

BLAISE APPLICATIONS IN STATISTICS NEW ZEALAND

Brett Martin

Statistics New Zealand

1. Introduction

Statistics New Zealand have recently developed a strategy to generalise and automate their survey processing activities. The Blaise system is a key component of this strategy. These new directions have resulted from a recent major internal reorganisation. In this paper I outline the changes to our organisational structures and discuss the new survey processing strategy. Some of our more interesting Blaise developments in the past year are also described.

2. The new organisation

Prior to the reorganisation there were over twenty separate survey sections. Each section was responsible for most of the development, operation and output activities relating to one or more surveys. A few services such as Sample Design, System Development and Publications were centrally managed and allocated as required to the survey sections.

As far as the individual sections were concerned this 'cottage industry' approach was reasonably effective. However the inefficiencies became obvious when viewed from a corporate perspective. Many separate systems performing similar functions were being built. Incompatible definitions and classifications were created as each section focused on its own requirements. The resulting survey data sets were difficult to compare or combine. The reorganisation took place in September 1992. Three groups were created, each with its own specific functional responsibilities that extend across all surveys.

Blaise applications in Statistics New Zealand

The 'Survey Management Group' responsibilities include development of classifications and frames, survey design and development, application system development and survey operations. They undertake respondent liaison and the collection/capture, coding and editing of all survey data.

The 'National and Regional Statistics Group' comprises twelve specialist subject matter teams who analyse and publish all 'Official Statistics' and associated explanatory, analytical or technical comment relating to their area.

The 'Information Services Group' is responsible for product development, marketing, client services and managing the Computing infrastructure.

Restructuring has brought into sharp focus the need to co-ordinate and standardise practices within and across groups. In the case of the Survey Management Group efficiencies are expected to help fund the expansion of the analysis area.

The changing external social and economic environment in New Zealand has increased demand for Statistical information. Traditional statistical outputs are being reviewed in order to better meet the needs of the Government and the Public. New technologies such as CAI and imaging are being employed to meet these challenges in a timely and cost effective manner. Our computing environment is also changing from a centralised mainframe environment to local and wide area networks of PCs and Servers. New processing systems are required for this environment.

The Blaise system was recognised to address these requirements and was adopted as a component of an overall strategy for Survey processing. This strategy is known as Generalised Automated Survey Processing (GASP).

3. The GASP strategy

The main objective of the GASP strategy is to standardise and integrate all survey processing functions and data: Activities that are common to a wide range of surveys will be automated through the use of generalised systems.

The survey and management data generated by such systems can be more simply integrated with other data and applied in a wider range of analysis and outputs because of the use of common definitions and data structures.

Initial work has involved applying Information Engineering methods to model the processes and the data required to conduct surveys. The resulting process model and logical data model provide a concise summary of the concepts and standard rules that underpin the survey processes from the initial design to the 'clean' unit record database. These models will be used to coordinate the acquisition or development of various software tools. It is envisioned that the processing system for any given survey will, in the future, be constructed using these tools.

4. The scope of GASP

There are five major subject areas covered in the GASP strategy:

- Classifications and Frames
- Respondent Liaison
- Data Capture, Coding and Editing
- Output Editing, Imputation and Estimation
- Survey Management

The Classifications and Frames area includes development of standard classifications and question definitions along with the operation of Survey frames such as the Business Directory, and Geographic databases containing Address, Street and Areas information. These components are probably the most important from the point of view of providing consistency and standardisation to the survey outputs. They are the foundation for all subsequent survey activities.

Respondent Liaison is an important function that is usually not specifically recognised as part of the survey process. Systems to minimise the number of surveys in which each respondent is asked to participate are planned. All communications with respondents including questionnaire postout and

reminders will be managed using co-ordinated systems and standard methods. Interviewer and Enumerator management systems are also included in this area.

The Blaise system covers the majority of our requirements in the Data collection/capture, coding and editing area. Its ease of use and modest startup requirements provide strategic advantages. There was some concern that Blaise would not seamlessly integrate with other GASP systems where relational databases and Windows software are expected to be used. The plans for Blaise version 3 provided some reassurance that these issues would be addressed in the future. Blaise systems mostly provide input to the database, so the main problem is one of interfacing and translating Blaise data into a database format. Fortunately the Setup generator and Manipula tools provide straightforward ways to do this. A direct relational database interface with Blaise, while desirable, was not deemed to be critical to its successful use. Our Executive Management Committee decided in November 1992 that Blaise would be used where feasible for all new or redeveloped data capture coding and editing systems.

Output editing, imputation and estimation processes are applied to the data once it is in a database format. This is to ensure that the processes can be repeated and adjusted as necessary. The Graphical Editing system described by Gary Dunnet at last year's IBUG (1) will be adapted to fit in with the GASP strategy. Bascula is currently being evaluated using data from our Household Labour Force survey. As Post Stratification weighting is currently used in this survey it was seen as a good starting point. The Linear weighting options will also be investigated. The results so far have been very encouraging.

Survey Management is a diverse function that covers all aspects of survey processing from initial design to the production of a 'clean' unit record database. By relating all the Survey functions to a common logical data model, management information can be generated during the survey processes and integrated with the survey data. The use of relational database technology is a key component of the GASP strategy. It will ensure that both

survey data and relevant management information can be easily accessed and used to manage our survey processing activities.

5. Implementation of GASP

The components of GASP that complement Blaise have not been implemented yet. Where suitable software can be found, it will be utilised. Any remaining components will be built over the next few years depending on our requirements and available resources. The restructuring and the development of an overall strategy has helped change attitudes within the organisation. The focus has changed from seeing the exceptions and differences to finding standard solutions. There are still many issues and practical problems to be resolved. However, much progress has been achieved and we are confident of finding appropriate solutions. The Blaise system has already addressed some of our immediate needs. Over the past year it has hastened the implementation or redevelopment of a number of surveys.

6. Recent Blaise developments

Statistics New Zealand has used Blaise for a number of years. Until the beginning of 1993 it was employed to capture and edit contract surveys or new surveys especially where the development time was short or there was a scarcity of development resources. A census of our surveys (collected with Blaise in March 1993) showed that 6.25% of our Surveys used Blaise which represented 9% of all questionnaires processed (excluding the Population Census). This year several major Blaise developments will substantially boost these numbers.

7. Migration Survey development

The processing of 4 million arrival and departure cards annually, was until recently, a significant task that involved many clerical operations to sample the cards, code, check, capture and edit responses.

A Blaise CAPI program with the ability to systematically sample cards for detailed capture is now in production. For each migration card the class of the traveller (e.g. 'Short Term Visitor'), is entered. This is passed to a Pascal procedure which inspects a file containing the sampling fraction and the number of cards processed within each class. The procedure determines if the card is to be sampled and indicates this to the Blaise program via a return code. If the card is sampled further questions are asked in order to capture and code information from it. The sampling procedure is quite complex as there are different sampling fractions for each of the three classes. Also a further sample of the sampled cards is taken to capture Names and Address for a follow-up survey of Returning New Zealanders.

The Pascal sampling procedure is called at two places to allow for the possibility of revising the class during capture. Firstly it is used to determine if the card is to be sampled. Then, after all entry, coding and editing is completed, it is called again to update the counts for the class. The same procedure is used in each case but with different parameters.

The Migration system has been phased in. Initially the clerically sampled cards were captured and coded, then online sampling was added. The same Blaise program was used but all the sampling fractions were set to 1 during the first stage. The main problem encountered has been the speed of execution. A 286 PC was too slow due to the amount of I/O required for the sampling procedure and the number of Coding fields involved. Throughput rates and accuracy have improved with the Blaise system.

8. Household Survey developments

The redevelopment of our Household Surveys is a strategic priority for the organisation. The planning for this started in 1992 and despite changes caused by the reorganisation it is progressing well. The first stage is to convert the current Household Expenditure and Income Survey (HEIS) questionnaires to a Blaise CAPI system. In addition to the current questions several new topics will be included. This stage is currently planned to be implemented in April 1994. Our Household Labour Force survey is scheduled to be implemented as a CAPI/CATI system in April 1995.

9. Interview management

One of the problems for interviewers administering a large questionnaire is that circumstances can change during the course of an interview. Respondents may leave or join in and sections of the interview may need to be postponed. To handle these changes the questionnaire routing needs to be dynamically configured. The difficulty achieving this was mentioned at last year's IBUG meeting (2).

A routing process has been developed for our Household surveys that allows the interviewer to select who is included in the interview at any point and also which sections of the questionnaire they wish to undertake. This mechanism includes Jump points to allow the interviewer to return at any stage to a screen showing the status of each part of the interview. From there the parameter questions which reconfigure the route can be set.

A simplified example program shows how this is achieved:

```
QUESTIONNAIRE EXAMPLE; {Interview Management example}
VAR
  I : INTEGER;
QUEST
  Seq_Num "Sequence Number of interview?" : 1..97 (KEY);
  HowMany "How many people usually live here?" : 1..3 (RESPONSE);
  Name    "Enter the Names of people who live here?"
          : ARRAY [1..3] OF STRING[20] (EMPTY);
  Part_1  "Status of Part 1" : STRING[8] (SCREEN, EMPTY, PROTECT);
  Part_2  "Status of Person in Part 2 "
          : ARRAY [1..3] OF STRING[20] (SCREEN, EMPTY, PROTECT);
  Part    "Do you want to run:-"
          : (All "all parts of the questionnaire?",
             Some "selected parts of the questionnaire?") (RESPONSE);
  Mode    "/1///Do you want to interview:-"
          : (Person "an individual?",
             Group  "a selected group?",
             Everyone "everyone together?") (RESPONSE);
  Who     "Select $Mode to interview."
          : Set of (Person_1 "$Name[1]",
                  Person_2 "$Name[2]",
                  Person_3 "$Name[3]") (RESPONSE);
```

Blaise applications in Statistics New Zealand

```
WhatPart "Select the part(s) you want to run"
      : Set of (Part_1 "Block_1",
              Part_2 "Table_1") (RESPONSE);
QuestEnd "To restart interview press ^F1 then 1 and press <ENTER>
//To finish type 9 and press <ENTER>" : 9..9 (RESPONSE, SCREEN);

BLOCK Nonresponse;
QUEST
  Reason "Reason why interview stopped" : STRING[20] (EMPTY);
  EndPoint "To stop interview type 9 and press <ENTER>"
      : 9..9 (RESPONSE, SCREEN);
ROUTE
  Reason; EndPoint;
CHECK
IF EndPoint = RESPONSE THEN
  COMPUTE Part := 1; COMPUTE Mode := 3;
ENDIF;
ENDBLOCK {Nonresponse};

BLOCK Block_1 (SUBFILE) {Block of questions any person may answer};
QUEST
  Q1 "Block_1 question one?" : (Yes);
  Q2 "Block_1 question two?" : (Yes);
  Done "This part complete?" : (Yes, No) (RESPONSE);
ROUTE
  NEWPAGE; Q1; Q2; Done;
ENDBLOCK {Block_1};

TABLE Table_1 {Table of questions for selected persons}
VAR
  J : INTEGER;
BLOCK D_Block (SUBFILE);
QUEST
  Q1 "Table_1 question one?" : (Yes);
  Q2 "Table_1 question two?" : (Yes);
  Done "This person complete?" : (Yes, No) (RESPONSE);
ROUTE
  Q1; Q2; Done;
ENDBLOCK {D_Block};
QUEST
  Person "" : array [1..3] of D_Block;
ROUTE
  FOR J := 1 TO 3 DO
```



```
        IF (J <= HowMany) AND ((Mode = Everyone) OR (J IN Who)) THEN
            Person[J];
        ENDIF;
    ENDDO;
ENDTABLE {Table_1};

QUEST {questionnaire}
  QPart_1 : Block_1;
  QPart_2 : Table_1;
ROUTE {questionnaire}
  Seq_Num; HowMany;
  FOR I := 1 TO 3 DO
    IF I <= HowMany THEN Name[I] ENDIF;
  ENDDO;
  Part_1;
  FOR I := 1 TO 3 DO
    IF I <= HowMany THEN Part_2[I] ENDIF;
  ENDDO;
  DUMMY[2]; Mode;
  IF Mode <> Everyone THEN Who ENDIF;
  Part;
  IF Part = Some THEN WhatPart ENDIF;
  IF (Part = All) OR (1 IN WhatPart) THEN QPart_1 ENDIF;
  IF (Part = All) OR (2 IN WhatPart) THEN QPart_2 ENDIF;
  NEWPAGE; QuestEnd;

CHECK {questionnaire}
  IF QPart_1.Done = Yes THEN COMPUTE Part_1 := 'Complete'ENDIF;
  IF QPart_1.Done = No THEN COMPUTE Part_1 := 'Not Done' ENDIF;
  FOR I := 1 TO 3 DO
    IF I <= HowMany THEN
      IF QPart_2.Person[I].Done = Yes THEN
        COMPUTE Part_2[I] := Name[I] + ' complete';
      ENDIF;
      IF QPart_2.Person[I].Done = No THEN
        COMPUTE Part_2[I] := Name[I] + ' not done';
      ENDIF;
    ENDIF;
  ENDDO;
  IF QuestEnd = RESPONSE THEN
    COMPUTE Part := 1; COMPUTE Mode := 3;
  ENDIF;
ENDQUEST;
```

Note that at the end of the questionnaire the interviewer has the choice of returning to the jump point or of ending the questionnaire. At the conclusion of the interview all the information collected is automatically saved when the parameter questions are computed to include all the answered questions. This results in a program that can be adapted by the interviewer to suit the interview situation.

10. A questionnaire management system

Due to the size of our HEIS questionnaires (3) the questionnaires have been split into separate Income and Expenditure programs. This split has made things more complicated as far as managing appointments and editing between questionnaires is concerned. Initially appointments were to be pre-loaded into the D02 file using the conversion routines. Unfortunately the conversion routines would not function with a program that is near the size of the Blaise system limits.

Now that the questionnaires have been split into smaller programs a simple Blaise program is used to manage the appointment list and start the execution of the Expenditure or Income questionnaire. Firstly the conversion routines load appointment data for this program. The interviewer selects from these appointments the household to be interviewed. The status of the Expenditure and Income questionnaires for that household are obtained using READBLAISE and displayed. The interviewer then answers a question about which questionnaire they wish to start. A DOS Batch file is written by a Pascal procedure and activated when the interviewer exits the management program. This Batch file starts the Expenditure or Income questionnaire with the key for the household and the applicable interview mode (Interview for the first time or Examine mode if some data has already been collected). At the conclusion of the interview the management program is started again to display the appointment list. This method can be applied to manage any number of related questionnaires.

11. Quarterly Business Survey development

In our quarterly business surveys respondent details are updated in conjunction with capturing the survey data. Also the survey data for up to five previous periods is accessed for editing and possible updating. These functions are implemented in the same Blaise questionnaire.

The key comprises two questions: the respondent number and the period for which the data is collected. If the period is EMPTY then questions about the respondent details are asked. The data from these questions is stored in a subfile. If a period is entered then the survey questions are accessed. The survey data for all periods is held in the D00 file. Having all the records in the same dataset allows the respondent information to be displayed on the capture screens and updated (via the Examine Form option) if required. Survey data from previous periods is also accessed and used in edit checks. READBLAISE is used to obtain the respondent details from the subfile and the previous periods data from the D00 file.

Manipula and the conversion routines load the respondent details from our Business Directory and extract respondent data for postouts and follow-up reminders. As only five periods of data are needed for editing or correction, the old survey data is archived to maintain the dataset at a reasonable size.

12. Setups

The Setup Generator has been invaluable in bridging between Blaise and other processing tools. It has proved to be one of the hidden strengths of Blaise.

One of our requirements, mentioned earlier, is to up-load Blaise datasets to relational databases. There are already some Setups such as for ORACLE which could be adapted for our databases. However one problem is that

Blaise Datetype fields are converted to 'ddmmyyyy' format. This format precludes direct loading of this data to most database datetype fields as this type of field expects a delimited date e.g. 'dd/mm/yyyy'.

To solve this problem a Setup was written to create Manipula programs, one for each subfile. These read the Blaise data and output ASCII data with dates in the 'dd/mm/yyyy' format. A description of this revised ASCII data is also output for creating the database tables along with programs to upload the data.

A Setup for the Graphical Editing system, described by Gary Dunnet at last year's IBUG, is planned. The future enhancement of the Graphical Editing system is being looked at and this will be influenced by the requirements of GASP. In theory there would be little difficulty in creating a Setup for it.

Statistics New Zealand in partnership with an Australian software development company, SpaceTime Research, have developed a Windows based tabulation package called Supercross. This system uses its own propriety data formats. A Blaise Setup is available to provide the majority of information required to generate the Supercross databases. The first version of this is currently being tested.

13. Survey development environment

A set of standards has been developed for those staff working with Blaise. Our Survey Development area has the major responsibility for writing Blaise applications. However the Application Development area is usually involved in writing Pascal procedures and interfacing Blaise datasets with other systems.

After our Local Area Networks and Wide Area Network are fully operational a library of standard Blaise code for common questions and answer types will be set up.

14. Conclusion

The Blaise system is being successfully used in our organisation and is a key component of an overall strategy aimed at integrating and standardising survey processing activities. Changing external and internal environments have given rise to demands for improved quality, timeliness and cost savings in the production of statistics. There is considerable strategic benefit with any technology that directly addresses the needs of a statistical organisation. Blaise is remarkable in that it provides this benefit in a very cost effective and robust way. New software and hardware products will continue to challenge and redefine the means by which statistical offices conduct their business. The continued development of the Blaise system is vital, if these challenges are to be met.

References

- (1) 'A Possible Future Development for the System'. *Essays on Blaise - First International Blaise Users Meeting* (1992).
- (2) 'The Development of Family Resources Survey'. *Essays on Blaise - First International Blaise Users Meeting* (1992).
- (3) 'Strategies for Developing Large Questionnaires'. *IBUG Newsletter 2* (Sept. 1992).