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

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 12:31:49:982 PM","Enter Form:1","Key:PCT1011          "
"5/2/2018 12:31:49:982 PM","Metafile name:C:\BlProj\BFY_PROD\InstrumentMain\Storage\2018-04-12,14,00,00\BFY_Bas
"5/2/2018 12:31:49:982 PM","Metafile timestamp:Thursday, April 12, 2018 1:59:46 PM"
"5/2/2018 12:31:49:982 PM","WinUserName:tumsms11"
"5/2/2018 12:31:49:983 PM","DictionaryVersionInfo:0.0.0.0"
"5/2/2018 12:31:50:008 PM","Enter Field:VerifyR","Status:Normal","Value:"
"5/2/2018 12:31:52:585 PM","(KEY:)1[ENTR]"
"5/2/2018 12:31:52:701 PM","Action:Store Field Data","Field:VerifyR"
```

- Blaise 5 data stored in SQLite or SQL Server database

SessionId	InstrumentId	TimeStamp	Content
{0002aeb7-f6c8-4d75-9bda-d4641cabce09}	{099382c8-f48d-4353-a95d-16175c6f2268}	2018-06-04 13:09:39.1730000	<StartSessionEvent Width="1382" Height="744" Device="WindowsDesktop" Language="ENG" KeyValue="5010010010" Platform="Windows" />
{0002aeb7-f6c8-4d75-9bda-d4641cabce09}	{099382c8-f48d-4353-a95d-16175c6f2268}	2018-06-04 13:09:42.2750000	<GotoUriEvent Uri="C:\TechSmith\Camtasia Studio 7\CamRecorder.exe" />
{0002aeb7-f6c8-4d75-9bda-d4641cabce09}	{099382c8-f48d-4353-a95d-16175c6f2268}	2018-06-04 13:09:42.2880000	<ActionEvent Action="CurrentPage()" ControlID="1" />
{0002aeb7-f6c8-4d75-9bda-d4641cabce09}	{099382c8-f48d-4353-a95d-16175c6f2268}	2018-06-04 13:09:50.4240000	<UpdatePageEvent LayoutSetName="HRS_Iwer" PageIndex="1" />
{0002aeb7-f6c8-4d75-9bda-d4641cabce09}	{099382c8-f48d-4353-a95d-16175c6f2268}	2018-06-04 13:09:50.4270000	<EnterFieldEvent FieldName="puSuspend" AnswerStatus="Empty" />
{0002aeb7-f6c8-4d75-9bda-d4641cabce09}	{099382c8-f48d-4353-a95d-16175c6f2268}	2018-06-04 13:10:22.4720000	<KeyboardEvent KeyStrokes="1" />
{0002aeb7-f6c8-4d75-9bda-d4641cabce09}	{099382c8-f48d-4353-a95d-16175c6f2268}	2018-06-04 13:10:23.3970000	<LeaveFieldEvent FieldName="puSuspend" Value="1" AnswerStatus="Response" />
{0002aeb7-f6c8-4d75-9bda-d4641cabce09}	{099382c8-f48d-4353-a95d-16175c6f2268}	2018-06-04 13:10:23.3970000	<ActionEvent Action="NextField()" />
{0002aeb7-f6c8-4d75-9bda-d4641cabce09}	{099382c8-f48d-4353-a95d-16175c6f2268}	2018-06-04 13:10:24.7400000	<UpdatePageEvent LayoutSetName="HRS_Iwer" PageIndex="652" />

- 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

CaseSummaryId	ADTSurveyTime	ADTCompleteDT	ADTSessions	SurveyQVisited	SurveyQAnswered
1	140.0351	2018-02-06 19:01:08.690	1	563	558
2	193.9663	2018-04-25 16:26:27.797	4	463	449
3	182.2263	2018-06-07 14:11:25.883	1	656	635
4	190.6610	2018-04-30 15:21:31.480	2	647	613
5	144.2035	2018-05-24 15:31:08.800	1	541	535
6	168.6878	2018-05-24 12:57:43.857	1	537	524
7	218.0365	2018-05-10 17:31:05.190	2	506	470
8	235.1607	2018-07-31 14:01:40.997	1	641	620
9	115.0860	2018-05-18 17:04:32.090	1	403	387
10	129.5034	2018-06-25 16:23:14.003	1	498	487

Audit Parser - ADTSession

ADTSession 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

adtSessionId	BlaiseSession	BegTS	EndTS	Completed	SessionEndStatus	SessionDuration	Device	Language	DisplayWidth	DisplayHeight	SurveyQVisited	SurveyQAnswered
1	1	2018-02-06 16:41:06.5860000	2018-02-06 19:01:08.6910000	1	EndQuestionnaire	8402.1055	WindowsDesktop	ENG	1382	744	563	558
2	1	2018-04-25 13:01:30.5500000	2018-04-25 14:17:47.3090000	0	Unknow	4576.7588	WindowsDesktop	ENG	1372	774	217	213
3	2	2018-04-25 14:23:42.6100000	2018-04-25 14:38:17.1770000	0	Unknow	874.5670	WindowsDesktop	ENG	1372	774	14	11
4	3	2018-04-25 14:39:59.4410000	2018-04-25 14:47:38.9950000	0	Unknow	459.5540	WindowsDesktop	ENG	1372	774	17	15
5	4	2018-04-25 14:51:00.6950000	2018-04-25 16:26:27.7970000	1	EndQuestionnaire	5727.1021	WindowsDesktop	ENG	1372	774	226	218
6	1	2018-06-07 11:09:12.3010000	2018-06-07 14:11:25.8820000	0	Unknow	10933.5811	WindowsDesktop	ENG	1382	744	656	635
7	1	2018-04-28 09:42:27.3920000	2018-04-28 11:38:17.4020000	0	InterruptSessionEvent	6950.0098	WindowsDesktop	ENG	1372	774	413	390
8	2	2018-04-30 14:06:41.8280000	2018-04-30 15:21:31.4790000	1	EndQuestionnaire	4489.6509	WindowsDesktop	ENG	1372	774	237	225
9	1	2018-05-24 13:06:56.5880000	2018-05-24 15:31:08.7990000	0	Unknow	8652.2109	WindowsDesktop	ENG	1372	774	541	535
10	1	2018-05-24 10:09:02.5870000	2018-05-24 12:57:43.8560000	0	Unknow	10121.2686	WindowsDesktop	ENG	1372	774	537	524

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

ADTPageId	BlaiseSes...	UserPageOrder	PageIndex	PageVisitNum	PageBegTS	PageEndTS	PageDuration	FirstField	LastField	LayoutSetName
37	1	37	564	1	2018-02-06 16:51:48.1750000	2018-02-06 16:51:53.8630000	5.6880	SecA2.HHMTab.HHMFollowup[3].A226ALiveWith	SecA2.HHMTab.HHMFollowup[3].A226ALiveWith	HRS_Iwer
38	1	38	565	1	2018-02-06 16:51:54.3620000	2018-02-06 16:52:01.6020000	7.2400	SecA2.HHMTab.HHMFollowup[3].X072AMoChgRe...	SecA2.HHMTab.HHMFollowup[3].X073AYrChgRes...	HRS_Iwer
39	1	39	567	1	2018-02-06 16:52:02.1120000	2018-02-06 16:52:04.4330000	2.3210	SecA2.HHMTab.HHMFollowup[3].X096AMarried	SecA2.HHMTab.HHMFollowup[3].X096AMarried	HRS_Iwer
40	1	40	565	2	2018-02-06 16:52:04.8800000	2018-02-06 16:52:09.2820000	4.4020	SecA2.HHMTab.HHMFollowup[3].X073AYrChgRes...	SecA2.HHMTab.HHMFollowup[3].X073AYrChgRes...	HRS_Iwer
41	1	41	567	2	2018-02-06 16:52:09.7680000	2018-02-06 16:52:13.1860000	3.4180	SecA2.HHMTab.HHMFollowup[3].X096AMarried	SecA2.HHMTab.HHMFollowup[3].X096AMarried	HRS_Iwer
42	1	42	568	1	2018-02-06 16:52:13.6100000	2018-02-06 16:52:16.9200000	3.3100	SecA2.HHMTab.HHMFollowup[3].X097APartnered	SecA2.HHMTab.HHMFollowup[3].X097APartnered	HRS_Iwer
43	1	43	571	1	2018-02-06 16:52:17.4320000	2018-02-06 16:52:23.9710000	6.5390	SecA2.HHMTab.HHMFollowup[4].A226ALiveWith	SecA2.HHMTab.HHMFollowup[4].A226ALiveWith	HRS_Iwer
44	1	44	572	1	2018-02-06 16:52:24.5710000	2018-02-06 16:52:31.9120000	7.3410	SecA2.HHMTab.HHMFollowup[4].X072AMoChgRe...	SecA2.HHMTab.HHMFollowup[4].X073AYrChgRes...	HRS_Iwer
45	1	45	574	1	2018-02-06 16:52:32.5380000	2018-02-06 16:52:35.0230000	2.4850	SecA2.HHMTab.HHMFollowup[4].X096AMarried	SecA2.HHMTab.HHMFollowup[4].X096AMarried	HRS_Iwer
46	1	46	575	1	2018-02-06 16:52:35.3280000	2018-02-06 16:52:38.4430000	3.1150	SecA2.HHMTab.HHMFollowup[4].X097APartnered	SecA2.HHMTab.HHMFollowup[4].X097APartnered	HRS_Iwer
47	1	47	692	1	2018-02-06 16:52:38.9470000	2018-02-06 16:52:41.1690000	2.2220	SecA2.Sec_A2_Info3.A112_A2Assist	SecA2.Sec_A2_Info3.A112_A2Assist	HRS_Iwer
48	1	48	808	1	2018-02-06 16:52:42.8240000	2018-02-06 16:52:44.9490000	2.1250	SecB.B084_IwMode	SecB.B084_IwMode	HRS_Iwer
49	1	49	809	1	2018-02-06 16:52:45.7190000	2018-02-06 16:53:00.6890000	14.9700	SecB.B000_	SecB.B000_	HRS_Iwer
50	1	50	810	1	2018-02-06 16:53:01.2520000	2018-02-06 16:53:18.9430000	17.6910	SecB.B132	SecB.B132	HRS_Iwer

Audit Parser - ADTField

Example information

1. Instrumentid
2. FieldName
3. UserFieldOrder
4. BlaiseSession
5. UserPageOrder
6. PageIndex
7. FieldVisitNumber
8. EnterTS
9. LeaveTS
10. LeaveTSError
11. FieldDuration
12. IsPrepopulated
13. isFinalData
14. EnterAnswerValue
15. AnswerValue
16. SpecialAnswer
17. Category
18. KeyStroke
19. KeyStrokeTS
20. KeyStrokeCount
21. NoKeyStroke
22. Action
23. ActionTS
24. LeaveFieldAction
25. LeaveFieldActionTS
26. LeaveFieldNav
27. ActionType
28. IsRemark
29. hasRemark
30. hasSuppressSignal
31. LaunchMedia
32. LaunchMediaTS
33. GoToUriEvent
34. GoToUriTS
35. ToggleVisEvent
36. ToggleVisTS
37. SelectedQbyQ
38. SelectedRemark
39. SelectedQInfo
40. SelectedDK
41. SelectedRefuse
42. SelectedQuit
43. FieldMediaSecs
44. FieldMediaTime
45. EnterRow
46. LeaveRow
47. SectionName
48. NextFieldBegTS
49. FieldDurationFull

ADTFieldId	FieldName	UserFieldOrder	BlaiseSession	UserPage...	PageIndex	FieldVisitN...	EnterTS	LeaveTS	FieldDuration	IsPrepopulated	isFinalData	EnterAnswerValue	AnswerValue	SpecialAnswer
199	SecE.KidStatus.E150_kids10milesFL...	199	1	165	1442	1	2018-02-06 17:17:53.9310000	2018-02-06 17:18:00.8370000	6.9060	0	1	NULL		DontKnow
200	SecE.KidStatus.E015_	200	1	166	1590	1	2018-02-06 17:18:02.1530000	2018-02-06 17:18:07.0800000	4.9270	0	1	NULL	1	
201	SecE.LongKidLoop.E175_	201	1	167	1611	1	2018-02-06 17:18:08.1820000	2018-02-06 17:18:17.6700000	9.4880	0	1	NULL	1	
202	SecE.LongKidLoop.childquestions_E0...	202	1	168	1612	1	2018-02-06 17:18:18.3320000	2018-02-06 17:18:26.7770000	8.4450	0	1	NULL	26	
203	SecE.LongKidLoop.childquestions_E0...	203	1	169	1613	1	2018-02-06 17:18:27.4510000	2018-02-06 17:18:33.5170000	6.0660	0	1	NULL	25	
204	SecE.LongKidLoop.hhmemberquestion...	204	1	170	1662	1	2018-02-06 17:18:34.1980000	2018-02-06 17:18:40.9960000	6.7980	0	1	NULL	38	
205	SecE.LongKidLoop.hhmemberquestion...	205	1	171	1663	1	2018-02-06 17:18:41.6500000	2018-02-06 17:18:47.7930000	6.1430	0	1	NULL	18	
206	SecE.LongKidLoop.hhmemberquestion...	206	1	172	1664	1	2018-02-06 17:18:48.4920000	2018-02-06 17:18:52.5380000	4.0460	0	1	NULL	7	
207	SecE.LongKidLoop.hhmemberquestion...	207	1	173	1665	1	2018-02-06 17:18:53.0360000	2018-02-06 17:18:56.5700000	3.5340	0	1	NULL	2	
208	SecE.LongKidLoop.E169_kidsinschool	208	1	174	1682	1	2018-02-06 17:18:57.3140000	2018-02-06 17:19:08.3710000	11.0570	0	1	NULL	1	
209	SecE.LongKidLoop.E170_E170MTemp	209	1	175	1683	1	2018-02-06 17:19:10.0130000	2018-02-06 17:19:20.1460000	10.1330	0	1	NULL	1	
210	SecE.LongKidLoop.childquestions[1].E...	210	1	176	1684	1	2018-02-06 17:19:21.3770000	2018-02-06 17:19:52.6760000	31.2990	0	0	NULL	15	
211	SecE.LongKidLoop.childquestions[1].E...	211	1	177	1686	1	2018-02-06 17:19:53.3620000	2018-02-06 17:20:03.6610000	10.2990	0	0	NULL		
212	SecE.LongKidLoop.childquestions[1].E...	212	1	178	1684	2	2018-02-06 17:20:04.3360000	2018-02-06 17:20:14.0670000	9.7310	1	1	15	17	

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

ADTFieldId	FieldName	AnswerValue	SpecialAnswer	IsRemark	hasRemark
446	SecN.NHomeStay.N114_OvernteNH	5		0	0
447	SecN.OutPatSurgery.N134_OutSurgLst2Yrs	1		0	0
448	SecN.OutPatSurgery.InsurancePayN135.N...	1		0	0
449	SecN.OutPatSurgery.InsurancePayN135.N...	1		0	0
450	SecN.DocVisit.N147_TimeSeeDoc	36		0	1
451	#SecN.DocVisit.N147_TimeSeeDoc	R said once per month, so answer should be abo...	NULL	1	0
452	SecN.DocVisit.InsurancePayN152.N433_In...	1		0	0
453	SecN.DocVisit.InsurancePayN152.N434_In...	1		0	0
454	SecN.DentalCare.N164_SeeDentPW	5		0	0
455	SecN.PrescpDrug.N175_TkMedsReg	1		0	0

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

WEB

CaseSummaryId	ProjectDataCollectionId	TimePerQ	ADTSurveyTime	SurveyQVisited	SurveyQAnswered	ADTSessions
16820	51	0.108616666666669	44.3156	434	408	1
11529	51	0.127247445255474	52.2987	441	411	1
17601	51	0.128491932059447	60.5197	502	471	1
11090	51	0.132093654266958	60.3668	482	457	1
13520	51	0				
17482	51	0				
13716	51	0				
10693	51	0				

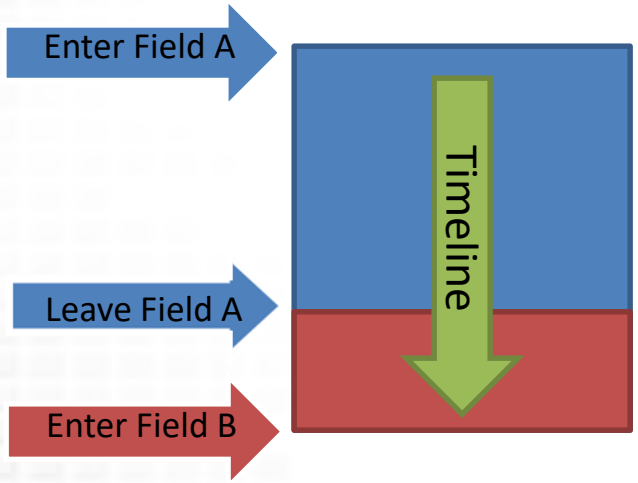
FTF

CaseSummaryId	ProjectDataCollectionId	TimePerQ	ADTSurveyTime	SurveyQVisited	SurveyQAnswered	ADTSessions
10362	6	0.089580294358135	109.5567	1234	1223	1
10380	6	0.094330343796711	63.1070	690	669	1
10363	6	0.098452369380315	81.0263	828	823	1
10403	6	0.098972851711026	130.1493	1354	1315	1
10341	6	0.100352727272727	93.8298	967	935	1
10330	6	0.107244316730523	83.9723	796	783	1
10384	6	0.107928216704288	143.4366	1378	1329	1

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



Section	FieldName	FieldDuration	FieldDurationFull	LoadingTime	Occurrence
SecF	SecF.ParentCare.F114Loop[2].F114AWhiSibAsstFN	29.37	35.11	5.74	4
SecQ	SecQ.IRAFollowup[3].UNFM_2Up1Down0.E2	8.00	13.40	5.40	1
SecN	SecN.PlanDetails[2].N050SWhoCov_S	7.82	13.20	5.38	5
SecT	SecT.EXITQUESTIONS.AMOUNT_10[2].T908_More10	1.86	7.05	5.19	1
SecE	SecE.KidTransMain.TransToKid[1].E078A	15.18	20.16	4.98	37
SecG	SecG.AskHelperInfoInOrder[4].G082A	12.17	17.08	4.91	2
SecE	SecE.KidStatus.E158_E023MTemp	14.87	19.72	4.85	20
SecN	SecN.OthHealthCare.UNFM_1up3down1.E3	11.00	15.85	4.85	2
SecG	SecG.AskHelperInfoInOrder[7].G071A	15.28	20.07	4.79	2
SecE	SecE.KidTransMain.TransFromKid[1].E089S	16.79	21.53	4.73	22
SecA2	SecA2.HHMTab.HHM[4].A186AX075ADupToWhom	14.84	19.56	4.72	2
SecA2	SecA2.HHMTab.HHM[3].A186AX075ADupToWhom	13.80	18.51	4.71	6
SecG	SecG.RSpendTime.G194_G100MTemp	24.39	29.02	4.64	575
SecT	SecT.EXITHOUSE.T054_T105MTemp	12.31	16.87	4.56	9
SecT	SecT.EXITQUESTIONS.AMOUNT_10[4].T907_More25	3.50	8.06	4.56	1
SecG	SecG.RSpendTime.G193_G099MTemp	19.37	23.92	4.55	4,322
SecE	SecE.KidTransMain.TransFromKid[1].E090_	14.99	19.51	4.53	9
SecE	SecE.KidStatus.E155_E013MTemp_S	10.04	14.57	4.53	64

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)

	WEB	TEL	FTF	FTF-E
IW average	119.43	103.18	103.90	145.57
IW median	106.55	103.36	100.69	143.82

Item	Web	CATI	DCAPI
Average Survey Time	119.0	109.6	128.2
Average Total Question Visits	396.2	469.0	510.8
Average Distinct Questions	395.0	455.5	510.3
Average Answered Questions	360.5	437.0	493.3
Average Revisited Questions	1.2	13.5	0.6
Percent Revisited	0.3%	2.9%	0.1%
Average Unanswered Questions	34.6	18.5	17.0
Percent Unanswered	8.7%	4.1%	3.3%

Thank You