The Challenges of Implementing the China Family Panel Study (CFPS)

Jiahui Yao, Peking University, Gina-Qian Cheung and Youhong Liu, University of Michigan

1. Introduction

The Chinese Family Panel Study (CFPS) is the largest and most comprehensive national panel study ever undertaken in China. It is designed to collect three levels of data: individuals, families, and communities. The survey topics include society, economy, population, education, and health changes. The data are used for academic research and policy decision-making in China.

The CFPS survey has features modeled after the Panel Survey of Income Dynamics (PSID), the National Longitudinal Study of Youth (NLSY) and the Health and Retirement Study (HRS). The study is hosted by the Institute of Social Science Surveys (iSSS) of Peking University.

CFPS preparation began in 2005. Two paper-and-pencil surveys were conducted in Beijing, Hebei, and Shanghai in 2007. In 2008, CFPS had a pilot for 2,400 households in Beijing, Shanghai and Guangdong. A follow-up CAPI survey was launched in 2009. During the 2009 pilot, CFPS fully tested all CAPI systems, including Blaise data collection, real-time sample management, real-time technical support, data quality, and real-time monitoring. CFPS started production in April, 2010, and completed data collection in September, 2010. The total completed cases are about 85,000!

In this paper, we will first describe the CFPS CAPI system architecture. Then we will discuss the challenges encountered during implementation of Blaise.

2. CFPS CAPI System Architecture

As shown in figure 1, the CFPS CAI system includes the following three subsystems (discussed in detail below):

- Data Collection system
- Data Synchronization System
- Systems At iSSS Center

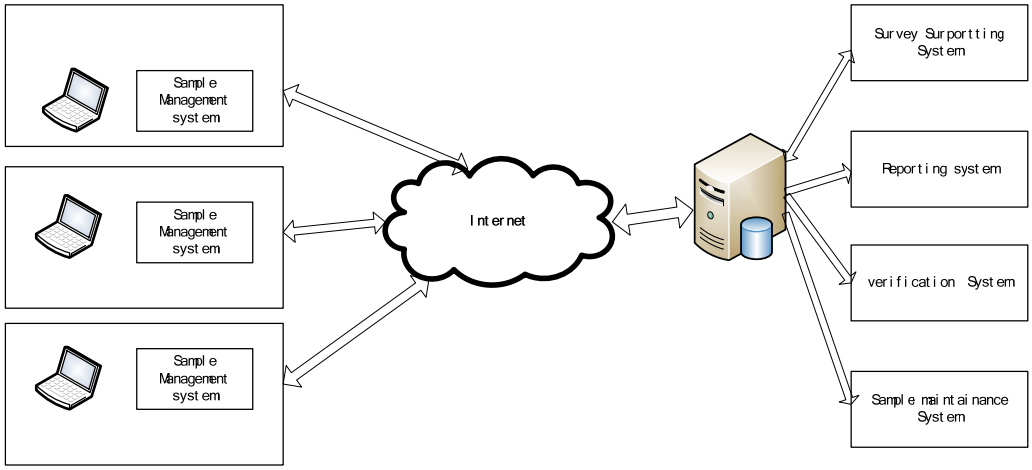
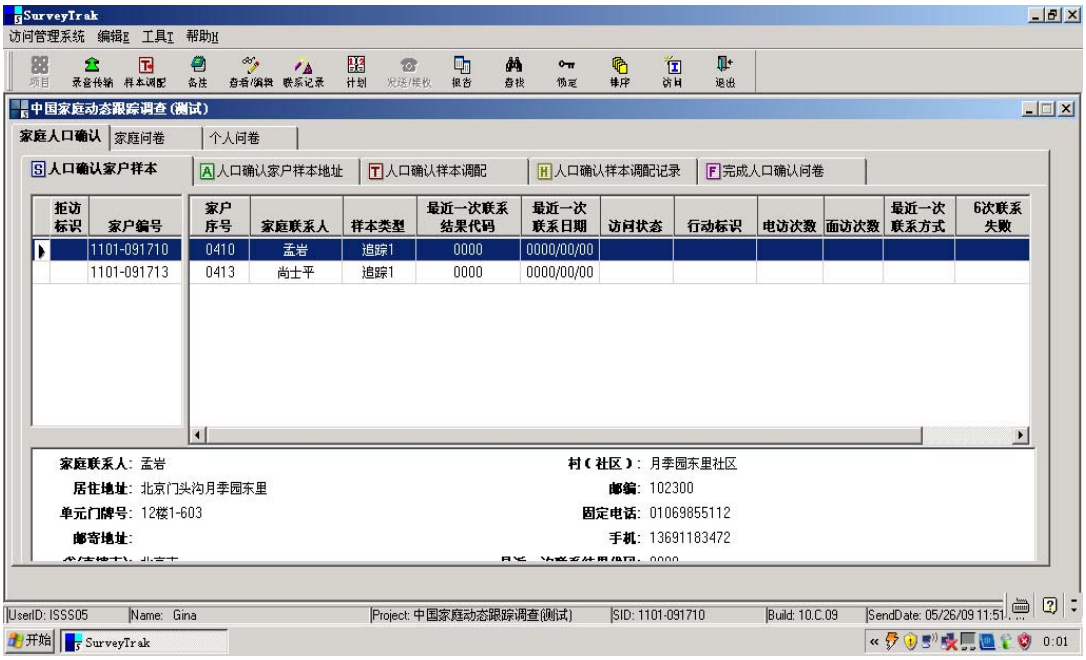


Figure 1 – CAI System

2.1 Data Collection System

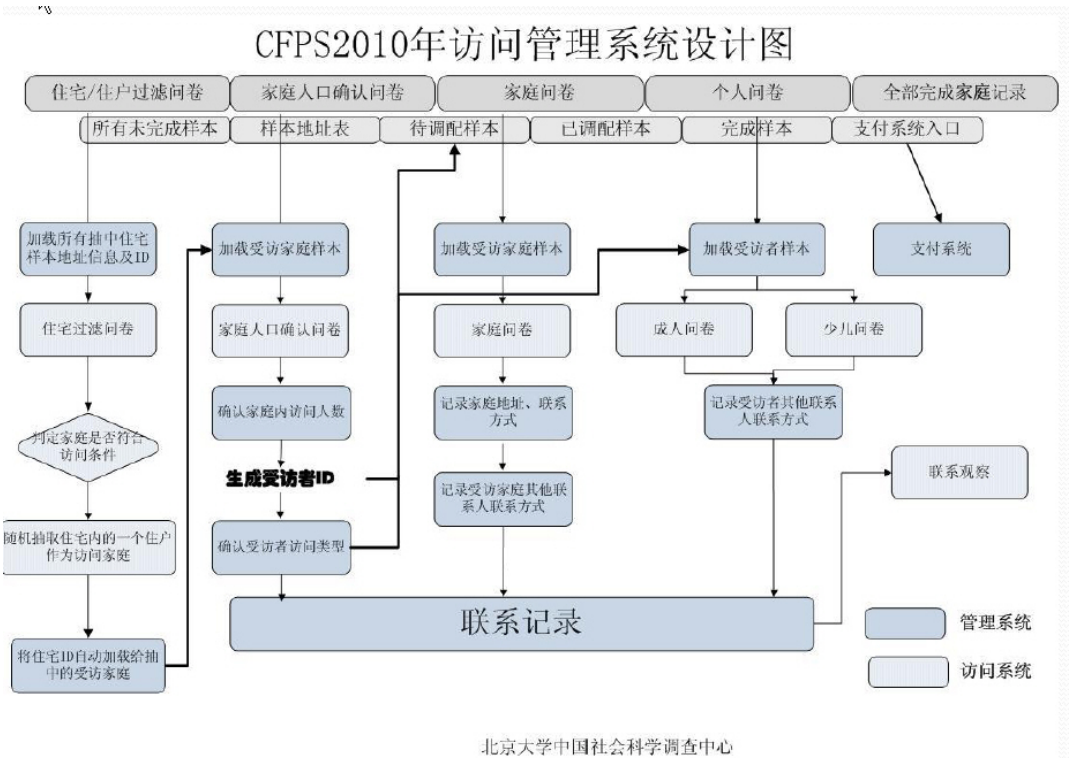
2.1.1 Sample Management System (SMS)

The CFPS Sample Management System evolved from the University of Michigan's Sample Management System, SurveyTrak. Staff from Michigan and iSSS worked closely together to develop the system for CFPS. The final product is all in Chinese.



2.2.2 Six Blaise Questionnaires

The six CFPS questionnaires are: Screening, Family Member Confirmation, Family, Community, Adult, and Child.



2.2 Data Synchronization System

The data synchronization system synchronizes the interview data and paradata between the interviewer's laptop database and the server database. The system has both server side and client-side databases. The basic principles are as follows:

- All database operations are recorded in the log table (such as add, delete, modify), and the last successful synchronization time (both upload and download) is recorded to another table;
- When uploading data, the data synchronization system compares the log table with the last successful upload time, and if there are differences, the latest updated data is uploaded to the server;
- When downloading data, the data synchronization system compares the log table with the last successful download time, and if there are differences, the latest updated data is downloaded to the client database.

2.3 Systems at the iSSS Center

To ensure production data collection runs smoothly and with high quality, several systems have been built to run at the iSSS Center with the server database.

2.3.1 Survey Supporting Systems

- Sample Transfer System - This is used by supervisors to transfer sample from one interviewer to another interviewer.
- Technical Support System - Supervisors use this system to submit technical issues raised by interviewers. They assign the problems for timely resolution to appropriate technical staff.
- Sample Help System – This is used to process special sample such as sample that were refused three times or that could not be contacted after six tries. In such cases, the iSSS Center contacts the respondent and assigns a special status code to the sample, and then informs the interviewer whether to further contact the respondent.
- Interviewer Management System – This is used to manage the interviewer's information, such as adding a new interviewer or modifying information for existing interviewers.

2.3.2 Reporting System

SAS Web Report Studio is used to report on daily survey processes, such as sample completion and sample distribution. The data are shown in dashboards, tables and curves.

2.3.3 Data Verification System

The data verification system selects samples to be verified according to the values of a

selected questionnaire's variables and other user backend database variables. It is a web system using BlaiseIS. The verification can be conducted by telephone, by review of sound recordings, and by face-to-face interviews.

2.3.4 Sample Maintenance System

This system is mainly used for sample tracking and maintenance after the completion of interviews. Similarly, this is a web system that uses BlaiseIS to complete questionnaires and to record the results of sample maintenance. The other function of the Sample Maintenance System is to send respondents customized holiday greetings through Short Message Service and E-mail.

3. Challenges

3.1 Implementation Challenges in SMS Development

3.1.1. Implementing family level of survey with five subprojects

- The entry point of the family level survey is the Household Screening Survey. Certain criteria, such as how long the family has lived in its current location, are used to determine if a family is eligible for further interviews.
- The second survey is the Family Member Confirmation Survey, in which all members in the household are listed in a table of the Blaise instrument. After the survey, SMS generates sample lines in the Adult or/and Child projects. The number of projects generated is based on the answers from the Blaise family list table. For example, if there are two people in the family who are older than 16, then two sample lines in the Adult project will be generated.

Both of these surveys are screening tools used to select families and individuals to be interviewed. The main surveys are the Family Overall, Adult, and Child surveys

3.2 Challenges for Blaise Development

In this section, we will describe details that are unique for implementing the CFPS in Blaise.

3.2.1 Setting up the Blaise Chinese Environment

In 2003, University of Michigan had a paper discussing how to display Chinese in Blaise. It mentioned that the NJ Star program was used to display Chinese for a multi-language Blaise instrument. Since CFPS is not a multi-language instrument, it is not necessary to use a third party program such as NJ Star to display Chinese. There are certain computer system settings that can be changed to enable Blaise to display Chinese properly.

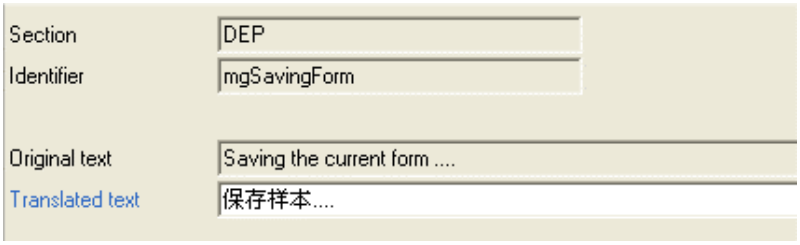
The following two changes to the Windows operating system are needed in order to display Chinese characters correctly:

1. Control Panel->Regional and Language Options->Advanced->Select a language to match the language version of the non-Unicode -> **Chinese (PRC)**
2. Control Panel->Regional and Language Options-Regional Options->Standards and formats->**Chinese (PRC)**

3.2.2 Translate Blaise System Text to Chinese

The above two settings in Regional and Language Options help to deal with Blaise questions and answers in Chinese language. To change Blaise System text to Chinese, we use the Blaise translation tool to set up a language INI file in Chinese. The INI file contains translated menus, dialogs, forms and error messages.

1. Translating: The basic object in the translation process is a Blaise database. *BlTxtTra.bdb* contains all the system text. The Blaise setup for this (*BlTxtTra.bla*) and the text database itself can be found in a subfolder of the Blaise system folder called *Translat*. To Translate: 1) open BlTxttra.bdb with Dep.exe; 2) **Setting Up the Blaise Chinese Environment** ocate the section and identifier that need to be translated; and 3) in the translated text field, input the Chinese text. See the figure below:

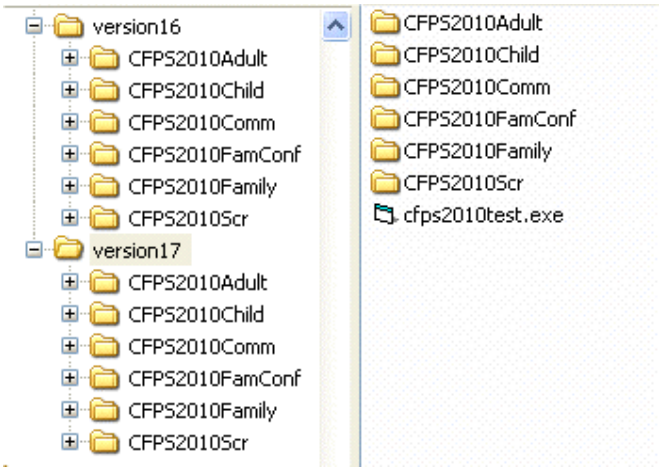


2. Transforming the translation to the INI file format: The Manipula setup BlTxt2IN.MAN in the Translate will conduct the transformation for us. The output of this Manipula program is a file called *Blaise.tra*.
3. Using the new language file: There are several ways to use the new language file. CFPS chose to use a DEP.exe parameter e.g. /#C:\blproj\CFPS\Blaise.tra.

3.2.3 Testing System for CFPS

There are six datamodels for CFPS 2010. As development evolved, it was necessary to build a testing system so that programmers could setup and update datamodels easily for the testers to test the latest changes. For quick development, VB6 was selected to construct the system.

The Testing system folder structure: When a new version is ready to test, first a version folder is created, e.g., version16. Second the VB6 cfps2010 is copied to the folder. Then, the datamodel files are copied to their corresponding directories.
By implementing this type of folder structure, a Blaise programmer can easily create and send the latest version to testers, and the older versions are still available for change verification.



3.2.4 Modify Blaise Preload in the Testing System

Some of the CFPS datamodels have complicated preload block structures. To ensure that different preload sceneries can be tested, the testing system must enable testers to modify preload values. The following steps in the Blaise datamodels and VB programs are set up for this purpose:

1. Define an Auxfield in the Datamodel: PreLoadForTesting
2. In the Datamodel's rules section, add the following code

```
PreLoadForTesting.KEEP
IF PreLoadForTesting = 1 THEN
    {To enable preload modification during test}
     Preload
ENDIF
Preload.KEEP
```

3. In the cfps2010test.exe VB program, the DEP parameters assign a value to this Auxfield. *i.e., /Q=PreloadForTesting=1*.

All three steps are defined so that in the testing environment, users can change preload field values freely because "PreloadForTesting=1" is fed to the database form. For the production environment, the preload block is locked. This is because the field assignment is not included in the DEP parameters. This is an elegant approach, because no code changes are required between testing and production to avoid potential programming errors.

The CFPS testing system described above is effective. However some system enhancements are required. For example, Bug Log is not implemented; users need to put their testing comments in a separate Excel or Word file. At the time development was taking place, CTT was not ready for CFPS, due to some conflicts between Chinese characters and CTT's backend SQL Server database design. With some database modifications and small code modification, CTT will be ready for testing Chinese projects.

3.2.5 Blaise Word and Math Tests in CFPS

CFPS2010's Adult and Child instruments include tests for math proficiency and word recognition. One of four sets of math tests and one of eight sets of word tests are randomly assigned to a selected respondent. The following steps are implemented in the Blaise instruments:

1. In the Family Confirmation instrument, we calculate which test group number a family would start with: RandomNum := Random (8) + 1. This number and the person's index number in the household list table are preloaded in the Adult and Child instruments.

2. In the Adult and Child instruments, the calculations below are used to determine which test group is used by the respondent. These calculations ensure that members in a family are assigned to different test groups:

   WhichMathList := (RandomNum + RespdentIndex-1) Mod 4 + 1
   WhichWordList := (RandomNum + RespdentIndex-1) Mod 8 + 1

3. For the word recognition level test, an external look up database was utilized to store the word list - see the structure and data below, note that there 34 words in each group:

\\SRC-Douvan\SRO\share\yliu\Blproj\CFPS2010\Dev\externalbdbs\

WordNumber	Group1	Group2	Group3	Group4	Group5	Group6	Group7	Group8
1	木	入	田	火	出	广	书	方
2	朵	色	比	真	乡	画	休	兔
3	贝	乐	桥	冬	父	刚	林	叫
4	百	丢	并	斗	丰	定	顾	连
5	如	甲	歌	际	折	铅	亚	景
6	航	农	虽	肯	科	史	困	欠
7	岛	末	鲜	股	瓦	冒	季	杜
8	扮	环	寿	芒	冈	滑	亏	壶

In the instrument, an array is used to store data read from the lookup database:

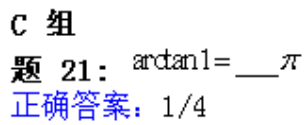
```
FOR I:=1 TO 34 DO
    WordData.Search      (I)
    WordData.Read
    IF WhichWordList = 1 THEN
        Words [I] :=    WordData.Group1
    ELSEIF WhichWordList = 2 THEN
        Words [I] :=    WordData.Group2
    ELSEIF WhichWordList = 3 THEN
        Words [I] :=    WordData.Group3
    ……..
```

ENDIF
xWordFill:=Words[I]
QWordTest [I]
ENDDO

The individual tests are represented in a field array. The fill xWordFill is determined with the above logic.

```
QWordTest (WordTest)
"@F&@F@I [出 示 字 表 ^GroupLetter[WhichWordList] ]  @I
@/@/字表^GroupLetter[WhichWordList]第@C^I@C个字: @I^xWordFill
@/@/@B访员注意: @B不能读出来!
@/@|请受访者用普通话读词. 对音调、前鼻音、后鼻音、平舌、翘舌不做区分.
@/@|不允许受访者查字典或询问他人.
@/@|如果受访人不知道或拒绝,输入""5"",F2输入理由  @I
 "  /"字词 - ^I" : ARRAY[1..34] OF (_1 "对", _5 (5) "错")
```

4. For the math proficiency test, html files are used. The files are named with a specific format, i.e., x_y.html, where x indicates the group number and y denotes the test number in the group. For example, 2_10.html is for test number 10 in group 2. The benefit of using html files is that we can link images into the html files. For example, in the figure below, the math representation cannot be presented by text. The technique here is to put the math image in an image file, then link the image file to the html file. The image files can be generated by a non-programming person to improve efficiency.



In the instrument, the file name is determined by the following code:

xQMathTest := str (WhichMathList)+ '_' + str(I) +'.html'

The field text in the math test is shown below. The xQMathText fill is the file name from the string concatenation above:

```
QMathTest (MathTest)
"@F&@F@I [出 示 数 学 表 ^GroupLetter[WhichMathList]]  @I
@/@/说出^GroupLetter[WhichMathList]组第@C^I@C题的答案:
@/<html><a href='c:/blproj/cfps2010/mm/^xQMathTest'>按这里看题目:</a></html>
@/@/@I@B访员注意: @B
@/@|记录受访者的答案!
@/@|测试过程尽量让受访者独立思考.
@/@|不允许受访者使用计算器或询问他人.
@/@|如果受访人不知道或拒绝,输入""5"",F2输入理由  @I
" / "数学-^I" : ARRAY [1..24] OF (_1 "对", _5 (5) "错")
```

Database

Just as at the University of Michigan, CFPS2009 used SQLAnywhere as our sample management database. Due to the availability of version 6 of SQLAnywhere in China, we had

to update the database to SQLAnywhere version 11 with the help of our colleagues at Michigan. There were many problems encountered due to the instability of SQLAnywhere version 11. The main issue is data synchronization.

With the above problems in CFPS2009, we decided to migrate the database to MSSQL2005. The major effort for this migration was to develop a new data synchronization function that integrates the new database platform and the sample management system.

The new synchronization function operated smoothly at the earlier stage of 2010 interview. After two months, synchronization became very slow and problematic when the database reached 3 Gigabytes. iSSS and Michigan worked together to find a solution – regular database backup and weekly database clean-up.

3.3 BlaiseIS

CFPS used BlaiseIS in two systems – Data Verification and Sample Maintenance. The major issue with our original design for BlaiseIS was preloading data into the Blaise database, which would lock the database, preventing data saving after completion of a BlaiseIS interview.

After consultation with the Blaise developer, we found that the preload operation needed to be performed before the sample is accessed from BlaiseIS's web pages. After implementing this protocol, the database-locking problem was resolved.

4. Conclusion

After just one and half years of using CAI and Blaise, the iSSS Center has gained tremendous experience. Of course, there are many areas of improvement needed, such as:

- Capturing Journal data in BlaiseIS;
- Testing the BlaiseIS instrument (this will be accomplished by modifying the University of Michigan's CTT system as mentioned above);
- Implementing a CATI capability in future surveys.

In the process of implementing CAPI, the iSSS Center encountered many challenges and difficulties. Working with colleagues at the University of Michigan, they met these challenges. The Blaise developer also answered many questions, whether they were simple or complicated. And above all, iSSS successfully completed the CFPS2010 survey. It is a landmark in Chinese social research history.