

Time Use Diary Data Capture System built in Blaise and Maniplus

Fred Wensing, Softscape Solutions, Australia

1. Introduction

Time use surveys generally make use of paper diaries that respondents fill in over a specified period. The activity descriptions in these diaries need to be captured in electronic form and coded so that data can be analysed.

This paper describes a time use diary data capture system built for Statistics New Zealand using Blaise and Maniplus.

2. The Time Use Survey

For every household selected in the Time Use Survey, up to two adults are selected to participate in the survey. These persons are required to complete a hand-written diary of all the activities they do within a 48-hour period. Different days are allocated to different persons to ensure that all days of the week are covered.

As well as a primary activity, participants may record up to three additional activities which they may be performing at the same time. The diary also has provision to record where the activity is taking place, for whom and with whom the activity is done.

When the completed diary is collected, a supplementary Blaise questionnaire is administered by an interviewer to collect other socio-demographic information which will later be used in the analysis.

The paper diaries are returned to the office for data capture.

3. System description

The main components of the Time Use Data Capture System are:

- A Blaise instrument that is used to record diary entries and assigned codes
- A central register that keeps track of the diaries and progress with coding
- A coding list of activities and synonym entries
- A high level interface to provide entry to the system
- A user interface that assists with the data capture and double entry functions
- An interface to assist with the marking of dual coded entries

For quality control purposes a selection of diaries is coded a second time by another coder. To support this dual coding function a parallel environment is set up using a duplicate set of programs. Dual coded diaries are compared and discrepancies checked. Provision exists to produce a corrected record from the marking process.

4. The Blaise data capture instrument

The Blaise diary data capture instrument starts with header information such as the identification of the respondent, the allocated days and dates for the diary. Basic demographic information obtained from the supplementary questionnaire is also recorded in the diary instrument.

Once the header information has been entered the episodes can be entered one at a time (in sequence). Each episode consists of one or more activities which are entered using literal text strings and a trigram lookup on the activity code list to locate a synonym of the literal string. Once a synonym has been selected then the corresponding code is inserted by the system into the episode. Other details such as the location of the activity, for whom and with whom the activity is done are also recorded for each episode.

The instrument makes provision for 24 or 48 hours to be captured based on the designated days and actual recorded diary days.

Operators have a choice of entering the data in one of two methods:

- The Form approach - one in which each episode requires a full page of entries (or form) to be completed.
- The Table approach - one in which each episode is recorded in a row in a a table.

For the Form approach (see Figure 1) the vertical split screen is used so that the whole episode can be readily seen. The identity and basic characteristics of the respondent are also visible in the Infopane where the corresponding activated question (and categories) is visible.

Figure1. Form approach to capture episode data

The screenshot displays the 'Time Use Survey Diary' application window. The interface is split into two main vertical panes. The left pane, titled 'Infopane', contains respondent information: Household: R289687003, Person: 1, #Members in HH: 1, Interviewer ID: 123, Respondent Name: Michael Palin, Age: 58, Sex: Male, Paid work: No, Occupation: (blank). Below this is a section for 'Activities for Episode[2]' with the question 'Where were you OR how were you travelling?' and a list of 11 radio button options: 1. At home, 2. At other people's home, 3. Work place or place of study, 4. Public or commercial area, 5. Bush, beach or wilderness, 6. Marae and other sites of cultural significance to Maori, 7. Private or non-commercial area, 8. Travelling by foot or bicycle, 9. Travelling by car, motorcycle, truck or van, 10. Travelling by bus, train, taxi, ferry, plane, 11. Other locations or modes of transport, 97. Response unidentifiable, 99. Not stated. The right pane contains a form for a single episode. It includes fields for 'Start' (Day: 1, Time: 8:00 AM), 'Stop time' (8:30AM), and 'Stop day' (1). A section titled 'Activities' lists four entries: 1. Literal: eat breakfast, Activity: eat breakfast; 2. Literal: Newspaper, Activity: Read newspaper; 3. Literal: (blank), Activity: (blank); 4. Literal: (blank), Activity: (blank). Below this are fields for 'Location', 'With?', and 'Codes' (11411, 44311). The status bar at the bottom shows '3/125' and 'DiaryData.e[2].Where Dirty'.

Short-cut keys are provided for common entries like sleep and work and other functions needed for coding such as clearing an entry or providing a new literal string.

The Table approach (see Figure 2) enables the operator to see more than one episode at a time and for table functions to be used to insert, split, merge or delete entries, which would otherwise be difficult to do. Short-cut keys are provided to activate these table functions.

The preferred method of data entry is to use the Form approach. No matter which method is used, however, the same information is recorded for each episode. Entries under the two methods are synchronised and changes made under one approach are immediately transferred to the other. The active method is noted on the parallel block tabs.

Summary data is produced within the instrument for each 24 hour period to assist with edits and to identify missing entries (eg. of eating or sleeping).

Edits are provided within the instrument to ensure that episodes are entered properly and that full days of entries are recorded. Problems identified by the edits are located using the **Navigate show all errors** menu item. The status of each edit check is also 'piped' into a series of flags to enable the system to produce a text report of the edit failures.

Figure 2. Table approach to capture episode data

The screenshot shows the 'Time Use Survey Diary' application window. At the top, there are menu options: Forms, Answer, Navigate, Options, Help. Below that is a toolbar with buttons: Method, Insert, Delete, Split, Merge, Clear (F6), Literal (F7), Sleep, Work, Avail for care, Education. The main window title is 'TUSDiary | Diary Data (Passive) | EQFields | Table Data (Active) | SummaryDay1 | SummaryDay2'. Below the title bar, there is a summary section: Household: R289687003 Person: 1 #Members in HH: 1, Interviewer ID 123, Respondent Name: Michael Palin, Age: 58, Sex: Male, Paid work: No, Occupation: (blank). Below this is a text prompt: 'Activities for Episode[1] On what day did the episode start?'. The main area contains a table with columns: Day, Time, Stop time, Stop day, 1. Literal, Activity, Code 1, Who for, Org(s), 2. Literal. The table has 11 rows, with the first two rows containing data: Row 1: [1] 1 4:00AM 8:00AM 1 Sleep Sleep 11211; Row 2: [2] 1 8:00AM 8:30AM 1 eat breakfast eat breakfast 11411 Newspaper. The status bar at the bottom shows '1/1 | DiaryTable[1].Start_Day: Dirty'.

5. Central register

In order to manage the data capture processes, the system has a central register which consists of a Blaise database and a management interface (see Figure 3).

The register is loaded with the details of households involved in the survey and whether there are diaries to be captured and coded. The register instrument starts with the Household identifier and a few summary attributes of the household such as its size. It then holds the identities of various persons involved with the household such as the interviewer, allocated coder, actual coder, dual coder and marker.

Figure 3. Register and management interface

The screenshot shows the 'Data Processing Manager Task List' interface. It features a table with columns: Household, Status, Status Date, Alloc. coder, Actual coder, Dual coder, Marker, Status Diary 1, Status Diary 2. The table lists 12 households with their respective statuses and dates. To the right of the table is an 'Allocation' section with buttons for Person, Coder, Dual coder, and Marker. Below the table is a search bar and a 'Key type' dropdown set to 'Primary key'. At the bottom, there are 'Actions' (Details, QA select, Non-Resp, Re-select), 'Filter' (Define, Refresh), and 'Reports' (HH_Status, DP_Workload) buttons. A status summary at the bottom reads: 'Total:12 Loaded:9 Assigned:1 Coding:2 Dualing:0 Marking:0 Done:0 Non- resp:0 Export:0'. An 'eXit' button is located at the bottom right.

The Register instrument also holds information about the one or two diaries that were completed by respondents from that household. The status values of each diary are updated directly by the programs which manage the diary coding, double entry and marking processes. The overall processing status of the household is then derived from the status values of the two diaries.

Most interactions with the Register data file are from the (Maniplus) programs which manage the various processes in the system. The main process which can be used to make selected changes to the status or allocation of persons is the Register management interface which has buttons that can be used to activate the various tasks.

6. Coding list of activities

The captured activities are coded to a structured 5-digit code list maintained as an external file to the diary instrument.

Once the literal string for an activity has been entered then the system looks-up the code list for that text string and assigns the corresponding code.

If the literal string is not found in the activity code list then the operator can do one of two things:

- Attempt to use a synonym description, which triggers a trigram search of the activity coding list to find a suitable result, or
- Lodge a new (uncoded) literal string into the record for subsequent decision by those who maintain the activity coding list

The Diary data capture system provides an Activity maintenance interface (see Figure 4) to enable various code maintenance functions to be carried out.

Figure 4. Activity code list maintenance interface



The Activity code maintenance functions include:

- Extracting a list of activity descriptions which have no code
- Importing a new list of activity descriptions and synonyms
- Identifying mismatches between old and new versions of the activity code list
- Running a program that will update previously uncoded activities by re-applying the look-up process against the new code list

7. Main interface to provide entry to the system

The system is entered through the main interface which gives access to the various processes that are to be carried out.

The main interface is supported by a list of authorised users which determines their role in the system (Data Processor, Quality Manager or Data Processing Manager). Depending on the role of an operator, the interface will provide access to more or less functionality (see Figures 8A and 8B).

Figure 8A. Main interface as seen by a data processor

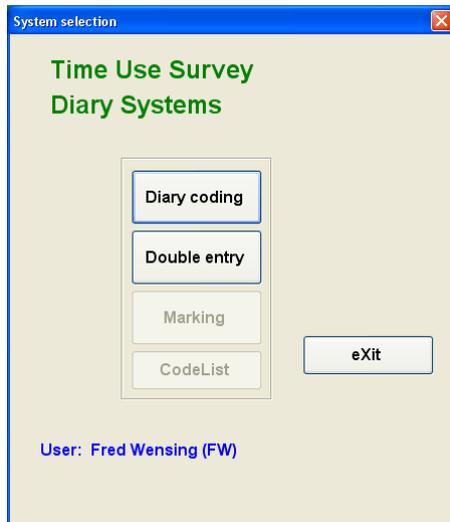
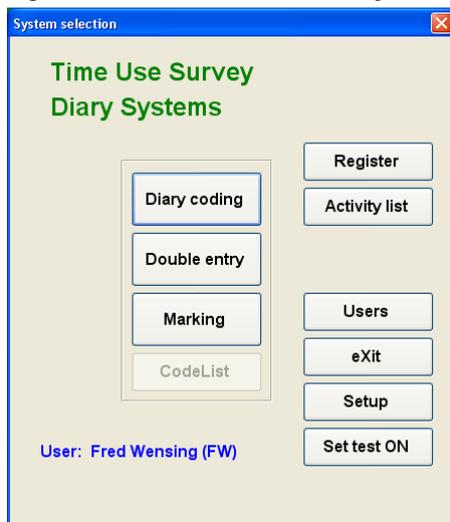


Figure 8B. Main interface as seen by a manager of the system



The list of users also contains a checking rate, for each user, which identifies the percentage of diaries that are to be double entered for quality assurance. These rates may be adjusted by the Data Processing Manager from time-to-time based on quality outcomes.

The main processes and sub-systems that are accessible through the main interface are:

- Installation of the local activity code list (whenever it needs to be updated)
- Diary coding
- Double entry
- Marking of dual coded entries
- User maintenance
- Activity code list maintenance
- Diary register maintenance
- Configuration of the system (set-up).

8. An interface to assist data capture functions

The data capture interface starts with a list of households that have been assigned to the operator for coding. To make it easier for an operator to locate a particular household in the system, the list can be filtered on the basis of diary status values and/or month of collection. Search on household number is also provided. It is also possible to enter a household which is not on the list, provided that it is in the system and not being used by another operator.

Once a household has been selected for data capture the operator is presented with a screen that contains the known information about the household (see Figure 9) and the status of diary data entry to-date. The interface also contains buttons that give access to other information and reports.

Figure 9. Data capture management interface

The screenshot shows a software interface titled "Details of selected household". At the top, it displays "Household ID: R289687006H001 (3 persons)" with a "View" button. Below this, there are two sections for "Selected person 1" and "Selected person 2". Person 1 is "Fred Jones, Male aged 28" with a "View" button. Person 2 is "Ted Jones, Male aged 18" with a "View" button. Underneath, there are two diary entries: "Diary 1" with status "Partially coded (<48 hours)", "BL status: Dirty", and "Episodes: 0", featuring "ED report" and "Edit" buttons; and "Diary 2" with status "Not started (but expected)" and a "Create" button. At the bottom of the window are "eXit", "Report", and "View" buttons.

You will notice from Figure 9 that the details of up to two selected persons and the status of diaries are shown. This is consistent with the survey design which required up to two persons to participate. The household diaries are always to be captured as a set because there may be interactions between the respondents that can affect the resultant coding.

To assist the coding staff to understand the context of activities that respondents have reported, access is provided to the full supplementary Blaise questionnaire and Household questionnaire information through the View buttons. Data from those sources cannot be changed by the operator. A Report button is also provided to produce a summary of the household situation (see Figure 10 for sample report) which can be printed if necessary.

9. An interface to assist with marking

For diaries that have been double coded the system compares the matching diary entries and identifies all the discrepancies. These records are then available to the quality manager to carry out a marking process where he/she checks each discrepancy and ticks off the correct or best entries.

The interface for marking is similar to that for data entry but there is more functionality (see Figure 13).

Figure 13. Marking management interface

Diary marking for this household

Household
ID: R100010002 (4 persons) View

Selected person 1
Name : Andrea Green
Male aged 74
Person 1 View

Selected person 2
Name : Bob White
Male aged 62
Person 2 View

Diary 1
Status: Partially coded (<48 hours)
Time diffs = 1
Code diffs = 1
Episode diffs = 1
Mixed: Edit Mark
Hybrid: Create Edit

Diary 2
Status: Not started (but expected)
Time diffs =
Code diffs =
Episode diffs =
Mixed: Edit Mark
Hybrid: Create Edit

eXit Report View

From this interface the operator can examine the matched episode records (using the Mark button). This brings up a screen which shows the discrepancies, one by one, and provides tick boxes to identify the correctly coded entry (see Figure 14).

Figure 14. Marking screen for one mismatched episode

Episode for scoring

Episode details
#2 8:00AM Day 1 8:10AM Day 1 8:30AM Day 1

Activity 1: Coder (A) Check washing Code: 3121 Dual coder (B) Read personal letter Code: 4231 A B N

Who for: 10 10

Activity 2:
Code:
Who for:

Activity 3:
Code:
Who for:

Activity 4:
Code:
Who for:

Previous eXit Next

For the entry shown in Figure 14, there is a discrepancy in the end-time (shown in red) and a discrepancy in activity 1. The operator can tick the corresponding boxes on the right-hand side to identify whether coder A, coder B or Neither are correct.

Once the marking has been completed it may be possible for the operator to prepare a hybrid record that combines the correct responses from the matched episodes. This will enable the most correct data to be used, particularly while the coding is in its early days.

10. Technical challenges

There were a number of technical challenges that needed to be handled in such a complex system. Some of these are discussed here.

10.1 Local deployment

The most efficient way to operate Blaise data capture, when there is a complex instrument and a fair amount of external look-up steps, is to place the Blaise data capture file and the external look-up files local to the operator.

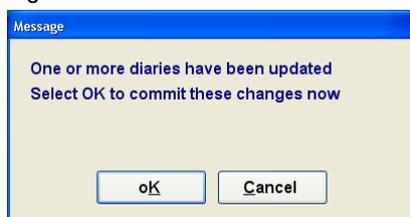
The two issues that needed to be handled as a result are:

- Ensuring that the activity code list is always up-to-date
- Transferring the diary data to and from the central data store

The consistency of the activity code list is managed by recording the version number of the current activity code list in the set-up settings for the system. The version number of the activity code list in the local environment is then checked whenever the system is started. If the version is not up-to-date then the main data entry access buttons are de-activated until the latest version is installed.

The Diary data capture system has a central data store for each process (diary coding, double entry and marking) which is placed on the network. Once a household has been selected then the system copies any existing data to the local disk drive before enabling the operator to change it. At the end of involvement with that household the system prompts the operator to commit the changes or go back and revert to the original version.

Figure 15. Exit screen when data has been changed



10.2 Filtered lists

The system makes use of a single central register of the households involved in the survey. That register is located on the network and is accessible to all managers and operators.

Once the survey gets underway, the list of households is expected to become large so a filter technique has been added to enable the system to target households that have been assigned to a particular user, have reached a particular status or from a particular month.

The filter works by applying a series of conditions to the list of entries in the main register. Those entries which match the filter conditions are then copied to a temporary file before displaying them on the screen.

Using a filter frees up the register for multiple shared access. The register data only gets updated with status information when the operator closes a diary coding, double entry or marking process.

Figure 16 shows the filter definition screen used for diary coding or double entry. A more comprehensive filter is available for register maintenance functions.

Figure 16. Filter definition screen for data capture

10.3 Tracking time over 48 hours

The Time Use diaries record episodes and activities over a 48 hour period commencing at 4AM on the first day and finishing at 4AM on the third day.

To make it easier to manage these time events, and check the sequencing of episodes, the diary instrument has been set up to convert each time point (consisting of a clock time and day number) into an integer with values from 0 to 2880. These codes then represent the 2880 minutes in a 48 hour period. Each start and end time is duly converted.

Using this simple conversion makes it much easier to check times against each other (to ensure episodes are consecutive) and to calculate time differences or sums.

10.4 Providing for alternative data entry methods

As mentioned in section 4, the diary data capture instrument provides support for two methods of data entry, the Form approach or the Table approach.

In order to synchronise the two methods the instrument uses two arrays, one for each method, containing all the episode fields. The first array is used in the ordinary flow of the questionnaire and the second is used in a table. Both arrays are initiated using the KEEP instruction then a single field called EntryMethod (with two possible values) is used to control which one is the active method and which becomes the passive method.

Depending on which method is active, the Rules in the diary instrument copies all the values of the corresponding active array elements to the other (passive) array.

It should be noted that duplicating the episode data in this way does add some load to the execution of the instrument. On the other hand, having the table method available makes it easier to handle episode level actions of insert, delete, split and merge.

10.5 Table operations

The special block commands of INSERT and DELETE, along with other assignments, are used to insert or delete rows in the episode table and move the remaining data around to achieve the desired outcome (being either insert, delete, split or merge).

These special operations are controlled by setting a Flag to allow the corresponding rules to be executed once. The Flag is reset to the inactive value at the end of that section of the rules to stop the actions from being repeated.

These special table events are triggered using corresponding buttons placed on the Menu bar. Once pressed the program behind the button activates the Flag value and the action type and runs the Rules.

10.6 Checking the dual coding of activity episodes

When comparing the dual coding of activity episodes it is not a simple matter of comparing the elements of two arrays. That is because one array or the other may contain more or less entries. This can happen where one operator considers that two or more episodes have taken place for a particular time period, and another operator considers that only one episode has taken place for the same time period. There are many other possibilities for time discrepancies to arise.

So, before the detail of two sets of activity episodes can be compared, they need to be processed and re-aligned based on start times and end times.

A procedure was developed in Maniplus that compares the start times and end times of the two arrays of episode information and makes corresponding re-alignments to all the entries when problems are detected. The re-alignment process involves inserting dummy episodes that have the same start and end times (effectively an episode with zero elapsed time) and contain no activities. The episodes effectively push the subsequent mismatched episode up by one. The whole process is then repeated across all the elements of the array until all entries match on start time or end time or both.

11. Conclusion

Development of this complex Diary Data capture system has been a challenge but Blaise and Maniplus has proven to be versatile and flexible. The posters will reveal more of the system design and illustrate some of the source code.