

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

Michael K. Mangiapane  
Technologies Management Office  
IBUC 2012, London, U.K.



# Introduction

- Interest in paradata has increased dramatically.
  - More efficient surveys.
  - Lowering costs.
- The U.S. Census Bureau has collected paradata for years.
  - CHI, CARI, CARMN, PANDA.
- New efforts to collect paradata.
  - pCHI, DCP, UTS.

# Contact History Instrument (CHI)

- Production: 2004, NHIS.
  - Now used for every household survey.
- Average 60-90 seconds to complete CHI.
- Fast review of CHI data for returning cases and reassigned cases in the field.

# CHI Process

- Two Points of Entry.
  - Automatic after exiting a survey.
  - Manually in Case Management (F12).
- Temporary BDB is created to collect data.
- Standard output created after completion.
- Dyer Jr. W. E. (2004). Contact History Instrument (CHI). 9<sup>th</sup> International Blaise User's Conference, Gatineau, Québec, Canada.  
<http://www.blaiseusers.org/2004/papers/03.pdf>

# Person-based CHI (pCHI)

- CHI wasn't compatible for NCVS.
  - Person-based survey instead of household.
- Request to collect CHI data for each eligible respondent.
- CHI process is the same with a few modifications.
  - Read in Household Roster.
  - Same CHI questions asked per respondent.
  - CHI Output file for each respondent.

• pCHI - CONTACT

• Select outcome that best describes this contact attempt.

LNO		NAME	AGE	BIRTHDATE	SEX
1	(C)	Dejuan Loe	46	03/22/1965	
2	(I)	Derrek Loe	31	02/20/1980	M
3		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** or **more** eligible person(s)
- 2. Made contact with only non-eligible persons
- 3. Noncontact

Continue CHI / Exit  **Attempt**

Immediate Attempt  **Yes**

Date

Time

Personal/Telephone  **Personal**

pContact

• **CONCERN / BEHAVIOR / RELUCTANCE**

- Select the categories that describe respondent concerns, behaviors, or reluctance during this contact attempt.
- Enter all that apply, separate with commas.

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

	Other Non-Contact Attempt	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 Strategy Attempt
(C) Dejuan Loe [1]											
(I) Derrek Loe [2]											
Larry Boe [3]		1						6,7,17			
Thomas Loe [4]											
(I) Sixfeet Under [5]											



### CONTACT STRATEGIES ATTEMPTED

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

- |   |  |
|---|--|
| <input type="checkbox"/> 1. Advance letter given  | <input type="checkbox"/> 13. Contacted other family members                        |
| <input type="checkbox"/> 2. Scheduled appointment                                       | <input type="checkbox"/> 14. Contacted property manager                            |
| <input type="checkbox"/> 3. Left note / appointment card                                | <input type="checkbox"/> 15. Visited county assessor / post office / permit office |
| <input checked="" type="checkbox"/> 4. Left promotional packet / informational brochure | <input type="checkbox"/> 16. On-line tracking database                             |
| <input type="checkbox"/> 5. Called household  | <input type="checkbox"/> 17. Sought help from SFR / RO                             |
| <input type="checkbox"/> 6. Left message on answering machine                           | <input type="checkbox"/> 18. Reassignment  |
| <input type="checkbox"/> 7. FR will request No One Home Letter                          | <input type="checkbox"/> 19. Offered incentive                                     |
| <input type="checkbox"/> 8. FR will request Refusal Letter                              | <input type="checkbox"/> 20. CED double placement                                  |
| <input checked="" type="checkbox"/> 9. FR will request Better Understanding Letter      | <input type="checkbox"/> 21. Used MAF or ALMI                                      |
| <input type="checkbox"/> 10. Called contact persons                                     | <input type="checkbox"/> 22. No Strategies   |
| <input type="checkbox"/> 11. Stake-out  | <input type="checkbox"/> 23. Other - specify                                       |
| <input type="checkbox"/> 12. Checked with neighbors                                     |  |

	Other Non-Contact Attempt	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 Strategy Attempt
(C) Dejuan Loe	[1]										
(I) Derrek Loe	[2]										
Larry Boe	[3]	1						6,7,17		4,9	
Thomas Loe	[4]										
(I) Sixfeet Under	[5]										

```
00000003.chi - N...
File Edit Format View
#caseid
00000003
#response_level
01
#line_nbr
0
#frdate
04/13/2012
#howcontacted
P
#contactstatus
P
#contactperson
1
#casestatus
2
#types
#reluctant
#strat
#chi_time
00143
#roster_ind
1
```

```
00000003-03.chi...
File Edit Format View
#caseid
00000003
#response_level
02
#line_nbr
003
#frdate
04/13/2012
#howcontacted
P
#contactstatus
C
#contactperson
1
#casestatus
1
#types
C01
#reluctant
R06$^R07$^R17
#strat
S04$^S09
```

```
00000003-04.chi...
File Edit Format View
#caseid
00000003
#response_level
02
#line_nbr
004
#frdate
04/13/2012
#howcontacted
P
#contactstatus
C
#contactperson
1
#casestatus
1
#types
C01
#reluctant
R22
#strat
S22
```

# Data Combing Process (DCP)

- Interest to capture paradata from NCVS.
  - Write values of variables in the Blaise Audit Trail.
  - Originally conceived to write data at any time during the interview.
    - Changed to write data every time the instrument is exited.

# DCP Requirements

- Use variables from the instrument based on a list provided by the sponsor.
  - List could be updated anytime.
- Fast.
- Run on FR's laptop.

# DCP via Current Processes

- Master Control System (MCS) has a Manipula script that combs for data for DSMD.
  - Requires Chameleon to create list of variables and Manipula Script.
    - Only located variables at the root datamodel level.
  - Requires extra processing time to create the script and output.

# DCP Redesign

- Manipula script to generate paradata for DSMD, UTS, or other sponsors.
  - One script for all surveys that want the DCP.
  - Sponsor provides variable list with fully qualified variable names.
  - Used in MCS.
  - Input of variable list and consolidated database.
  - Separate output file instead of appending the audit trail.

# The New DCP

- Compiled and run in Blaise 4.8.2.1606.
- Prints non-response or value.
  - ConsolidatedDB.GETVALUE(InputFields.VarName, UF)
    - UF to properly print enumerated fields.
- Character delimited ASCII output
  - | (pipe character).

# DCP Commands

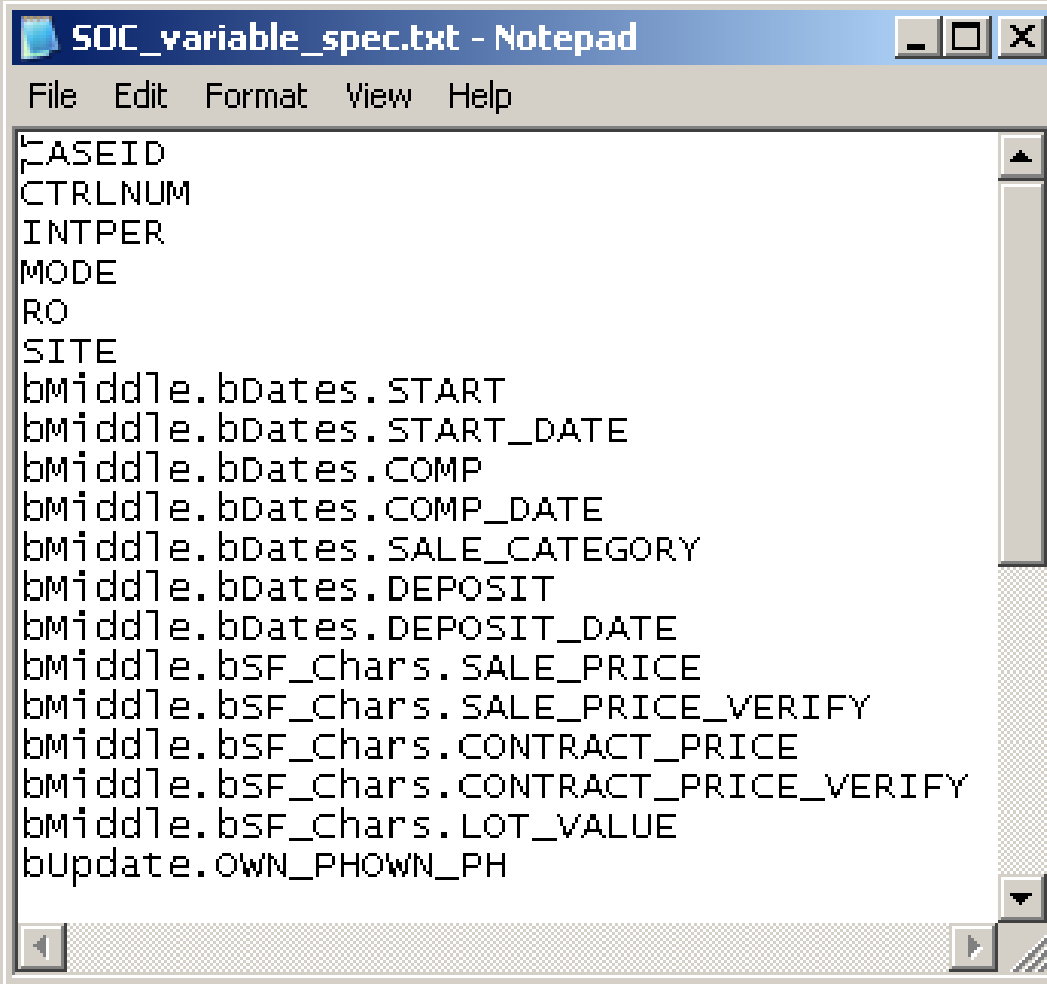
- *Compile:* b4cpars  
c:\scripts\data\_combing\_process.man  
/Hc:\survey /Oc:\survey
- *Run:* manipula  
c:\survey\data\_combing\_process.msu  
/Wc:\survey  
/Icaseids,paradatalist.txt,consdb.bdb  
/OZ999\_paradata.uts /q



# DCP Cycle

- Read (next) Case ID from list.
- Retrieve Case from consolidated database.
  - Read in variable name.
    - ConsDB.GETFIELDINFO(InputFields.VarName,'FIELDSTATUS')
  - Print Output
    - DK, RF, ParadataOut.VarValue + ConsDB.GETVALUE(InputFields.VarName,UF)
  - Repeat until variable list EOF, then reset list.
- Repeat until Case ID list EOF.

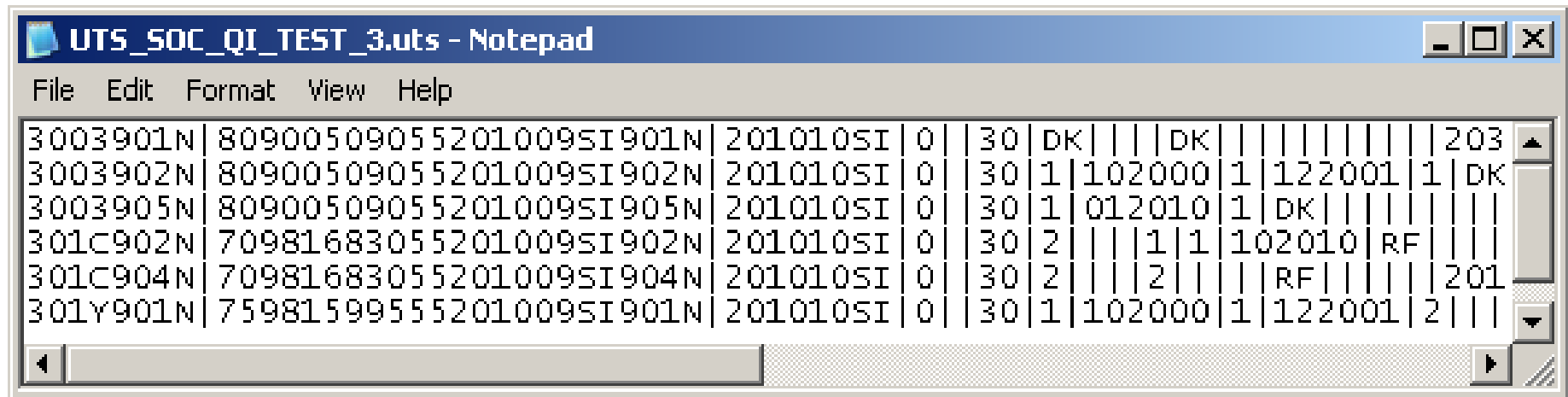
# Example Variable List



The image shows a Notepad window titled "SOC\_variable\_spec.txt - Notepad". The window contains a list of variables, each on a new line. The variables are: 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, and bupdate.OWN\_PHOWN\_PH. The window has a standard menu bar with File, Edit, Format, View, and Help. The text is displayed in a monospaced font.

```
SOC_variable_spec.txt - Notepad
File Edit Format View Help
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
```

# DCP Output



```
UTS_SOC_QI_TEST_3.uts - Notepad
File Edit Format View Help
3003901N|80900509055201009SI901N|201010SI|0|30|DK|||DK|||203
3003902N|80900509055201009SI902N|201010SI|0|30|1|102000|1|122001|1|DK
3003905N|80900509055201009SI905N|201010SI|0|30|1|012010|1|DK|||
301C902N|70981683055201009SI902N|201010SI|0|30|2|||1|1|102010|RF|||
301C904N|70981683055201009SI904N|201010SI|0|30|2|||2|||RF|||201
301Y901N|75981599555201009SI901N|201010SI|0|30|1|102000|1|122001|2|||
```

# DCP Now and In the Future

- The current DCP design fulfills all current requirements.
  - Processes 270 cases with 306 variables per case in ~35 seconds.
- Replaced DSMD process. Used by UTS and other sponsors.
  - Will be used for all surveys that require data combing.
- Researching counting DK, RF, and Empty variables.
  - Request to count fields on-path but left empty.

# Unified Tracking System (UTS)

- Central paradata storage and reporting from various sources in the Census Bureau.
  - Includes CHI, pCHI, DCP, ROSCO, CARMN, WebCATI, payroll, and financial data.
- Oracle Data Warehouse and SAS Enterprise Business Intelligence.

# UTS Test Portal

Persistent: Portal Users

- Home
- UTS Home
- ACS
- CPS
- NCVS
- NHIS
- SOC

**Progress and Outcome Reports** Shared: Portal Users

[Click here to refresh collection.](#)

- 1001 NCVS Daily Cumulative Progress Report for CAPI.srx
- 1005 Case Level Progress Report for CAPI.srx
- 1008 Total FR CAPI Workload.srx

**Data Quality Reports** Shared: Portal Users

[Click here to refresh collection.](#)

**Data Collection Effort Reports** Shared: Portal Users

[Click here to refresh collection.](#)

- 3101 NCVS Total Contact Attempts by Type, per hour CAPI.srx
- 3102 NCVS Total Contact Attempts for all Cases for CAPI.srx
- 3103 NCVS Completed Interviews by Time of Day for CAPI.srx
- 3104 NCVS Interviews by Num of Contact Attempts for CAPI.srx
- 3105 NCVS Open Cases by Num of Contact Attempts for CAPI.srx
- 3106 NCVS Open Cases Experiencing Reluctance for CAPI.srx
- 3107 NCVS Attempts by CHI Strategy for Open Cases CAPI.srx
- 3501 NCVS Hours, Miles and Costs Report.srx
- 3525 NCVS Pick your Own Task Code.srx

**Official Costs Reports** Shared: Portal Users

[Click here to refresh collection.](#)

- 4000 Budget and Actual Cost for all Surveys.srx



>> Daily Cumulative

### 3102 NCVS Total Contact Attempts for all Cases for CAPI

Time Hierarchy > 2011 > August

Total Cases Assigned at the Start of the Interview Period

	Total Cases Assigned
All ROs	
<input checked="" type="checkbox"/> National	64

Contact Attempts For Checked In Cases

Day of Survey	31									
Calendar Date	31AUG2011									
	Total Cases with CHI Records	Cases with No CHI Records	Cases with CHI Records completed by PV	Cases with CHI Records completed by phone	All Attempts	Attempts by PV	Attempts by phone	Attempts per case	Attempts per case Excludes Cases with No CHI records	
All ROs										
<input checked="" type="checkbox"/> National	27	3	15	12	27	15	12	0.90	1.00	

The formula used to calculate 'Attempts per case' is All Attempts / (Total Cases with CHI Records + Cases with No CHI Records).

The formula used to calculate 'Attempts per case Excludes Cases with No CHI records' is All Attempts / Total Cases with CHI Records.

Contact Attempts For Not Checked in Cases

Day of Survey	31											
Calendar Date	31AUG2011											
	Cases	Cases Open	Cases Open	Attempts for Cases	Attempts by PV	Attempts by phone	Attempts	Attempts	Attempts Per	Attempts Per	Attempts Per	Attempts Per Open

Daily **Cumulative**

### 3102 NCVS Total Contact Attempts for all Cases for CAPI

[INTERVIEWER](#) > National

Total Cases Assigned at the Start of the Interview Period

Region	Total Cases Assigned
<input type="checkbox"/> <input type="checkbox"/> Charlotte	4
<input type="checkbox"/> <input type="checkbox"/> Dallas	34
<input type="checkbox"/> <input type="checkbox"/> Philadelphia	26

Contact Attempts For Checked In Cases

Year		2011								
Calendar Date		31AUG2011								
Region	Total Cases with CHI Records	Cases with No CHI Records	Cases with CHI Records completed by PV	Cases with CHI Records completed by phone	All Attempts	Attempts by PV	Attempts by phone	Attempts per case	Attempts per case Excludes Cases with No CHI records	
<input type="checkbox"/> <input type="checkbox"/> Charlotte	0	2	0	0	0	0	0	0.00	.	
<input type="checkbox"/> <input type="checkbox"/> Dallas	15	1	9	6	15	9	6	0.94	1.00	
<input type="checkbox"/> <input type="checkbox"/> Philadelphia	12	0	6	6	12	6	6	1.00	1.00	

The formula used to calculate 'Attempts per case' is All Attempts / (Total Cases with CHI Records + Cases with No CHI Records).

The formula used to calculate 'Attempts per case Excludes Cases with No CHI records' is All Attempts / Total Cases with CHI Records.

Contact Attempts For Not Checked in Cases

Year		2011									
Calendar Date		31AUG2011									
Region	Cases	Cases	Attempts for	Attempts	Attempts by	Attempts	Attempts	Attempts	Attempts	Attempts	Attempts
<input type="checkbox"/> <input type="checkbox"/> Charlotte											
<input type="checkbox"/> <input type="checkbox"/> Dallas											
<input type="checkbox"/> <input type="checkbox"/> Philadelphia											



# Conclusions

- The U.S. Census Bureau has collected and analyzed paradata for years, but the tools to enable us to do more with the data we have are now becoming available.
- Blaise has and will continue to have a significant role in creating and collecting paradata.

# Questions?

Michael K. Mangiapane

U.S. Census Bureau

+1 301 763 7736

[michael.k.mangiapane@census.gov](mailto:michael.k.mangiapane@census.gov)