INSTITUTE FOR SOCIAL RESEARCH • SURVEY RESEARCH CENTER
SURVEY RESEARCH OPERATIONS
UNIVERSITY OF MICHIGAN

Laura D Yoder, MPH

DATA PROCESSING AND DATA DELIVERY

- Data Manager

Overview

- Many ways to access and adapt old processes
- Modification of well defined procedures and the challenges and successes we had along the way
- Some challenges due to our design and some due to software constraints

Five pieces of data processing

1. Survey Preload: information known from the start that may be used throughout the instrument (i.e. age, gender, family composition, previous employment, etc.)
2. Auxiliary System Preload: preload to sample management systems, authentication services
3. Migration: non-harmless changes to the instrument require data to be migrated from the old to the new version
4. Merge: combine multiple survey databases into a single master database
5. Data Delivery: delivery of survey data and survey paradata

1. Survey Preload

- Offline
  - Caret delimited string pushed in by sample management system
    - tCapi and tShared separated into correct SurveyTrak table
    - Loaded when Iwer initiates survey
- Online
  - Loaded directly into the instrument on the server
    - All preload is present before the R/Iwer initiates survey
  - Manipula script to load the instrument (for HRS, created by programmer)
    - Load caret delimited string (or xml file) with individual and shared preload concatenated
- Considerations
  - Recompile manipula with each data model change, even if no preload change
  - Updating preload can be error prone and require extra testing

Preload Examples

Offline – table in the sample management system

vProjectId	vSampleLine	lvPreloadData
SRC.SRO.HRS2018.PROD	0100010010	0100010010^0100010^^^^^^^^^^^^011^010^1^1^1^005^6^^0100010010^3^1^5^5^2016^^1^^5^5^
SRC.SRO.HRS2018.PROD	0100030030	0100030030^0100030^^^^^^^^^^^^031^030^1^1^1^005^5^1^0100030030^2^5^1^3^2016^^1^^5^5
SRC.SRO.HRS2018.PROD	0100040040	0100040040^0100040^^^^^^^^^^^^041^040^1^1^1^005^5^1^0100040040^2^5^5^5^2016^^1^^5^5
SRC.SRO.HRS2018.PROD	0100131040	0100131040^0100131^^^^^^^^^^^^041^040^1^1^1^005^4^^0100131040^3^1^5^3^2016^^1^^5^5^
SRC.SRO.HRS2018.PROD	0100380010	0100380010^0100380^2^1^0^^0^2^5^041^1^^^011^010^1^1^1^005^1^^0100380010^2^5^1^4^201
SRC.SRO.HRS2018.PROD	0100380040	0100380040^0100380^^^^^^^^^^^^041^040^1^1^1^005^1^5^0100380040^2^5^5^4^2016^^1^^5^5
SRC.SRO.HRS2018.PROD	0100500010	0100500010^0100500^^^^^^^^^^^^011^010^1^1^1^005^5^^0100500010^3^5^5^12^2016^^1^^5^5
SRC.SRO.HRS2018.PROD	0100590020	0100590020^0100590^^^^^^^^^^^^021^020^1^1^1^005^1^5^0100590020^2^5^1^6^2016^^1^^5^5
SRC.SRO.HRS2018.PROD	0100590030	0100590030^0100590^2^1^0^^0^2^5^021^1^^^031^030^1^1^1^005^1^^0100590030^3^1^5^12^20
SRC.SRO.HRS2018.PROD	0100630010	0100630010^0100630^^^^^^^^^^^^011^010^1^1^1^005^4^^0100630010^1^5^5^4^2017^^1^^5^5^

Online – excerpt from preload Manipula script

```
DATAMODEL InHrs2018Data

 FIELDPROPERTIES
   Remark: Open
   IsVisited: TIsVisitedFieldProperty
   AlienActionEvent : string

ATTRIBUTES = DONTKNOW, REFUSAL

  INCLUDE "HRS18SpecialAnswers.incx"
  INCLUDE "HRS18_Type.incx"
  INCLUDE "HRS18_SCV.incx"
  INCLUDE "HRS18_Basis_Tables.incx"
FIELDS
   SampID /"SAMPLE ID"     : STRING[10]
   HHID   /"HOUSEHOLD ID"  : STRING[10]
  {tShared}
   Preload_RTab          : ARRAY [1..2]  OF B_RTab
   Preload_HH            : B_HOUSEHOLD
   Preload_Respondents   : ARRAY [1..3]  OF B_People
   Preload_Children      : ARRAY [1..50] OF B_People
   Preload_HHMembers     : ARRAY [1..20] OF B_People
  {tCAPI}
   Preload_SCV           : B_SCV
   Preload_RVARS         : B_RVARS
   Preload_PastPens      : ARRAY [1..10]  OF B_PastPens
   Preload_Job           : ARRAY [1..10]  OF B_Job
   Preload_Hlth_Plan     : ARRAY [1..3]   OF B_HlthPlan
   Preload_RSiblings     : ARRAY [1..20] OF B_Siblings
AUXFIELDS
   FLJ535, FLJ005 : STRING
ENDMODEL {InHrs2018Data}
```

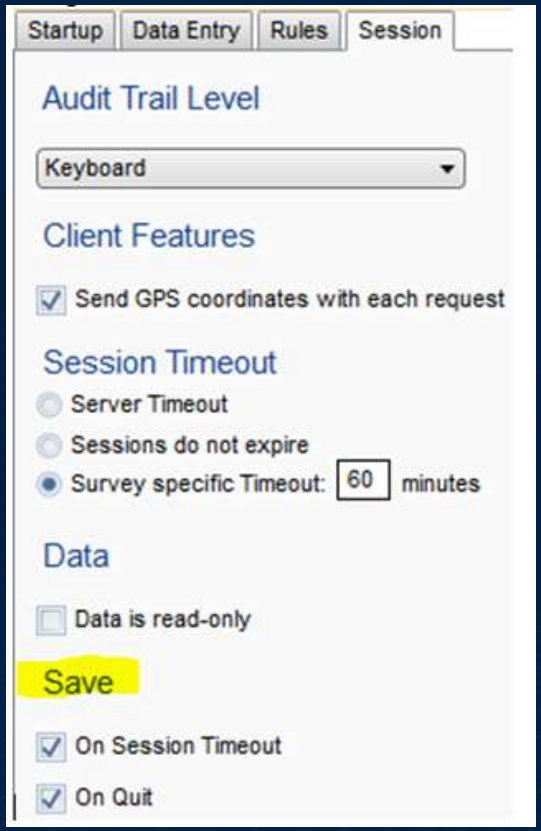
2. Auxiliary System Preload

- Offline
  - Sample Management System (SurveyTrak)
    - 4 main tables: tSample_Line, tSample_Line_Address, tCapi, tShared
- Online
  - Sample Management System (MSMS)
    - 10 csv files loaded into 6 objects: SampleLine, ContactPerson, Name, Address, Phone, Email
    - Additional protocol files
  - Authentication Services
    - Since online respondents do not go through the sample management system, we load login credentials into a separate databases
- Considerations
  - Format of two Sample Management Systems are different
  - Can’t reuse Login/Password combinations across projects

3. Migration

- Offline
  - Blaise to Blaise script sent to laptops, runs the next time the instrument is launched
  - Must have each script (for multiple migrations) on laptop
    - Example: Start in V2 and don't touch again until V5 - have to go from V2 to V3, V3 to V4, V4 to V5 before resuming interview
- Online
  - Same Blaise to Blaise script
    - All preload is present before the R/Iwer initiates survey
  - Migrate on the server, do whole database at once
- Considerations
  - Program instrument to save to main database on suspend
  - Delete session database after migration
  - Custom programming to return R to correct location

Session setting needed for migration



4. Merge

- Offline
  - Use Interview Data Merge application
    - Combine single BDBx/SQLite database into master BDBx/SQL Server database
    - Set merge criteria and create merge Manipula script
    - Audit data are converted from SQLite to SQL Server data
- Online
  - No merge required as all data are stored on the server in one master BDBx and SQL Server database
- Considerations
  - Need correct ODBC connections to SQL Server databases
  - Need SQLite database reader
  - Create and compile Manipula script for each instrument and each version
  - Ensure all data storage locations are set up correctly

Merge Manipula script

```
SETUP HRS2018_Merge
 SETTINGS
   DESCRIPTION = 'BLAISE to BLAISE'

 USES
   InputMeta 'HRS18'
   OutputMeta 'HRS18'

 INPUTFILE InputFile1: InputMeta ('\\...\Storage\2018-08-03,09.35.00\HRS18', BLAISE)
 OUTPUTFILE OutputFile1: OutputMeta ('\\...\MasterSurveyData\HRS18', BLAISE)

 SETTINGS
   MAKENEWFILE = NO
 MANIPULATE
   OutputFile1.WRITE

ENDSETUP//HRS2018_Merge
```

5. Data Delivery

- Main Data
  - Client requests a single BDBx with completed cases from both offline and online
    - Offline data already contain only completes (set up in merge criteria)
    - Online data require removal of incomplete cases (Manipula scripts)
    - Merge the two databases together for delivery
  - Remarks
    - Run data out Manipula script - get .fps file
    - Limit .fps to include only “Remark” field property
    - Have to run in version 5.4 even though instrument is in version 5.3 due to manipula bug
- Audit Data
  - Raw audit data from both online (completes and partials) and offline (only completes)
  - Parsed audit data and timings (additional information in Transforming Survey Paradata presentation Thursday afternoon)

Output examples

PrefMode=WEB

	A	A2	A2E	B	C	D	E	F
Mean	3.76	4.63	0.00	1.49	14.09	13.91	9.34	4.79
Median	2.97	3.62	0.00	1.10	11.86	12.26	4.92	0.00

PrefMode=TEL

	A	A2	A2E	B	C	D	E	F
Mean	3.22	4.46	1.06	1.67	13.85	11.24	7.90	2.63
Median	2.75	3.79	0.00	1.42	12.70	10.78	5.06	0.07

PrefMode=FTF

	A	A2	A2E	B	C	D	E	F
Mean	4.34	4.96	1.23	1.81	14.75	12.11	8.03	0.28
Median	3.36	4.14	0.00	1.47	13.88	12.01	6.23	0.06

PrefMode=FTF-E

	A	A2	A2E	B	C	D	E	F
Mean	3.61	4.31	1.20	1.72	13.61	11.17	7.71	1.96
Median	2.92	3.65	0.00	1.41	12.62	10.83	5.45	0.07

Sessions to Complete	N	%
1	512	42.6
2	311	25.9
3	168	14.0
4	76	6.3
5	57	4.7
6	22	1.8
7	23	1.9
8	15	1.2
9	11	0.9
10	2	0.2
11	1	0.1
12	2	0.2
13	1	0.1
23	1	0.1
27	1	0.1
Total	**1,203**	**100**

	WEB	TEL	FTF	FTF-E
IW average	120.07	103.19	102.80	143.48
IW median	107.57	102.65	99.53	141.97

Summary

- While transition of the HRS from Blaise 4.8 to Blaise 5 is still in progress, we have found many ways to adapt existing processes to the new world
- It has been frustrating at times, often because of both our own decisions and the limitations of working with a new product
- We hope that the processes we put in place for the HRS can be used across other future projects