Transforming Survey Paradata

Andrew Piskorowski, Mark Simonson, Laura Yoder, Cheng Zhou

TSG
Background

• Lots of previous work around Blaise audit data
• New tool that can parse both Blaise 4 and Blaise 5 formats
• Output in standard relational database format
• Enough similarity between 4 and 5 to create one set of tables
• In general, Blaise 4 has a subset of Blaise 5 information (e.g. No Page level for BL4)

• Blaise 5 structure
  – Case level (CaseSummary table) → Audit Session(s) (ADTSession) → Audit Pages (ADTPage) → Audit Fields (ADTField)

• Blaise 4 structure
  – Case level (CaseSummary table) → Audit Session(s) (ADTSession) → Audit Fields (ADTField)
Native Blaise Formats

• Blaise 4 data stored in delimited text file

```
5/2/2018 12:31:49:962 PM, "Enter Form:1", "Key:RCT1011"
5/2/2018 12:31:49:962 PM, "Metafile name:C:\BFL\BFL\BFY_PROD\InstrumentMain\Storage\2018-04-12,14,00,00\BFY_Bas
5/2/2018 12:31:49:962 PM, "Metafile timestamp:April, 12, 2018 1:59:46 PM"
5/2/2018 12:31:49:962 PM, "WinUserName:tnmsi11"
5/2/2018 12:31:49:962 PM, "DictionaryVersionInfo:0.0.0.0"
5/2/2018 12:31:50:000 PM, "Enter Fields:Verify", "Status:Normal", "Value:"
```

• Blaise 5 data stored in SQLite or SQL Server database

```
<table>
<thead>
<tr>
<th>SessionId</th>
<th>InstrumentId</th>
<th>TimeStamp</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>0002e0b7f6c8-4d75-9d0a-d4f4-1cabc019f0</td>
<td>099382c84d8d4353c95d1647c82d68</td>
<td>2018-06-04 13:09:33:173000</td>
<td>&lt;StartSessionEvent Width=&quot;1382&quot; Height=&quot;744&quot; Device=&quot;WindowsDesktop&quot; Language=&quot;EN&quot; KeyValue=&quot;501001010&quot; Platform=&quot;Windows&quot; /&gt;</td>
</tr>
<tr>
<td>0002e0b7f6c8-4d75-9d0a-d4f4-1cabc019f0</td>
<td>099382c84d8d4353c95d1647c82d68</td>
<td>2018-06-04 13:09:27:500000</td>
<td>&lt;GotoURL Event Uri=&quot;C:\TechSmith\Camtasia Studio 7\CamRecorder.exe&quot; /&gt;</td>
</tr>
<tr>
<td>0002e0b7f6c8-4d75-9d0a-d4f4-1cabc019f0</td>
<td>099382c84d8d4353c95d1647c82d68</td>
<td>2018-06-04 13:09:28:800000</td>
<td>&lt;ActionEvent Action=&quot;CurrentPage()&quot; ControlID=&quot;1&quot; /&gt;</td>
</tr>
<tr>
<td>0002e0b7f6c8-4d75-9d0a-d4f4-1cabc019f0</td>
<td>099382c84d8d4353c95d1647c82d68</td>
<td>2018-06-04 13:09:50:424000</td>
<td>&lt;UpdatePageEvent LayoutName=&quot;HR интерес&quot; PageIndex=&quot;1&quot; /&gt;</td>
</tr>
<tr>
<td>0002e0b7f6c8-4d75-9d0a-d4f4-1cabc019f0</td>
<td>099382c84d8d4353c95d1647c82d68</td>
<td>2018-06-04 13:09:50:472000</td>
<td>&lt;EnterField Event FieldName=&quot;Pu.Suspend&quot; AnswerStatus=&quot;Empty&quot; /&gt;</td>
</tr>
<tr>
<td>0002e0b7f6c8-4d75-9d0a-d4f4-1cabc019f0</td>
<td>099382c84d8d4353c95d1647c82d68</td>
<td>2018-06-04 13:10:22:472000</td>
<td>&lt;KeyboardEvent KeyStroke=&quot;1&quot; /&gt;</td>
</tr>
<tr>
<td>0002e0b7f6c8-4d75-9d0a-d4f4-1cabc019f0</td>
<td>099382c84d8d4353c95d1647c82d68</td>
<td>2018-06-04 13:10:23:370000</td>
<td>&lt;LeaveField Event FieldName=&quot;Pu.Suspend&quot; Value=&quot;1&quot; AnswerStatus=&quot;Response&quot; /&gt;</td>
</tr>
<tr>
<td>0002e0b7f6c8-4d75-9d0a-d4f4-1cabc019f0</td>
<td>099382c84d8d4353c95d1647c82d68</td>
<td>2018-06-04 13:10:23:740000</td>
<td>&lt;ActionEvent Action=&quot;NextField()&quot; /&gt;</td>
</tr>
</tbody>
</table>
```

• Parser needs to be flexible
  – Read and write to a variety of formats
  – Code that can easily manipulate and extract string data
Audit Parser - How It Works

The UM-SRO Audit Parser application was designed to do the following:

– Understand both Blaise 4.8 and Blaise 5 audit trail formats.
– Be able to read and write to and from text based delimited files, SQL Server tables and SQL Anywhere tables.
– Allow a “manual” mode, where a user selects a case to parse and can view the results.
– Allow a “batch” mode, where the program runs on a scheduled basis and processes multiple cases.
– Do all this with a minimum number of application dependencies, so the program can run on a user’s PC or on a Server.
  • We have a standalone version, which can read from one SQL server and write back to the same or different SQL server without any other dependency.

Finding Audit Data

– Leverage Sample Management System to get case status, etc
– “Self discovery” based on instrument id and location
Audit Parser - How It Works

“Parsing” part is somewhat complex, the key is the output data model
– Trade-offs between capturing all detail and having database that is easy to use, efficient, and queryable
– SQL Server database but generic enough for most relational database systems

Database Tables
– **CaseSummary**: one record for each respondent and each instrument
– **ADTSession**: one record for each session
– **ADTPage**: one record for each page for each visit
– **ADTField**: one record for each field it was visited
– Have a parent-child relationship from CaseSummary to ADTSession (for each case, one or more sessions), to ADTPage (for each session, one or more pages) to ADTField (for each page, one or more fields)
Audit Parser - CaseSummary

CaseSummary table contains 3 types of information

– Audit data processing and summary info
– Sample Management System (SMS) info
– Interviewer Quality Control (QC) info

Example information

– SampleId, Instrument Name, Instrument Id
– Audit: complete, number of sessions, timestamp of complete
– SMS: complete, number of sessions, timestamp of complete
– Iw Length, number fields visited, number questions answered
– QC: Iwer id, interview order, media files
Audit Parser - ADTSesstion

ADTSesstion table contains session level information based on audit data
- Each survey must have at least one session
- Can determine mode, device, etc.

Example information
- Session number, complete flag (last session)
- Device type, browser, operating system, width and height
- Mode, layout set name
- Last field entered, last field answered
Audit Parser - ADTPage

ADTPage table contains information at the page level from the audit data
- Can have multiple fields on a single page

Example information
- User page order, page visit number
- Page Index, Begin/End Time
- First/Last field
- Page duration, layout set name
### Audit Parser - ADTField

#### Example information

<table>
<thead>
<tr>
<th>ADTFieldId</th>
<th>FieldName</th>
<th>Instrumentid</th>
<th>FieldVisitNumber</th>
<th>LeaveFieldActionTS</th>
<th>LeaveFieldNav</th>
<th>LeaveFieldAction</th>
<th>LeaveTS</th>
<th>LeaveTSError</th>
<th>FieldDuration</th>
<th>IsPrepopulated</th>
<th>isFinalData</th>
<th>EnterAnswerValue</th>
<th>AnswerValue</th>
<th>SpecialAnswer</th>
<th>TYPOGRAPHIC</th>
<th>NextFieldBegTS</th>
<th>FieldDurationFull</th>
</tr>
</thead>
<tbody>
<tr>
<td>199</td>
<td>SecE KidStatus: E150_kids/10minVisFL</td>
<td>199</td>
<td>1</td>
<td>155</td>
<td>162</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>NULL</td>
<td>NULL</td>
<td></td>
<td></td>
<td>199</td>
<td>1</td>
</tr>
<tr>
<td>200</td>
<td>SecE KidStatus: E05_</td>
<td>200</td>
<td>1</td>
<td>156</td>
<td>159</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>NULL</td>
<td>NULL</td>
<td></td>
<td></td>
<td>200</td>
<td>1</td>
</tr>
<tr>
<td>201</td>
<td>SecE Long Kid Init MS2</td>
<td>201</td>
<td>1</td>
<td>167</td>
<td>161</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>NULL</td>
<td>NULL</td>
<td></td>
<td></td>
<td>201</td>
<td>1</td>
</tr>
<tr>
<td>202</td>
<td>SecE Long Kid Init MS2 chezquestions</td>
<td>202</td>
<td>1</td>
<td>168</td>
<td>161</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>NULL</td>
<td>NULL</td>
<td></td>
<td></td>
<td>202</td>
<td>1</td>
</tr>
<tr>
<td>203</td>
<td>SecE Long Kid Init MS2 hmember_reque</td>
<td>203</td>
<td>1</td>
<td>159</td>
<td>161</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>NULL</td>
<td>NULL</td>
<td></td>
<td></td>
<td>203</td>
<td>1</td>
</tr>
<tr>
<td>204</td>
<td>SecE Long Kid Init MS2 hmember_reque</td>
<td>204</td>
<td>1</td>
<td>170</td>
<td>162</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>NULL</td>
<td>NULL</td>
<td></td>
<td></td>
<td>204</td>
<td>1</td>
</tr>
<tr>
<td>205</td>
<td>SecE Long Kid Init MS2 hmember_reque</td>
<td>205</td>
<td>1</td>
<td>171</td>
<td>163</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>NULL</td>
<td>NULL</td>
<td></td>
<td></td>
<td>205</td>
<td>1</td>
</tr>
<tr>
<td>206</td>
<td>SecE Long Kid Init MS2 hmember_reque</td>
<td>206</td>
<td>1</td>
<td>172</td>
<td>164</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>NULL</td>
<td>NULL</td>
<td></td>
<td></td>
<td>206</td>
<td>1</td>
</tr>
<tr>
<td>207</td>
<td>SecE Long Kid Init MS2 hmember_reque</td>
<td>207</td>
<td>1</td>
<td>173</td>
<td>165</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>NULL</td>
<td>NULL</td>
<td></td>
<td></td>
<td>207</td>
<td>1</td>
</tr>
<tr>
<td>208</td>
<td>SecE Long Kid Init MS2 E18_kids/school</td>
<td>208</td>
<td>1</td>
<td>174</td>
<td>162</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>NULL</td>
<td>NULL</td>
<td></td>
<td></td>
<td>208</td>
<td>1</td>
</tr>
<tr>
<td>209</td>
<td>SecE Long Kid Loop E170_E170IMtemp</td>
<td>209</td>
<td>1</td>
<td>175</td>
<td>163</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>NULL</td>
<td>NULL</td>
<td></td>
<td></td>
<td>209</td>
<td>1</td>
</tr>
<tr>
<td>210</td>
<td>SecE Long KidInitNoMedia questions</td>
<td>210</td>
<td>1</td>
<td>176</td>
<td>164</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>NULL</td>
<td>NULL</td>
<td></td>
<td></td>
<td>210</td>
<td>1</td>
</tr>
<tr>
<td>211</td>
<td>SecE Long KidInitNoMedia questions</td>
<td>211</td>
<td>1</td>
<td>177</td>
<td>166</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>NULL</td>
<td>NULL</td>
<td></td>
<td></td>
<td>211</td>
<td>1</td>
</tr>
<tr>
<td>212</td>
<td>SecE Long KidInitNoMedia questions</td>
<td>212</td>
<td>1</td>
<td>178</td>
<td>168</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>NULL</td>
<td>NULL</td>
<td></td>
<td></td>
<td>212</td>
<td>1</td>
</tr>
<tr>
<td>213</td>
<td>SecE Long KidInitNoMedia questions</td>
<td>213</td>
<td>1</td>
<td>179</td>
<td>166</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>NULL</td>
<td>NULL</td>
<td></td>
<td></td>
<td>213</td>
<td>1</td>
</tr>
<tr>
<td>214</td>
<td>SecE Long KidInitNoMedia questions</td>
<td>214</td>
<td>1</td>
<td>180</td>
<td>166</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>NULL</td>
<td>NULL</td>
<td></td>
<td></td>
<td>214</td>
<td>1</td>
</tr>
</tbody>
</table>

© 2018 by the Regents of the University of Michigan
Audit Parser - Timings

Derived from Session, Page, and Field level information

– Raw timestamps as well as calculations done by parser application

Example information

– Total survey time in minutes (sum of session durations)
– Time spent in each session
– Time spend on each page (UpdatePageEvent to next UpdatePageEvent)
– Field duration (time in seconds between EnterField and LeaveField)
– Field duration full (time in seconds between EnterField and next EnterField)
Uses - 1

- Main uses of paradata
  - Backup of survey data and remarks
  - Quality control
  - Troubleshooting and intervening on potential issues
  - Queries and reporting
Uses - 2

• Main uses of paradata
  – Backup of survey data and remarks
  – Quality control (e.g., survey time, # of question answered for each R)
  – Troubleshooting and intervening on potential issues
  – Queries and reporting

Average is ~0.3 min or 18 seconds
Uses - 3

- Main uses of paradata
  - Backup of survey data and remarks
  - Quality control
  - Troubleshooting and intervening on potential issues (e.g., check loading time)
  - Queries and reporting
Uses - 4

• Main uses of paradata
  – Backup of survey data and remarks
  – Quality control
  – Troubleshooting and intervening on potential issues
  – Queries and reporting (e.g., interview time/answer rates by different modes)

<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>119.43</td>
<td>103.18</td>
<td>103.90</td>
<td>145.57</td>
</tr>
<tr>
<td>IW median</td>
<td>106.55</td>
<td>103.36</td>
<td>100.69</td>
<td>143.82</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item</th>
<th>Web</th>
<th>CATI</th>
<th>DCAPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Survey Time</td>
<td>119.0</td>
<td>109.6</td>
<td>128.2</td>
</tr>
<tr>
<td>Average Total Question Visits</td>
<td>396.2</td>
<td>469.0</td>
<td>510.8</td>
</tr>
<tr>
<td>Average Distinct Questions</td>
<td>395.0</td>
<td>455.5</td>
<td>510.3</td>
</tr>
<tr>
<td>Average Answered Questions</td>
<td>360.5</td>
<td>437.0</td>
<td>493.3</td>
</tr>
<tr>
<td>Average Revisited Questions</td>
<td>1.2</td>
<td>13.5</td>
<td>0.6</td>
</tr>
<tr>
<td>Percent Revisited</td>
<td>0.3%</td>
<td>2.9%</td>
<td>0.1%</td>
</tr>
<tr>
<td>Average Unanswered Questions</td>
<td>34.6</td>
<td>18.5</td>
<td>17.0</td>
</tr>
<tr>
<td>Percent Unanswered</td>
<td>8.7%</td>
<td>4.1%</td>
<td>3.3%</td>
</tr>
</tbody>
</table>
Thank You