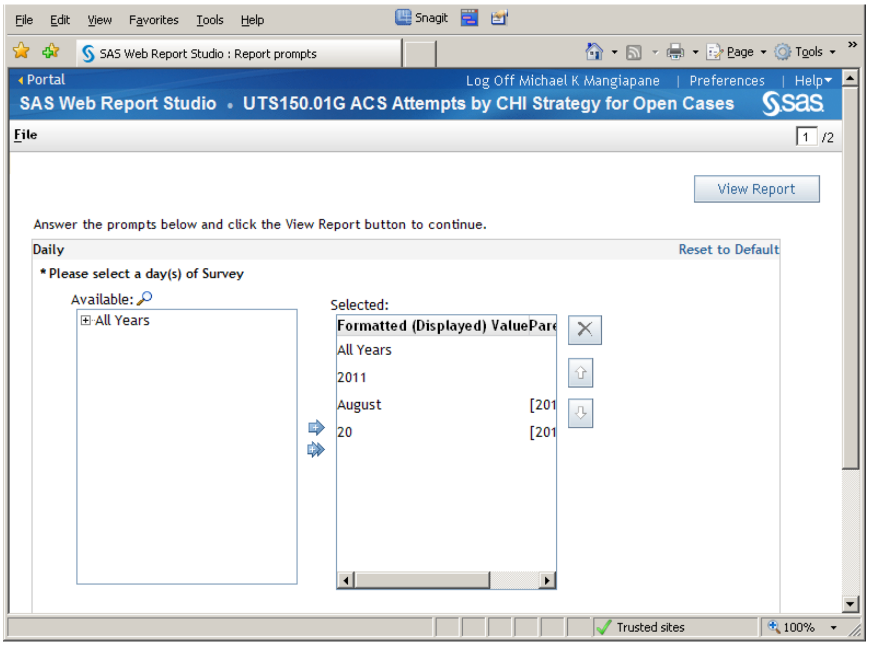
Click here to refresh collection.

- UTS 650.01 Item Level NonResponse Report.srx

Fig. 11 - UTS Portal Page

Not all users will have access to all of the data saved inside the cube. Field Supervisors may only see cost, progress, and quality reports related to their assigned geographic areas, while a survey sponsor may see the same reports for all geographic areas.

Selecting a report brings up a screen to filter which data they want to see in the report.



File Edit View Favorites Tools Help
Snagit
SAS Web Report Studio : Report prompts
Page Tools
Portal
Log Off Michael K Mangiapane | Preferences | Help
SAS Web Report Studio • UTS150.01G ACS Attempts by CHI Strategy for Open Cases
File
1 /2
View Report
Trusted sites
100%

Answer the prompts below and click the View Report button to continue.

Daily — Reset to Default

* Please select a day(s) of Survey

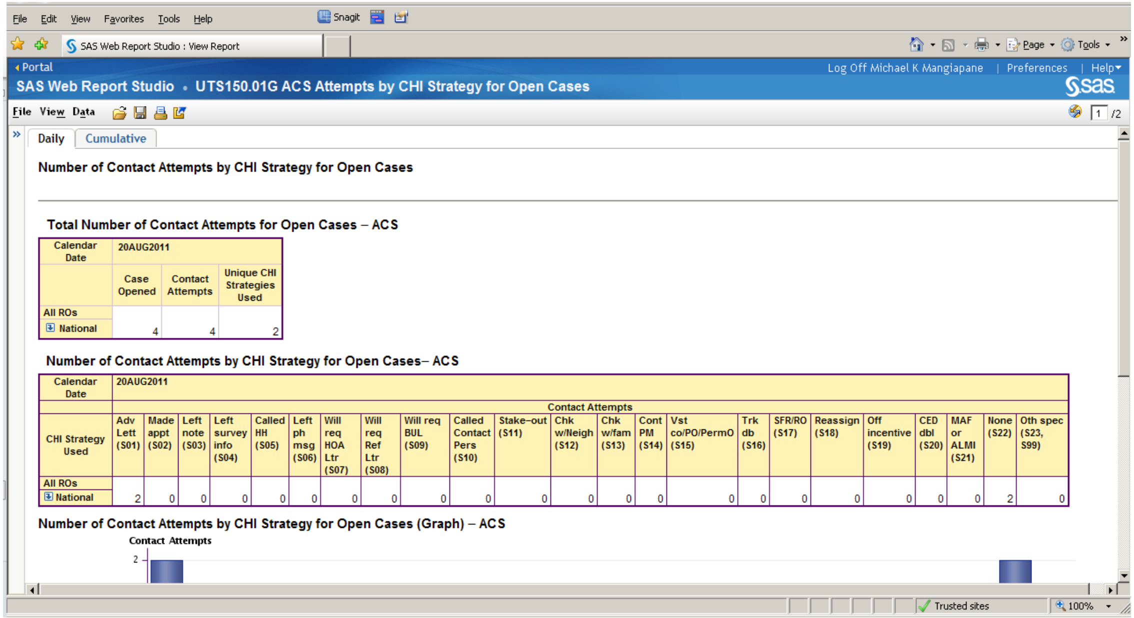
Available: ⊞ All Years

Selected:

Formatted (Displayed) Value	Pare
All Years	
2011	
August	[201
20	[201

Fig. 12 - Report Filter Screen for the CHI Strategy Report

The view report button will then display the report with the selected options.



File Edit View Favorites Tools Help
Snagit
SAS Web Report Studio : View Report
Page Tools
Portal
Log Off Michael K Mangiapane | Preferences | Help
SAS Web Report Studio • UTS150.01G ACS Attempts by CHI Strategy for Open Cases
File View Data
1 /2
Daily
Cumulative
Trusted sites
100%

Number of Contact Attempts by CHI Strategy for Open Cases

Total Number of Contact Attempts for Open Cases – ACS

Calendar Date	20AUG2011		
	Case Opened	Contact Attempts	Unique CHI Strategies Used
All ROs ⊞ National	4	4	2

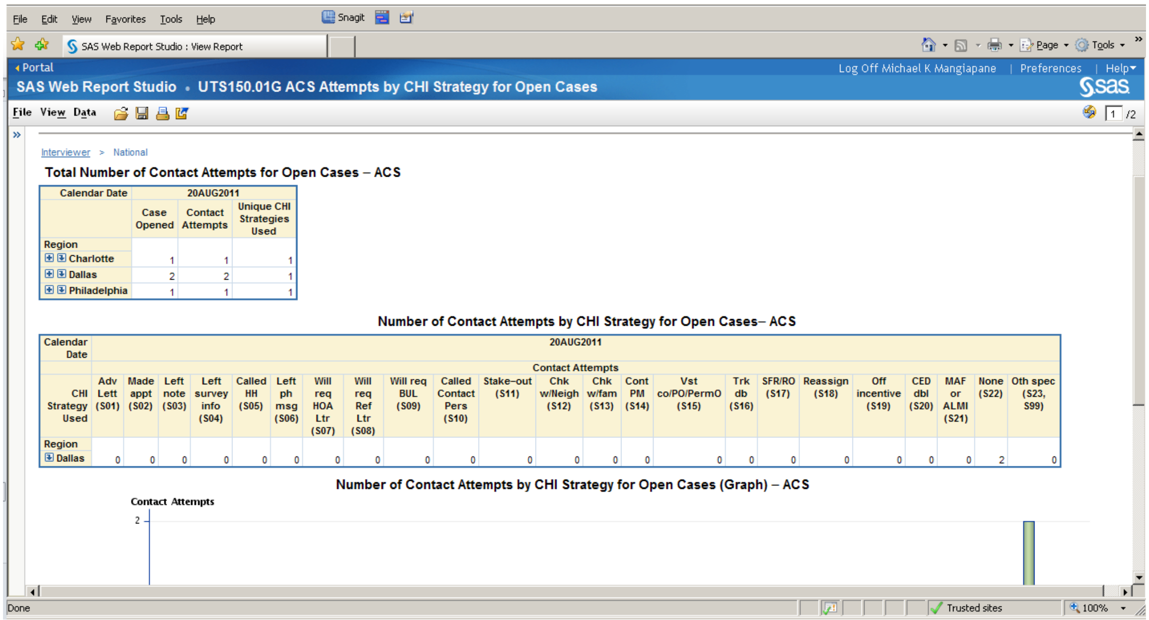
Number of Contact Attempts by CHI Strategy for Open Cases– ACS

Calendar Date	20AUG2011																						
	Contact Attempts																						
CHI Strategy Used	Adv Lett (S01)	Made appt (S02)	Left note (S03)	Left survey info (S04)	Called HH (S05)	Left ph msg (S06)	Will req HOA Ltr (S07)	Will req Ref Ltr (S08)	Will req BUL (S09)	Called Contact Pers (S10)	Stake–out (S11)	Chk w/Neigh (S12)	Chk w/fam (S13)	Cont PM (S14)	Vst co/PO/PermO (S15)	Trk db (S16)	SFR/RO (S17)	Reassign (S18)	Off incentive (S19)	CED dbl (S20)	MAF or ALMI (S21)	None (S22)	Oth spec (S23, S99)
All ROs ⊞ National	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0

Number of Contact Attempts by CHI Strategy for Open Cases (Graph) – ACS

Fig. 13 - Example report generated after selecting all of the options in the filter screen:

With the default options selected, the UTS displays all of the retrieved data collectively as a national sample. However, the user can use the "drill down" icon to the left of National and separate the data by RO. They can also right-click in the chart and set up a filter or sorting to organize the data.


File Edit View Favorites Tools Help
Snagit
SAS Web Report Studio : View Report
Page Tools
Portal
Log Off Michael K Mangiapane | Preferences | Help
SAS Web Report Studio • UTS150.01G ACS Attempts by CHI Strategy for Open Cases
File View Data
1 /2
Interviewer > National
Done
Trusted sites
100%

Total Number of Contact Attempts for Open Cases – ACS

Calendar Date	20AUG2011		
	Case Opened	Contact Attempts	Unique CHI Strategies Used
Region			
Charlotte	1	1	1
Dallas	2	2	1
Philadelphia	1	1	1

Number of Contact Attempts by CHI Strategy for Open Cases– ACS

Calendar Date	20AUG2011																							
	Contact Attempts																							
CHI Strategy Used	Adv Lett (S01)	Made appt (S02)	Left note (S03)	Left survey info (S04)	Called HH (S05)	Left ph msg (S06)	Will req HOA Ltr (S07)	Will req Ref Ltr (S08)	Will req BUL (S09)	Called Contact Pers (S10)	Stake–out (S11)	Chk w/Neigh (S12)	Chk w/fam (S13)	Cont PM (S14)	Vst co/PO/PermO (S15)	Trk db (S16)	SFR/RO (S17)	Reassign (S18)	Off incentive (S19)	CED dbl (S20)	MAF or ALMI (S21)	None (S22)	Oth spec (S23, S99)
Region																							
Dallas	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0

Number of Contact Attempts by CHI Strategy for Open Cases (Graph) – ACS

Fig. 14 - Report after user clicks the Drill Down under "Total Number of Contact Attempts for Open Cases - ACS" and applies a filter to "Number of Contact Attempts by CHI Strategy for Open Cases – ACS"

In the above example, the "Number of Contact Attempts by CHI strategy for Open Cases – ACS" was filtered to only show regions where the CHI strategy of "None" was greater than zero. If a user wanted to take it further, they could view and filter strategies by Field Supervisor (FS) and by the FRs that report to the FS. This is just one of many ways that the UTS can be customized to display data for analysis by the user.

6.3 Plans for UTS

The ultimate goal of the UTS is to formulate dynamic response design for our surveys and help bring down the costs of conducting surveys at the Census Bureau. Since the user will be able to see how much each case is costing them, they can remove any outliers that are well above the average cost per case with little response or useful data. It will also allow a survey sponsor to identify when a survey has reached a statistically significant sample and no more interviews are needed for that interview period. In terms of responsive design, the sponsors will be able to immediately measure how a change to a survey is having an impact in the field rather than waiting until all of their data comes back and is processed.

Besides storing the CHI data and DCP data, the UTS is also developing a process for parsing Blaise audit trails to satisfy a request for calculating the amount of time that an FR spends in each section of the instrument. We have tried in the past to perform this calculation within Blaise, but have found that turning the appropriate section timers on and off presents a difficult challenge if a user backs up from one section into another.

7 Conclusion

For a number of years the Census Bureau has collected and analyzed paradata from their surveys through CHI, case level notes, Blaise audit trails, and critical items in the instrument data. Now our systems are evolving to effectively collect, track, and use more paradata for responsive survey design. Even with all of the data that we will collect from a number of different systems to put together in the UTS, we anticipate that Blaise will have a significant role in collecting paradata for our surveys.

8 References

Bates, N., Dahlhamer, J., and Singer, E. (2008). Privacy Concerns, Too Busy, or Just Not Interested: Using Doorstep Concerns to Predict Survey Nonresponse. Journal of Official Statistics, 24, 591-612.

Maitland, A., Casas-Cordero, C., Kreuter, F. (2009). An Evaluation of Nonresponse Bias using Paradata from a Health Survey. Section on Government Statistics – JSM 2009.

Tan, L. (2011). An Introduction to the Contact History Instrument (CHI) for the Consumer Expenditure Survey. Consumer Expenditure Survey Anthology, 2011, 8-16.

9 Acknowledgements

The author would like to acknowledge the following people for input into this paper and input on some of the features and usability of the UTS and CHI.

Nancy Bates, U.S. Census Bureau
John Wilen, U.S. Census Bureau
William E. Dyer Jr., U.S. Census Bureau
Malcolm Robert Wallace, U.S. Census Bureau
Tom Spaulding, U.S. Census Bureau
Karen Bagwell, U.S. Census Bureau

The views expressed in this paper are those of the authors and not necessarily those of the U.S. Census Bureau.