

What users want from a tool for analysing and documenting electronic questionnaires: the user requirements for the TADEQ project

Maureen Kelly

Abstract

The TADEQ project is funded under the European Commission's Esprit programme to develop a tool for documenting and analysing electronic questionnaires. It is led by Statistics Netherlands and the other partners are the Office for National Statistics, UK; Statistics Finland; Instituto Nacional de Estatística, Portugal; Max Planck Institute, Saarbrücken, Germany.

National Statistical Institutes (NSIs), research institutes, and commercial marketing research organisations are making an increased use of computer-assisted interviewing (CAI) systems for collecting survey data. The growing possibilities of computer hardware and software have made it possible to develop very large and complex electronic questionnaires. As a consequence, it has become more and more difficult for developers, interviewers, supervisors, and managers to keep control of the content and structure of CAI instruments. The TADEQ project proposes to develop a flexible tool for documenting and analysing electronic questionnaires. As a documentation tool, it must be able to produce a human-readable presentation of the electronic questionnaire.

The paper will describe the first stage of the project: the user consultation and translation of the results into a user requirement to serve as a basis for the remainder of the project.

Keywords: TADEQ; user requirements; documentation; flexibility; electronic questionnaires

Introduction

Questionnaire documentation is a core tool within the survey process. Throughout the life of a survey there are many different users of the documentation who use it to perform a number of different tasks. Computer assisted interviewing (CAI) systems are increasingly replacing paper questionnaires to collect survey data. The documentation of the CAI instruments used to collect data, electronic questionnaires¹, has been problematic and time-consuming. The increasing capabilities of computer hardware and software have made it possible to develop very large and complex questionnaires. It has become increasingly difficult to comprehend these complex questionnaires in their entirety. The TADEQ project was set up to develop a tool for the automatic documentation and analysis of electronic questionnaires.

1. Background

Computer Assisted Interviewing

National Statistics Institutes (NSIs), research institutes and commercial market research organisations are increasingly using CAI instruments used instead of paper questionnaires to collect survey data. CAI has made many quality improvements to survey data that have been well-documented (see for example W L Nicholls II et al. or M Couper and W L Nicholls II).

However, one area where CAI instruments have not brought improvements is questionnaire documentation. The paper questionnaire, as well as being the interviewing tool, also served as its own documentation; it contained the information required by documentation users – the questions and the routing instructions. The edit checks were carried out separately in the office and these were usually available in a separate document.

With the introduction of CAI, questionnaires became more complex. At the same time, it became more difficult to provide documentation to understand them. The documentation of CAI instruments became a separate task, one that was not necessary when paper questionnaires were used. At first, this task was not recognised. In the early days of CAI, when electronic questionnaires were developed for computers of limited size and speed the computer programs were seen by the pioneers as 'self-documenting': the programs contained all the information about the questions asked and the routing. However, to a non-programmer, using the actual program could be confusing. Users were nevertheless forced to accept it, as it was the only documentation available. This situation was always unsatisfactory and soon became wholly untenable: as CAI instruments grew

¹ Throughout the paper CAI instruments and electronic questionnaires are used to refer to computer programs which are used for data collection. CAI programs refer to the source code used to produce the electronic questionnaires.

more complex users were unwilling and unable to use the CAI programs. Therefore users required other forms of documentation of the electronic questionnaires.

Current solutions to the documentation problem

At present there are a variety of ways in which users have attempted to solve their documentation problems: producing separate questionnaire specifications independent of the CAI program; manual editing of the program; and semi-automated documentation of the electronic questionnaire.

Some surveys produce a questionnaire specification that is wholly independent of the CAI program. This type of documentation can be useful for testing the instrument. However, these documents are produced manually, which can be time consuming and error-prone. As the questionnaire is produced independently, there is no guarantee either that it matches the specification exactly or that the specification accurately documents it at the end of the process.

On other surveys, there is an attempt to document the CAI instrument once it has been written, in part or in full. Manual editing of the CAI program has formed the basis of much recent documentation of this kind. However it is a large and laborious clerical task to get the documentation into a presentable format. Again, as it involves manual editing it can be error-prone. Moreover, it is difficult in practice to ensure that any amendments to the CAI program are carried over into the documentation, particularly if these amendments are made under extreme time pressures.

Consistent and accurate documentation can only be obtained if it is generated automatically from the electronic questionnaire itself. In 1996 the Social Survey Division of the Office for National Statistics (ONS) developed a tool, in co-operation with Statistics Netherlands, which could automatically produce documentation from programs in Blaise. This documentation tool met many of the user needs for ONS's documentation: it has produced the documentation for such large and complex surveys carried out by ONS as the Family Expenditure Survey, the General Household Survey, and the Family Resources Survey. It has greatly improved the speed with which the documentation is produced: for example, it is fast enough to be used by the ONS Omnibus Survey, which has rapid turnaround from module request to fieldwork and results. However, the output still requires some manual editing so it can still be subject to errors and the updating issues that other documents have - albeit to a much reduced extent.

Many continuous surveys have a different kind of problem when updating their documentation, in which the majority of the questions asked are the same as in the previous version. It is sometimes quicker to manually update the previous version of the document with the recent changes, rather than to start the documentation from scratch again, and repeat all the editing done on the previous version. Over time, there is a risk that documentation and CAI instrument may diverge significantly, as in the more general case mentioned above where documentation and instrument are developed separately.

The TADEQ project

There is a need to produce documentation for CAI instruments that is comprehensive and accurate. It needs to be produced quickly and requires no or little manual editing. This can only be obtained if it is generated automatically from the CAI program, as that contains virtually all the information required for the documentation. What is needed is a software tool capable of automatically translating the CAI program into a human readable form.

The TADEQ project is funded under the European Commission's Esprit programme to develop a tool for documenting and analysing electronic questionnaires to meet this need. It is led by Statistics Netherlands and the other partners are the Office for National Statistics, UK; Statistics Finland; Instituto Nacional de Estatística, Portugal; Max Planck Institute, Saarbrücken, Germany.

The objectives of the project are to produce a neutral tool, which can be used by different CAI systems. The existing ONS tool is limited to documenting questionnaires developed in Blaise, and it does not attempt to meet all the user needs which have been identified by the user consultation and which TADEQ will incorporate.

The TADEQ tool will be able to produce both paper and electronic documentation. The tool should be able to analyse the structure of the questionnaire and report statistics and possible problems with the structure.

There are several stages to the project:

- A survey among the current users of documentation from electronic questionnaires and development of a user requirement for the TADEQ project which will form the basis for the remainder of the project.
- The development of a neutral file format for storing questionnaire contents and questionnaire logic in an efficient machine-readable form.
- Development of an interface to CAI systems. The interface must allow CAI systems to deposit metadata in a way that can be read by the neutral file format.

- Development of a software module that is able to read the neutral file format, produce paper and electronic documentation, and carry out an analysis of the structure of the questionnaire.
- Testing of the developed tools by users of CAI systems.

This paper discusses the first of these stages: the user requirements for documentation of electronic questionnaires.

2. The users of documentation from electronic questionnaires

There are many different users of the documentation from electronic questionnaires, spanning the whole of the survey process. In the first quarter of 1999, the project team consulted a purposive sample of more than one hundred users in twenty-four organisations throughout Europe. The respondents completed a paper questionnaire or its Web equivalent. The survey's aim was to provide ideas about the different types of user and their different documentation requirements. As well as the survey, a number of discussions took place to get a more in-depth view of what users required.

Conducting a survey involves a large number of people, who perform different tasks and they use their questionnaire documentation in a variety of ways to help carry out these tasks. Different users therefore have different requirements for their documentation. TADEQ has to produce documentation from CAI instruments that has the flexibility to meet the needs of all these users.

The responses received from users in the user consultations covered the whole spectrum of survey work: from commissioning the work at the beginning to archiving the data at the end. Fifteen main roles were identified within the survey.

Users:

- commissioned survey research;
- were survey customers;
- were survey managers;
- designed surveys and survey questions;
- were subject matter specialists;
- developed electronic questionnaires;
- developed and use other CAI applications;
- developed CAI software;
- interviewed respondents;
- trained and managed interviewers;
- processed or edited the data;
- analysed data from their own survey;
- analysed and used data from other surveys;
- were survey methodologists; or
- archived and disseminated survey data and results.

Most users carried out more than one role within a survey. For example, many survey customers also commissioned surveys and acted as managers for part of the survey process. Sometimes customers were also subject matter specialists or data analysts. The division of responsibilities for carrying out these tasks can vary from organisation to organisation, survey to survey. For example, on one survey the people from the organisation responsible for the data collection may also analyse the results in the first instance. On another survey, the people commissioning the survey may do the analysis themselves, or a third party, not involved in the any other part of the survey, may be commissioned to do the analysis.

Users perform a number of tasks and can use the documentation in a variety of ways. They therefore have different requirements for their documentation. Users also have different levels of knowledge about the survey and its contents. Some users involved in the day-to-day running of the survey have an intricate working knowledge of the questionnaire contents and need documentation to check on minute detail. Others, such as managers and customers, may not have the information they require first hand. They may rely on the documentation rather than their own knowledge to provide this information. Policy makers and academics analysing the data once it has been made public are not involved in the main survey process, but still require information, albeit at different levels. The questionnaire documentation is their only source of information about the survey.

Below are listed some of the main tasks for which questionnaire documentation can be used, and the types of users that perform the different tasks.

Designing surveys and survey questions

Survey designers often wish to use questions from previous surveys, to build on good practice or to facilitate the comparison of results.

Approval of questionnaire design

The people who are responsible for the design of the CAI instrument and the data collection are not usually the people who commission or design the survey. The commissioner may wish to approve the CAI instrument to ensure that the instrument is collecting the information they require. A paper questionnaire used to provide a ready form of documentation. However, with CAI this approach is no longer feasible – commissioners can rarely understand the CAI program and, even if they can, it does not usually provide the information structured in a way that they can use quickly and accurately. Tailored documentation is therefore an essential tool in gaining approval for the content of a survey.

In some organisations the developers of the CAI instruments are specialist computer programmers. Specialist survey researchers design the questions and also need to approve the design of the electronic questionnaire. Like survey commissioners, the survey researchers need to approve the design of the CAI instruments developed by the programmers, so they have the same needs for documentation.

Providing information on the content of the questionnaire

Different users require different levels of detail from questionnaire documentation. Some users, particularly data analysts, methodologists and survey designers, require the exact wording, routing and answer categories for each question. Other users such as policy makers and survey managers may require more of an overview of the questionnaire and less detailed information about individual questions.

Interviewers also like some sort of documentation of the electronic questionnaires, even though they have access to the questionnaires. Although CAI instruments have dynamic routing (that is, the next question that appears is dependent on previous answers and is automatically controlled by the program), it is sometimes difficult to get a sense of the overall structure and flow of the questionnaire as a whole. Alternative documentation can provide this. Interviewers find it useful as they can use it to gauge the length of an interview and which questions may appear in certain circumstances.

Public documentation

Many survey organisations and customers, particularly those in the public sector, have an obligation to make their survey data publicly available. Documentation has to be produced which can be archived or published (in survey reports). People who have had no connection with the process of producing the survey data will be the main users of this documentation. It is essential that the documentation is self-explanatory and comprehensive enough to meet their needs. These people may use the documentation for a variety of purposes: for example, to analyse the data from the survey, to help design another survey or to carry out a methodological study.

Users' requirements

Users require (and therefore TADEQ will need to produce) documentation to be:

- accurate – the information displayed in TADEQ should exactly reflect the CAI instrument;
- automatic – the documentation should require no or little manual editing and therefore be quick to produce;
- comprehensive – all the information that the user may require from the electronic questionnaire should be available (even if it is not all displayed at the same time);
- flexible – users need to be able to easily change the information shown by the tool to meet their individual requirements;
- available electronically and, subject to the constraints of the medium, on paper; and
- easy to use for a non-technical user who has no knowledge of CAI programs.

3. Paper and electronic documentation

One of the objectives of TADEQ is to produce both electronic and paper documentation. Almost all the users consulted use paper documentation. Some users had electronic documentation although none stated they used it exclusively. However, people

report that their use of electronic documentation is limited by a need to mimic the paper documentation they currently use (for example, a Word document). Improvements in the quality of electronic documentation available would mean users would be able to utilise the many benefits of using electronic documentation such as: quick navigation using hyper-text links; computerised search and help facilities; and being able to copy information into other computer packages. Electronic documentation could be used interactively and the user could alter the documentation to suit their own needs.

However, even with these advantages, there will still be a need to produce paper documentation. Almost all the users that were consulted stated they would still require a paper version. The reasons users gave were:

- it was what they were used to;
- they could still use it when they did not have access to a computer;
- to show clients, policy makers and to take to meetings;
- they could easily annotate the document;
- they could use it as a reference when using other computer packages (for example, some data analysts prefer a paper document containing the information on variables rather than switching between screens on a computer; and
- they would need it for publication in a paper report.

TADEQ needs to produce both types of documentation.

Good usability of the documentation produced is vital in both modes. A document can contain all the necessary information but if it is not well presented users will not be able to find the information they require. The documentation will have failed in its prime purpose.

There are different considerations for documentation in the two types of media. The paper document is static, whereas the electronic version can be interactive. The best layout of a paper document is likely to be different from a document viewed on a screen. In a paper document all the information needed has to be shown on the page, whereas on a screen, items can be displayed or hidden as necessary.

Although some surveys produce different paper documents for different users, most produce one document that attempts to meet the requirements of the majority of their users. Electronic documentation can be tailored to suit individual requirements. The navigation and searching for information is different in the different media: in an electronic environment, users can use hyper-text links, clicking on icons; paper documents require other sign-posts such as page numbers, contents page, section headings, and an index so that users can find the information they require.

4. Information required from the electronic questionnaire

Users require a wide range of information from the questionnaire. The key to TADEQ will be the flexibility to show information as and when required, and allowing the user to have control, as far as possible, over what the documentation looks like.

The key information that users need is: what questions were asked; what the possible answers were; to whom they were asked; and in what order.

Although most users want the question text, answer categories and a description of the routing, there are various ways this information can be presented. Users may not wish to see all the possible information on all the variables at one time. It may actually be counter-productive to show too much at the same time; essential information could be lost in a forest of detail. Users need to be able to select which variables they require information for and what information they require for these variables.

As well as information for specific variables there is other information that a user will need to know:

- A title to identify the document. Users may also want a version number, the filename, the time and date of production and who (person or organisation) produced the documentation.
- The questionnaires that TADEQ will be used to document may be long, with a huge number of questions. For example, Manners and Deacon report an instrument with some 30,000 questions. Many of the questions were in loops for all household members, but some 3,500 were unique. In addition there were 2,500 consistency checks. In order for users to navigate their way around the documentation it needs to be split into meaningful sections, with signposts and labels. This is particularly important for paper documentation. Some CAI systems allow questionnaires to be built from blocks of questions. If these exist they can be used to split the questionnaire into sections or sub-questionnaires.
- Indexes and variable lists are also essential, in order for users to find the variables they are looking for. Indexing should be possible by description or by keywords as well as by variable name so those users without detailed knowledge of the survey can find the relevant variables.

- Some CAI systems offer the facility to use external data files to read data in or code data contained within the questionnaire. Users may wish to have references to these files, and a description of their function. They may wish to have a list of all the files used within the questionnaire and also state at a particular question which file was used. The CAI instrument may also use external programs that are not part of the CAI system. Again, these programs need to be documented with a description, if available, of what the function of each program is.
- Some users may wish to add external information to the electronic version of TADEQ, that is information not produced from the CAI instrument itself. This further information may be about specific variables, such as frequencies; or a history of changes to the questionnaire; copies of external coding frames or more general information about the survey, such as dates of field work, sample sizes etc. There may be a need to add information to meet data documentation and archiving standards set by customers, National Statistical Institutes or international agreements.

5. Information required for each variable

As previously mentioned the information about the variables contained within the questionnaire is the fundamental part of any documentation. A wide range of information is required for each variable. There are different ways of presenting each piece of information as well as issues surrounding how all the information fits together.

Which variables are shown

It may be the case that a user does not want to display all the variables within an electronic questionnaire. Users need to be able to select the variables they require individually. However, for large and complex questionnaires this could be a time-consuming task. Users need to select different groups of variables according to different criteria:

- Many CAI systems allow instruments to be built from blocks of questions. If these exist, then the user could select/deselect all the questions within a block.
- Not all variables in a questionnaire are questions that are asked to respondents. As well as variables that are *asked*, some variables are computed and *shown* for information and others are computed and *kept* in the questionnaire without being seen. Also some variables can be asked in some circumstances and computed in others. Users may want to see all the variables, only those that were asked or shown or kept, or any combination of these.
- Some CAI instruments use variables that are temporary and do not appear on the final database. These are sometimes used in computations, routing and other functions within the questionnaire, which are not required afterwards. Some users may wish to see all the variables used in the program others may only want to see those that are permanent variables.

Not showing all the variables can cause problems if any of the variables that are not shown are referenced elsewhere in the documentation, such as in the routing specification. To users unfamiliar with the CAI program these references will be meaningless. TADEQ needs to find a way of dealing with such variables. One solution would be to provide a list of variables that are not shown but are mentioned elsewhere in the documentation.

Variable identifiers

These are essential, however different users need to identify the variables in different ways. In Blaise, variables are identified using names that include the blocks they belong to, using dot notation. For example, QTHComp.QHComp.NumAdult. Some users, particularly developers of CAI instruments, require these full names so that they can find the question within the program itself. These may also be needed to identify a variable with the same name in several blocks. Other users may prefer to have a short variable name (for example, NumAdult), as this is how the variable is likely to appear in the final database. However, using these short names could result in more than one variable having the same name. Many users also thought that the variables should be numbered as it would give an easy reference point. Other users felt that a description of the variable was important, particularly if the user was unfamiliar with the data and documentation.

Question text

Documentation of the question text is essential. However, many CAI instruments use text substitution where the text of a question can be altered depending on individual circumstances of the respondent. Although it usually improves the flow of the interview, it makes documentation of the question text difficult. TADEQ substitution can use either the values of another variable directly to a temporary string variable into which the appropriate text is computed depending on the answers to previous questions. In extreme cases, a question may be no more than a sequence of text substitutions.

The documentation tool needs to identify the parts of the text where substitution is used. In many instances the user will not know what the variables refer to (and the names may not be intuitive) and therefore some indication what the text substitution is necessary to understand the question. There are several possibilities: a label for the text substitution for the variable; a link, reference, or box showing where all the text alternatives (and the conditions under which they apply); or a conventional simplification, such as a pronoun is always shown as 'you', 'yours' etc. The main requirement from users is for a description or a conventional simplification. However, in many cases where the text varies substantially, all the alternatives will be required.

Answer types

The type of answer required from the questions is essential information for the majority of users. There are different types of answers to questions: numeric (integer and real) values; strings, where a text (usually of specified length) is entered; enumerated types, where there are specified response categories; multiple response enumerated types, where more than one answer can be recorded; and other special types, for example for recording times, dates or verbatim answers.

Enumerated types have similar presentational problems as variable identifiers, since the answer categories can be referenced in a number of different ways (by number, a short name that is used in the CAI program, the full answer, or a descriptive label). Some CAI systems allow text substitution to be used in the response categories and therefore the answer categories can have the same problems with text substitution as occur in the question text.

Some questions can be left blank, or special keys can be used to enter 'don't know' or a 'refusal' at a specific question. Many users need to know when these keys can or cannot be used or when the questions can be left empty.

Routing

Documentation of the routing was essential for the majority of users. However the routing for many of the surveys for which TADEQ will be designed is complex and difficult to show in a comprehensible format. One problem with routing (as has already been mentioned and can make routing incomprehensible to users) is that the routing can contain variables that are not displayed anywhere in the documentation. Also, the references to the answer categories from an enumerated type need to be consistent with what is shown in the main documentation. For example, if the short names as used in the CAI program are used in the routing but are not listed in the documentation then the routing can become meaningless.

As the routing within large CAI instruments can get very complex, many users require an explanation in natural language for what the routing is doing. This is particularly useful for customers, policy maker or other users who are unfamiliar with programming languages. Although some users would find this sufficient, many users still require the more 'technical' and exact routing. Some users would like the option to display both. These descriptions of the routing would need to be incorporated within the CAI instrument so that they can be used by TADEQ.

There are two different concepts that CAI programs can use to control the routing of the questionnaires: 'conditional' routing, which specifies under which conditions the questions will be asked, and 'goto' routing which states which questions will be appear next once certain answers are given. Electronic questionnaires use only one type of routing, users want both types of routing displayed. The majority of users require the conditional routing and although many users also wanted the goto routing, none stated they could use this type of routing on its own.

Both types of routing have problems associated with clearly displaying the information. Some of the issues and problems in showing the routing information required by users are described below. The paper by Jelke Bethlehem elsewhere in this publication discusses the detail of displaying routing structures of electronic questionnaires.

Go-to routing

For complex questionnaires 'go-to' routing is difficult to display in an easy-to-read format, as the 'go-to' instructions themselves may be dependent not only on the answers to the current question but also answers to previous questions. For example:

Marstat

Are you:

- | | |
|--|---|
| (1) single, that is, never married | -> if (DMSIZE > 1) goto LiveWith
else goto HHldr |
| (2) married and living with your husband/wife | -> goto HHldr |
| (3) married and separated from your husband/wife | -> if (DMSIZE > 1) goto LiveWith
else goto HHldr |
| (4) divorced | -> if (DMSIZE > 1) goto LiveWith
else goto HHldr |
| (5) or widowed ? | -> if (DMSIZE > 1) goto LiveWith
else goto HHldr |

This is a simple example. Many of the conditions used in CAI programs are much more complicated. In complex instruments, go-to routing which creates the conditions for a variable from routes from several different parts of the instrument makes it very difficult for an analyst to reconstruct the conditions accurately. Many of the users stated that they would like to have documentation which shows 'goto' routing for their questionnaire (irrespective of the CAI software used to collect the information). However, many users were worried about the complexity of their surveys and that this would make the documentation unusable.

One option is to display goto routing as a graphical representation as a 'flow diagram'. Many users thought a diagram would be useful to get an overview of the questionnaire as a whole or to look in detail at the flow of questions in particular sections. However, they felt that a diagram showing the detailed routing for the entire questionnaire would be too complicated to use.

Users stated that they needed some sort of overview of content and structure of the questionnaire. Although many users stated that this would be extremely useful very few had documentation that provided one. In the electronic version of TADEQ this diagram could be used interactively, as the basis of navigation throughout the questionnaire; to move up and down hierarchies and between blocks to see the flow of questions at a particular section. However the user must be able to know where they are within the questionnaire as a whole, as it would be easy for a user to get lost within a large and complex questionnaires.

Users would also like the option to alter the information shown in the diagram, such as showing just the variable names, or names plus the description and the answer categories. To keep the diagram simple many users may wish to display the variable names and would therefore need links and references to the other information they require.

Conditional routing

The conditional routing of questions in complex questionnaires is usually built up from a number of nested statements. For example:

```
IF (LDMINDINTERVIEW[LDM1] = now) THEN
  Natural
  IF Natural = Yes THEN
    NatNum
    IF (Filling = yes) THEN
      NumFill
    ENDIF
  ENDIF
ENDIF
```

Some users need to see the full conditions under which each variable applies. This is particularly useful for data analysts or other users looking at individual questions in isolation. However in complex questionnaires the number of nested conditions can get very large, resulting in a number of complicated conditions being repeated for each variable. This can make the documentation hard to read and many users do not find all this information useful; the most useful information is lost in the forest of detail.

Many users will only be interested in the last few nested conditions, as these are the most likely to change in a related sequence of questions. Users need to be able to control the levels of routing that are displayed. This could be done either by hiding the routing which is common to a selected set of questions or they may wish to hide all the nested conditions. Whenever any routing is hidden there would need to be some indicator to the user that they were not seeing the full conditions of the routing for the variables.

In many electronic questionnaires questions are repeated, such as asking the same questions for each adult in the household. These questions can be nested within each other. This repetition, although it makes the instruments more efficient, can cause problems when trying to make the routing transparent for users. In general, the questions that are repeated should be shown only once (otherwise the documentation would be extremely lengthy). Users need to know under what conditions and the number of times the questions are repeated. In many instances repeated questions require the use of temporary variables as counters that may not appear in the documentation. Sometimes the number of times the question is repeated is not fixed and depends on the answer to another question within the questionnaire. The ability to add natural language descriptions to describe the routing and to hide some of the nested conditions would make the routing for repeated questions more comprehensible.

How the question appeared on screen

Some users, particularly methodologists, stated that they would like to know how the questions looked on screen. This was something that was not available on their current documentation. However, the information required to produce this is not usually part of the individual questionnaire, but is controlled by the CAI software itself. With the increasing complexity of

computer software the interviewers can change the settings on their computer. That could affect how the questions look, such as changing the background colour, or the colour, size and font of the question text. Also, with the use of text substitution the questions themselves can change in different circumstances and this can affect how the questions appear on the screen.

Edit checks and computations

One of the major advantages of using CAI is the ability to conduct consistency checks on the data during the interview. Some users want to see what the checks were, others do not. Again, TADEQ will need to be flexible enough to provide the information in the format that is required. The users may wish only to display the checks for certain variables. They may wish to display the checks as part of the information for each variable or they may want all the edit checks to be displayed together. They may wish just to see the conditions under which the questions are checked or they may also wish to see the message that appears to the interviewer.

Likewise, some users want to see the computations that are carried out within the CAI instrument, others do not. There are many uses for computed variables: to calculate variables that are subsequently used in routing or checks; or for text substitution; or be used in the analysis of the data, instead of calculating them at a later stage. Some users may wish to know that the variables were computed without knowing the specific derivation, others wish to see the exact conditions under which all the possible values are computed. Some wish to see the computations with the main documentation for the variables, others want them to be listed separately.

Many of the issues and problems of documenting conditional routing also apply to the documentation of edit checks and computations. For example, there may be a repetition of nested conditions if a number of different variables are computed or checked under the same conditions, which could greatly add to the length of the documentation.

6. Analysis tools

The aim of TADEQ is to produce a tool that can analyse the structure of a CAI instrument as well as document its content. The first prototype will concentrate on the documentation tools and the analysis tools will be added later. However during the user consultation some users did state some user requirements:

- Questions which can never be asked due to impossible conditions within the routing;
- Basic information on the questionnaire, such as the number of questions, the maximum, minimum and average number of questions asked etc.
- Which questions apply given certain conditions, for example, people aged over 16 and not working;
- A theoretical simulation of the number of questions and therefore the length of the interview when fieldwork is carried out. This could be done by estimating of the average number of questions asked calculated from the average number of questions for separate subgroups of the sample and weighting them according to external information (such as population figures).
- Comparison and documentation of the changes between two versions of the same survey.

7. Display templates

As the user will be able to control what information is displayed, they should be able create templates on how the information will look on the screen or page (such as fonts, sizes, colours for different information). It should be possible for the user to save these options so that the documentation will appear in the same way the next time it is opened. Users could then use the same options to look at the documentation from another questionnaire. This would enable users to produce a 'house-style'. They may also want a different template for producing documentation for internal use and one for general dissemination. Users may also want to create a special template for producing a paper document that could include the creation of page numbers, contents page and an index etc.

8. Conclusion

Large and complex questionnaires require a large amount of information to be documented. The TADEQ tool needs to be flexible to allow the user to display as little or as much information as they require. The means of achieving this aim will be different for electronic and paper modes of documentation. This paper has discussed a number of the different requirements and problems that the TADEQ project has to solve in order to produce documentation that can be used by a wide range of users to help perform a wide range of tasks.

The TADEQ project will produce a tool that will automatically document electronic questionnaires. It will provide an accurate and comprehensive representation of an electronic questionnaire. The project will make full use of the facilities available when documentation is itself electronic, including those such as hypertext links which are not available for paper versions of the documentation. However, it will be guided by the user requirement for paper documents for the foreseeable future. The project will ensure that a wide range of options to print out the documentation is available.

The paper has concentrated on what the TADEQ tool may do for users - on its outputs. As with any tool, the inputs (in this case, from survey designers and their customers) will be crucial to its success. The quality of any automatic documentation from electronic questionnaires can only ever be as good as the quality of electronic questionnaires themselves, irrespective of the quality of the documentation tool. If the questionnaire is badly structured so will be the subsequent documentation. Many of the features requested by users, and which will be available in TADEQ, will require the addition of information to the electronic questionnaires, such as natural language descriptions of routing. There will be ways to add this information at the later stages of a survey, but, to realise the full potential of the TADEQ tool most cost-effectively, users will need to plan and input documentation at the design stage of the survey work. In this, we see another example of the way that CAI integrates the formerly sequential processes of survey work.

References

Anderson, S (1997) Automated Paper Documentation of Blaise III. *Actes de la 4e Conférence Internationale des Utilisateurs de BLAISE*, INSEE, Paris, 1-20.

Bethlehem J and Manners T (1998) TADEQ: A tool for Analysing and Documenting Electronic Questionnaires. *Proceedings of the 5th International Blaise Users Conference*, Statistics Norway

Couper M P and Nicholls W L (1998) The History and Development of Computer Assisted Survey Information Collection Methods *Computer Assisted Survey Information Collection Methods*, ed. Couper M P et al., Wiley, New York

Manners T and Deacon K (1997) An integrated household survey in the UK *Actes de la 4e Conférence Internationale des Utilisateurs de BLAISE*, INSEE, Paris

Nicholls W L et al (1997) The effect of new data collection technologies on survey data quality *Survey Measurement and process quality*, ed. Lyberg L et al., Wiley, New York

About the Author

Maureen Kelly is a survey researcher in the Social Survey Division of the Office for National Statistics. She has worked on a number of surveys within the division and also provides advice and support for the development of Blaise questionnaires. She can be contacted at ONS, D2/01, 1 Drummond Gate, London, SW1V 2QQ; tel 020 7533 5308; fax 020 7533 5499; e-mail: maureen.kelly@ons.gov.uk.