The IASSIST QUARTERLY represents an international cooperative effort on the part of individuals managing, operating, or using machine-readable data archives, data libraries, and data services. The QUARTERLY reports on activities related to the production, acquisition, preservation, processing, distribution, and use of machine-readable data carried out by its members and others in the international social science community. Your contributions and suggestions for topics of interest are welcomed. The views set forth by authors of articles contained in this publication are not necessarily those of IASSIST.

Information for Authors:
The QUARTERLY is published four times per year. Authors are encouraged to submit papers as word processing files. Hard copy submissions may be required in some instances. Word processing files may be sent via email to jstratford@ucdavis.edu. Manuscripts should be sent to Editor: Juri Stratford, Government Information and Maps Department, Shields Library, University of California, 100 North West Quad, Davis, California 95616-5292. Phone: (530) 752-1624.

The first page should contain the article title, author's name, affiliation, address to which correspondence may be sent, and telephone number. Footnotes and bibliographic citations should be consistent in style, preferably following a standard authority such as the University of Chicago press Manual of Style or Kate L. Turabian's Manual for Writers. Where appropriate, machine-readable data files should be cited with bibliographic citations consistent in style with Dodd, Sue A. "Bibliographic references for numeric social science data files: suggested guidelines". Journal of the American Society for Information Science 30(2):77-82, March 1979. Announcements of conferences, training sessions, or the like, are welcomed and should include a mailing address and a telephone number for the director of the event or for the organization sponsoring the event.

Editors
Karsten Boye Rasmussen, Eckerbergsgvej 56, 5230 Odense M, Denmark. Phone: +45 6612 9811 Email:kbr@sam.sdu.dk

Juri Stratford, Government Information and Maps Department, Shields Library, University of California, 100 North West Quad, Davis, California 95616-5292 Phone: (530) 752-1624. Email:jstratford@ucdavis.edu

Production
William Block, 25 Blegen Hall, 269 19th Ave S, University of Minnesota, Minneapolis, MN 55455. Phone: 612-624-7091. Email:block@secsci.umn.edu

Walter Piovesan, Research Data Library, Simon Fraser University, Burnaby, B.C. Canada V5A 1S6. Phone: (604) 291-5869. Email:walter@sfu.ca

Features
4 Development of a Health Data Archive for Bangladesh
D. Leadbetter & L. Gainness

18 Text, Sound and Videotape: The Future of Qualitative Data in the Global Network
Louise Corti

26 Computer Assisted Personal Interviewing: A Method of Capturing Sensitive Information
E. Forster & A. McLeekey

Content
Volume 23 Number 2 Summer 1999

Title: Newsletter - International Association for Social Science Information Service and Technology
ISSN - United States: 0739-1137 © 1999 by IASSIST. All rights reserved.
Development of a Health Data Archive for Bangladesh - An Example of a Cost Effective and Sustainable Approach to Information Sharing in a Developing Country.

by Deana Leadbeter & Lorna Guinness

Office 97 software, thus requiring no extra capital investment in the initial development period.

The creation of an operational Archive in a short space of time and at minimal cost has allowed potential users to see the immense benefits of such a tool. The flexibility of the Archive design will allow it to expand to meet the demands of more databases and users with few technical problems. The next steps will see wider dissemination so that more databases related to the health sector will be entered on the Archive, and users will expand to a wider audience in the GOB, donors, NGO’s and research institutions. The process of institutionalisation and mechanisms for cost-recovery are now being addressed, to ensure the maintenance and sustainability of the Archive.

Background

There are a large number of organisations working in the health sector within Bangladesh, including aid organisations, Government of Bangladesh (GOB) and a variety of non-governmental organisations (NGOs). Many of these organisations are involved in data collection and all of them need relevant and up to date information to carry out their work. However, although a considerable amount of data has been, and continues to be, collected, it is not always easy to find out what information has been collected or to gain access to this information. Information can be over-protected, located in numerous sites and difficult to track down. These problems are exacerbated by limited computing facilities.

Problems in data access, and lack of information exchange and co-ordination between organisations carrying out research, often lead to a duplication of data collection efforts and can limit the opportunities for improving the process of information gathering through collaboration and dialogue. These difficulties can slow the decision-making and planning process or cause it to be based on less reliable information. However in Bangladesh, as in some other developing countries, there is a current trend to move towards involving all stakeholders in a health sector wide approach to policy-making, planning and programme implementation. This means that the need for co-ordination in information gathering and access is therefore

Abstract

Gaining access to information on the health sector in Bangladesh, and in many other developing countries, can sometimes be very hard. Although a considerable amount of data is collected by government departments, non-governmental organisations (NGOs) and other agencies, it is not always easy to find out what information has been collected or to gain access to this information. These difficulties can reduce the potential value of the information, slow the decision-making and planning process or cause it to be based on less reliable information. With the current trend towards involving all stakeholders, in developing countries, in a health sector wide approach to policy-making, planning and programme implementation, the need for co-ordination in information gathering and access is greater than ever.

The Health Economics Unit, of the Ministry of Health and Family Welfare, has initiated the development of a Health Economics Data Archive (HEDA) for Bangladesh, which aims to address the problems of access to information for policy-makers, planners, researchers and others involved in the health sector. Amongst the aims of the project are: providing a tool for dissemination of research results; a standardised approach from which to improve methods of data collection; the development of a health data dictionary for Bangladesh; encouraging data security; and fostering a culture of information sharing. Use of the Archive can also prevent duplication of research activities and encourage improved or standardised methodologies.

The needs and suggestions of the potential users and holders of an Archive were obtained through a process of workshops, seminars and consultation. The Archive itself was then started as a small entity holding the databases, and supporting documentation, for Health Economics Unit studies. A user-friendly front-end screen was designed in Access 97 software, enabling searches by subject area, keyword, geographical area and free text to identify databases held on the Archive. At present, it is possible to hold and use the Archive on a standard PC computer using Microsoft

Example

Development of a Health Data Archive for Bangladesh - An Example of a Cost Effective and Sustainable Approach to Information Sharing in a Developing Country.

by Deana Leadbeter & Lorna Guinness

Office 97 software, thus requiring no extra capital investment in the initial development period.

The creation of an operational Archive in a short space of time and at minimal cost has allowed potential users to see the immense benefits of such a tool. The flexibility of the Archive design will allow it to expand to meet the demands of more databases and users with few technical problems. The next steps will see wider dissemination so that more databases related to the health sector will be entered on the Archive, and users will expand to a wider audience in the GOB, donors, NGO’s and research institutions. The process of institutionalisation and mechanisms for cost-recovery are now being addressed, to ensure the maintenance and sustainability of the Archive.

Background

There are a large number of organisations working in the health sector within Bangladesh, including aid organisations, Government of Bangladesh (GOB) and a variety of non-governmental organisations (NGOs). Many of these organisations are involved in data collection and all of them need relevant and up to date information to carry out their work. However, although a considerable amount of data has been, and continues to be, collected, it is not always easy to find out what information has been collected or to gain access to this information. Information can be over-protected, located in numerous sites and difficult to track down. These problems are exacerbated by limited computing facilities.

Problems in data access, and lack of information exchange and co-ordination between organisations carrying out research, often lead to a duplication of data collection efforts and can limit the opportunities for improving the process of information gathering through collaboration and dialogue. These difficulties can slow the decision-making and planning process or cause it to be based on less reliable information. However in Bangladesh, as in some other developing countries, there is a current trend to move towards involving all stakeholders in a health sector wide approach to policy-making, planning and programme implementation. This means that the need for co-ordination in information gathering and access is therefore

Abstract

Gaining access to information on the health sector in Bangladesh, and in many other developing countries, can sometimes be very hard. Although a considerable amount of data is collected by government departments, non-governmental organisations (NGOs) and other agencies, it is not always easy to find out what information has been collected or to gain access to this information. These difficulties can reduce the potential value of the information, slow the decision-making and planning process or cause it to be based on less reliable information. With the current trend towards involving all stakeholders, in developing countries, in a health sector wide approach to policy-making, planning and programme implementation, the need for co-ordination in information gathering and access is greater than ever.

The Health Economics Unit, of the Ministry of Health and Family Welfare, has initiated the development of a Health Economics Data Archive (HEDA) for Bangladesh, which aims to address the problems of access to information for policy-makers, planners, researchers and others involved in the health sector. Amongst the aims of the project are: providing a tool for dissemination of research results; a standardised approach from which to improve methods of data collection; the development of a health data dictionary for Bangladesh; encouraging data security; and fostering a culture of information sharing. Use of the Archive can also prevent duplication of research activities and encourage improved or standardised methodologies.

The needs and suggestions of the potential users and holders of an Archive were obtained through a process of workshops, seminars and consultation. The Archive itself was then started as a small entity holding the databases, and supporting documentation, for Health Economics Unit studies. A user-friendly front-end screen was designed in Access 97 software, enabling searches by subject area, keyword, geographical area and free text to identify databases held on the Archive. At present, it is possible to hold and use the Archive on a standard PC computer using Microsoft
greater than ever.

In the United States and Europe overcoming these information problems is assisted with the use of a depository of information, called a Data Archive. This is a database that holds metadata i.e. information about the data that is held on other databases, as well as holding the actual data for some of these databases. The wealth of data concerning the health sector in Bangladesh continues to grow. Although these data are of potentially no less value than that in Western Archives, central archiving has not been a common practice and so the location of these data remains dispersed and difficult to access.

In a series of workshops, held by the Health Economics Unit of the Ministry of Health and Family Welfare, Bangladesh, in 1996, a serious problem in awareness of past and present research activities and also of the location of key health sector databases was identified. This situation is of concern due to increased costs of accessing information, duplication of research efforts and limited opportunities for improving information collection process. In addition, if there is a lack of knowledge concerning data or difficulties in accessing data, the potential benefits are not fully realised. The data are used only for their primary purpose and then either discarded, or stored but not re-used. Since data collection is usually very costly, if the data can be used for other analytical work (secondary data analysis), or if they can be used to inform the design of future studies or routine data collection exercises, then considerable savings and additional benefits could occur. In Bangladesh, this potential is currently not being fully realised which results in a resource waste, that resource-poor Bangladesh can ill-afford.

To address this problem and facilitate collaboration and information sharing amongst researchers and stakeholders, it was suggested, at the HEU training workshops, that Bangladesh should begin to develop a central depository for health economics relevant data, in the form of a Database Archive. The Health Economics Data Archive (HEDA) was therefore proposed to bring these data together to a central location, providing the similar functions to Database Archives operating in the U.S. and Europe, thus allowing for data to be re-used and wider dissemination of their key findings. In other words, adding value to the primary research carried out.

This paper describes the process of the development of HEDA in Bangladesh and the particular method used, which was low cost and user-friendly. Thus, it suggests a possible model for other resource-poor nations, where the full value of the wealth of primary data and generated research information may not be fully realised at the moment.

Aims of the Health Economics Data Archive (HEDA)
In the initial phases of development, the establishment of HEDA for Bangladesh had two primary aims:

1. Improving accessibility to data

By documenting health sector databases and using a standardised approach to documentation, as well as providing electronic searching facilities, HEDA provides easy access to data.

2. Dissemination of Health Economics research findings

An Archive makes the work of any research activity or organisation available to a wide audience in more detail than is possible through the publication of research papers or other reports. In the case of the Health Economics Unit, it was felt that HEDA would widen access to the primary and secondary research findings, within and outside the Ministry of Health and Family Welfare.

In addition to the primary aims of HEDA, there are several important additional benefits that arise from the process of developing HEDA that were considered as critical outputs of the project:

3. Improving study designs and methods of data collection

Before a study can be entered on an Archive, a study description form has to be completed. The completion of this form requires the lead investigator to describe the study design in a clear and consistent manner. Experience has shown that this not only provides valuable information for anyone wishing to carry out secondary analyses on the study data, but it is also a useful checklist of the issues that need to be considered when designing a study. Using the study description form can therefore serve as an ongoing training exercise in study design for all staff involved in the process. Completion of the form at the start of each study, rather than when the database is complete and ready to be entered on to an Archive, is helpful in ensuring high quality study designs. Having a clear and well-documented design is also likely to be useful when collaborators in several different organisations are involved in a study.

4. Development of a Data Dictionary

The study description also requires the studies to have clear definitions for all data items for the study, which are easily available to anyone wishing to access the data. A Data Dictionary should therefore be set up, within or in parallel to the Archive, which includes...
data definitions across all studies in the Archive. This helps to identify where data from different studies can be combined or compared because the same definitions have been used and also where different definitions have been used for similar data items, and hence direct comparisons are not valid. As with the study descriptions, if this process of documenting data definitions in full is carried out at the start of the study it will lead to improved study procedures, particularly in the area of data collection and analysis. Setting up this Data Dictionary helps users, who wish to carry out secondary analyses on the data or to combine data from different studies. It is also a valuable resource when designing future studies, particularly where these follow on from, or need to be compared with, the results of other studies. Further, it can enable improved data quality by allowing cross validation between databases.

5. Fostering data security

Another challenge, when storing data, is that of data security - both in terms of not allowing unauthorised access or inappropriate use - and in terms of ensuring that the data are maintained in good condition. Data can be lost at the flick of a switch, or may get corrupted because of problems with the power supply or physical environment, and databases can be manipulated without permission. The updating, maintenance, back-up and security problems usually faced with storing data can be placed under the responsibility of those responsible for the Archive, thus saving time and money for the original data producers.

6. Enabling bibliographic citations of databases

By establishing databases as bibliographic entities and "publishing" them as such, as well as offering advice on citation, Archives play a major role in extending research and scholarship, giving recognition and acknowledgement in the same way as printed piece of research work.

7. Fostering information sharing

Experience in other countries has shown that initiatives for information sharing can lead to greater

---

**Figure 1: Answering a Health Economics Query**

**Without Archive**

- Secretary's query: what is government health expenditure relative to household expenditures on health?
  - Consult HEU
    - Government expenditure obtained from Public Expenditure Review
      - Consult BBS: obtain dataset but only one year of data available
        - Locate and consult other organisations with similar data
          - Compile data and analyse results
            - Report to Secretary

**With Archive**

- Secretary's query: what is government health expenditure relative to household expenditures on health?
  - Consult Health Economics Data Archive
    - Pre-prepared data, analysis and results obtained
      - Report to Secretary
understanding and collaboration between organisations across all their activities not just those related to data collection and analysis. It also facilitates more comprehensive data analyses by linking data collected by different departments or agencies. This is of value at any stage of health sector development but is particularly relevant in Bangladesh or those countries introducing a sector wide approach, which requires a greater co-ordination within the Ministry of Health and Family Welfare and between all those involved in the programme, including the Donors.

How HEDA can add value
The value of a resource such as HEDA can be demonstrated by two examples. In the first example, the Secretary of the Ministry of Health and Family Welfare may make an ad hoc request to his assistant for information on the current level of household expenditures on health as opposed to government expenditures. What does the assistant do? The required processes of data collection and analysis are shown in Figure 1, both with a Data Archive available and without a Data Archive.

A second example of the value of the Archive can be shown by the steps taken by the Ministry of Health and Family Welfare, Bangladesh, to develop a new approach to revenue generation in the health sector. The government officer designated to assist in the gathering of background information will require data on the income levels, health expenditures and health seeking behaviour of the population, other health services provided by NGOs and the private sector and methods and current levels of revenue generation. How does the government officer do this?

The officer may have to locate and approach a number of different sources:

- The Bangladesh Bureau of Statistics (BBS) for household incomes and levels of health expenditures;
- Find and consult or carry out a surveys on health-seeking behaviour, a survey of NGOs active in the health field, a survey of private clinical services;
- Locate and consult MOHFW financial information on current levels of revenue generation.
- Health care facilities and MOHFW for unit costs of health services

A phone call or visit to a central Archive would establish whether this information was available and, if so, could provide the officer with the necessary information, saving both time and money for the officer.

An Archive also prevents duplication of research activities and encourages improved or standardised methodologies.

Under the second scenario, the government officer could have consulted the Archive to discover that a survey of private clinics had been completed and therefore the planned private clinic survey was not necessary. Alternatively, it could be that a survey of private clinics had been completed but was out of date. Using the instruments and results of the old survey available on the archive, the officer could make improvements based on the problems encountered in the first survey and collect information in a standardised way to create a time series.

Development of the Health Economics Data Archive
Identification of need and appropriateness
In June 1996 a series of workshops, meetings and training sessions regarding health databases were organised by the Health Economics Unit (HEU). The purpose of these sessions was:

- to promote greater knowledge of existing databases in Bangladesh,
- to help create a programme of co-operation among different data providers and data users and
- to agree a way forward for developing a metadatabase or Data Archive for health data in Bangladesh.

The participants in these workshops were asked the following specific questions:

1. What and where are the existing databases in Bangladesh?
2. What are the means of access to these databases?
3. How, and how often, are the data collected/updated for each of the databases?
4. How can the needs of consumers be fed into future database design and management?
5. What areas of mutual collaboration could be pursued, and how can this collaboration best be carried out?

The training sessions that were linked to these workshops were designed to orient non-specialists in the utility of databases and to explain best practice, associated with their design, handling and use. The training was aimed at mid-level managers in Government and those with a specific interest in databases and the use of information.

It was proposed at the workshops that a database could be set up to hold information about data already collected which could potentially be of use in health economics and other health related studies. This would function as a Data Archive holding data as well as ‘metadata’ concerning a particular study. Holding ‘metadata’, rather than the data itself, allows the data holder the option of retaining control over the specific purposes for which the data can be released, where the data held are confidential or sensitive.
It was expected that both the workshop and the training sessions would help to assess the feasibility of setting up an Archive, and to identify the data items that could be included in such an Archive. It was planned that participants of the workshops and training sessions could pilot the collection of these data items. A programme for setting up a Health Economics Data Archive (HEDA) for Bangladesh could then be drawn up, taking into account the information provided and views expressed at the workshop and training sessions and also the experience with the proposed pilot.

Since the participants at the workshops and the training sessions were mainly from Government Departments, individual meetings were also held with other organisations to discuss the HEDA project and obtain their views.

Once the pilots were completed and development was underway, potential users and HEU staff was consulted about the design of HEDA and the methods of access. Based on this a specification was drawn up for the technical support required and the need for both a computer programmer and a database manager with extensive experience in the health sector to join the HEDA was identified. In consultation with the HEDA technical team and potential users, the structure of the front-end screens and the search criteria were agreed. The design of HEDA was tested with examples of user queries, given by potential users of HEDA identified by the HEU, then considering how the use of the HEDA might assist in answering these queries. This ensured that the development of the HEDA was following a model appropriate both to the future users and holders of the Data Archive.

Issues that were considered in the design and development of HEDA
Since the concept of a Data Archive was new to many people within the health sector in Bangladesh, some basic principles were agreed at the outset. These principles are outlined below.

The term ‘Archive’ can be used for a repository or store of any material, although it is most commonly used for a store of information. The purpose of keeping anything in store is so that it is available for use when required. There is no point in keeping anything in any type of store unless:

- You know it is there
- You can get access to it when you need to use it
- It is kept in a good state

An important feature of a Data Archive is therefore to facilitate use of the information as well as holding the data. This means that a Data Archive should provide:

- A secure place to hold data, and also to hold information about the data or information derived from the data
- Information on what is held on an Archive
- Mechanisms to find the information or data when needed
- Mechanisms to access the information or data when needed

This requires, when archiving data, that sufficient and accurate data documentation is provided both on the data background, including sampling methods, sources of data, investigators etc, and the data characteristics, including data definitions, data relationships and coding systems. This is discussed further in the section on database documentation below.

In addition to these guiding principles, there were some key criteria, which were agreed on for HEDA in order to address the two common causes of developmental failure:

- when projects are over-ambitious so that they often fail to deliver within agreed deadlines
- when users, whose future participation is essential, do not see any benefits for themselves after the initial promises, and so they lose confidence and interest in the project

The key criteria were as follows:

1. **Timeframe**
   The initial stage of the project should be designed so that it is manageable within a short timeframe and produces a product that can be demonstrated to users within that timeframe. This means that the first phase should cover only a limited set of databases. Priority for inclusion in this initial set should be given to those databases that are already well structured and documented so they can be brought in to the Archive with a minimum of effort in order to demonstrate benefits within the short timeframe.

2. **Maintenance of user interest**
   To maintain user interest the databases selected for inclusion at the initial phase should be those that are considered to be of most interest to potential users.

3. **Limited technical resource requirements**
   Although specialist technical skills are needed for the initial development, the Data Archive should be designed so that it can be easily updated and developed by an in house team after the initial phase has been completed.

4. **Ease of access**
   The Archive should be accessible on a user-friendly front-end screen that requires minimal training and should be located in a central location with easy physical access.
Documentation of databases

The concept of two types of data documentation were agreed – macro or *data background* and micro or *data characteristics*.

The *data background* covers the context within which the data were collected and issues relating to how the information can be used, including:

- Supplier and user documentation
- Original forms and instructions
- Reports on data collection and usage
- Original output
- Minutes of meetings or policy documents
- levant to the collection and use of the data
- Information on data quality and usefulness

The *data characteristics* cover details about the data items and how they are held, including:

- Data types, field descriptions, data ranges etc.
- Data relationships
- Coding schemes
- Missing values
- System information

Without this level of documentation it would be impossible to achieve many of the aims of an Archive, including the basic premise of using the databases. A standardised tool for the documentation of each database to be held on any Archive is required to facilitate this process.

Technical Requirements

In the initial stages of development, support from a computer programmer is essential. However, the Archive should be designed such that once the software has been developed, it can be easily maintained and updated and Archive queries easily answered by in-house staff with a minimum of computer skills. In the long run, as the Archive expands it would be expected that a health information specialist will be required to maintain and update the database as a permanent member of staff. This is discussed further in the section on future directions.

Clearly Staged Development Process

Experience with projects involving the development of software has shown that a critical factor in the long-term success of such projects is ensuring that the software is available at the same time as users are made aware of its potential uses. Raising expectations before the product is available can be counter-productive. In addition, starting small but with flexibility can encourage a stronger foundation and demand for the project and allows the project to grow with the demands placed upon it. For these reasons, a clearly staged development process is necessary.

Design of HEDA

The HEDA was designed to provide users with access to:

- data documentation for establishing the history of data collection and analysis process for any data held on HEDA
- raw data
- a data dictionary, and
- some key results from analyses of the studies, where available.

It is also planned to distribute lists of the databases, key findings of newly acquired databases and developmental news of HEDA, both on a regular basis to regular users and on request. This will be provided on either floppy disks (probably containing Excel tables), or hard copy. HEU will also provide additional results tables in response to ad hoc requests from users.

Following discussions, it was decided that the first set of databases to be included should be drawn from studies carried out by HEU. The rationale for this was that these databases were likely to be more easily and quickly accessible to the HEU staff involved in the development, and HEU staff would be more familiar with the data structures and the results. This approach would allow HEU staff to test out the procedures for documenting the databases, using their own data. Any problems could then be identified and rectified before other organisations were asked to complete the documentation. Another advantage of this approach was that potential contributors would be able to see HEDA in operation before being asked to complete the documentation for their studies. This should help them to understand more clearly some of the requirements of the documentation, and also to see the value of contributing information to HEDA.

At present HEDA itself contains the databases, tables of key results from those HEU databases and tabulations created in analysis of secondary sources of data. It is planned that, at later stages, contents will include databases containing health financing and expenditure data (e.g. National Health Accounts) and socio-economic information (e.g. from the Bangladesh Bureau of Statistics), as well as information from research and NGO projects.

HEDA also contains documentary information about each database, and this is described in more detail below. It is expected that some of this information will be of interest in itself e.g. study design, as well as providing background information to aid the selection and use of data within HEDA. The design of HEDA includes the provision of facilities to search not only using pre-specified lists, but also using the keywords in the study design and, if necessary, in the text within the documentary information, including the data dictionary.
Documentary information for each database

As stated above HEDA contains documentary information for each database held on HEDA. In order to record this documentary information and create the searching facilities within HEDA, it was necessary to use a standardised questionnaire. At the training sessions, in 1996, a questionnaire was presented, covering the metadata collected by the UK's National Economic and Social Research Council (ESRC) Database. The data items on this questionnaire were discussed and participants agreed that, with only a couple of exceptions, which could easily be amended, all the questions were suitable for use in Bangladesh. All the participants agreed that this information could be collected about their databases for inclusion in HEDA. The ESRC Data Archive was approached to check that they had no objections to the use of their form, and to obtain advice on the use of the data documentation. The response from the ESRC was very positive. They were happy for their form to be used and made some useful comments in relation to the proposed development, and thanks are due to them for their help and encouragement throughout this project.

Thus, for each database the following information should be included in HEDA:

a) Study Description

This is based on the study description questionnaire used by the ESRC Data Archive in the UK, amended to suit local circumstances. This includes summary information such as study name and topic areas, as well as more details on the study design. These details will need to be known by any future user of the study data, as well as being used for searching for studies within HEDA which cover the user's particular area of interest.

b) Data Dictionary

For each data item (or each variable, in statistical terminology) a set of information is required. This is specified at the end of the study description questionnaire.

For each data item included in a database the following are needed:

i) Data item identifier (probably a summary name)
ii) Full name of data item
iii) Description of data item
iv) Data type
v) Field size
vi) Coding system (if used)

vii) Any particular comment about the data item
e.g. parts of the list of responses and codes may only be relevant in specific organisations.

viii) Whether this is being used as a proxy for another data item

Other information that is expected to be included, either within the data description or within the comments, relate to whether it is a computed data item and, if it is, how it was calculated, whether it is raw (primary) data or derived data, and any relationships between data items.

c) Survey Form

Where data were collected by survey, a copy of the questionnaire form used on the survey will be available.

It is possible that, for some databases, some of the information required for completion of the questionnaire may not be immediately available. Therefore, to collect this information, it was suggested that the questionnaire forms should first be sent out to the participating organisations, and then a visit should be arranged to review the forms completed and clarify any points of confusion. A specific appointment would be made for this visit to ensure that the person with the knowledge of the database is available to answer any queries.

Entry of Data Documentation

User-friendly electronic data entry forms were created within the HEDA software, for entry of data background and data characteristic information on the Archive.

Accessibility

It was agreed that HEDA should be easily accessible and comprehensible to a range of different users. This requires the use of software that is readily available and user-friendly. After discussions with IT experts familiar with database programmes, it was decided that HEDA should be developed using Access 97 which comprises both a programming language for development purposes and user-friendly facilities for use by non-specialists.

The design of the front-end screens and the search facilities has made full use of the facilities already available within the Access software. This has allowed for rapid, flexible and cost effective development of the system. The approach has been to provide a mixture of using user-friendly menus and 'buttons' for selecting the chosen options for searching and/or viewing the contents of HEDA, as well as using standard Access or Windows features. All the features used will already be familiar to Archive users who use other Windows based software, and an instruction sheet will be provided for those unfamiliar with Windows.

In these initial stages, HEDA is stored on one central
computer where it can be accessed both by HEU staff and by other users. Two options were considered for the final location: remaining within the offices of the HEU and moving to the National Resource Centre for Health Economics. This is discussed further later in this paper. Once a run time version of HEDA is available it should be possible to hold a version at both sites, providing easy access for both GOB officials and external researchers.

**Search Facilities**

HEDA design includes the provision of a user-friendly interface. This 'front-end' should help direct the user to the most useful databases according to his or her work. This is through clear subject categorisation, plus an index and explanation of the databases contained within. In developing search procedures a compromise had been reached between the speed and efficiency of the search and the amount of freedom the searcher is allowed in specifying the search criteria. If the search is restricted to using pre-selected terms then the database can be set up to allow this to be carried out quickly and easily. The drawback of this approach is that the terms selected by those setting up the database may not cover the specific interests of the full range of potential users, or be suitable for categorising additional study databases when these are added to HEDA.

An alternative approach is a 'free text' search, which allows the user to enter any word, or combinations of words, and search for a mention of these in the study documentation. This gives the user freedom to pick terms that reflect their area of interest. However searching through the full documentation for all studies can be slow, particularly as more studies are included in HEDA. Also there may be different ways of describing a particular topic. If the terms chosen to describe the topic of interest, for the purposes of the free text search, are not those used within the study description then the search will not pick up this study.

The approach taken to deal with these issues was to provide a mixture of search options. The front-end system to the Data Archive provides for several different searches

These use:

(i) Study name where this is already known.
(ii) Pre-selected topic areas
(iii) Geographical areas
(iv) Keywords or key topics within the study description
(v) Free text within supporting documentation
  e.g. any mention of immunisation

For example, using the pre-selected topic areas (option ii above) provides a quick route to finding studies covering general areas. Within the study description there is also the opportunity for the investigator to specify key topics and keywords that describe the areas covered by the study. These are then available for the user of HEDA to use in their searches (option iv above). The user can either type in a topic or keyword describing an area in which they are interested, or can select from a 'pick list' which gives all the topics or keywords recorded in the study descriptions held on HEDA. This is a slightly slower search method than using the pre-selected topic areas, but is much quicker than free text searching, and still allows considerable flexibility and specificity in the search criteria. Also the 'pick lists' of topics and keywords can be automatically updated every time a new study is entered on to HEDA.

For very specific queries these search methods may not be sufficient and so the user would then need to use the free text search facilities (option v above). This searches for the occurrence of a specified word, or combination of words, within the study description and also within the data definitions.

The study description questionnaire asks whether the study is national or district specific, and asks for the districts covered by the study. This allows the user of HEDA to search for studies relating to a specific district within Bangladesh. In this case, instead of selecting from a pick list of the districts, the selection is made using an annotated map of Bangladesh.

It is expected that, as well as using the search facilities to find studies satisfying a specifically defined search criteria, users will find it helpful to use the front-end system for browsing the information on HEDA. At any stage the user can browse through the information relating to all studies, or to a selected set of studies, to learn more about their design, and to view some of the results of analyses on the study database. Browsing through some of the lists created from the study descriptions, such as the lists of key topics and key words that have been recorded, or browsing through the data definitions will also be helpful in identifying studies of interest.

**Queries**

Because of the different needs and technical capabilities of users, a flexible approach is needed in terms of methods of access. This includes the need for a user friendly front-end to give direct access to HEDA, and the issuing of both short bulletins in hard copy and copies of key results on floppy disk. A query service will also be offered, where HEU staff will access HEDA on behalf of users. It is expected that this query service will be required where key results tables available through the front-end system do not provide sufficient information to answer the queries therefore requiring additional analysis. These key results tables will initially be tables giving the results of the analyses that were carried out when the study was first analysed. As HEDA develops, the range of these tables will be increased to cover the more common queries.
To run these queries the user will use the searching facilities to identify the appropriate database and extract the data for analysis, to create tables and print or save to floppy disk. It will also be possible to cross-reference the databases and, where compatibility allows, create links.

**Technical Support Required**

For the development of HEDA in Phase 1 technical support was required in two main areas:

(i) in Access97 programming
(ii) in system design and standards for data definitions and coding

The staff contracted to provide technical support was given the opportunity to discuss, and comment on, the draft specification and programme before this was finalised. This was important since it gave them an understanding of the overall aims of the project, and they were therefore able to participate in ensuring that the work they were carrying out provided the best technical solution to the requirements of the project.

The Data Archive was designed so that it could be easily updated and developed by an in house team after Phase 1 had been completed. The technical specialists provided an element of training to the in house staff during Phase 1 (mainly through advice and support on tasks that in house staff will be carrying out). This should ensure that routine updating can be carried out by in house staff with additional support being required for ad hoc technical inputs. However, once the scale of the project requires it, a database manager will be required as a permanent member of staff. Information requests in the form of queries from outside HEU will also require HEU staff input, particularly where synthesis or evaluation of the results is needed.

**Security**

HEDA has been developed to enable all those who are familiar with Windows environments to gain access to databases for downloading and to examine the techniques used in data collection. In order to prevent misuse on HEDA, security passwords have been built in for different levels of users, and access to original data files for general users will be ‘read only’ which will allow copying but not amendments. To prevent data loss, there is a CD-ROM back up system, and back ups will be taken on a regular basis.

**Updating and Development**

The effect of taking the in-house, staged development approach was that the initial phase of the project focused mainly on the inclusion, within HEDA, of databases available within the HEU itself. The further development of HEDA therefore includes adding additional databases as they become available as well as providing enhancements to the front-end screen and the pre-prepared analyses to meet user needs.

In the early stages HEU staff will be responsible for updating and developing HEDA to meet these and other user requirements as they arise. These developments will include modifications to the front-end screens and to the pre-prepared analyses available, as well as additions to the databases themselves. It is expected that some assistance will be needed, even in these early stages, from a health information specialist on a part time basis, for example on the development of the Data Dictionary. It is recognised that, as HEDA grows and develops, some more dedicated support will be required for updating, maintenance and continued development.

**Staged Developmental Process**

It was important to ensure that the timetable for the HEDA activities were agreed upon before any further approaches were made to the potential contributors and users of the proposed HEDA. The developmental process was planned in a series of stages, to allow review and dissemination and key points. This was planned to create demand for HEDA and obtain maximum input from potential users. The expected outputs of the three planned development stages are as follows:

**Phase 1: Design and programming of HEDA**

- HEDA, held on a single desktop computer, including database background information for all HEU research activities as well as the data characteristics and the database itself for at least three HEU studies.
- Key findings of HEU research documented and held in electronic distribution form
- HEU Personnel trained in using HEDA
- Full documentation of HEDA development
- Draft GOB-approved protocol covering rights of access to HEU databases
- Launch seminar for HEDA for potential users and contributors

**Phase 2: Development of HEDA contents and long term plan**

- HEU personnel trained in updating and developing HEDA
- Instruction manual
- Full set of HEU databases on HEDA
- Final GOB-approved protocol covering access to HEU and other GOB databases held on HEDA, or for which study descriptions are held on HEDA.
- Agreed programme for inclusion of selected non-HEU databases
- Agreed plan for institutionalisation, cost recovery, ongoing maintenance and support, and future developments, of HEDA
- Further dissemination seminars
Phase 3: Institutionalisation and implementation

• Implementation of institutionalisation activities planned in phase 2
• Staff recruited to provide ongoing maintenance and support to HEDA
• Promotion of HEDA use for MOHFW, NGO, university and other research personnel through training, a newsletter and briefing seminars.

Analytical Tools
In addition to the databases available, HEDA also provides access to a suite of statistical and modelling packages. This allows the user to carry out analyses or modelling using the data they have selected and copied from Archive databases. It is expected that this will be of use where the packages provided are not available on the user’s own systems. Also, even if a user is intending to take a copy of the data for analysis on their own systems, they may wish to carry out some preliminary investigations using the software available on HEDA. This may be helpful in case the results of these preliminary analyses suggest that some amendments may be required to the data selected, for examples including some additional data items.

Hardware and software requirements
During the process of specifying the database structure and the costing of the development, it was suggested that the ESRC Data Archive should be approached to see if the software used in running their metadatabase could possibly be transferred for use in Bangladesh. This was discussed with the ESRC Archive and it was found that their software was not suitable for transferring, although it appeared that the structure proposed could be set up fairly quickly in Microsoft Access 97. The advantages of Access are that it is a user friendly package and it is easy to transfer extracts from an Archive written in Access into other Windows based packages e.g. for inclusion in reports written in Word. It is also easy to view tables previously prepared in Word or Excel from such an Archive.

In addition, MS Access is available as part of MS Office97 packages and already available at the HEU. As a result, in the first phase, no software or hardware upgrading was necessary, thus keeping initial development costs to a minimum.

Progress to date
The programming of the front-end screens and the search facilities were completed in 1998, within three months of the start date for development. Initial testing was carried out and any necessary amendments made to the software. Testing of all the data and facilities continues and will be an ongoing process. Documentation of the software is available within the programme but an information sheet for users has yet to be developed.

The pilot study, subsequent to the 1996 HEU workshops, found that the ESRC documentation form was, subject to minor amendments, suitable for use in Bangladesh. Data International, a consultancy group working closely with the HEU, was then asked to complete the study description questionnaires for the HEU studies, in liaison with the HEU staff responsible for the individual studies (usually the principal investigator for the study). This information was entered on to HEDA using the electronic data entry forms. Experience with completing these forms led to some minor amendments being made to the questionnaire, but no major problems were found with the questionnaire content or design. A list of potential keywords was prepared to assist the principal investigators in identifying keywords relevant for their studies, although the principal investigators were free to specify whatever keywords or topics they felt best described their study.

The study descriptions were entered on to HEDA for all completed HEU studies, even if the databases and other supporting information were not yet available for the study. In order to prepare the HEU databases for inclusion in HEDA a list of all HEU studies was prepared, together with their current location and state of documentation. A list of outputs (tables of results) available for each of these databases was also prepared. These outputs will be made available to users via HEDA, and will also be available on request on either floppy disk or hard copy.

The demonstration of HEDA at a formal Launch seminar at the end of the three month development period in 1998 (phase 1), and in a series of individual demonstrations following the Launch, showed that HEDA was already a product that is both easily accessed and useful. Subsequent activities have focused on adding more databases to HEDA, and on plans for institutionalisation and appointment of staff which are discussed later (Phases 2 and 3).

Lessons learnt
Data documentation
Apart from the difficulties in actual physical access to information, the major problem in sharing data is that of complete and comprehensible data documentation. This is usually one of the greatest challenges in developing any Archive and the development of the HEU Data Archive has been no exception.

As well as documentation for users of the output tables, documentation is also needed for those wishing to use the databases held on HEDA for secondary data analysis. In order to carry out an analysis on any database it must be clear exactly what is the meaning of the terms that are used in the study. Often there is a lack of common understanding on data items. For example, if one wants to talk about bed capacity within a hospital, what is a bed? Alternative views of the definition of a bed could be:
• A fully functional bed in a hospital
• A space in a hospital that is available for a bed or mattress in a hospital
• A broken bed lying in the hospital storeroom

Or, to what does revenue refer?
• Are we talking about the revenue allocations of the GOB?
• Are we talking about the revenue budget of the Ministry of Health?
• Are we talking about revenue collected from user fees?

In addition, data items are often used as a proxy for another data item, which can cause confusion if this is not clearly specified. For example, allocations are sometimes used as a proxy for expenditure, or utilisation as a proxy for demand.

Ideally, these data definitions should be decided and documented before the study can be carried out and possible proxy data items identified, but as with the other supporting documentation that has been discussed, this is not always the case. Retrospectively completing documentation involves interviews with principal investigators and examination of survey questionnaires and codebooks. This is a time consuming task but the benefits are multiple, leading to the ability to re-use data and process or learning of the investigators contacted, therefore resulting in value added on research and improvements in methods of work.

The work so far on data definitions has focused mainly on clarifying and documenting data definitions within individual studies. However, if the data are to be linked or compared across studies, a common Data Dictionary is needed which covers the data across all the studies and which uses a common name for data items that are used in more than one study. At the moment the data definitions are held in one Data Dictionary, but no work has been done to identify common or proxy data items.

This requires additional work to review those data items that appear to be similar, and to identify whether the same data definition has in fact been used and whether the data items can therefore be linked. Once this review has been carried out then a linking table can be set up which includes the study data item name and the data item name from the common Data Dictionary. This can be used to list all data items within a study, or to look at all studies that include a particular data item in the data dictionary. Having a common Data Dictionary is an essential part of HEDA, so the steps that need to be taken to achieve this will need to be agreed. As with other parts of HEDA development, the technical task of merging the individual dictionaries into one is likely to be more straightforward than the non technical issues i.e. the task of checking across studies for consistency of definitions and clarifying the situation where different definitions have been used.

Search facilities
During Phase 1 a question was raised as to whether the search criteria should link only to studies, or whether it should be possible to identify individual output tables within a study. It was decided that, as far as the documentation and search procedures are concerned, the study description (in particular the keywords and topics), should give sufficient indication of the areas addressed by a particular study and its related tables. There should therefore be no need to have additional search criteria linked to individual tables within the file of output tables. Also any derived data items in the output tables should be in the data dictionary, and so searching on a particular data item will identify the studies (although not the individual output tables) in which it has been used. Once a study has been selected using the various search criteria, then the user can scan, ‘by eye’, the supporting documentation, including the list of descriptions of the output tables. It is expected that this will, in most cases, be sufficient for the user to select the tables of interest.

If HEDA grows and develops then it may be possible to consider introducing some more sophisticated search procedures although, as has already been mentioned, the introduction of more complex and flexible search procedures can lead to slower, more cumbersome searching. It was therefore felt that, at least in the short to medium term, the best approach would be to keep the search procedures relatively simple by linking the search criteria to studies, and not to individual tables. It is possible that, on reviewing the output tables, it may be felt that a user would not be able to find the table they want. In this case, the table title in the ‘pick list’ could be made a little clearer, and the topics covered by the table could be included in the study description.

Presentation of key findings and preset queries
One issue that needed to be addressed during Phase 1 was the format in which the output tables should be held. The outputs held on HEDA include tables giving the results of analyses already carried out using the study data. There are two methods of holding these outputs. The first option is to hold a specification of the calculations that were carried out, and use this to recalculate the results from the data whenever the tables are required. The other option is to hold the results of these calculations i.e. the actual outputs. This second option saves having to spend time recalculating the results each time they are required, but may need more space to hold the tables.

For the initial databases entered on to HEDA the decision was taken to hold the actual outputs, usually in Word or Excel tables, rather than re-calculate the results each time. In many cases what can be viewed (and copied for further
manipulation if required) is just an electronic copy of the
tables of results from the published reports. This means
that entering the tables of results does not involve any new
data entry, just taking a copy of the electronic version of
the existing documents and then setting up the necessary
references and linkages within HEDA.

One of the strengths of the design of HEDA is that different
approaches can be taken for different studies or for
different sets of tables within a study, and so a different
approach can be taken for future studies if this is preferred.
This includes the option of having hard copies available if
the tables of results are not available electronically. In this
case asking to view these tables within HEDA would
simply lead to a message indicating where and how the
hard copies can be viewed.

An important issue also raised was the supporting
documentation required for the output tables, including
definitions of the derived items. This documentation
should be available to anyone wishing to use the output
tables. Ideally this text should have been prepared at the
time the tables were produced. However, if the principal
investigator for the study had not prepared their report with
a more general readership in mind, then the existing
explanatory notes may not be sufficient for the purposes of
HEDA. It was therefore found that some additional work
was needed to enhance the existing documentation.

The output tables for each study can be held in one file or
held in a series of files – one for each table or subject
related group of tables. Each file containing a set of tables
is listed separately in the ‘pick list’ that is viewed when a
particular study has been selected. Thus, if each table is set
in a separate file their identification is more immediate than
if all tables for one study are held in a single MS Word or
Excel file. If it was felt necessary for a particular study,
every table could be held on a separate file, but this could
involve a considerable amount of additional work in setting
up these individual files. For each study entered on to
HEDA, the benefits of having more detailed listings of
individual output tables will therefore need to be weighed
against the extra work involved in setting this up before a
decision is made on how the tables should be held.

Training
Training for HEDA users will be carried out through a
process of in-house workshops and learning by doing.
However, HEDA can also act as a training tool itself. By
providing a series of databases on various health economics
related issues, along with the full and detailed
documentation of the data and the data collection processes
and analytical software packages, it provides a facility for
training in:

- Questionnaire design
- Sampling methods
- Statistical analysis
- Health economics analysis

Success of following the basic principles
The basic criteria followed for the development of HEDA
were to start small, limit the timescale and ensure users
were involved and could recognise the need and relevance
of the project at all stages. Following these criteria meant
that in a short space of time and with very limited resources
HEDA was able to stand alone as a useful and technically
easily accessible package. The potential users and
contributors have shown interest in its further development
as they can see results at this early stage and visualise the
benefits in the future. Continued involvement with these
users is essential towards maintaining the momentum
already created.

Future directions
Expanding the HEDA user population
The driving forces behind the development of HEDA have
been the issues of co-ordination and information sharing. It
has started small, but this is not for want of ambition.
Smallness has given greater flexibility to adapt and make
amendments during the development phase. It is hoped
that, unlike many initiatives that have started big, the
enthusiasm and interest will not wane after the first phase
since the project can already demonstrate benefits. There
have been many incidental benefits during the development
phase, as has been discussed above, but the main benefit is
that the HEDA database is useful as it stands. It already
provides access to both modelling and analysis tools and a
series of comprehensive databases on health economics in
Bangladesh, with the room and flexibility for growth and
expansion at low cost.

In the future, it is expected that there will be further
development of the model and a steady increase in the
number of databases held on it. The developments will be
based on feedback from all potential users, in particular
those who attended the Launch and other demonstrations of
HEDA. As well as developing the model, the aim is to
ensure that HEDA is used by all those with an interest in
information about health and health services whether
government, donors, and researchers, and to see the
numbers of users growing over the years.

Further workshops, to demonstrate HEDA, are planned.
These will focus on the donor community, who are
expected to be both contributors to, and users of, the
information held on HEDA. In addition to formal group
sessions such as this, and informal individual
demonstrations during the early stages of the project, it is
important that users are kept up to date on progress with
HEDA. It is therefore planned to issue a Newsletter to let
interested individuals or organisations know what new
features or new data are available on HEDA. The
frequency of this publication will depend on the speed of

Summer 1999 15
development of HEDA but it is expected that it will be issued quarterly.

Addition of further databases
The initial phases of the project within Bangladesh involved including only HEU databases in HEDA. However, the way HEDA has been designed means that it is relatively easy to bring in data from other organisations, and the inclusion of databases from other organisations is now underway. The results of workshops and discussions with those involved in collecting or using health data, have indicated that several organisations would be interested in having their data included in HEDA. This will improve dissemination of results from many different sources. Also the wider the range of data included in HEDA the more useful it will be in providing answers to users’ ad hoc queries.

Towards this end, study description questionnaires were made available to all those attending the Launch, and are also being sent to other organisations who hold health sector relevant databases and may be interested in contributing data to HEDA. It is planned that the completed questionnaires will be entered on to HEDA even if the database itself is not to be held on HEDA. HEDA users will then have a reference to sources of information in addition to those held on HEDA.

Hardware and software requirements
The technical requirements of HEDA will need to be reviewed in the light of the proposed developments. It is expected that the computer currently being used for HEDA will need to be upgraded in terms of memory, speed and disk space available as more databases are added to HEDA. Also increased security facilities would be available if the operating system was changed from Windows 95 to Windows NT. The front end systems which have been written in Access 7 would not need to be changed since they will run under Windows NT and so this change would only involve minor programming amendments.

Staffing
As has already been mentioned, existing HEU staff have been responsible for updating and developing HEDA in the early stages with some clerical and data input support from staff at Data International. If HEDA is to build on its successful initial phase and develop as planned, then some more dedicated support will be required for updating and maintenance, and for liaison and support for user and data producers wishing to enter databases on to HEDA. It is suggested that this person should be a health information specialist who can carry out analyses to support user queries as well as being responsible for maintaining and developing the database. The recruitment process is currently underway.

Multi-user networks and dial-up access
If the Windows NT operating system were used, this would also allow dial up access if it were wished to include this in later developments. Alternatively a new Microsoft product has just been launched for multi-user access. This may be more appropriate for use with HEDA since it is designed to need less powerful facilities at the remote sites. It should be noted however that both these options do require reliable and high quality telephone connections. It is also not clear how well remote access, without an HEU staff member available to answer queries, will work in practice. It is therefore suggested that dial up access is not considered until HEDA has been in use for some time and there has been an opportunity to assess the level of support required by users.

One way of making information available to users on what is held on HEDA, and also possibly giving access to some of the data or results tables, is via a Web site. One of the benefits of developing an open web site is that the information is then easily available to anyone with access to Internet, whether in Bangladesh or elsewhere. This could be particularly useful in developing collaborative links between health sector researchers and analysts in Bangladesh and those working in other countries, particularly in the Asian region. However there are several difficulties with this sort of development. There is no control over who can access the information, and whether they are then using the information appropriately. Also specific technical skills are needed both to set up and to maintain the site.

It is therefore suggested that, if a web site is to be developed, a staged approach should be taken to this development. The first stage should focus on providing textual information summarising the activities of the HEU, and the databases that it has available, plus e-mail contact details. Summary tables of results or other relevant study information could then be e-mailed to interested enquirers. E-mailing information on request would be much simpler to do than setting up a front-end system to access information via a web site. It would also allow records to be kept of all those who have received specific study information and provide some control over access.

Institutionalisation
Consideration needs to be given to the most suitable place within the organisation for housing and maintaining HEDA. Although the HEU has been responsible for setting up HEDA, and will be supporting it in the short term, this may not be the most suitable option in the longer term. The statutes for an Institute of Health Economics at Dhaka University have recently been drawn up, and it is proposed that this Institute should house a National Resource Centre for Health Economics. It is expected that HEDA will play a central role in the development of any Health Economics Resource Centre. It has therefore been agreed that the Resource Centre that is being set up should house and take
on the responsibility of the running of HEDA in the longer term. Plans for setting this up are currently underway.

The issue of accessibility by the main users of HEDA needs to be taken into account when making any recommendations regarding the future siting of HEDA. Since it is not expected that dial up access will be available for some time then ease of physical access will be an important factor. Housing HEDA at the University will make it more accessible to academics, researchers and other interested organisations outside the GOB, such as donors. However this option would create barriers to access by GOB staff and so may have the effect of reducing the use of HEDA, and the valuable information held in it, by policy makers on the GOB for their decision making.

It should be noted however that HEDA has been designed to run on any reasonably powerful PC, and both the software and the data will be copied onto CD-ROM on a regular basis for back up purposes. It is therefore be relatively straightforward to house HEDA at the University and carry out any updating there, but to have a copy of HEDA also running at a site within the MOHFW. This could be regularly updated via CD-ROM. If this option was followed then an information analyst, based at the MOHFW, could be responsible for supporting GOB users of HEDA and liaising with GOB data producers, as well as carrying out analyses using HEDA to answer ad hoc queries from policy makers.

Sustainability
If HEDA is to be properly maintained, suitable funding arrangements need be agreed both in the short term and in the longer term. One of the main aims of HEDA is improved dissemination of information, and this will be achieved by encouraging data collectors to deposit information about their databases in HEDA and by encouraging use of the information and databases held on HEDA. Any fees introduced for either depositing information, or for using HEDA, will therefore need to be carefully considered to ensure that they are not creating barriers to effective development and use of HEDA. Funding mechanisms used elsewhere usually include some ‘block’ funding to cover the basic cost of maintaining and developing the Archive, with only a proportion of the overall cost being recovered through user fees. These user fees are usually linked to a registration fee for an organisation or individual wishing to sign up as an ‘Archive user’, rather than being linked to amount of use which can be difficult to monitor. An additional charge is usually made for use of Archive staff time to access and analyse information on behalf of a user, unless it is a routine query, which can be dealt with quickly. Opportunities for obtaining ‘block’ funding from different sources are currently being considered and will be explored further once the institutionalisation process has been completed.

Acknowledgements
Thanks are due to all the participants at the workshops, training sessions, development meetings and launch for HEDA for their time and helpful comments on the design and development. In particular thanks are due to the following for their commitment and assistance to the project:

Mr Shafiq Hussein, Computer Programmer, HB Consultants,
Dr Indrani Haque, Independent University of Bangladesh,
Farrah Hannan, Azizur Rahman and Tahmina Begum of Data International.

We would also like to express our appreciation to Professor Sushil Howlader, Head of the Institute of Health Economics for his vision and commitment in the setting up of this Institute, and in taking the Data Archive forward as a key part of the Institute’s planned Resource Centre.


Abstract
We are currently seeing a new culture emerging in the social sciences, of a new form of secondary analysis - that of primary qualitative data. It has come about largely as a result of the moves by British social science funding organisations towards formalising archiving policies of data created in the course of research they fund. Funders want added value from research and believe in sustaining a solid research base for the future, in the form of the preservation of empirical findings. Now, this includes qualitative data in addition to quantitative.

However, not only is this a new methodological approach for traditional qualitative researchers it is also challenging the way qualitative researchers view ownership of their raw data. New ideas about sharing and providing access to qualitative data are emerging - and in the UK, this is being championed by the Qualidata Resource Centre at the University of Essex.

This paper seeks to address a number of issues. From an archival point of view, how do qualitative data differ from quantitative data? Second, what might the implications be for the acquisition, preservation, dissemination and re-use of qualitative data archives for Data Archives? Thirdly, I want to discuss the kinds of procedures required to document and provide access to qualitative data. Inherent in this are the special problems relating to confidentiality of some qualitative materials, and I will suggest ways of overcoming these. Finally, I want to raise a number of questions relating to how the traditional Data Archives might want to consider acquiring, storing and disseminating qualitative data. Is it in their interest to acquire them? What kind of infrastructure needs to be in place to accomplish this?

Background to archiving qualitative data in the UK
The ESRC Qualitative Data Archival Resource Centre (Qualidata) is supported by the Economic and Social Research Council (ESRC) and is located in the Department of Sociology at the University of Essex. The Centre was established in 1994 in order to redress the balance in the bias towards archiving quantitative data from British social science research. It currently has funding up until the end of September 2000. Our relationship to the UK Data Archive is one of a younger sibling.

The Data Archive was set up in 1967 by the Economic and Social Research Council (ESRC) in order to retain the most significant machine-readable data from the research, which it funds. In order to achieve this, ESRC instigated a 'Datasets Policy' whereby all machine-readable data generated from ESRC awards should be offered for archiving. There was, however, a significant loophole in this policy. Although the advances of word processing now mean that most research of any kind is machine-readable, until recently most machine-readable data was statistical, based on surveys. Qualitative research was paper-based. Thus the Data Archive received only a proportion of the raw research data funded by the ESRC.

As Paul Thompson, Director of Qualidata, stated in his 1991 pilot report to the ESRC: 'there was no intellectual reason for this'. Qualitative and quantitative research are equally based on comparison. Classic re-studies include not only Rowntree's three surveys of poverty in York, and Llewellyn Smith's repeat of Booth's poverty survey in London, but also the two successive multi-method community studies of Banbury, or, to take an anthropological instance, the controversial restudy and reinterpretation by Oscar Lewis of Redfield's Tepoztlan in Mexico.

It is not therefore clear why the early Social Science Research Council (SSRC) did not feel the need to provide for the archiving of non-machine readable research data. Perhaps it was simply felt that there were enough existing archives to ensure that significant material was saved. But in practice, that was certainly not the case. Some qualitative material was archived, but usually in special temporary deposits. Thus the interviews on which Professor George Brown's notable studies of the social origins of depression, are based, were for many years held at his Medical Research Council Unit, of which the long-term future remained until very recently uncertain. Similarly, the material from Paul Thompson's national study of 'Family Life and Work Experience before 1918', a unique and unrepeatable set of 444 interviews with men...
and women born before 1918, were kept on a short-term basis in a special room at the Sociology Department at Essex, and consequently became the basis of a series of books and articles by visiting scholars, but had no secure future. More generally, little attempt of any kind was made to archive research material.

When a small pilot study commissioned by the ESRC was carried out in 1991, it was revealed that 90% of qualitative research data was either already lost, or at risk, in researchers’ homes or offices. Even with the 10% ‘archived’, it turned out that many of the so-called archives had none of the basic requirements of an archive, such as physical security, public access, reasonable catalogues, or with recorded material, listening facilities. It was estimated that to create a resource on the scale of that at risk would cost at least £20 million. For the older material, moreover, the risk was acute, and the need for action especially urgent.

**Qualidata’s mission**

Qualidata was set up by the ESRC with a dual mission. The first was a rescue operation aiming to seek out the most significant material created by research from past years. The second was to work with the ESRC and the Data Archive to ensure that for current and future projects the unnecessary waste of the past does not continue. Qualidata is not an archive itself; it is both a clearinghouse and an action unit. Its role is to locate and evaluate research data, catalogue it, organise its transfer to suitable archives, and publicise its existence to researchers and encourage re-use of the collections. We maintain a catalogue, Qualicat, located on the World Wide Web, which provides information both about qualitative datasets archived by the Centre and those identified by the Centre as having already been archived. The catalogue structure follows that of CESSDA very closely, with some new and modified fields to suit the characteristics of qualitative data.

The Centre consults with the ESRC and other funding bodies on, the now explicit, qualitative aspects of the Datasets Policy and provides advice to researchers on the implications of archiving for research, both through organised workshops and through individual consultations. The Centre also aims to provide a general stimulus to the practice and standards of qualitative research, especially in documenting social science research in Britain, as well as encouraging a more active interface between qualitative and quantitative research.

**How do we define qualitative data?**

Qualidata is concerned with research data arising from the range of social science disciplines, including sociology, social policy, anthropology, social and economic history, political science, social and human geography and social psychology. We define qualitative data as data collected using a qualitative methodology, which contrasts markedly to the traditional quantitative approach. Qualitative research is defined by openness and inclusiveness, aiming to capture participants’ lived experiences of the world and the meanings they attach to these experiences from their own perspectives. Moreover, a qualitative perspective encompasses a diversity of methods and tools rather than a single one. Our definition of qualitative includes in-depth or unstructured interviews, field and observation notes, unstructured diaries, personal documents, photographs and so on, in typed, hand-written, images, audio and video format and either as a digital or non-digital representation.

**Where do we put the data?**

One of Qualidata’s ongoing objectives is the selection of public repositories suitable and willing to receive research material. Given that a high proportion of archives used by earlier researchers had proved to be inadequate, a proper evaluation of each potentially suitable archive is essential. A programme of visits to key national archives took place during the first six months of the project, and one of our on-going activities is to liaise with new repositories which have special collecting priorities. Meeting with traditional archivists raises a number of interesting points about how these professionals view the acquisition and cataloguing of qualitative data collections, and about their relationships with traditional librarians. Although we did have a professional archivist on the team at the beginning, essential for gaining credibility with traditional archivists, we are now, primarily a team of social scientists who have adopted a cross-fertilised approach of data archiving and traditional archiving.

Repositories willing to accept qualitative deposits from Qualidata include:

- The Data Archive, University of Essex
- Renowned University archival repositories across Britain
  - British Library of Political and Economic Science, London School of Economics
  - The Modern Records Centre, University of Warwick
  - National Social Policy and Social Change Archive, University of Essex
  - British Library (Sound Archive and Manuscripts)
  - Specialist Institute Libraries
    - Institute of Criminology, University of Cambridge
    - Contemporary Medical Archives Centre, Wellcome Institute, London
    - British Universities Film and Video Council, London
  - National Museum Archives
  - Imperial War Museum, London
  - Labour History Archive, Manchester
  - Science Museum, London
Each repository specialises in a number of fields of research. Some had not acquired qualitative research data before, but were very keen to begin. Furthermore, some have now acquired valuable collections of qualitative social science data and wish to keep acquiring data from us in their particular areas of interest.

Evaluating qualitative data for archiving
Qualidata is a small unit: two full-time and two part-time senior staff; and four part-time processing officers. Masses of data are out there, and the suitability of data for archiving is assessed according to a set of criteria developed by Qualidata. Potential depositors are first invited to submit a sample of data, such as a transcript, to Qualidata, together with some documentation about the project. This includes the following requirements for datasets:

• Of a sufficiently qualitative nature
• In good physical condition, e.g. good quality recordings, abbreviations explained etc.
• Can be made freely accessible for academic use
• Perceived as having potential for secondary analysis
• Be able to fit in with existing collections
• Sufficient documentation to enable informed re-use
• Copyright, confidentiality and informed consent situation is satisfactory
• Resources needed to make material available do not outweigh potential for re-use (If the requirements of archiving are taken into consideration from the outset of a project, it is possible to keep extra work to a minimum. For example, ESRC applicants are now encouraged to include in their schedule and budget the necessary resources by which to prepare data for archiving)
• A suitable repository can be found (although if the materials are considered very high priority then Qualidata will house them temporarily).

Processing the data
The Centre undertakes processing work necessary both to ensure that data archived conform to legal and ethical guidelines, for example to abide by commitments of confidentiality given to research participants, and to achieve the greatest practicable accessibility and usability for the data.

Any acquiring organisation will know that some collections of data arrive in a very disorganised state whereas others will be immaculately filed, indexed and labelled. The amount of time and resources required to document material from a previous qualitative study very much depends on how old the material is and much there is. Qualidata does accept hand-written material, such as field notes, but where totally illegible, may need to be retyped. This is an expensive process and is only done in the most exceptional circumstances e.g. where the material is felt to be particularly valuable. We also encounter problems with audio-recordings without summaries or transcripts, as transcripts are almost always requested by researchers. In extreme cases, summaries may be carried out by Qualidata. Digitisation is also sometimes undertaken to give greater accessibility of datasets.

Preservation of confidentiality and informed consent in qualitative data
Since the archiving of qualitative data is fairly recent in terms of the history of social science, I would like to outline some of the procedures we have set up for safeguarding the anonymity of informants. The research community has long recognised the importance of respecting the rights of research participants. These rights take two principal forms: the right to have their identity protected (if so desired); and the right to make an informed decision about the uses made of the data that they provide. Personal information should be kept confidential, whether or not a pledge of confidentiality has been given to research participants, and should be stored in a secure manner according to the provisions of the UK Data Protection Act (1998).

Various professional and commercial organisations within the field of social science research have their own ethical guidelines and rules of conduct. Whilst some offer more detail with regards to issues like interviewing in difficult circumstances and preservation of anonymity, all present issues regarding the kind of ethical judgements researchers must make when embarking on a research project. The principal for preserving privacy, as articulated for example, in the British Sociological Association (BSA) statement, is that of the anonymisation of data. However, only one set of guidelines discusses issues relating to the sharing of research data.

Qualidata has undertaken considerable consultation within the research community, as well as liaising with potential depositors of data, concerning the issues of confidentiality and informed consent. These have undoubtedly been the most frequent causes of concern in the archiving of data. Qualidata has a deep concern both for the rights of participants and the professional integrity and peace-of-mind of researchers, and therefore both the issues of confidentiality and informed consent must be addressed in the context of archiving qualitative material. However, in many ways, adhering to guarantees of anonymity is always problematic. The very nature of qualitative data lends itself to descriptions of the interviewees, their lives and their surroundings, and in doing so, presents a dilemma to the researcher in how much to reveal. Is it really possible to completely disguise a workplace or a village or the central characters in the drama? I believe that future re-users of a qualitative dataset are presented with similar, if not the
same, issues as the first authors, concerning respecting the rights of participants.

We have produced information sheets relating to the issues of Confidentiality and Informed Consent and Confidentiality, Consent and Copyright in the Interviewing of Children, both available from the Centre upon request or via its WWW site. These information sheets describe the current legal and ethical situation and suggest solutions by which to respect the rights of participants. Of course, Qualidata recognises that some datasets cannot be ethically archived, particularly those that address sensitive issues.

The options used by Qualidata for preserving confidentiality, where appropriate are:

- **Anonymisation** of material is just one option available for helping make qualitative data accessible as a future research resource. It can include the removal of identifiers: the use of pseudonyms; and the use of other techniques for disguising the link between individual identifiers and data. It is, of course, important to arrive at an appropriate level of anonymisation to ensure that the data is not distorted to a degree, which devalues their potential for reuse.

- **A period of closure.** Where appropriate, a specified period of closure can be applied, although some archives are naturally resistant to accepting material that cannot be used for a long period of time. The saving grace for extremely sensitive materials is that time is of the essence. In 50 years time any tensions should have dissipated, and the information will become history.

- **Restricted access (operated by the archive).** Access to the data can be restricted to bona fide researchers for genuine research purposes.

- **Restricted access (operated by the depositor).** It is possible to make it a condition of deposit whereby all potential secondary researchers must liaise with the depositor to discuss their intentions for secondary analysis. The depositor may choose to only give access when satisfied that the data will be used in an appropriate manner in each case. Traditional archivists are well used to this approach.

- **User undertaking not to disseminate any identifying information.** Most archives operate user undertakings not to breach confidentiality by using identifiable information in published work. This condition is, of course, more effective if used in conjunction with restricting access to bona fide researchers. Such a written undertaking does have contractual force in law. Furthermore, the good reputation of a secondary user depends upon abiding by these undertakings.

- **Re-contacting participants.** It is possible for investigators to go back to research participants to obtain consent for deposit in a public archive, this being something with which Qualidata can sometimes assist. This is very time consuming but usually productive.

- **Gaining informed consent in writing** for material to be placed in an archive (at the time of fieldwork, but usually after an interview). Qualidata has a sample Informed Consent form, which is also available upon request. This also allows for transfer of copyright Depositors have absolute control in setting the terms and conditions for access. An agreement is then set up between the deposit and recipient repository to implement these terms and conditions. Secondary users given access to the data must be made aware of such terms and conditions, and should abide by them. In this respect, as data archivists, we place much emphasis on the responsibility of the secondary user.

**Why are qualitative researchers sceptical about sharing and re-using qualitative data?**

I would like to digress for a moment or two and consider why qualitative researchers show such scepticism towards archiving. This is simply because there has not been an established culture in social science for re-using someone else's qualitative data. Oral historians do use other sources, but this is because they are primarily social historians.

To establish why sociologists have not used colleagues' data, we must first recognise that qualitative researchers are a different breed from the ranks of the quantitative brigade. Some, but not all, see the concept of secondary analysis as purely about number crunching, and others feel very threatened by the idea of sharing or making data accountable. There are a number of reasons for this doubt and worry.

1. It is far more interesting to do your own fieldwork, even if it is extremely costly and possibly may be replicating previous studies of similar populations (at the expense of the taxpayer!)

2. Generally, qualitative social 'scientists' are just not used to making their findings accountable. They are worried about others seeing their data, and possibly picking holes in them. Some argue that certain approaches used in qualitative research, for example, grounded theory (Glaser and Strauss 1967) which opposes the scientific paradigm of testing hypotheses, do not lend themselves to verification.

3. Many researchers we have spoken to feel very strongly that, through fieldwork, they have established a special bond with their interviewees. Many also have promised informed consent at the time of interview
which precludes the use of the participants’ contributions for anything other than their own eyes or, at least, the current piece of research.

4. Some researchers are concerned that their material cannot be used sensibly without the accumulated background knowledge which they have acquired during its collection. This is particularly so with longitudinal studies of a group where the researcher feels that a special rapport has been developed without which the material may be meaningless. Thus the essential contextual experience of ‘being there’ cannot be shared.

We believe there is a solution to each of the negative points raised above:

1. To gain a more informed approach and to stop the proliferation of repetitive work, new studies should make more attempts to delve into earlier related research and try to include some comparative element. In order to be able to accomplish this, a firm bedding of archives across the UK needs to be cultivated on a regular basis and nurtured thereafter.

2. If we are to accept the label ‘scientist’, then we should adopt the scientific model of accountability, reliability and validity. The quality of social research is highly variable, and in the UK there are no quality control standards for qualitative studies (except for market research). We believe it is bad practice for raw data not to be available for future scholars and, furthermore, detrimental to the progress of history. As far as I am aware, it is unheard of for a social science journal to cite access to the original source of data, as is necessary in most natural scientific journals.

3. Interestingly enough, the complete protection of anonymity that researchers sometimes offer their participants is untenable - a first publication which a journalist then seizes upon may undermine this promise with a misguided stroke of a pen. In essence it is impossible to promise total anonymity. In contrast, we have found that when recontacting participants to gain permission for archiving, the majority seem to be in favour, even though this wasn’t mentioned at the time of fieldwork. Our experiences tell us that, providing their contribution is not abused, for example, their identifying characteristics are not cited (if they choose them not be), they are happy for serious scholars of the future to look at the raw materials. Most people do believe that research is for the public good, and that their contribution will be used in some way to create a better informed society, and even go some way towards implementing policy changes.

Contractual archival policies mean that investigators must now either rethink negotiations about informed consent and be prepared to discuss with their participants, at some stage, access to data beyond their own team.

4. The ‘being there counts’ argument is understandable but also an easy opt out of being prepared to share data. Indeed, there are instances where research data are, in a sense ‘re-used’, by the investigator themselves. For example, some principal investigators who write the final articles resulting from a project have employed research staff or a field force to collect the data. Similarly, for those working in research teams, sharing one’s own experiences of the research is essential. Both rely on the fieldworkers and co-workers documenting detailed notes about the project and communicating them to each other. Of course, audio and videotape recordings enhance the capacity to re-use data without having actually been there. For archives, documentation of the research process provides some degree of the context, and whilst it cannot compete with being there, field notes, letters and memos documenting the research can serve to help aid the original fieldwork experience.

What about the format of data?

We deal with all formats. Much qualitative data nowadays is digital in the sense that the text is word-processed or hand-written material is scanned, or audio-visual material is in digitally recorded form. Qualidata has developed standards for the documentation of qualitative digital data in liaison with the UK Data Archive. Generally materials are reduced to their simplest form, ASCII, TIFF4, but the Data Archive also accept Rich Text Format (RTF) and Adobe’s Portable Document Format (PDF).

We put digital data alongside paper-based materials in repositories or, where possible, offer it to the Data Archive at Essex. Data from mixed methods studies are usually offered first to the Data Archive, for example, so those in-depth interview transcripts sit alongside the statistical dataset. The Data Archive are experienced in handling, storing and disseminating textual data, and presently have the advantage over some traditional repositories in being able to keep up with changing media and storage technologies. However, for acquisition by the Data Archive, textual data must be, as far as possible, anonymous. Preservation of confidentiality is addressed below.

Far more qualitative researchers are now using digital data. The last three years have seen a huge growth in the use of computer-assisted qualitative data analysis software (CAQDAS) packages in qualitative research. CAQDAS software, such as NUDIST and ATLAS-ti, is rapidly becoming the accepted tool for handling the description and interpretation of qualitative data. For Qualidata issues about preservation of data from these packages is something we
have had to address with some urgency. These are proprietary software packages and in the past it has not been possible to import and export data from one package to another. Qualidata has developed guidelines on what to keep for archival purposes - i.e., reducing the data to its simplest form - ASCII text or RTF. As expected, in the past year we have seen software developers taking steps to encourage sharing between packages, for example adding export and import facilities to their programmes, and even beginning to build XML export features.

Digitisation - where are we going and what are we keeping?
The Data Archive in the UK archives primarily numerical, textual data: documentation for datasets is now stored in image format mostly in the form of PDF files; and more recently they have also begun to acquire image based datasets.

Qualidata is currently working on a large-scale digitisation project. This is the preservation of Professor George Brown’s life’s collection of research data. The major focus of the work is on the role of psychosocial factors in the onset, course and chronicity of, and recovery from, clinical depression (a major public health problem). The distinctive feature of George Brown’s approach has always been the ability to combine both qualitative and quantitative aspects of the same data. The publications resulting from Brown’s team reflect this duality in combining a host of statistical tables with a wealth of case history material. Thus the surveys above are all coded and the statistical data for each project will be archived with the Data Archive here at Essex.

Qualidata is image scanning the paper schedules, many of which contain a great deal of annotation in hand written form. The original TIFF files and a final PDF file for each case (patient) will still archived. PDF has been chosen by the Data Archive as a suitable archival format, as have many other institutions in Britain. However, we can never be sure whether this format may become extinct, and at the very least we would hope if it did, that conversion to the new formats would be an option. Perhaps we can allow ourselves to relax just a little, as we move into a climate of technological sharing and interoperability.

But when do we throw away the paper? A number of options come to mind, in no particular order:

- When our physical storage space is full up
- When we are confident we have a permanent representation
- When the paper starts degrading

In my own experience, thinking back to the forty four filing cabinets worth of George Brown’s data, I am terrified of getting rid of any of them! They are going to be available in electronic book form, and as safe as they could be in a prestigious Data Archive, but what if...? To avoid this panic and to appease our sense of sentimentality, our current strategy is to keep samples of original data, so for each project we will select about ten cases and these will be placed with a suitable academic (paper based) repository. If scholars still want to set eyes upon the original documents, they can!

Can traditional archives cope with digital non-numerical data?
Well, in short, some can and some can’t! Some of our host repositories have the facilities to provide copies of, say transcripts on disk, whereas others just can’t provide that service. This is usually simply a case of under resourcing. It is not uncommon in the traditional British archive world to see one, or at best two, archivists responsible for sorting, cataloguing, housing, and providing access to archives. This leaves little time for digitisation programmes and resources may not stretch to obtaining high-powered computers for storage. Reviews of electronic documents in personal papers and organised records held by archival repositories in Britain highlight problems of staffing, software, hardware, expertise and dissemination.

The other side of the picture, and of course, an ironic one, is the increasing lack of physical storage space for paper-based archives. Many archives are full up with paper documentation, and those with inadequate storage facilities are using hot or damp basements for storage. Microfilming and digitising saves on storage space, but does not necessarily represent a cheaper option: filming and scanning are expensive operations and the maintenance of electronic records in the long-term involves periodic transfers of data to new media and software. Technological changes - and the ever-reducing cost of computer storage - will undoubtedly mean that digitisation becomes a more attractive option over time, not least because it allows the records themselves to be disseminated electronically.

With the dawning of the Age of the Digital Library, and closer relationships being forged by academic libraries and archives with IT departments, and new centrally funded programmes, I don’t imagine archivists will turn away machine-readable versions of transcripts for much longer.

Problem areas for archiving qualitative data
Video recording and other image (such as photos), and to a lesser extent audio data, all present added difficulties for archiving and it is preferable that participants play a key role in the decision to archive.

Audio-tape recordings
Tape recordings of interviews are almost always used in qualitative studies. These may be individual interviews, focus groups, observation and naturally occurring conversation. For some projects, full transcription is
essential, for others summaries may suffice. Methods of transcription also vary: sociologists generally want to capture the words, whereas linguistics are more concerned with recording other contextual features of the interview, such as pauses, laughter, tears etc.

In terms of re-use potential of data, the ideal is to retain the original tape recordings. There is really no substitute for listening to people’s own words; a transcription is a subjective interpretation of the real-life conversation. In reality, it is often not possible to archive audiotapes where the material is ‘sensitive’, without either restricted access, a period of closure and/or retrospective permission from participants.

Anonymising tape recordings in the same way as for the transcripts is vastly time-consuming and prohibitively costly. Blanking out of identifying information on analogue media is also rather pointless as it distorts the data. Perhaps digital audio data may be less problematic. New software is now available for which researchers can edit, anonymise, label and copy their own data with far more ease. Again, this is still labour intensive and in the UK there is still no consensus about what the best audio format is for archival purposes. Current popular options are Minidisc, R-DAT and CD-R, but there is still no consensus on the relative longevity of these media.

The even more problematic case of video-data
Everything discussed with reference to audio data is worse for video data, with the added complexity of faces. We have not yet been able to archive much interview video data, as researchers have been very anxious about the possibility of identifying participants. There is no way around seeking permission to archive video data, and we are advising that permission is sought either before or after interview, depending on the sensitivity of the research and context of the interview setting. However, it is still evident, at least in the UK, that only a few branches of social science have taken on board the use of video methods: social anthropologists; socio-linguists and discourse analysts and educationalists.

So, should the traditional data archives acquire and store audio, video and multi-media data?
As technology moves forward many Data Archives across the world will have to begin considering the storage of digitised and indexed data from audio, video and multimedia data. We will see great improvements in storage options and indexing facilities for audio and video data. DVD is an exciting but volatile format and surely will replace audio and video CD. Since all Windows operating systems will be supporting it, it looks likely to dominate the market. Whilst it is still very expensive, inevitably costs will drop.

But I would like to pose the question: is it in the interest of the traditional social science data archives to take this route? For example, accepting and storing digitised audio and video of qualitative data creates serious issues regarding confidentiality and access, and also indexing. Whilst most Data Archives do not accept photos, audio or video-tapes, there are other specialist archives in Britain set up to receive and deal with these formats of data (although not with a social science remit). These have established standards and have dedicated working groups e.g. the Digital Archiving Working Group run by the BL, PRO and JISC, and Research Libraries Group. We are seeing guidelines emerging for the preservation on each and every kind of media.

New types of data clearly require specialist staff for evaluation, processing and documentation. The reason that the UK Data Archive is able to acquire textual and image qualitative material is that Qualidata acts as the front-line, engaging in evaluation, processing and documentation of these data. Thus the staff time and expertise to deal with qualitative data are not required of the Data Archive’s own personnel, who are busy enough with their own specialist roles. With this infrastructure in place, the Data Archive can provide access to a greater range of social science data.

An alternative model might be for the social science Data Archives’ to act in the role of brokers, where storage and access of social science data in say, audio and video formats, can be negotiated through Data Archives established systems, but not necessarily either processed or stored there.

There are now smaller embryonic "qualidatas" growing across Europe. However they are typically run by academics based in sociology departments, and usually have no links with their own country’s Data Archive Community. I am helping to build a network of these Centres and hope that the Data Archive Community will begin to take on board the contemporary and historical significance of qualitative data. To do this we all need to communicate and debate the issues I have addressed in this paper.

3. BS 7911 is the trademark for the standard adopted by the Market Research Society in 1988 for ‘Specification for organizations conducting market research’. This came about partly as a result of the hugely varying quality of qualitative studies in this arena.
4. The archive will include twelve collections, based on distinct projects dating from 1969 to the present. The earliest and probably best known study to many social scientists and clinicians is the Camberwell Study, conducted from 1969-75 and providing the basis for the eminent book, 'Social Origins of Depression', by Brown and Harris. The team pioneered the Life Events and Difficulties Schedule (LEDS), a survey instrument used to record stressful experiences and significant life events.


Louise Corti, Deputy Director and Manager, Qualidata, Department of Sociology, University of Essex, Colchester CO4 3SQ, UK. E-mail: cortl@essex.ac.uk Tel: +44 1206 873058 Url: www/essex.ac.uk/qualidata/
Computer Assisted Personal Interviewing: A Method of Capturing Sensitive Information

Abstract
This paper will discuss how Computer Assisted Personal Interviewing (CAPI) coped with collecting sensitive data in a difficult interview situation. A recent research project, funded by Scottish Homes (a UK Government Housing Agency for Scotland) used the CAPI technique to collect information on home ownership at the margins of affordability. The project used an innovative joint approach between the academic sector and a leading UK survey consultancy. It could be argued that a more sensitive method of collecting this sort of information would have been in-depth interviews, which could then have been analysed using qualitative research methods. The paper will discuss the outcomes of using CAPI and quantitative research methods in such a sensitive project.

by Emma Forster and Alison McCleery*

It is suggested that the use of CAPI has achieved a better response rate on sensitive questions than other techniques would have. The use of CAPI has a number of well-known advantages, such as improvements in data quality and turnaround times. This paper will assess whether CAPI can deliver in a number of interview conditions or if its potential benefits will be realised only under certain conditions. It will critically review how the quantitative method worked in this specific situation before placing the discussion in its wider research methodology and research environment context.

Introduction
This paper firstly will review the literature with a view to describing the current state of the art of CAI, secondly it will describe a Scottish housing project which used CAPI and consider the quality of data output, and thirdly, it will draw conclusions about the use of CA(P)I on the basis of the authors’ own findings but also placing the discussion in its wider research methodology and research environment context.

The questions which this paper seeks to explore are broadly:

* Is CAPI suitable only for large projects?
* Is CAPI good with sensitive data?
* Does use of CAPI improve data quality?
* Can CAPI substitute for qualitative research in a contract research environment?

These types of questions have not been asked before as, arguably, studies on data quality have been too restricted. Furthermore, it is suggested that the potential of computer-assisted data collection methods has not been fully utilised. This paper explores whether quantitative research could be regarded as a universal solution.

Part 1: Literature: Previous users of CAPI
Compared to even a few years ago, Computer Aided Interviewing (CAI) is now relatively widespread and mature. A move to CAPI can, for example, lead to improvements in data quality and turnaround times; it can even make possible surveys that would not otherwise be contemplated. For these and other reasons, many survey organisations and clients have been persuaded that CAI is where the future of survey research lies.

Indeed for almost every traditional approach to survey data collection there is now a computer assisted alternative. The two most widespread are Computer Assisted Telephone Interviewing (CATI) and Computer Assisted Personal Interviewing (CAPI) (Collins & Sykes, 1998). Just because CAPI/CAI is so widespread these days, there is no longer much research

IASSIST Quarterly
which compares them as modes to the more traditional modes. Instead the focus of the literature has shifted to improving CAI, as for instance Bulmer et al. (1998), away from questioning the intrinsic comparative value of the mode itself. However, due to the shift which has occurred in the way survey data are collected with telephone surveys and, to a lesser degree, mail surveys now being more extensively used, this has stimulated a limited amount of empirical research on the influence of the data collection method on data quality. De Leeuw et al. (1996) compare a mail, a telephone, and a face-to-face survey and found that the different data collection methods did have an effect. There is less research however on the question of whether to adopt CAI but more on what effect it has and how this might be mediated.

Disadvantages of CAPI and CAI in general

There is conflicting evidence on the impact of CAPI on data quality. Collins & Sykes (1998) find little evidence of clear improvement as yet. Furthermore, despite general acceptance of CAI in the literature, it is recognised that paper and pen interviewing (PAPI) still has its role. The promise of CAI will be realised only under certain conditions and CAPI is not a panacea.

Face to face interviews and CAPI as a method of data collection also have weaknesses associated with their usage. For instance Blyth (1998) lists the hours needed to be worked by the interviewers; falling response rates obtained by this method; and personal safety considerations amongst others, all factors that would militate against the increased use of this method. To this list Blyth (1998) adds equipment cost and software/hardware/data interaction as well as the need for batteries, as major issues to be considered in CAPI.

Although, as will be seen later, there are ways round the capital investment and overheads barriers to the use of CAI, nevertheless these types of hardware and software issues associated with CAPI must be fully taken on board before the decision is taken to use CAPI as a data collection method. Finally, the quality and accessibility of the output data from CAI has been questioned. This is because CAI does not eliminate human error, which where CAPI is concerned, simply manifests itself in a different way as compared with PAPI. Routing mistakes may mean that whole questions are missed in every interview, whereas in PAPI the interviewer may inadvertently turn over two pages together and may miss a whole page but not in every case or at least not the same error in every case i.e. PAPI is likely to be associated with a series of random inconsistent errors, while CAPI is likely to be associated with a single consistent serious error affecting every interview schedule. However, it is suggested that the latter is highly responsive to improvement as a result of piloting the questionnaire and training of interviewers.

Currently one of the disadvantages of CAPI is that the non-technical reader finds the content, structure and workings of a CASIC questionnaire much more difficult to understand (Bulmer et al., 1998). The growing possibilities of computer hardware and software associated with technological advance 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. Various attempts have been made to render these CAI questionnaires intelligible to non-specialists. Recently a more comprehensive attack on the documentation problem was launched. This project has been named TADEQ (A Tool for Analysing and Documenting Electronic Questionnaires), Manners & Bethlehem, 1999. The TADEQ project proposes to develop a flexible tool for documenting and analysing electronic questionnaires. This tool aims to be neutral so it will be possible to use it in combination with existing computer assisted interviewing (CAI) systems. As a documentation tool, it must be able to produce a human-readable presentation of the electronic questionnaire. TADEQ will produce two types of output: a paper version of the documentation, which can be used e.g. by either interviewers or managers; and an electronic version of the documentation, in some kind of hypertext format, allowing developers or researchers to scrutinise the contents and structure of the questionnaire.

This development poses a conundrum for data archivists: should they adjust their systems to include the tool to convert CAI questionnaires into this standard format once they have received them, or should they get data depositors to do this before they submit their questionnaire? The result of such a tool is that it produces standardisation of survey documentation and so is advantageous for data archivists in the long run.

Advantages of CAPI

Dent (1999) does recognise that the well-known advantages of technology are that they provide an improved research quality, a quicker turnaround and easier integration with other activities. For instance management reporting and marketing approaches. He further states that the main reasons for the use of CAI include managing complexity of data, of samples and of reporting and maintaining records and files. These have been comprehensively illustrated elsewhere. Bulmer et al. (1998) in their finding that the introduction of computer assisted survey information collection (CASIC) has made technically
feasible a much higher level of questionnaire complexity than was possible in paper and pencil days. This confirms it is literally the case that much of today’s more ‘serious’ work could not be otherwise achieved without CAI.

De Leeuw et al. (1995) reviews the evidence of the effect of computer assisted interviewing on data quality and find that there are clear advantages of CAI in two main areas: survey data quality; and acceptance of the computer by respondents and interviewers. Their main conclusions are that computer-assisted data collection methods are accepted by both respondents and interviewers, and that survey data quality improves, especially when complex questionnaires are used. By 1999, in a discussion of the widespread market penetration of CATI and to a lesser extent CAPI, De Leeuw has added lower costs to the previously identified advantage of improved data quality.

Although a study by Hox & De Leeuw (1994) shows that response to mail surveys has been improving recently, nevertheless face-to-face surveys continue to achieve the highest response rates. Given Dillman et al.’s (1993) finding that asking potentially difficult and/or objectionable questions lowers the response rate, a CAPI approach which combines good results on sensitive questions with the traditional high response rate of face-to-face interviews should now be the method of choice. Furthermore, the use of technology gives a professional image, and CAPI is generally liked by respondents, while it intrigues many elderly respondents. Research should therefore concentrate on further reducing human error associated with CAPI. This present paper offers a step along the way to producing more research on face-to-face methods, producing additional research into non-response in face-to-face surveys.

Already a major advantage of CAPI is the way in which it is able to reduce potential interviewer and respondent error. Routing errors are eliminated because the script automatically routes to the correct questions. This ensures that data are generally more complete, can considerably reduce the number of ‘non-responses’ and, correspondingly, the need for corrective editing (or even re-contacting respondents) later on. CAPI also has the ability to ‘range-check’ data and carry out logic and consistency checks during the interview. Either ‘hard’ or ‘soft’ checks are set, to query or confirm key pieces of data with the respondent. In this project for Scottish Homes CAPI was used to check financial data during the interview and clarify whether figures relate to pounds or pence. Ranges were set for key pieces of financial data and discrepancies queried with the respondent if they fell outside certain ranges. Although we did not use this facility in this study, CAPI can also be used to calculate and provide derived variables during the course of the interview (which can then be fed back to the respondent or queried with them, as appropriate).

Returning to the matter of sensitive questions, there is limited evidence that CAPI is better than PAPI for sensitive questions. Few studies have looked at this. Those that have compared the same questions and different collection modes have tended to compare between CAI methods and rather than between PAPI and CAPI.

When looking at various modes of survey data collection which had to ask difficult or sensitive questions, it was found by Tourangeau & Smith (1996) that the mode of data collection did indeed affect the level of reporting of sensitive behaviours. This study compared three methods of collecting survey data about sexual behaviours and other sensitive topics: computer-assisted personal interviewing (CAI), computer-assisted self-administered interviewing (CASI), and audio computer-assisted self-administered interviewing (ACASI). It was found that the three mode groups did not differ in response rates, but both forms of self-administration tended to reduce the disparity between men and women in the number of sex partners reported. Self-administration, especially via ACASI, also increased the proportion of respondents admitting to the use of illicit drugs. This study also highlighted the importance of the closed answer options in determining the response. Thus it is suggested that open answers, although time-consuming to code, may produce more unbiased answers in answering difficult or sensitive questions.

Other evidence (De Leeuw, 1999) points to that perceived confidentially playing a role in obtaining higher response rates. Earlier research by De Leeuw et al. (1995) concluded that this is an under-researched area. The authors’ own survey, while it does not give definitive evidence as it does not use a comparison of methods, nevertheless does lend weight to the argument that use of CAPI gets a good response rate on sensitive questions. De Leeuw (1999) established from her review that there is a greater willingness to report extreme views using CAPI.

MORI Scotland have found, when comparing data from paper based and CAPI surveys, that the problem of non-response to specific questions is significantly reduced, and that some sensitive information is collected more fully (such as household income) with CAPI. For example, in transferring the national MORI Omnibus survey from paper to CAPI, it was found that the proportion agreeing in principle to being re-contacted rose from 77.4% to 80.8%, and the proportion refusing to declare a household income declined from 16.8% to 14.1%. However, there is no evidence that, overall, using CAPI has a significant
effect one way or another on respondent willingness to participate in surveys.

What is, however, certain is that keying errors associated with data entry are very much reduced since data entry is done once, during the interview, rather than by coding onto paper and subsequently transferring responses to computer. Even with 100% verification, errors of 0.05% on some variables may be expected where one is interpreting hand-written numbers; experience to date suggests that fewer errors are associated with CAPI.

**Summary of advantages and disadvantages of CAPI**

Thus it appears from the literature that CAPI is good for capturing sensitive data, has a fast turnaround time (good for the short deadlines that are associated with contract research) and generally improves data quality (although there is conflicting evidence on this). One of the main disadvantages is cost: the hardware and software is expensive and then there is the interviewer training over and above. This is likely to limit the use of CAPI to commercial survey companies with a large throughput which can afford overheads by spreading the capital investment.

Furthermore, it is worth stressing at this point that CAPI and more generally CAI do not in themselves address many of the common problems of collecting survey data for the quantitative studies and furthermore cannot pretend to emulate the painstaking detail of qualitative interviewing. Nevertheless, it is also, paradoxically, time to say that the conclusion reached from the available literature is that many projects/surveys could not be done with such accuracy, and some not done at all, without CAI. In particular those surveys which require complex routing but need to be carried out in a fairly short interview time.

**Part 2: Project description and results**

*Brief description of project*

The purpose of this research was to:

1. develop an understanding of the issues concerning poorer owner occupiers, and identify what greater role information could play in attaining more successful housing outcomes;

2. undertake a detailed case study of owners in South Rogerfield on Glasgow’s eastern periphery. Specifically the aims of the case study were to:

   * analyse the factors, both immediate and contextual, which influence the decision to buy;
   * produce an evaluation of the retrospective understanding of the owner’s responsibility in respect of common repairs;
   * determine what information had been used in the process of buying, and of that identify information deemed to have been useful and information considered misleading or unhelpful at each stage in the decision;
   * assess the role that information or the lack of it played during any difficulties regarding repair bills, financial difficulties, redundancy; and
   * identify examples of best practice where provision of information helped avoid common pitfalls and where lack of information obscured common pitfalls which happen to those buying on limited budgets.

3. Catalogue the available information for homebuyers before and after they buy their home.

The actual results of the project are reported elsewhere in Forster and McCleery (1999a, 1999b).

South Rogerfield was identified as the survey area by Scottish Homes on account of the high number of repossessions: between 1987 and 1993 49 houses in the estate were repossessed by lenders: this is equivalent to 20% of the total number of properties. Together South and North Rogerfield form one of the fourteen neighbourhoods which make up the sprawling peripheral housing estate of Easterhouse on Glasgow’s north-eastern perimeter, as shown on location maps 1 and 2. Comprising circa 300 housing units, South Rogerfield consists of 3-storey tenement blocks built in the late 1950s and arranged in either rows or quadrants, predominantly with continuous frontages. On the instructions of Glasgow District Council, the properties were sold off and improved in about 1985 by two firms of property developers, Crudens and Barratt, although twenty or so properties pepper-potted throughout are still council-owned, while six are in housing association shared ownership. The modernisation carried out by the developers included storey height reduction, new roofs, double glazing, new bathrooms and central heating.
Map 1: Location of survey area within Scotland

Map 2: Location of Easterhouse estate, Glasgow, Scotland

Establishment of the sample was not straightforward, but the eventual response rate was very satisfactory for an area of this type. After initial confusion as to how many flats there were in South Rogerfield, 138 full interviews with owner-occupiers were obtained. However, 48 more households, all occupying rented housing, were asked a much more limited list of questions.

If the 41 sheltered and vacant properties are subtracted from the total count, the number of ‘valid and occupied’ flats comes to 308. From these 186 interviews were obtained, 138 with owners and 48 with renters, giving a total response rate of 60%. A fuller breakdown of response types is provided in Table 1.

<table>
<thead>
<tr>
<th>Table 1: Response rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total addresses identified</td>
</tr>
<tr>
<td>Sheltered or vacant</td>
</tr>
<tr>
<td>‘Valid and occupied’ (Used as base)</td>
</tr>
<tr>
<td>Interview with owner</td>
</tr>
<tr>
<td>Interview with tenant</td>
</tr>
<tr>
<td>Overall response rate</td>
</tr>
<tr>
<td>Refusal (including not interested and too busy)</td>
</tr>
<tr>
<td>Too ill to take part</td>
</tr>
<tr>
<td>Insufficient English</td>
</tr>
<tr>
<td>No contact after 4 calls</td>
</tr>
</tbody>
</table>

The data from the owner-occupiers was collected by MORI Scotland using Computer Aided Personal Interviewing (CAPI).

Choice of method
This section begins by defending the choice of CAPI as the survey method. Thereafter the experience of using CAPI is described, prior to a discussion of the extent to which the reality of using the method matched expectations.

First the choice was made between qualitative and quantitative. Sensitive questions are often dealt with in in-depth interviews - qualitative methodologies. The following section explains why qualitative was considered unsuitable for us in this project. A method was needed that produced a quick turnaround. The whole project ran only for 3 months and the data collection phase was allocated only 3.5 weeks, with a draft of the survey results needed shortly after the data collection was finished. Qualitative interviews would have surely discovered more in-depth information on the topic in question but could not have been collected nor analysed in the time allowed. Quantitative method was chosen primarily because of the time-scale. Due to the constraints of research work in a research contract environment, the future of qualitative research in that setting is questionable. Increasingly today, the nature of contract research demands quantitative data which can be analysed statistically and held in data archives for future comparability.

Moving on to discuss the specific choice of CAPI. CAPI has a good response rate on sensitive questions (Tourangeau & Smith, 1996; De Leeuw, 1999). Yet sensitive data is in the past thought to be better collected in a self-completion method. Yet, CASI has a lower response rate than face-to-face, and has a slower turnaround time. As is seen above in the literature review, the fast turnaround time is one of CAPI’s main advantages.

Our expectations in choosing CAPI for this survey were that:

* It would catch the client’s imagination to win the contract in the first place
* CAPI would help achieve first time round a 60-70% response rate needed in connection with the blanket coverage in a small area, (it would not be possible to re-sample to improve the response rate - therefore it needed to be possible to get a fairly high response rate straight off.)
* Fast turnaround time, in particular it would help speed up the stage between the end of the survey and obtaining the data
* would cut down on data input errors
* Allow complex routing through debt, income, mortgage/endowment payments and repair sections yet it was envisaged that each interview should not last more than 30 minutes.

In summarising the aims of chosen survey method, the nature of the survey demanded a high response rate. The area under study and therefore size were very small and 100% blanket coverage sought. Furthermore, an important objective of the study was to collect as much data as possible on sensitive questions relating to income, mortgage and other house-related payments, debt etc.

**Choice of joint approach**
A collaboration was chosen between a survey company and academic institution, primarily because it was felt that this difficult interview situation asking sensitive questions would need very highly trained and very experienced interviewers. The tight deadlines necessitated this approach, as there was no time for training of interviewers. Not only professionalism and timesaving but cost was an important factor too. This strategic approach avoided the initial expensive investment required in hardware and software i.e. the up-front costs mentioned earlier which are problematical for resource-lean UK academic institutions.

**Outcomes of choice of CAPI**
In fact, our expectations of use of CAPI as a data collection method were exceeded as we got better response rates, and lower refusal rates than we expected. CAPI proved successful in terms of data quality and a high response rate on sensitive question. Our survey differs from the common experience that asking potentially difficult and/or objectionable questions lowers the response rate as found by Dillman et al. (1993).

The refusals to answer sensitive questions did not rise above 6.5%. The question that had the biggest refusal rate was the income question with 9 households refusing to give this information. Income was asked only in bands so as to improve the response rate and this seems to have worked as there was a high level of response to this question.

An examination of the pattern and number of refusals reveals that it was the same few households that account for most of the refusals. A full list of the questions where refusals occurred is in appendix 1 and the breakdown of the number of refusals is found in appendix 3. A further examination was made of the characteristics of these people and it was found that they did not fall into one demographic or economic group but were scattered between these. Full results of the testing of characteristics of the people who refused to answer some of the questions are to be found in appendix 2. Overall, it was shown that there were very limited refusals in the income and debt questions.

So, in summary, the total refusal rate remained low throughout the survey. Possible reasons for this include professionalism of the highly trained interviewers, the perceived higher confidentiality that CAPI gives the respondents. However, there was not a question on the interviewees attitude to CAPI in the questionnaire and so it is only possible to speculate on this.

While the fairly high response rate overall (Table 1) was pleasing for the researchers, it was not on the whole unexpected and was in conjunction with the fast turnaround time the reason that this method was chosen. However, what was thought to be unusual and not mentioned elsewhere in the literature was that the survey was a fairly small-scale one. Normally CAPI is presented only as cost-effective in large-scale data surveys. Also unexpected was the high response rate to the sensitive questions.

**Lessons to learn/Unavoidable problems**
As the literature made clear, CAPI is not foolproof and is only as good as the interviewers who administer it and the programmers who set up the routing. In this particular survey two questions were completely missed from every interview due to a fault in the routing. Because of the nature of contract research, with limited budgets and tight time-scales there was no pilot of this survey and only limited testing of the routing before the interview schedule went into the field.

CAPI can arguably improve on PAPI by improving turnaround without any loss of data quality or even an improvement in data quality, in particular with sensitive questions. But for reasons stated earlier use of CAPI has to date been favoured for large studies only. Blyth predicted in 1998 that in the future CAPI would be most relevant for big international players and for very large surveys. He felt that non-CAPI will gravitate more quickly to telephone and that this will result in only a small number of field-only CAPI companies. However, our survey offers a less limited future for CAPI in that it proved successful for a smallish study in which an academic institution sub-contracted the data collection to a commercial concern.
Conclusions
In summary:

* CAPI not panacea (although better than PAPI)
* Not a replacement for qualitative methods
* Appropriate for certain data in certain contexts e.g. can achieve good results with sensitive data
* Further research into how even better advancements in CAPI is now appropriate
* Involves high-level of investment - up-front costs - and so is not for everyone - although strategic studies to this problem are possible (as we have shown).
* Previously better for large-scale studies, but especially when handling sensitive data, possible for small-scale studies too (as we have shown).

CAPI was the most suitable method for use in this contract research project primarily because it gets data quickly. However, additional advantages also emerged. The surprising finding was CAPI was useful and cost-effective in small-scale data collection situations too. The combination between academics and market research survey company worked well. The resultant high response rate and low refusal rates commends this method for use in other similar interview situations. But of course, without conducting the same study again using PAPI, it is impossible to conclusively say whether it was CAPI alone that led to high response rate and low refusal rate. Blyth (1998) points out that with increasingly fragmented populations and busier lifestyles we need to use a mix of technologies to obtain maximum response from our survey population. As Blyth (1998) rightly points out the focus should be on the answers and not the media, however, it is important to consider media in the context of which media gives the most (highest quality) answers. It is also necessary to adjust the media according to the research environment context, although in doing so we must not lose sight of issues of research philosophy and quality.

Acknowledgements
The authors would like to thank Scottish Homes, the funders of this research work ‘Housing Information and Advice and Home Ownership at the Margins’, and also MORI, Scotland who carried out the survey work and to Simon Braunholtz in particular.

Glossary

Computer Assisted Personal Interviewing                  CAPI
Hand-Held Assisted Personal Interviewing                 CATI
Computer Assisted Telephone Interviewing                 CAI
Computer Assisted Self Interviewing                      CADI
Computer Assisted Interviewing                           PAPI
Computer Assisted Data Input                             CADAC
Computer-Assisted Data Collection Methods                ACASI
Audio Computer-Assisted Self-Administered Interviewing  CASIC
Computer Assisted Survey Information Collection          ASCAIC

Appendix 1: The questions where refusals occurred:

QE1. What was the purchase price of this property?
QE3. And roughly how much do you think you could sell this property for now, if you put it on the market?
QE4. From which of the sources on this card did your household get the money to buy this property?
QE8. Apart from money to move in, how much have you borrowed since you moved into the flat for costs to do with the house - for instance repairs, improvements, cookers/fridges etc.?
QE13. At the moment, how much does your household pay each month in mortgage or loan payments?
QE14. How much, if anything, does your household pay in the additional separate endowment part of the mortgage each month?
QF3. How easy or difficult is it for your household to pay the mortgage payments?
QF4. Have you been more than two months behind with your mortgage payments at any time in the past two years? Please look at this card and tell me the letter next to the band in which you would place your total household income for the year from employment or benefits.
QSV. At the moment do you (or your partner) have any money saved or invested?
QSV2. SHOWCARDS: How much do you (and your partner) have saved together?
Please tell me the letter on this card for the group in which you would place your total savings?
Appendix 2: Who refused to answer?

In total 33 refusals to 11 questions by 14 households.

Case number 113 refused 7 questions
Case number 41 refused 5 questions
Case number 72 refused 4 questions
Case number 94 refused 3 questions
Case number 26 refused 2 questions
Case number 55 refused 2 questions
Case number 99 refused 2 questions
Case number 130 refused 2 questions
Case number 68 refused 1 question
Case number 69 refused 1 question
Case number 70 refused 1 question
Case number 100 refused 1 question
Case number 109 refused 1 question
Case number 123 refused 1 question

| Household occupations of respondent households who refused to answer any questions. |
|-------------------------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|       |
| Part time and full-time worker                  | Housewife and full-time worker | Two retired people | Four full time workers | Three full time workers | One full time worker | One full-time worker and one LT sick | One part-time worker only | 2 full-time, 1 part-time and 1 student | Total |
|------------------------------------------------|-------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-------|
| **Total**                                      | **3**                         | **3**           | **1**           | **1**           | **1**           | **2**           | **1**           | **1**           | **1**           | **14** |

<table>
<thead>
<tr>
<th>Household sizes of respondent households who refused to answer any question</th>
</tr>
</thead>
<tbody>
<tr>
<td>HH Size</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td><strong>Total</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of children in respondent households who refused to answer any question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of children</td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td><strong>Total</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Household types of respondent households who refused to answer any question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Household type</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td><strong>Total</strong></td>
</tr>
</tbody>
</table>
### Appendix 3: The number of refusals

#### QE1 What was the purchase price of this property?

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Refused</td>
<td>2</td>
<td>1.4</td>
<td>66.7</td>
<td>66.7</td>
</tr>
<tr>
<td>Don’t know</td>
<td>1</td>
<td>.7</td>
<td>33.3</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>3</td>
<td>2.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>135</td>
<td>97.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>138</td>
<td>100.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### QE3 how much do you think you could sell this property for now?

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Refused</td>
<td>1</td>
<td>.7</td>
<td>9.1</td>
<td>9.1</td>
</tr>
<tr>
<td>Don’t know</td>
<td>10</td>
<td>7.2</td>
<td>90.9</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td>8.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>127</td>
<td>92.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>138</td>
<td>100.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### QE4 From which of the sources on this card did your household get the money to buy this property?

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>137</td>
<td>99.3</td>
<td>99.3</td>
<td>99.3</td>
</tr>
<tr>
<td>Refused</td>
<td>1</td>
<td>.7</td>
<td>.7</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>138</td>
<td>100.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### QE8 Apart from money to move in, how much have you borrowed since you moved into the flat for costs to do with the house

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Refused</td>
<td>2</td>
<td>1.4</td>
<td>1.7</td>
<td>1.7</td>
</tr>
<tr>
<td>Don’t know</td>
<td>4</td>
<td>2.9</td>
<td>3.4</td>
<td>5.0</td>
</tr>
<tr>
<td>None</td>
<td>113</td>
<td>81.9</td>
<td>95.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>119</td>
<td>86.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>19</td>
<td>13.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>138</td>
<td>100.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### QE13 how much does your household pay each month in mortgage or loan payments

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Refused</td>
<td>2</td>
<td>1.4</td>
<td>33.3</td>
<td>33.3</td>
</tr>
<tr>
<td>Don’t know</td>
<td>4</td>
<td>2.9</td>
<td>66.7</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>6</td>
<td>4.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>132</td>
<td>95.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>138</td>
<td>100.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### QE14 How much, if anything, does your household pay in additional endowment

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Refused</td>
<td>2</td>
<td>1.4</td>
<td>5.4</td>
<td>5.4</td>
</tr>
<tr>
<td>Don’t know</td>
<td>4</td>
<td>2.9</td>
<td>10.8</td>
<td>16.2</td>
</tr>
<tr>
<td>Nothing, it is included in figure quoted</td>
<td>31</td>
<td>22.5</td>
<td>83.8</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>37</td>
<td>26.8</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>101</td>
<td>73.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>138</td>
<td>100.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### QF3 How easy or difficult is it for your household to pay mortgage payments

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy</td>
<td>109</td>
<td>79.0</td>
<td>79.0</td>
<td>79.0</td>
</tr>
<tr>
<td>Sometimes have difficulties</td>
<td>23</td>
<td>16.7</td>
<td>16.7</td>
<td>95.7</td>
</tr>
<tr>
<td>Often have difficulties</td>
<td>3</td>
<td>2.2</td>
<td>2.2</td>
<td>97.8</td>
</tr>
<tr>
<td>Always have difficulties</td>
<td>1</td>
<td>.7</td>
<td>.7</td>
<td>98.6</td>
</tr>
<tr>
<td>Don’t know</td>
<td>1</td>
<td>.7</td>
<td>.7</td>
<td>99.3</td>
</tr>
<tr>
<td>Refused</td>
<td>1</td>
<td>.7</td>
<td>.7</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>138</td>
<td>100.0</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

### QF4 Have you been more than two months behind with your mortgage payments at any time in the past 2 years

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>10</td>
<td>7.2</td>
<td>7.2</td>
<td>7.2</td>
</tr>
<tr>
<td>No</td>
<td>126</td>
<td>91.3</td>
<td>91.3</td>
<td>98.6</td>
</tr>
<tr>
<td>Refused</td>
<td>1</td>
<td>.7</td>
<td>.7</td>
<td>99.3</td>
</tr>
<tr>
<td>Don’t know</td>
<td>1</td>
<td>.7</td>
<td>.7</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>138</td>
<td>100.0</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

### Income Group

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Under £5,000</td>
<td>3</td>
<td>2.2</td>
<td>2.2</td>
<td>2.2</td>
</tr>
<tr>
<td>B £5,000-£10,000</td>
<td>11</td>
<td>8.0</td>
<td>8.0</td>
<td>10.1</td>
</tr>
<tr>
<td>C Over £10,000-£15,000</td>
<td>36</td>
<td>26.1</td>
<td>26.1</td>
<td>36.2</td>
</tr>
<tr>
<td>D Over £15,000-£20,000</td>
<td>35</td>
<td>25.4</td>
<td>25.4</td>
<td>61.6</td>
</tr>
<tr>
<td>E Over £20,000-£25,000</td>
<td>19</td>
<td>13.8</td>
<td>13.8</td>
<td>75.4</td>
</tr>
<tr>
<td>F Over £25,000-£30,000</td>
<td>9</td>
<td>6.5</td>
<td>6.5</td>
<td>81.9</td>
</tr>
<tr>
<td>G Over £30,000</td>
<td>9</td>
<td>6.5</td>
<td>6.5</td>
<td>88.4</td>
</tr>
<tr>
<td>Don’t know</td>
<td>7</td>
<td>5.1</td>
<td>5.1</td>
<td>93.5</td>
</tr>
<tr>
<td>Refused</td>
<td>9</td>
<td>6.5</td>
<td>6.5</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>138</td>
<td>100.0</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>
### QSV Have you (or your partner) savings?

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>42</td>
<td>30.4</td>
<td>30.4</td>
<td>30.4</td>
</tr>
<tr>
<td>No</td>
<td>88</td>
<td>63.8</td>
<td>63.8</td>
<td>94.2</td>
</tr>
<tr>
<td>Refused/unsure</td>
<td>8</td>
<td>5.8</td>
<td>5.8</td>
<td>100.0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>138</td>
<td>100.0</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

### QSV2 How much do you (and your partner) have saved?

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Under £1,000</td>
<td>8</td>
<td>5.8</td>
<td>19.0</td>
<td>19.0</td>
</tr>
<tr>
<td>B £1,000-£2,999</td>
<td>10</td>
<td>7.2</td>
<td>23.8</td>
<td>42.9</td>
</tr>
<tr>
<td>C £3,000-£4,999</td>
<td>8</td>
<td>5.8</td>
<td>19.0</td>
<td>61.9</td>
</tr>
<tr>
<td>D £5,000-£9,999</td>
<td>1</td>
<td>0.7</td>
<td>2.4</td>
<td>64.3</td>
</tr>
<tr>
<td>E £10,000-£16,000</td>
<td>3</td>
<td>2.2</td>
<td>7.1</td>
<td>71.4</td>
</tr>
<tr>
<td>F Over £16,000</td>
<td>6</td>
<td>4.3</td>
<td>14.3</td>
<td>85.7</td>
</tr>
<tr>
<td>Don’t know</td>
<td>2</td>
<td>1.4</td>
<td>4.8</td>
<td>90.5</td>
</tr>
<tr>
<td>Refused</td>
<td>4</td>
<td>2.9</td>
<td>9.5</td>
<td>100.0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>42</td>
<td>30.4</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>96</td>
<td>69.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>138</td>
<td>100.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### References


1 The TADEQ project is funded under the European Commission’s Esprit programme. 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, Saarbrucken, Germany. More information on the TADEQ project can be found at: http://www.blaiseusers.org/tadeq/abouttdq.htm

2 The survey was both of renters and owner-occupiers but only the owner-occupiers were asked a full interview using CAPI. The responses of the owner-occupiers only will be dealt with in this paper.

Emma Forster and Alison McCleery, Dept of Psychology and Sociology, Napier University, Redwood House, 66 Spylaw Road, Edinburgh, EH10 5BR, Scotland, U.K. Telephone: +44 131 455 5139 Fax: +44 131 455 5141 e-mail: e.forster@napier.ac.uk

DATA IN THE DIGITAL LIBRARY: Charting the Future for Social, Spatial and Government Data

June 7-10, 2000
Northwestern University

The Twenty-Sixth (26) Annual Conference of the International Association for Social Science Information Services and Technology (IASSIST) will be held on the campus of Northwestern University in Evanston, Illinois on June 7-10, 2000. This year’s conference Data in the Digital Library: Charting the Future of Social, Spatial and Government Data emphasizes the strengthening relationships between archives and libraries in managing, preserving and providing access to "digital collections".

IASSIST conferences bring together data professionals, data producers, and data analysts from around the world who are engaged in the creation, acquisition, processing, maintenance, distribution, preservation, and use of numeric social science data for research and instruction.

http://www.spc.uchicago.edu/DATA LIB /ia2000/
The International Association for Social Science Information Services and Technology (IASSIST) is an international association of individuals who are engaged in the acquisition, processing, maintenance, and distribution of machine readable text and/or numeric social science data. The membership includes information system specialists, database librarians or administrators, archivists, researchers, programmers, and managers. Their range of interests encompasses hard copy as well as machine readable data.

Paid-up members enjoy voting rights and receive the IASSIST QUARTERLY. They also benefit from reduced fees for attendance at regional and international conferences sponsored by IASSIST.

Membership fees are:
- Regular Membership: $40.00 per calendar year.
- Student Membership: $20.00 per calendar year.

Institutional subscriptions to the quarterly are available, but do not confer voting rights or other membership benefits.

- Institutional Subscription: $70.00 per calendar year (includes one volume of the Quarterly)

I would like to become a member of IASSIST. Please see my choice below:
Options for payment in Canadian Dollars and by Major Credit Card are available. See the following web site for details:
http://datalib.library.ualberta.ca/iassist/mbrship2.html
- $40 (US) Regular Member
- $20 Student Member
- $70 Subscription (payment must be made in USS)
- List me in the membership directory
- Add me to the IASSIST listserv

Please make checks payable, in US funds, to IASSIST and Mail to:

IASSIST,
Assistant Treasurer
JoAnn Dionne
50360 Warren Road
Canton, MI 48187
USA

Name:
Job Title:
Organization:
Address:

City: State/Province:
Postal Code: Country:
Phone: FAX:
E-mail: URL: