

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

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.

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

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, key word, 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.

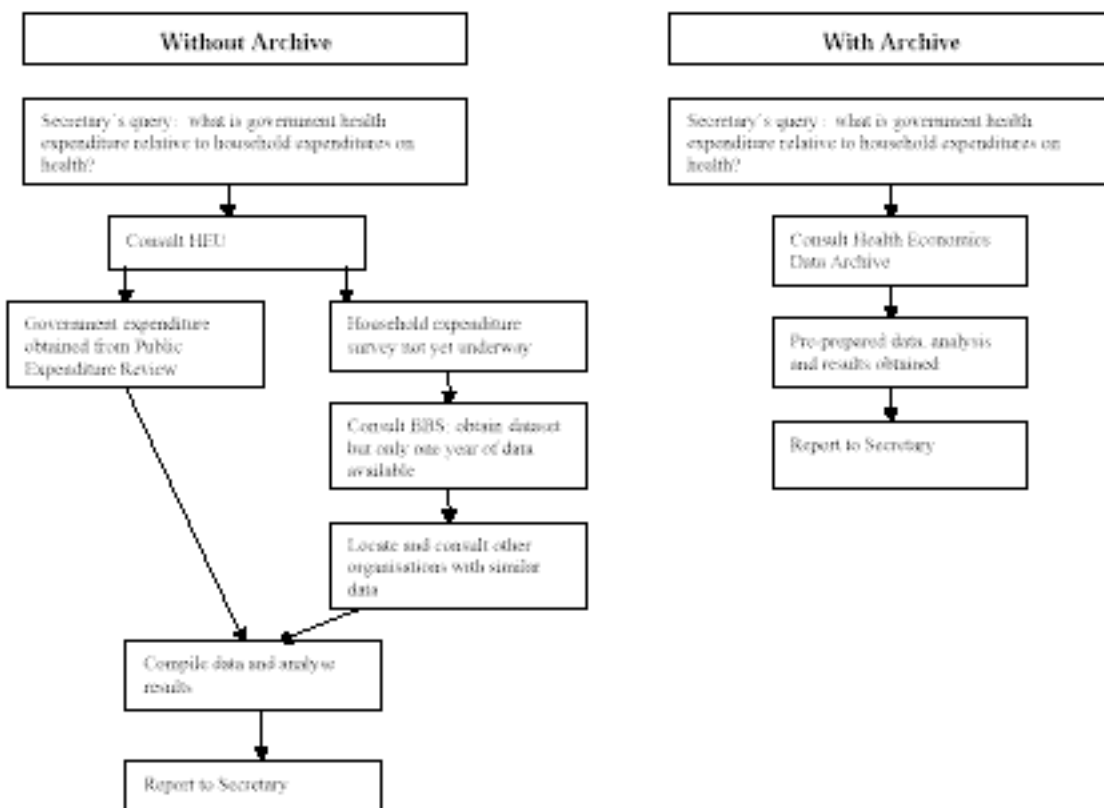
6. Enabling bibliographic citations of databases

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



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 services. 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
- Relevant 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 to be 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 re-calculating 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 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 Access7 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.

* Paper presented at: International Association for Social Science Information Service & Technology, Building Bridges, Breaking Barriers: the future of data in the global network, Toronto, May, 1999.

Deana Leadbeter, International Health Information Specialist, South East Institute of Public Health, Tunbridge Wells, UK. Lorna Guinness, Maxwell Stamp plc, London, UK, and Associate Economist, Health Economics Unit, Ministry of Health and Family Welfare, Government of the People's Republic of Bangladesh.