



**INSTITUTE FOR SOCIAL RESEARCH  
SURVEY RESEARCH CENTER**  
UNIVERSITY OF MICHIGAN

# Session Data Preservation and Migration – Problems and Solutions

Jason Ostergren and Helena Stolyarova

Health and Retirement Study, University of Michigan Institute for Social Research

# HRS Chronology

- The Health and Retirement Study (HRS) at the University of Michigan Institute for Social Research is a longitudinal study that originated in 1992, switched to conducting interviews using Blaise 4 in 2002, and Blaise 5 in 2018.
- HRS has conducted 3 “waves” of Blaise 5 interviews so far at two-year intervals (2018, 2020, and 2022, which is almost complete.) HRS is currently gearing up for the 2024 “wave.”
- The switch to Blaise 5 was heavily influenced by needing to add a web self-interview mode, after experimenting with other web options.

# Preserve Session Data

- What is session data?
  - Session Data refers to the contents of the the Blaise 5 working database that maintains the state of the instrument, including the values of temporary data and properties of fields, in addition to regular data.
- Why does session data need to be preserved?
  - HRS needs session data to be preserved between sessions and across new data model releases (within a "wave," e.g. HRS 2020 for use in:
    - Unfolding sequences (question series for estimation) in Blaise Procedures
    - Signal suppression
    - Resuming at the field where suspended (HRS interviews are long and this is not uncommon)
    - Delivering session data along with other data for processing
- What are the consequences of not preserving session data?
  - Loss of data (e.g. children-in-law erased); loss of time (manually re-entering answers)

# Types of Session Data Persistence

- Between Sessions of Same Datamodel
- After Completion
- Between Sessions of Different Datamodels

# Persist Between Sessions (Same Datamodel)

- Session data is preserved between sessions (i.e. suspending a case and resuming it later) as part of the normal operation of Blaise 5
- However, in HRS 2018 SurveyTrak was not preserving session data between suspend and resume; this was caught and fixed before HRS 2020
  - What caused this? SurveyTrak packs up files after each suspend and in 2018 it was not including session data files
- Because normal operation of Blaise assumes that session data persists between sessions, it is especially critical that it not be lost in this circumstance

# Persist After Sessions (Same datamodel)

- Session data is deleted after the official completion of a case as part of the normal operation of Blaise 5
- This was not caught as a major problem until well into the HRS 2020 field period (remember HRS wants to keep session data for later analysis)
- We have a plan in place to preserve session data upon completion for HRS 2022
- Cheng Zhou enacted a plan to use a SQL Server trigger to save session data right before Blaise normally deletes it
- There have been discussions about having this sort of capability as an option in a future release of Blaise 5

# Persist Between Different Datamodels

- HRS has a history of releasing multiple versions of its datamodel each wave (an average of one per month over a year is not surprising)
- These are very similar datamodels, usually with logic or wording bugfixes, Spanish translations, or added question sequences arising from unexpected events (e.g. covid)
- When a case is suspended in one datamodel version, but must be resumed in a later one, we need to preserve session data as usual, but this was not working in HRS 2020
- New CBS updates (a special unsupported 5.10.10 release) helped us accommodate this need for HRS 2022
- This capability should be a normal part of the Blaise 5 feature set in 5.13 and later

# Persist Between Datamodels – Requirements

- The key requirement for preserving session data between sessions in different datamodel releases is that the new datamodel must not have harmful changes



# Harmful/Harmless Changes

- What are harmful and harmless changes?
- Visit the Blaise 5 help pages for details:
  - <http://help.blaise.com/Blaise.html?datafilecompatibility.htm>
- “As a rule of thumb, extension of the data definition such as an additional field, enlarging a string or expanding answer categories are fairly harmless.”
- There are ways to make almost any proposed change harmless with enough attention to detail

# Session Data

The screenshot shows the Blaise Server Manager interface with the 'Install survey - Configure installation' dialog box open. The dialog box contains the following information:

- Package:** C:\SVN\instdev\Blaise5Examples\IBUC2023\IBUC2023\IBUC2023.bpkg
- Survey:** IBUC2023
- Installation:** Overwrite data:  Yes
- Clear sessions:**  Yes

The background interface shows a list of surveys with the following data:

Instrument	Status
HRS24	Active
HS_SandBox	Active
IBUC2023	Active

At the bottom of the dialog box, there are buttons for 'Help', 'Cancel', '< Back', 'Next >', and 'Install'.

The screenshot shows the Blaise Server Manager interface with the 'Session Viewer' icon circled in red. The main window displays a table of survey data:

Instrument	Status	Install Date	Cati Role	Web DE Client	Version	Can Upgrade	ID
HRS24	Active	9/25/2023 10:40 AM	Mvc	5.13.5.344!	False	17f54858-6e2e-46f6-bc34-e530	
HS_SandBox	Active	9/27/2023 8:38 AM	MvcDebug	5.13.5.344!	False	c1d5ea3c-9110-4ad4-87c3-18fb	
IBUC2023	Active	9/27/2023 10:49 AM	MvcDebug	5.13.5.344!	False	cbda145c-9ad0-4c7e-a041-f0e0	

Below the table, there is a 'Settings' tab and a 'Monitoring' tab. The 'Settings' tab is active, showing a list of settings and their values:

Setting	Value
ActionsSetupFileName	
AggregateDefinitionsFileName	
AllowDownloadOverMeteredConnection	False
AppSupport	On
BlaiseVersion	5.13.5.3445
CariFileName	
CatiRole	
CatiSpecificationFileName	

The version number '5.13.5.3445' is visible in the bottom right corner of the window.

# Sample Datamodel

```

1  DATAMODEL IBUC2023 "HRS Questionnaire"
2  FIELDPROPERTIES
3  Remark : OPEN
4  IsVisited : IIsVisitedFieldProperty
5  Mode : STRING
6  MODES = SELFADMIN DESCRIPTION ENG "taken by respondent" SPN "tomada por encuestado",
7  IWERADMIN DESCRIPTION ENG "administered by interviewer" SPN "administrado por el entrevistador"
8  LANGUAGES = ENG "English", SPN "Spanish"
9  PRIMARY Sampid
10 TYPE
11 IIsVisitedFieldProperty = (No (0), Yes(1))
12 FIELDS
13 SampID
14 /"SAMPLE ID": STRING[50], NODK, NORF, NOEMPTY
15 {HARMLESS_CHANGE
16 ENG "This is a new field, so the change should be HARMLESS when it is added."
17 / "More info specify"
18 : STRING }
19 HARMFULL_CHANGE
20 ENG "This is a existing field, so the change should be HARMFULL when it is removed."
21 / "More info specify"
22 : STRING
23 TEST_OPEN_FIELD
24 ENG "This is the permanent open field in the database. When it is visited,
25 the IsVisited field property for this field is set to YES"
26 / "TEST open field"
27 : OPEN
28 TEST_IS_VISITED
29 ENG "Previous field was visited: ^{FLIsVisited}"
30 : ( CONTINUE (1) ENG "Continue" SPN "Continúe")
31
32 LOCALS FLIsVisited : STRING
33 AUXFIELDS TEST_AUXFIELD : STRING
34 RULES
35 Sampid.KEEP
36 {HARMLESS_CHANGE}
37 HARMFULL_CHANGE
38 TEST_OPEN_FIELD
39 IF TEST_OPEN_FIELD.IsVisited = YES THEN
40 FLIsVisited := 'Yes, this field was visited'
41 TEST_AUXFIELD := 'Auxfield has value'
42 ENDIF
43 TEST_IS_VISITED
44 ENDMODEL
45

```

The screenshot shows a control panel with the following elements:

- Language :** ENG
- Settings :** CAWI\_Settings
- Parallel :** (empty)
- Active Field :** TEST\_IS\_VISITED
- User Name :** Anonymous

On the right side of the panel, there are buttons for Refresh, Export, Watch, and Fields Sorted.

Below the panel, a table displays field values:

Field	Value	Key Type
SampID='123'	123	Primary Key
HARMFULL_CHANGE='This is the existing field'		
TEST_OPEN_FIELD='Type something here or don't'		
TEST_AUXFIELD='Auxfield has value'		

# Harmful/Harmless Changes

```
File Home Source
Cut Find and Replace Options Server Manager Para Data Viewer Source Converter 4
Copy Tool Windows Snippets Editor Browse Database Dashboard Session Viewer Windows Explorer
Paste Start Page Reset IDE Database Data Viewer Resource Database Editor Command Prompt
Clipboard Views Options Database External Commands

Start Page IBUC2023.blax X
1 DATAMODEL IBUC2023 "HRS Questionnaire"
2 FIELDPROPERTIES
3 Remark : OPEN
4 IsVisited : TIsVisitedFieldProperty
5 Mode : STRING
6 MODES = SELFADMIN DESCRIPTION ENG "taken by respondent" SPN "tomada por encuestado",
7 IWERADMIN DESCRIPTION ENG "administered by interviewer" SPN "administrado por el entrevistador"
8 LANGUAGES = ENG "English", SPN "Spanish"
9 PRIMARY Sampid
10 TYPE
11 TIsVisitedFieldProperty = (No (0), Yes(1))
12 FIELDS
13 SampID
14 /"SAMPLE ID : STRING[50], NODK, NORF, NOEMPTY
15 HARMLESS_CHANGE
16 ENG "This is a new field, so the change should be HARMLESS when it is added."
17 / "More info specify"
18 : STRING
19 HARMFULL_CHANGE
20 ENG "This is a existing field, so the change should be HARMFULL when it is removed."
21 / "More info specify"
22 : OPEN
23 TEST_OPEN_FIELD
24 ENG "This is the permanent open field in the database. When it is visited,
25 the IsVisited field property for this field is set to YES"
26 / "TEST open field"
27 : OPEN
28 TEST_IS_VISITED
29 ENG "Previous field was visited: ^{FLIsVisited}"
30 : ( CONTINUE (1) ENG "Continue" SPN "Continúe")
31
32 LOCALS FLIsVisited : STRING
33 AUXFIELDS TEST_AUXFIELD : STRING
34 RULES
35 Sampid.KEEP
36 HARMLESS_CHANGE
37 HARMFULL_CHANGE
38 TEST_OPEN_FIELD
39 IF TEST_OPEN_FIELD.IsVisited = YES THEN
40 FLIsVisited := 'Yes, this field was visited'
41 TEST_AUXFIELD := 'Auxfield has value'
42 ENDIF
43 TEST_IS_VISITED
44 ENDMODEL
45
```

File Home

Language : ENG

Settings : CAWI\_Settings

Parallel :

Active Field : TEST\_IS\_VISITED

User Name : Anonymous

Refresh Export

Select fields

Watch

Fields Sorted

Data Entry State Search Commands

F SampID='123'	Primary Key
F HARMLESS_CHANGE='This is the added field'	123
F HARMFULL_CHANGE='This is the existing field'	
F TEST_OPEN_FIELD='Type something here or don't	
A TEST_AUXFIELD='Auxfield has value'	

# Detecting Harmful Changes

```
1 DATAMODEL IBUC2023 "HRS Questionnaire"
2   FIELDPROPERTIES
3     Remark : OPEN
4     IsVisited : TIsVisitedFieldProperty
5     Mode : STRING
6   MODES = SELFADMIN DESCRIPTION ENG "taken by respondent" SPN "tomada por encuestado",
7           IWERADMIN DESCRIPTION ENG "administered by interviewer" SPN "administrado por el entrevistador"
8   LANGUAGES = ENG "English", SPN "Spanish"
9   PRIMARY Sampid
10  TYPE
11    TIsVisitedFieldProperty = (No (0), Yes(1))
12  FIELDS
13    SampID
14      /"SAMPLE ID": STRING[50], NODK, NORF, NOEMPTY
15    HARMLESS_CHANGE
16      ENG "This is a new field, so the change should be HARMLESS when it is added."
17      / "More info specify"
18      : STRING
19    HARMFULL_CHANGE
20      ENG "This is a existing field, so the change should be HARMFULL when it is removed."
21      / "More info specify"
22      : (No (0) "NO", Yes(1) "YES")
23    TEST_OPEN_FIELD
24      ENG "This is the permanent open field in the database. When it is visited,
25        the IsVisited field property for this field is set to YES"
26      / "TEST open field"
27      : OPEN
28    TEST_IS_VISITED
29      ENG "Previous field was visited: ^{FLIsVisited}"
30      : ( CONTINUE (1) ENG "Continue" SPN "Continúe")
31  LOCALS FLIsVisited : STRING
32  AUXFIELDS TEST_AUXFIELD : STRING
33  RULES
34    Sampid.KEEP
35    HARMLESS_CHANGE
36    HARMFULL_CHANGE
37    TEST_OPEN_FIELD
38    IF TEST_OPEN_FIELD.IsVisited = YES THEN
39      FLIsVisited := 'Yes, this field was visited'
40      TEST_AUXFIELD := 'Auxfield has value'
41    ENDIF
42    TEST_IS_VISITED
43  ENDMODEL
44
45
```

The screenshot shows the Blaise Server Manager interface with a dialog box titled "Install survey - Install" open. The dialog box displays the following information:

**Install**  
Installation of surveys

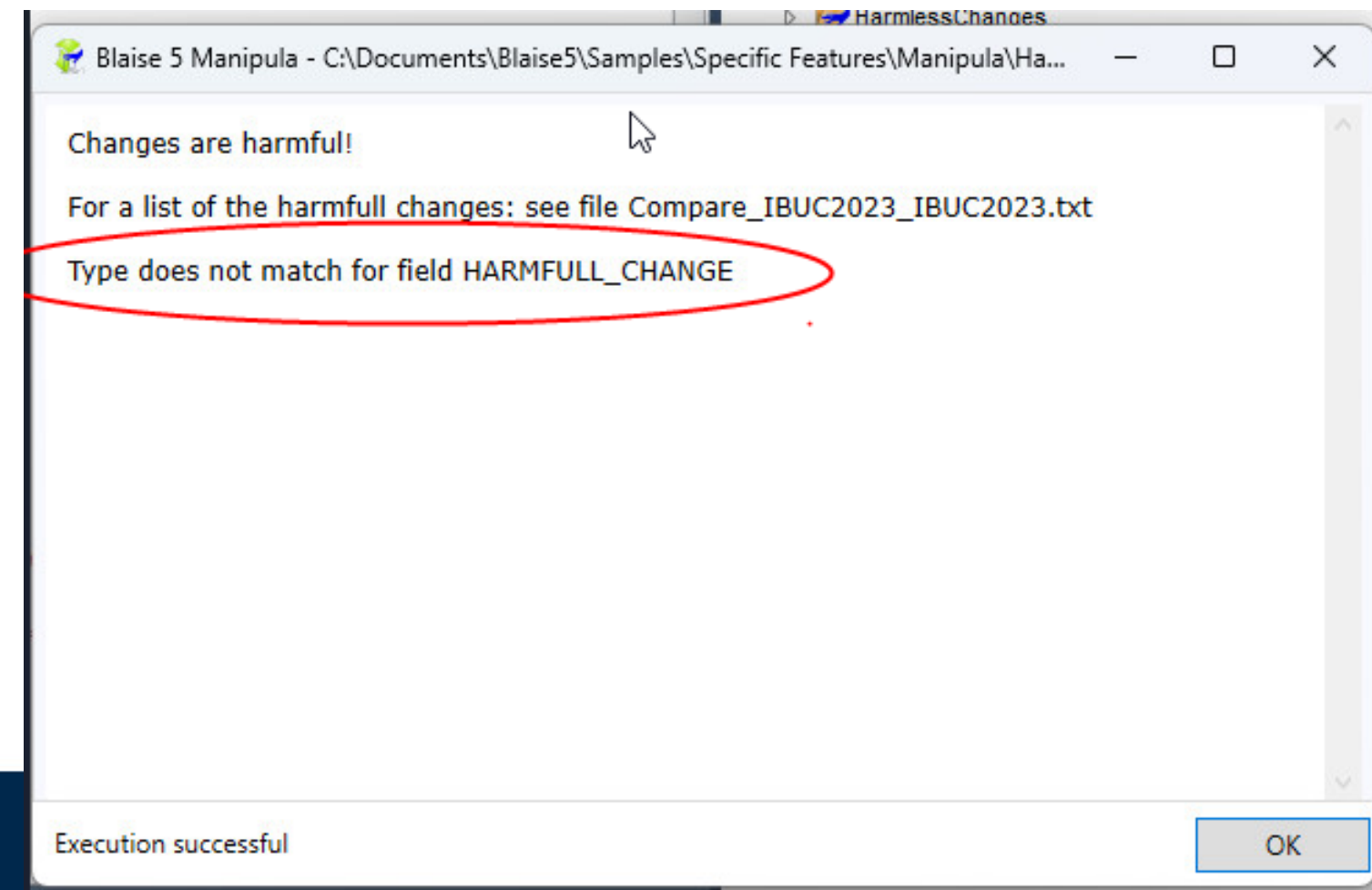
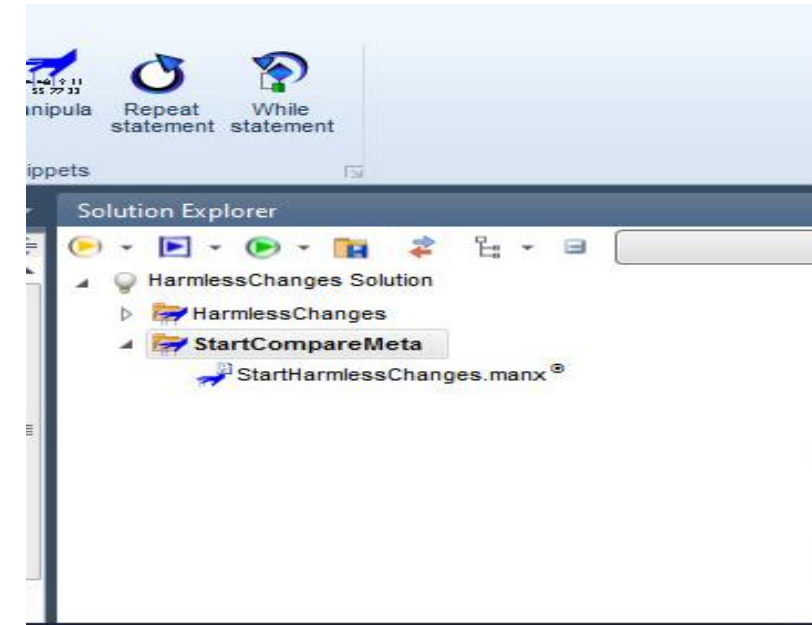
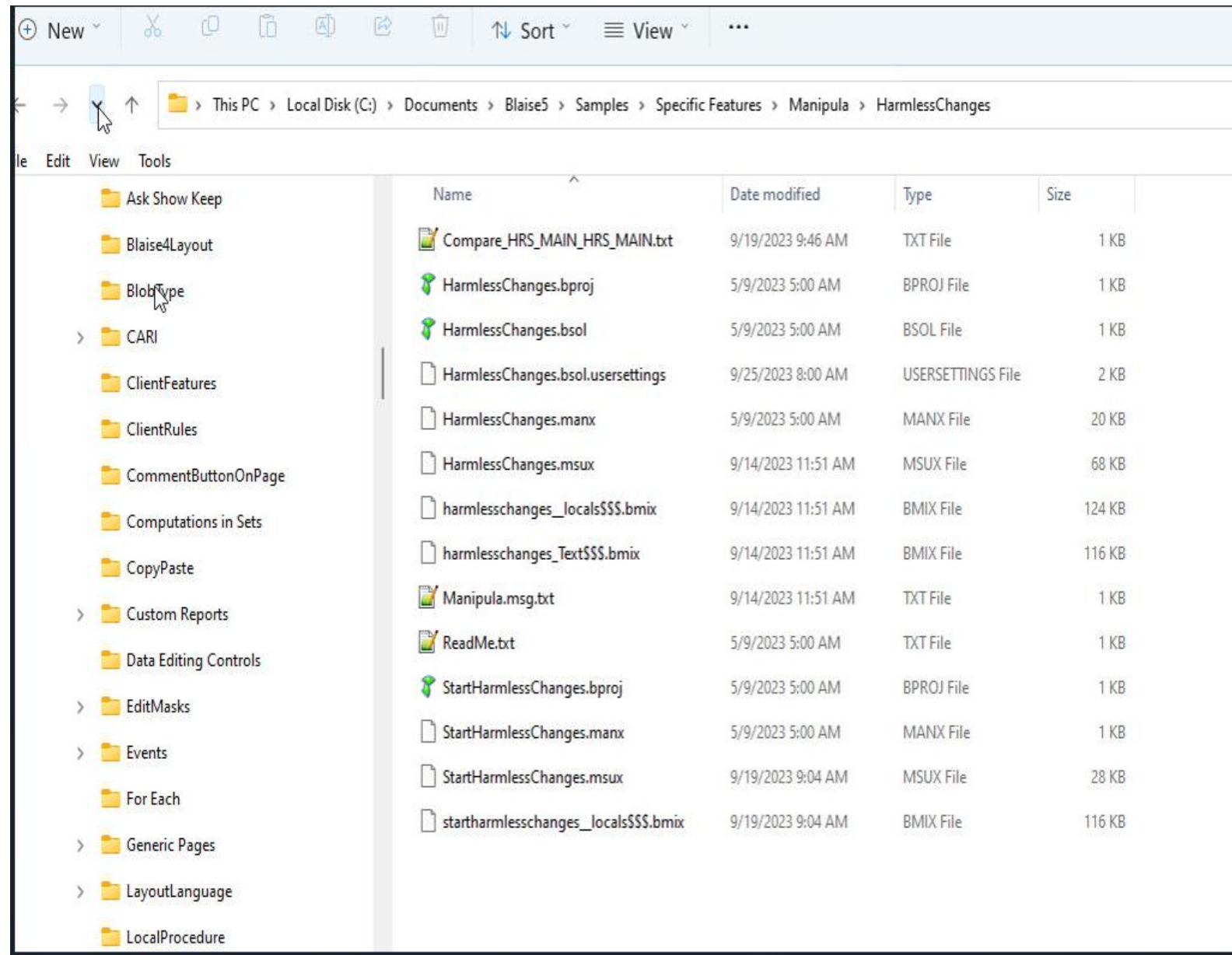
Installation completed with 1 errors

**LocalDevelopment - Surveys:**

Name	Status	Package
IBUC2023	Install failed :Data interface is incompatible with existing data. Please remove existing survey first or allow overwrite of datafile.	C:\SVN\instdev\Blaise5_Examples\IBUC2023\IBUC2023\IBUC2023.bpkg

The background interface shows a tree view of the server structure under "LocalDevelopment" with "Surveys (3)" expanded, listing HRS24, HS\_SandBox, and IBUC2023, all with an "Active" status.

# Harmless Change Detection Tool



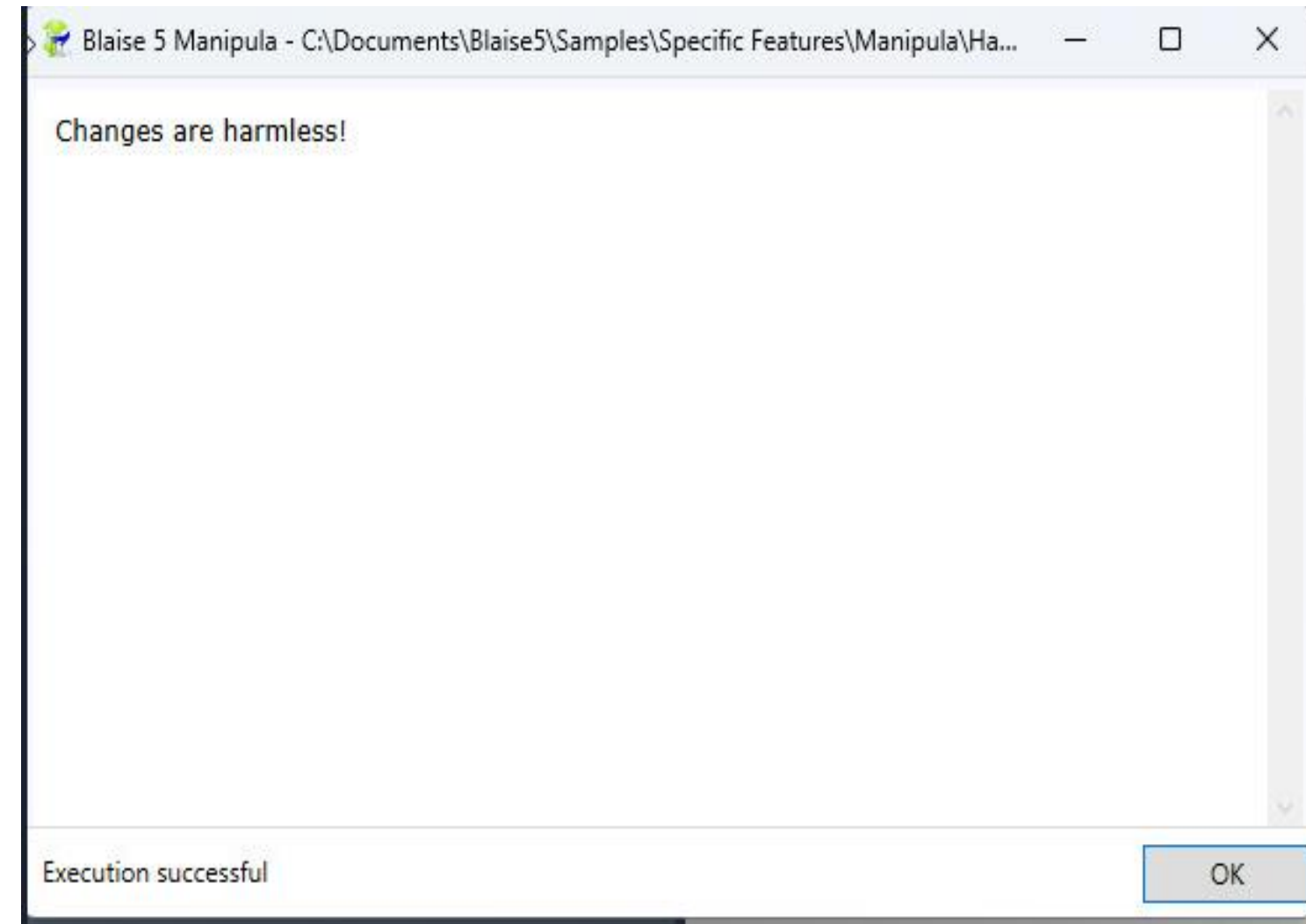
# Handling Harmful Changes between Datamodels

1. We create a new field that has the desired type without commenting out the old one.

```
(keep the old field in the datamodel)
HARMFULL_CHANGE
  ENG "This is a existing field, so the change should be HARMFULL when it is removed."
  / "More info specify"
:OPEN
(Create a new field with the desired type)
HARMFULL_CHANGE_NEW
  ENG "This is a existing field, so the change should be HARMFULL when it is removed."
  / "More info specify"
  : (No (0) "NO", Yes(1)"YES")
```

2. We keep the old field on the route, but put an IF -THEN statement that is never TRUE around it.

```
RULES
  Sampid.KEEP
  (Keep the old field on the route, but so that it is never asked)
  IF 1 = 2 THEN
    HARMLESS_CHANGE
  ENDIF
  HARMFULL_CHANGE_NEW
  HARMFULL_CHANGE
  FROM APPY FIELD
```



# Detecting harmful changes

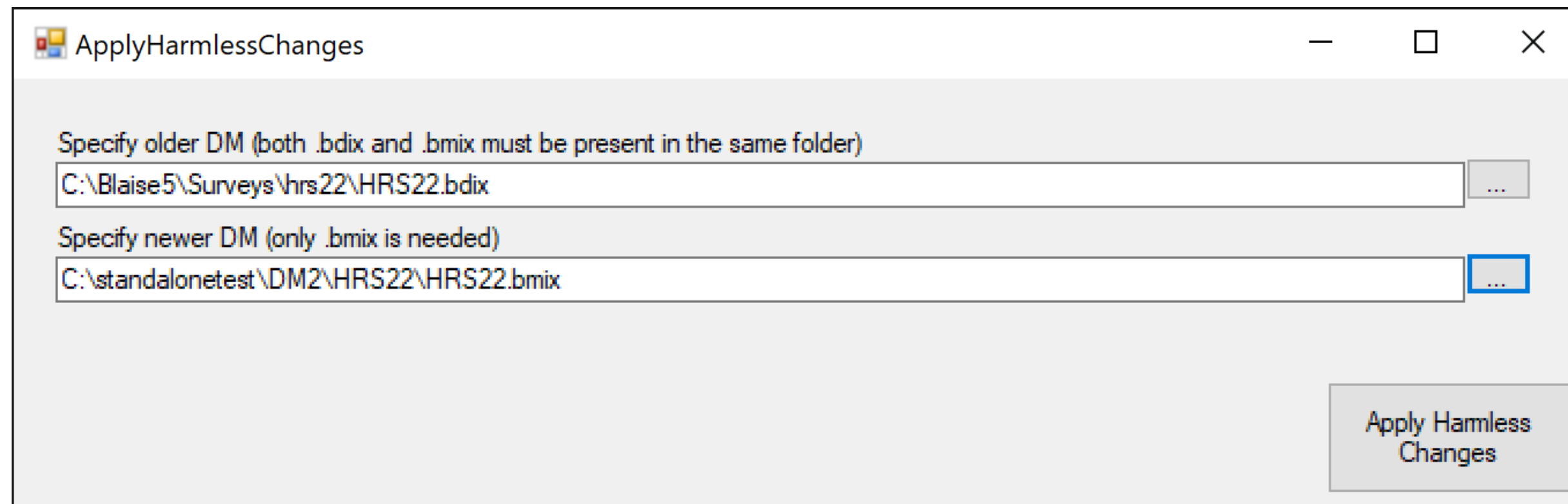
- Blaise 5 ships with a sample Manipula script for detecting harmful changes, which can be found in the samples folder, for example:
  - `\Documents\Blaise5\Samples\Specific Features\Manipula\HarmlessChanges\HarmlessChanges.bsol`
- This script compares two datamodels and reports whether it detects either 1) no changes, or 2) only harmless changes, or 3) it lists all harmful changes
- This manipula sample did not work well for HRS, due to a bug causing a listing of more than a thousand false positives
- As a result of discussions with CBS, a small code change was made to make this script work as intended for HRS



# Harmless Change Update Tool

```
private void ApplyHarmless(string oldDMbdixFilename, string newDMbmixFilename)
{
    DataLinkAPI.IDataLink5 dl = DataLinkAPI.DataLinkManager.GetDataLink(oldDMbdixFilename) as
DataLinkAPI.IDataLink5;
    dl.ApplyHarmlessChanges(newDMbmixFilename);
}
```

## Simulating a server install



The screenshot shows a Windows application window titled "ApplyHarmlessChanges". The window contains two text input fields for file paths, each with a browse button (three dots) to its right. The first field is labeled "Specify older DM (both .bdix and .bmix must be present in the same folder)" and contains the path "C:\Blaise5\Surveys\hrs22\HRS22.bdix". The second field is labeled "Specify newer DM (only .bmix is needed)" and contains the path "C:\standalonetest\DM2\HRS22\HRS22.bmix". A button labeled "Apply Harmless Changes" is located at the bottom right of the window.

# HRS Session Data Persistence Between Datamodels Use Cases

- Offline Standalone
- Normal Server Based
- Standalone Sync (also a 'between sessions' issue)

# Persisting Between Datamodels - Standalone

- Here is an illustrative process for datamodel migration with harmless changes to preserve session data which works in Standalone mode:
  - 1. copy a preloaded bdbx into the bpkg of the initial datamodel (or equivalent process to handle preload)
  - 2. copy this bpkg of the initial datamodel into the same folder as the special dep (and its associated .dlls)
  - 3. run the special dep with the commandline argument -RunMode:ThickClient (and other normal arguments) --- and test some unfoldings and signals, and suspend
  - 4. run the "apply harmless changes" tool against the deployed .bdix and the .bmix from the new datamodel with the harmless changes
  - 5. copy the 3 files that are updated by the "apply harmless changes" tool from the deploy folder into the same preloaded bpkg from step 2
  - 6. run the special dep with the commandline argument -RunMode:ThickClient (and other normal arguments) --- it will resume in the correct place in the new datamodel

# Persisting Between Datamodels – Server

- Process for datamodel migration with harmless changes to preserve session data -- Server-based (not incl. DIM)
  - 1. copy a preloaded bdbx into the bpkg of the initial datamodel (or equivalent process to handle preload)
  - 2. in Server Manager install the bpkg of the initial datamodel
  - 3. in Server Manager start in browser with normal arguments --- and test some unfoldings and signals, and suspend
  - 4. run the "apply harmless changes" tool as administrator against the deployed .bdix in the \Blaise5\Surveys folder and the .bmix from the new datamodel with the harmless changes
  - 5. in Server Manager start in browser with normal arguments --- it will resume in the correct place in the new datamodel

# Persisting Between Datamodels -Sync

- This concerns transferring cases between the server and the laptop via the DIM (for mode switching)
- Previously, CBS had indicated a variety of obstacles to this working
- Discussed this topic with CBS at various meetings
  - Ralph from CBS did some tests and reported the results in a series of emails in JIRA that this should work now
  - Ralph: “That is, that syncing new datamodels with the same checksum (where you only fix a typo in question text) or any harmless updates (wider answer-types or new fields) should be pushed to the client.”
- We hope Blaise 5.14 will support this capability

# Importance of Session Data

- Why does using session data matter to HRS?
- HRS relies heavily on complicated logic to avoid unnecessary question sequences in what is already a very long interview (e.g. unfoldings)
- The relatively recent addition of CAWI complicates this further
  - the need to allow backing up means less “gates” spread throughout for R sanity
  - without gates, sequences that are reliant on auxiliary fields may be recalculated with data loss on resume
  - inability to force answers (e.g. press 1 to continue) means our usual “gates” are not practical anyway -- have to use something like isVisited to trigger actions when you can never be sure any given question will be answered
  - isVisited is normally only stored in session data, but it is now possible to persist it by re-declaring it (we are not yet confident about relying on this).

# Going Forward

- One clear trend in all of this is that each set of Blaise fixes have unmasked new deeper issues with HRS and session data over time, but it is also the case that the magnitude of the problem keeps shrinking wave-on-wave.
- That is, whereas the first issue impacted potentially thousands of cases, the subsequent issue may have affected hundreds of cases, and the most recent is impacting perhaps less than a hundred.
- Additionally, new versions are bringing more features to address these issues. For example, the Blaise 5 Server Manager in Blaise 5.13 presents more options when installing surveys over previously existing ones.
- Overall, our hope would be that questions about what happens to session data will eventually permeate every part of the system.