Laura D Yoder, MPH

DATA PROCESSING AND DATA DELIVERY

• Data Manager

© 2018 Regents of the University of Michigan
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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>010001010</td>
<td>010001010</td>
<td>010001010</td>
<td>010001010</td>
<td>010001010</td>
</tr>
<tr>
<td>010001010</td>
<td>010001010</td>
<td>010001010</td>
<td>010001010</td>
<td>010001010</td>
</tr>
<tr>
<td>010001010</td>
<td>010001010</td>
<td>010001010</td>
<td>010001010</td>
<td>010001010</td>
</tr>
<tr>
<td>010001010</td>
<td>010001010</td>
<td>010001010</td>
<td>010001010</td>
<td>010001010</td>
</tr>
<tr>
<td>010001010</td>
<td>010001010</td>
<td>010001010</td>
<td>010001010</td>
<td>010001010</td>
</tr>
</tbody>
</table>

### Online – excerpt from preload Manipula script

```plaintext
DATAMODEL InHrs2018Data

FIELDPROPERTIES
  Remark: Open
  Invisted: InvistedFieldProperty
  AlienActionEvent: string

ATTRIBUTES = DONTKNOW, REFUSAL

INCLUDE "HRS18SpecialAnswers.incx"
INCLUDE "HRS18_Type.incx"
INCLUDE "HRS18_SCV.incx"
INCLUDE "HRS18_Basis_Tables.incx"

FIELDS
  SampleID / SAMPLE_ID / STRING[10]
  HHID / HOUSEHOLD_ID / STRING[10]

  (Shared)
  Preload_RTab : ARRAY [1..2] OF B_RTab
  Preload_HH : B_HOUSEHOLD
  Preload_Respondents : ARRAY [1..3] OF B_People
  Preload_Children : ARRAY [1..20] OF B_People
  Preload_HHMembers : ARRAY [1..20] OF B_People

  (LOC)
  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_HhhPlan : ARRAY [1..3] OF B_HhhPlan
  Preload_Rsiblings : ARRAY [1..20] OF B_Siblings

AUXFIELDS
  FL3545_FL005 : 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

Audit Trail Level

Client Features
- Send GPS coordinates with each request

Session Timeout
- Server Timeout
- Sessions do not expire
- Survey specific Timeout: 80 minutes

Data
- Data is read-only

Save
- On Session Timeout
- On Quit
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

```plaintext
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)
### PrefMode=WEB

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>A2</th>
<th>A2E</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>3.76</td>
<td>4.63</td>
<td>0.00</td>
<td>1.49</td>
<td>14.09</td>
<td>13.91</td>
<td>9.34</td>
<td>4.79</td>
</tr>
<tr>
<td>Median</td>
<td>2.97</td>
<td>3.62</td>
<td>0.00</td>
<td>1.10</td>
<td>11.86</td>
<td>12.26</td>
<td>4.92</td>
<td>0.00</td>
</tr>
</tbody>
</table>

### PrefMode=TEL

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>A2</th>
<th>A2E</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>3.22</td>
<td>4.46</td>
<td>1.06</td>
<td>1.67</td>
<td>13.85</td>
<td>11.24</td>
<td>7.90</td>
<td>2.63</td>
</tr>
<tr>
<td>Median</td>
<td>2.75</td>
<td>3.79</td>
<td>0.00</td>
<td>1.42</td>
<td>12.70</td>
<td>10.78</td>
<td>5.06</td>
<td>0.07</td>
</tr>
</tbody>
</table>

### PrefMode=FTF

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>A2</th>
<th>A2E</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>4.34</td>
<td>4.96</td>
<td>1.23</td>
<td>1.81</td>
<td>14.75</td>
<td>12.11</td>
<td>8.03</td>
<td>0.28</td>
</tr>
<tr>
<td>Median</td>
<td>3.36</td>
<td>4.14</td>
<td>0.00</td>
<td>1.47</td>
<td>13.88</td>
<td>12.01</td>
<td>6.23</td>
<td>0.06</td>
</tr>
</tbody>
</table>

### Output examples

#### Sessions to Complete

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

<table>
<thead>
<tr>
<th></th>
<th>WEB</th>
<th>TEL</th>
<th>FTF</th>
<th>FTF-E</th>
</tr>
</thead>
<tbody>
<tr>
<td>IW average</td>
<td>120.07</td>
<td>103.19</td>
<td>102.80</td>
<td>143.48</td>
</tr>
<tr>
<td>IW median</td>
<td>107.57</td>
<td>102.65</td>
<td>99.53</td>
<td>141.97</td>
</tr>
</tbody>
</table>

© 2018 Regents of the University of Michigan
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