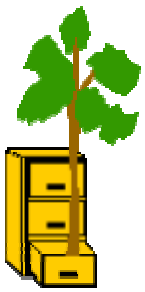
FORESTRY RESEARCH PROGRAMME

Project R7277/ZF0095
1 November 1998 – 30 April 2000

Documentation of UK holdings of growth and yield, inventory and other data from tropical forests

Final Technical Report
June 2000

Howard Wright[1]
Mark Atkinson[3]
Nell Baker[1]
John Healey[3]
Sandro Leidi[2]
Julie Smith[1]
Jenny Wong[3]



ATROFI-UK

Archive of TROpical Forest Inventory

[1]Oxford Forestry Institute
South Parks Road
Oxford
OX1 3RB

[2]Statistical Services Centre
The University of Reading
Whiteknights Road
Reading RG6 6FN

[3]School of Agricultural and Forest Sciences
University of Wales
Bangor
Gwynedd LL57 2UW

TABLE OF CONTENTS

ACKNOWLEDGEMENTS 2

ABBREVIATIONS 3

EXECUTIVE SUMMARY 4

1. BACKGROUND 5

2. PROJECT PURPOSE 8

3. RESEARCH ACTIVITIES 8
   3.1 Organisation of project work 8
   3.2 Attendance at CATIE meeting 8
   3.3 Survey of other initiatives 9
   3.4 Choice of datasets to be included in the meta-database 10
   3.5 Determination of relevant meta-data 11
   3.6 Design of information collection forms and procedures 11
   3.7 Design of the database ATROFI-UK 12
   3.8 Collection of datasets 17
   3.9 Development of web page and interactive search procedures 17
   3.10 Archival policy 18

4. OUTPUTS 18
   4.1 The ATROFI-UK meta-database 18
   4.2 Volume functions compilation 21
   4.3 Workshop 21
   4.4 Recommendations on archival policy 24

5. CONTRIBUTION OF OUTPUTS 26

REFERENCES 28

ANNEX 1 30
   Project Logframe 30
ANNEX 2 33
   List of project staff 33

ANNEX 3 35
   Preliminary questionnaire - NOTES 35

ANNEX 4 37
   Natural forest inventory information collection form 37

ANNEX 5 40
   PSP study information collection form 40

ANNEX 6.....45
Volume study information collection form.....45

ANNEX 7.....49
WORKSHOP PROGRAMME 49

ANNEX 8.....51
DRAFT REPORT OF WORKSHOP.....51
1. Acknowledgements.....53
2. Acronyms.....54
3. List of participants 55
4. Workshop report 56
4.1 Introduction/background.....56
4.2 Aims of the workshop.....56
4.3 Structure of meeting.....56
4.4 Papers presented.....56
4.5 Discussion.....71
4.6 Conclusions.....77
5. Annexes.....79
Annex 1 Agenda 79
Annex 2 Database structure 79

ACKNOWLEDGEMENTS

The project team is grateful to all those holders of datasets who cooperated in the completion of questionnaires and the provision of information. It is also grateful for the time given by those who participated in the workshop both in the preparation of their presentations and in the discussion.

This publication is an output from a research project funded by the United Kingdom Department for International Development (DFID) for the benefit of developing countries. The views expressed are not necessarily the views of DFID. R7277 Forestry Research Programme.

ABBREVIATIONS

ATROFI	Archive of TROpical Forest Inventory
CATIE	Centro Agronómico Tropical de Investigación y Ense anza
CIFOR	Center for International Forestry Research
CITES	Convention on International Trade in Endangered Species
DFID	Department for International Development
FAO	Food and Agriculture Organisation (of the United Nations)
FC	Forestry Commission
FROGGIE	Forest Reserves of Ghana Graphical Information Exhibitor
FRP	Forest Research Programme
IIED	International Institute for Environment and Development
IUFRO	International Union of Forest Research Organisations
NERC	Natural Environment Research Council
NRI	Natural Resources Institute
OFI	Oxford Forestry Institute
TREMA	Tree Management And Mapping software
TROPIS	Tree growth and permanent plot information system
UNCED	United Nations Conference on Environment and Development
UNEP	United Nations Environment Program
WCMC	World Conservation Monitoring Centre

EXECUTIVE SUMMARY

Over the last 50 years many datasets that arose from work in tropical forests have been kept in UK organisations and institutes, and by individuals. The security and documentation of much of this data is often minimal and there is a real danger that some of it may be lost. This project set out to compile a catalogue of such datasets in the form of a meta-database. The datasets of interest fell into four categories: tree volume data, including all tropical plantation and natural forest species; plantation yield tables, *i.e.* permanent sample plot data from plantations; growth and yield of natural forest, *i.e.* permanent sample plot data from natural forest; and static inventories in natural forest.

A meta-database (ATROFI – UK, Archive for TROpical Forest Inventory for data residing in the UK) was set up using Access. ATROFI-UK is a meta-database and therefore does not contain raw data but rather a summary of what is in the listed datasets and, most importantly, a contact address of the holder of the raw data. The purpose of a meta-database is to publicise and share the whereabouts and availability of datasets across institutions. A project web page was created (http://www.atrofi-uk.com) from which the database can be searched by type of data, country and species. Other outputs from the project are also available on the web site. Currently the database contains a total of 122 records of studies comprising 30 permanent sample plots studies, 22 natural forest inventories and 70 sets of species volume functions. A total of 23 tropical countries are involved. Further datasets have yet to be entered. It is emphasised that, although a number of datasets have been documented and catalogued, the actual raw data is not yet secure.

A two-day workshop, entitled “Maintaining Forest Data for Future Use”, was held at Reading on 30-31 March 2000 to consider archival policy and practice for historic and current tropical forest inventory data. The workshop consisted of two parts: first, presentations on the form and functioning of archival systems and, second, a discussion on archival policy for tropical forest data. A full transcript of the workshop is given in Annex 8. The main recommendation to arise from the workshop was the need for all projects, whether research or bilateral, to be proactive concerning the management of data collected. Projects should have a data management plan incorporated into the project memorandum which would detail not only the management of data during the life of the project but the arrangements to be made for its archival after the end of the project.

1. BACKGROUND

Around a hundred years ago formal forestry was established in most tropical countries in which the UK had an interest. Inventory for the purpose of determining the stocking of the forest (static), tree growth rates from dynamic permanent sample plots (PSP) and for determination of timber value (volume tables) was generally undertaken by the early Forestry Departments. During the colonial period all reports and much of the data were often copied to UK libraries at the Oxford Forestry Institute (OFI) or the former Land Resources Division, now the Natural Resources Institute (NRI). After independence, forestry links with the UK were generally maintained and assistance was often given for later inventory and computational work (latterly through the ODA, now DFID, bilateral forestry projects). During the period 1965-1990 over 100 volume tables were prepared by OFI on behalf of many national Forestry Departments. Data and reports generated by ODA assisted inventories were also copied to the UK and considerable holdings of data and grey reports from colonial times to the present have accumulated in the UK. In addition, extensive plot data from tropical forests has been collected over the past 30 years from ecological research by individual scientists. In many countries original data and reports have been lost and the UK holdings represent the only extant copies of much of the information. Unfortunately the UK holdings are spread among many institutes and are virtually unknown and generally inaccessible to forest researchers and national forest authorities. However, these data are potentially invaluable as a resource for research, monitoring and forest management. In the words of Vanclay (*pers. comm.*) when working at CIFOR 'while new plots can be established, we can never buy back the past, so once these plot records are lost, they are gone forever'. When forest authorities are aware that there are significant data on their forests held in the UK they are often very keen to have these analysed and repatriated. For example, the Ghana Forestry Department was able to request that the ODA Forest Inventory Project commission a study of PSP data sent to OFI for computer entry and analysis in the 1970's (Alder 1989). More recently the OFI has had requests for data from Bangladesh (volume table data not available in country) and Tanzania (Permanent Sample Plot (PSP) data).

In many countries large scale static inventories were undertaken to determine the national stocking and condition of the forest for policy and strategic purposes (e.g. Ghana, Nigeria, Guyana, Belize). Data from these historic inventories are in an even more parlous state than that from PSPs and very little has been done to assess their availability and usefulness. Where the forest has been logged or even lost the data are still informative as they at least give a measure of the previous state of the forest and provide an ecological overview of the forests (e.g. Nigeria, Hall 1977). Such information is not just of historical interest as it could provide the basis of an assessment of forest structure, biodiversity change and productive potential.

The growth and yield potential of a forest are generally determined using data derived from periodic re-enumerations of permanent, marked plots (PSPs). Any retrievable data from such plots is potentially very valuable especially if the plots can be re-located and enumeration continued. This interest has resulted in the development of a number of international initiatives intended to make PSP data more widely accessible. The CIFOR maintained TROPIS (Tree growth and permanent plot information system) meta database, established in 1995/96, is the largest of these. TROPIS now contains over 25,000 plots covering 3,078 species from 66 countries. TROPIS has had

at least 100 requests for information, with around 90% of these turning up some useful information. Other initiatives are the holdings maintained by CATIE, the DFID Kalimantan project and the Asian Institute of Forest Management in Malaysia.

As a result of the complexity of forest growth and population dynamics the analyses of PSP data are often presented in the form of computer-based growth models (Alder 1995, Vanclay 1994). Unfortunately this approach, where every study develops a new model, has meant that it is difficult to collate simple, comparative statistics across the tropics (Johns 1997). In particular it would be useful to be able to derive universal indicators of health and vitality that are not based on complex modelling. Such indicators could be basal area stocking and accumulation rates, post-logging mortality rates or novel statistics based on the stand table such as those proposed by Vanclay (1996). The development of generic models for natural forest (e.g. R7278 Humid and semi-humid tropical forest yield regulation with minimal data) will require the estimation of various parameters including growth. The increasing need to show that potential exploitation of forests is sustainable means that data on the growth rates of forests is essential. Frequently there may be static inventory data available but no information on the likely rates of growth.

Accessible forest inventory data over wider areas can be an invaluable tool for increasing knowledge of the distribution and ecology of tree species. As an example; static inventory data in Ghana has been successfully used to develop species distribution maps (Hawthorne 1995a) and ecological profiles of 210 high forest tree species (Hawthorne 1995b). Species mapping from inventory data is an area of considerable potential and the early work on FROGGIE has been developed by Hawthorne into the TREMA program (Hawthorne *et. al.* 1999). The development of an interface between GIS and inventory databases is also a component of the Bangor DFID FRP funded project to generate generic tools for assessment of biodiversity in tropical montane ecosystems. The FRP project ‘Evaluation and development of methods of rapid biodiversity assessment in relation to the conservation of biodiversity in tropical moist forests’ researched the potential utilisation of disparate data sources for biodiversity assessment.

With the exception of Philip (1994) and Lanly (1981) standard inventory texts do not specifically refer to the highly diverse mixed species and ages typical of tropical forests. There is however, recent work that deals with PSP protocols specifically for tropical forests (Alder and Synnott 1992). Despite the long history of inventory in the tropics there has never been a comprehensive review of the application of different designs nor an evaluation of their relative efficiency in different forest types or for various purposes. Standard designs are based on the assumption that the trees are randomly located and evenly distributed across the forest and large plots are used to minimise between plot variance. There is also concern that smaller, more numerous plots would be more representative of the forest (Alder 1992, Wong 1993). Since inventory is very expensive and time consuming it needs to be designed as efficiently as possible. Increased access to historical inventory would provide statisticians with sufficient data and case studies to be able to improve the efficiency of sampling design and provide guidance on the optimal protocols for different forest types and objectives.

Static inventories generally involve large numbers of plots spread at low intensity across extensive areas of forest. Generally plot sizes are relatively large and huge numbers of trees may be enumerated (e.g. > 600,000 records in the Ghana inventory). This means that the field sheets can fill several filing cabinets and, unless the data were transcribed onto computers, archive copies were not generally made. In many cases these original field sheets will be lost and all that remains are summary stand tables by strata (be these Forest Reserves, compartments or biogeographical regions). Both Hall (1977) and van Rompaey (1993) successfully used stand tables for ecological research and there is obvious value in retrieving such datasets. However, much more analysis is possible if the raw plot data are available. For example, plot data was successfully used to perform a detailed analysis of the timber yields resulting from the application of timber harvesting rules in Ghana (Wong 1995). Statistical issues arising from comparison between data derived from different inventory designs (e.g. transects compared to plots) also need to be examined. Rogers (1998) produced a forest type map for Guyana using NOAA satellite data; the training set for the images was obtained by use of the data from the FAO inventory carried out over 25 years previously.

The intention of many forestry departments was to establish periodic static inventory to inform revisions of policy and management plans and in many countries more than one inventory may be available. Repeat inventories in Ghana have been used to test the sustainability of previous logging regimes (e.g. Adam 1989) but generally little use has been made of such datasets. Although no analysis of repeat inventories can provide the same level of detail or certainty as PSPs, they should be able to yield useful information to complement PSP analyses or to plug gaps, and some methodology for dealing with them is required (Wong 1996). Repeat inventories could potentially provide much useful information on the rate of broadscale change in forests and may be particularly useful in assessing biodiversity change as the greater number of plots may make it more likely that rarer trees will be sampled.

Demand for information on the availability and sources of growth and yield data has been recognised by CIFOR in the setting up of TROPIS. The improvement of information systems is generally recognised as a global priority and IUFRO has convened a consortium of international, regional, and national organisations for the purpose of developing a strategy for, and implementing, an Internet-based meta-data service to provide coordinated world-wide access to forest information. They point out that this improvement of access was formally recognised as a priority by UNCED in Agenda 21, Chapter 40:

> *"Countries and organisations should exploit various initiatives for electronic links to support information sharing, to provide access to databases, and other information sources, to facilitate communication for meeting broader objectives, such as the implementation of Agenda 21."*

This IUFRO initiative of a Global Forest Information System was supported by the International Forum on Forests III in Geneva in May 1999.

The topic was specifically identified in the FRP call for proposals in 1997.

2. PROJECT PURPOSE

The complete logical framework for the project is given in Annex 1. The purpose is given as:

> *"Appropriate indicators of ecological and economic sustainability developed and promoted"*

which was Output 1 of Purpose 1 for the 1995 FRP strategy for the Tropical Moist Forest production system. This Purpose would contribute to the Goal of "Tropical moist forests sustainably managed".

Significant archives of historical static and recurrent inventory data are held by UK forestry and other institutions. These holdings are disparate and largely undocumented and there is a real danger that they could be permanently lost. The project was set up to catalogue and database details of data holdings and to report on potential uses of the data with particular regard to the development of baseline information for forest management.

To contribute to achieving this purpose a number of outputs were proposed:

(1) A catalogue of UK inventory, both static and recurrent, data holdings. There will be a master database of information on each data set (this will include a brief description of the data, ownership, archival status, computer format and potential value) which will be publicised and freely distributed on floppy disk and on an internet web page.

(2) A compilation of all tree volume/biomass functions developed at OFI and other UK organisations. This will be available on the internet and as a hardcopy publication.

(3) A report reviewing existing data archival systems and formats and making recommendations on the potential for standardising the data sets identified.

3. RESEARCH ACTIVITIES

3.1 Organisation of project work

To assign and coordinate the various activities of the project that involved a number of staff from different organisations a series of regular planning meetings were held in Oxford for the duration of the project. These meetings, six in all, were essential in order to review the progress of the project and to assign tasks for the next period. Minutes were kept and these provide a record of decisions and responsibilities.

A list of the staff involved in the project and their main responsibilities is given in Annex 2.

3.2 Attendance at CATIE meeting

A IUFRO workshop, held at CATIE, Costa Rica in February 1999, on "Long-term observations and experiments in forestry – tropical forests" was attended by Bob Burn, a member of the project team. This provided a good opportunity to publicise the work of the project and to make useful contacts. Overall the presentations were a little disappointing as they consisted mostly of case studies with little discussion of

statistical issues. The session on data management, however, was good and included a demonstration of the MIRA system which is based at CATIE.

3.3 Survey of other initiatives

Early in the project a review of existing initiatives, similar to this project, was undertaken. A report was produced which is available on the project website (Leidi, 1999). Table 1 gives a brief summary of some of the characteristics of the different systems located. These initiatives are divided into three types.

Firstly, catalogues storing mainly meta-data, intended to contain all ancillary information available about a dataset. Examples are location [country, nearest town], type of study [experiment, inventory], type of data collected [diameters, counts, species richness] but mainly a contact address for potentially interested users. They hardly store any actual raw data, their purpose being to publicise and share the whereabouts and availability of datasets across institutions rather than to share the raw data itself. They normally use Access software on a PC platform without a tailor made front-end for interrogation. Thus enquiries by the public are normally directed to an individual curator. TROPIS by CIFOR is an example of such a catalogue.

Secondly, databases proper which store actual raw data, with the purpose of unifying and pooling them within and across research institutions for integrated, combined, longitudinal [over time] analyses of integrated datasets for, say, the same species under different growing conditions. They also contain meta-data as descriptors of the stored raw data, but to these descriptors they add *fields* that contain the raw data. Either kept at a single institution or shared as stand-alone software, they often have a tailored front-end that puts pre-established queries in Windows dialog boxes that make interrogation and data retrieval feasible to users with no programming knowledge. MIRA by CATIE is such an example. Due to the different format and protocols adopted when dealing with natural forests, these systems need an intelligent way of understanding the data structure and combining them. CATIE is developing such a system called SCIBOS.

Compared to the above, ‘forestry only’ databases, usually integrating very large sets of raw data within the same institution, are easier to set up. They have the advantage provided by a rigid data collection procedure or protocol that can be enforced for the enumeration at the national level. Examples are national forest services, like that of Canada (NFDP) and the USA (FIADRS), which have made public their inventories through very advanced database systems.

There is a continuum between the two types; catalogues might store some raw data in different formats (TROPIS has some), and a database proper for sharing within research institutions may store raw data in summarised form only, such as TREDAT by CSIRO.

There are also compendia, providing summaries at the national level of other summaries supplied by global organisations. They are more of a census of the areas classified as ‘forest’ and ‘protected forest’ by ‘forest type categories’ as defined by IUFRO. An example is FIS by WCMC.

Finally, some environmental databases are included as examples of efficient information systems that offer a public on-line search service of data retrieval. This type of database is made possible by a rigid and unified data collection protocol adopted by hundreds of collaborating sites.

3.4 Choice of datasets to be included in the meta-database

Early in the project it was necessary to define clearly the types of data that were to be included. Although it was acknowledged that all data related to natural forest in the tropics is of potential interest, the project limited itself to data pertaining to *trees*[1] occurring in *forests*[2]. During the course of the project there were several datasets that were offered for inclusion that originated from more general ecological studies and also for trees outside forests. It was decided that, at least initially, the project should concentrate efforts on the collation of the priority *forestry*[3] datasets identified in the project document. Therefore, agroforestry data and surveys of trees outside forests (i.e. on farms or forest gardens) were specifically excluded. Ecological studies were accepted if they enumerated large trees but not if they were only concerned with non-tree plants or animals.

In order to clarify the project interest the following four categories of forestry data were identified:

1. **tree volume data** to include all tropical plantation and natural forest species;
2. **plantation yield tables**, *i.e.* permanent sample plot data from plantations;
3. **growth and yield of natural forest**, *i.e.* permanent sample plot data from natural forest;
4. **static inventories in natural forest**.

Static inventory of plantations was omitted as they were not considered of lasting value since they were often carried out prior to felling and it is possible that both the trees and forest may no longer be in existence.

In summary the criteria for accepting a dataset into ATROFI were:

Criteria	Included	Excluded
Land use	Primary forest Managed natural forest Plantations	Agroforestry Agriculture
Organisms	Trees	Studies with only non-tree plants Animals
Data type	Quantitative	Qualitative
Data quality	Raw data, summary tables for relatively small areas	General descriptions of forest over large areas
Documentation	Protocols and sources available	No protocol available
Geographical area	Tropical and sub-tropical: arid, montane and humid environments	Temperate

[1] Trees – perennial woody plants with a single stem more than 3m tall.
[2] Forests – land covered by trees. These can be naturally occurring = natural forest or planted = plantation .
[3] Forestry – the practice of forest management usually for timber.

3.5 Determination of relevant meta-data

Since the database would not contain raw data it was necessary to consider what information a potential user of a dataset would require in order to make a judgement of its utility. It was envisaged that the meta-data held on each dataset would be provided by data holders on a questionnaire. In order to keep the questionnaires as amenable as possible it was necessary to try and ascertain the minimum set of meta-data that would be required to facilitate value judgements of the dataset. This was achieved by considering how different datasets would be valued for a range of potential uses. For example, if the data was to be used to assess changes in tree biodiversity then a user would need to know what proportion of the trees were given scientific names. Likewise, someone interested in biomass studies would need to know if the sizes of trees were measured and what the lower size threshold for inclusion of a tree might have been. On top of questions related to the protocols used to collect the data, it is also important for a user to be able to ascertain the completeness, integrity and likely availability of the datasets themselves. The criteria and associated value judgements are summarised in Table 1 below. The particular nature of the three categories of data type also meant that there are specific items of meta-data that are relevant to each, these are summarised in Table 2.

Tables 1 and 2 attempt to indicate the types of judgements that can be made of the potential utility of a dataset for secondary use. There was a lot of discussion about where the cut off should be for entry into the database. In order to ensure that the highest value datasets were prioritised for attention it was decided that datasets for which there was no extant protocol, with incomplete data and for which only general summaries were available would not at this time be entered.

3.6 Design of information collection forms and procedures

The design of the information collection forms, the development of the list of required meta-data and the database itself were closely linked. They were developed iteratively using test data for the personal holdings of the project members.

Attempts to cover all the meta-data needed for the four categories of forestry data in a single questionnaire resulted in an excessively long and complex design. It was felt that the complexity of the questionnaire would be a serious disincentive for people attempting to complete it. A two-stage process was therefore developed to minimise the unnecessary completion of lengthy questionnaires. A short screening questionnaire was therefore prepared. The first stage was a mail-shot of a short screening questionnaire. The information requested on this preliminary questionnaire is designed to enable a rapid assessment of the location, type and general quality of the candidate datasets. The layout of this questionnaire is given in Annex 3. The information on the returned questionnaires was used to screen for datasets for entry into ATROFI using the criteria outlined above. Data holders were then requested to complete the lengthier information collection forms relevant to each of the datasets which were of interest to ATROFI.

After several iterations and tests, three information collection forms were developed:

- Natural forest inventory information collection form – see Appendix 4

- PSP study information collection form – see Appendix 5
- Volume study information collection form – see Appendix 6

The plantation yield studies and natural forest PSPs were dealt with on one form because they both deal with repeated measures on plots. The differences between them are the scale and duration of the studies and the number of species being considered.

Volume studies were treated differently to the other data types as the main interest is usually in the volume function or results of the study rather than the raw data. Therefore the coefficients of the derived volume functions were requested as well as the meta-data required (e.g. number of trees in study etc.) to make an evaluation of its potential applicability to the present or different sites. Since it is generally the case that families of functions (overbark, underbark to a range of top heights) can be derived for each species in a volume study the information collection form was designed to accommodate multiple functions.

Data entry from each of the information collection forms was undertaken in Reading directly into ATROFI using ACCESS 'forms' which were designed to look exactly like the printed forms to facilitate easy entry. Data entry routines were incorporated into the data entry front-end to minimise the risk of incorrect or illogical entries.

3.7 Design of the database ATROFI-UK

Based on the survey of existing meta-databases, it was decided that the database format should have a very similar structure to TROPIS, using Access as an application to set up the database structure. It was decided that as many of the TROPIS fields should be included as possible plus additional ones as the information the project will collate comes from different sources of data collection procedures, not only PSPs. However, it was also decided that ATROFI would use the study as the basis for records rather than the individual PSP plots used by TROPIS. Individual plots would be entered if available.

The database was named ATROFI-UK as it represented an Archive for TROpical Forest Information for data residing in the UK. ATROFI-UK is a meta-database and therefore does not contain raw data but rather a summary of what is in the listed datasets and, most importantly, a contact address of the holder of the raw data. The purpose of a meta-database is to publicise and share the whereabouts and availability of datasets across institutions.

The structure of the database consists of 13 tables: eight main tables containing the information collected in the detailed questionnaires, three junction tables to represent many-to-many relationships and two lookup tables (see Figure 1).

There are eight main tables in the database, centred around the core **Study** table. This table links to three separate tables, one for each study type *i.e.*: Permanent Sample Plots, Natural Forest Inventories and Volume Functions. These subset tables were necessary given that the information requested for each study type differs substantially. The questionnaire layout is also different for each of the three study types.

Database	Location	Has a web site?	Data stored as	On-line search?	Software and platform	Method of query	Speed of search	Clarity of search
FORESTRY CATALOGUES								
Sepasal	Kew Gardens	√	NA			Write to curator		
TROPIS	CIFOR, Bogor Indonesia	√	NA		Access on PC	write to curator		
DATABASES PROPER								
GYCH at Samarinda	ITFMP – Jakarta Indonesia		Raw, but models supplied		Visual FoxPro 5	Write to SCH		
TREDAT	CSIRO Australia	√	Summary		Access on PC	write to curator		
MIRA	CATIE, Costa Rica	√	Raw		Knowledge Man / 2 (?)	Can be installed		
CNDF	Canadian Forest Service	√	summary	√	Some GIS since maps	Download	NA	Only report provided
FIADRS	US Forest Service	√	Raw and summary	√	Oracle…	Download	Quick	Restricted, very clear
FIS	WCMC Cambridge	√	summary	√	ArcInfo GIS Dbase files	CD and download	NA	Only report provided
ENVIRONMENTAL DATABASES								
CTFS	Smithsonian Inst. USA	√	Raw	√				
EMAN	Canadian Govt	√	Raw?	√		Download	fair	unrestricted, fairly clear
LTER	Nat Academy Sci. USA	√	Raw	√		download	slow	unrestricted, confusing
ECN	ITE at Merlewood	√	Raw	√	Oracle on Unix	Download	quick	restricted, very clear

Table 1: Features of known similar initiatives, catalogues and databases proper.

Criteria	Dataset value		
	Low	**Moderate**	**High**
Location	Only general description	Maps of plot locations	Latitude, longitude plot references
Protocols	Not available	General description of protocol	Field instructions
Data type	Only general tables	Summary data for small areas	Raw data
Species	Use of non-specific trade names	Only commercial species	All species
Tree size	Only largest trees	Only commercially sized trees	All sizes from low threshold
Enumeration	Only presence/absence	Counts of trees in size classes	Measurements for each tree
Completeness	Incomplete	Partial	Complete
Integrity[+]	No cleaning, obvious 'errors'	Uncertain, but data looks good	Evidence of through cleaning
Media	Paper records	Digital records only	Paper and digital records
Accessibilty	Difficult to ascertain owners	Owners known, permission difficult to obtain	Owners known, permission obtainable and likely to be given

[+] Whether data has been checked for integrity before analysis, especially important for PSP data and any digital data after data entry.

Table 2. Criteria for judging value of forestry datasets

Data type	Criteria	Dataset value		
		Low	**Moderate**	**High**
Yield / PSP	Number of measurements	Only one measurement	Several measurements, short time interval, no recent measurement	Several measurements, long time interval, recent measurements
	Treatments	Not recorded	Recorded treatments, no replicates	Recorded treatments, several replicates
Static	Coverage	Coverage of small area	Coverage of several reserves or district	National coverage
	Number of plots	Low numbers (< 20)	Moderate numbers (20-100 plots)	Large numbers (>100 plots)
Tree volume	Number of trees	Few	Moderate (20-100)	Many (>100)
	Number of functions	Only one per species	Several	Complete set (u/b and o/b to a range of top heights or range of site types)
	Species	Existing volume tables common	Existing volume tables uncommon	Existing volume tables rare to non-existent

Table 3. Criteria for specific dataset types

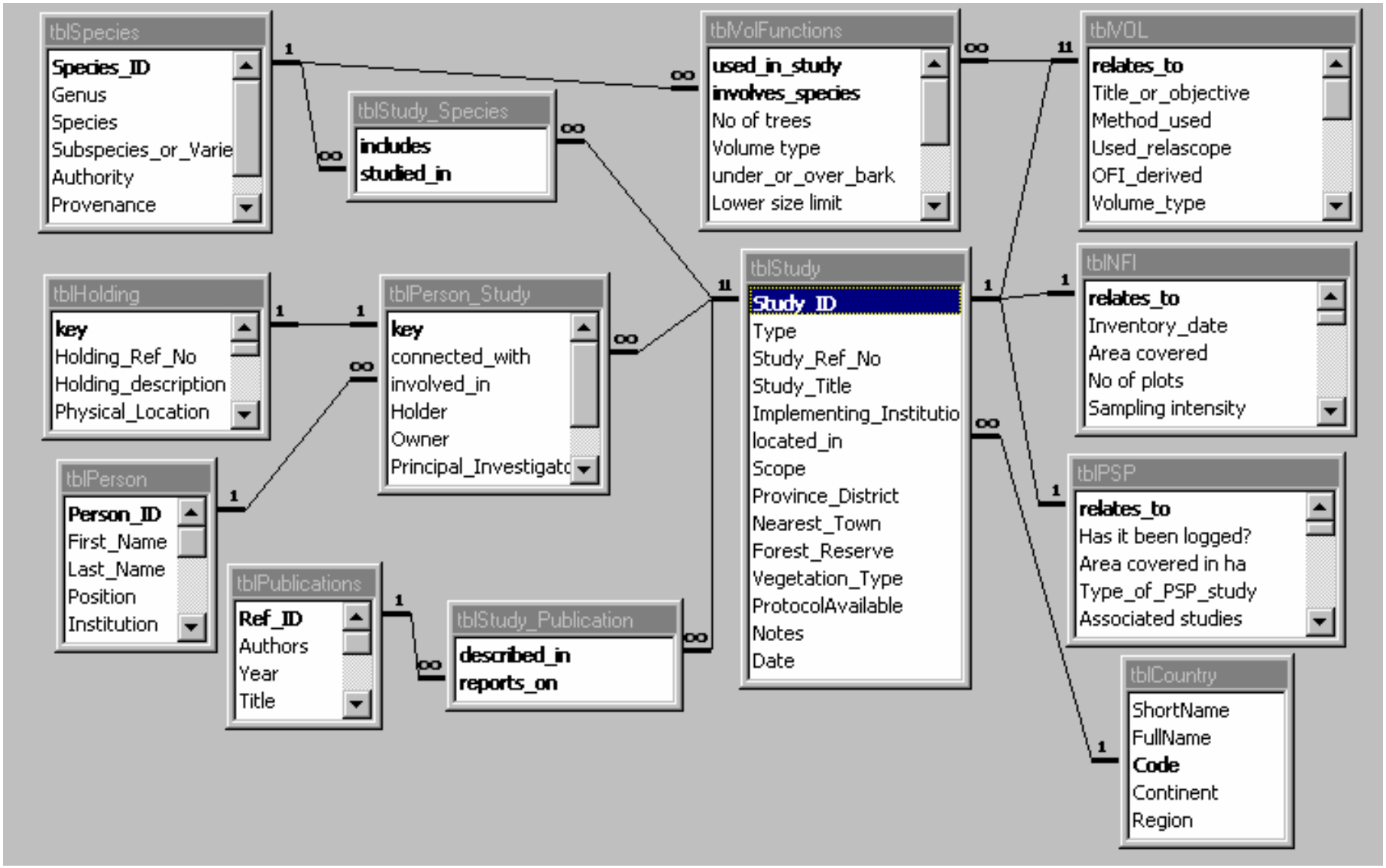


Figure 1. Structure of the ATROFI meta-database

There are eight main tables in the database, centred around the core **Study** table. This table links to three separate tables, one for each study type *i.e.*: Permanent Sample Plots, Natural Forest Inventories and Volume Functions. These subset tables were necessary given that the information requested for each study type differs substantially. The questionnaire layout is also different for each of the three study types. To preserve referential integrity and so avoid duplicates, a record cannot be entered in the study type subset tables if it does not already exist in the main **Study** table.

The **Person** table contains the names, details and address, when known, of anyone connected with a study. Also the person's role is indicated, whether informant / holder, owner of the data, principal investigator or Intellectual Property Rights contact.

The **Holding** table is separate from the study table as there may be more than one holding of the same data set within the UK. It is linked to the **Study** table by a many-to-one link via the **Person_Study** junction table. The **Holding** table describes the physical location of records in the UK and their archival status.

The **Publications** table contains references, whether published or not, that describe any aspect of a study. For example, the protocols or methodology used to collect data. Papers that present analyses and findings based on the data are also included whenever possible.

The **VolumeFunctions** table contains details of each volume function. This table is also a junction table because a Volume study can contain functions for many species and the same species can be studied in many Volume studies. Both the title of each function and the equation itself are stored as text fields so that reports can easily be generated to list the equations for each species. This avoided the need to specify all the different types of function that have been used.

The **Country** table is a lookup table containing 235 countries with the standard two-letter abbreviation from the ISO list. The other Lookup table is the **Species** table, which has over 4,000 botanical names of tropical tree species, including synonyms.

Three junction tables define the many-to-many relationships between pairs of tables:

- **Study_Species** because a study can cover several species and a species may be represented in several studies;
- **Person_Study** because the same person can be involved in several studies (in the same or in a different role) and a study can have many people connected with it;
- **Study_Publication** because the same study can be mentioned in many publications and several studies can be described in the same publication.

The relationships between these tables are shown in Figure 1. Each box represents a table and shows some of the fields. A line linking two boxes represents a relationship. Fields in bold denote the "key fields" that maintain the integrity of the relationships. At each end of the relationship line is a symbol: "1" indicates this table is at the "one" end of a one-to-many relationship; "∞" indicates the table is at the "many" end.

3.8 Collection of datasets

The location and capture of potential datasets followed a four-stage process:

a) preparation of a list of potential organisations and individuals that might hold datasets;
b) an initial letter describing the project;
c) the preliminary single page questionnaire to enable assessment of the holding;
d) detailed questionnaires completed either by the holder of the data or by a visit from project staff.

The preliminary letter was sent out to 66 organisations/institutions and individuals that potentially could hold data. Responses were limited. Eighty five of the preliminary questionnaires were sent out of which about 50% were returned. Further opportunities were taken to make contacts and solicit returns, such as the large meeting of the Tropical Forest Forum at Kew in May 1999 and the inclusion of a notice in the Commonwealth Forestry Association's newsletter. A system of reminders was employed particularly for those individuals and organisations that the project was reasonably sure held data. The membership list of the Commonwealth Forestry Association was also consulted.

Visits were made to a number of organisations and individuals. Those in Scotland were visited by Tim Baker from Aberdeen under a sub-contract. Visits in England were undertaken by Nell Baker and Jenny Wong.

3.9 Development of web page and interactive search procedures

It was decided that access to the database would be primarily online. A rather rigid search system was proposed similar to that in place for the ECN and FIADRS database retrieval systems. It is a very efficient method for information retrieval as it forces the user to search within what is held in the database and not to specify a vague query that may yield no result.

In December 1999 a dot com domain name was registered for the project and a home page set up. This had the address ***http://www.atrofi-uk.com***. The database was made available online and searches can be done separately for each of the three study types. Alternatively, selections can be restricted by the country where the study was carried out, by vegetation type and/or by the name of any person involved in a study.

The searching is carried out in a number of stages:

1. The initial results of a search gives a listing, for the study type selected, of study title, country and vegetation type of all available studies that meet the search criteria. Clicking on a specific study leads to:
2. A summary that characterises the study: its design and protocol, geographical location, implementing institution, scope and year of implementation. Two further options are then given;

3. It is possible to retrieve detailed information about the archival condition of the dataset and, secondly, the contact details of the data holder, to whom questions concerning the release/use of such data should be directed.

The search facility on the web page does not link directly to the ACCESS database. A series of text files is created from the ACCESS database and placed on the server. These files are searched using a search engine written in the Perl language. This is useful because it means that data can be held on ATROFI without necessarily being automatically publicly available. For example, data can be checked or held back until full permission is granted. Users are also more likely to re-visit the site on announcement of a number of new records added rather than a continual dribble of new records.

The project home page also includes an enquiry form that may be completed and submitted.

3.10 Archival policy

As the project progressed it became clear that there was frequently no defined policy concerning the archiving of data collected under various bilateral projects both those managed directly by DFID and those where the management was contracted out. This was a failing not only of DFID projects but many of the data holders expressed similar concerns. No policy usually means that there is no mechanism for archiving data. The project team therefore decided that the workshop at the end of the project should be concerned with this topic. The original plan had been to hold a workshop to brief those concerned with the analysis and archival of these data in the DFID Forestry Partner countries with the aim of determining country interest in the repatriation or further analysis of the data. It was decided that this was impracticable.

A two-day workshop was planned for the end of March 2000 at the University of Reading. From the start the intention was to bring together as wide a range of expertise and experience in the area of data archival as possible. The aim of the workshop was twofold. Firstly, to publicise the ATROFI-UK database and secondly, to reach a consensus on the main points to be included in a realistic archival policy and practice for tropical forestry data.

4. OUTPUTS

4.1 The ATROFI-UK meta-database

As of 22 May 2000 the ATROFI-UK catalogue contains a total of 122 records of studies made up as follows:

Study type	Number
Permanent sample plots	30
Natural forest inventory	22
Volume studies	70
Total	122

An example of the map that appears on the web page when requesting a search of ATROFI is shown in Figure 2. This map illustrates the world wide geographical distribution of the datasets and is automatically updated as new records are added to the database. Currently, a total of 23 tropical countries are involved:

Country	Number of studies	Country	Number of studies
Bahamas	1	Malawi	5
Bangladesh	4	Malaysia	1
Belize	1	Nepal	1
Botswana	2	Nigeria	9
Cameroon	4	Peru	3
Costa Rica	1	Solomon Islands	2
Ecuador	1	Sri Lanka	1
Gambia	2	Sudan	1
Ghana	5	Swaziland	1
Grenada	2	Tanzania	3
Indonesia	3	Thailand	2
Jamaica	6	Uganda	54
Kenya	1	Zambia	3
Lesotho	1	TOTAL	122

There are still more data to be entered into the database, particularly for the volume studies. Entered volume studies cover about 57 tree species, both grouped and ungrouped, and detail 142 volume and conversion factor equations. These functions are stored as text fields; some examples:

equation titles: Bole volume underbark (m3) for dbh (cm) and bole length (m)
Conversion factor for total volume to 20cm top diameter volume

equations: V=-0.007822+0.00004186D*D+0.0001320D*H+0.00004363D*D*H
F=1.0/(1.0762+53.6079exp(-0.1665D))

The following species have at least one volume function:

Species or groups of species	Number of functions
Avicennia africana	4
Albizia coriaria	2
Albizia glaberrima	2
Albizia spp	4
Alstonia boonei	4
Aningeria altissima	2
Antiaris toxicaria	2
Azadirachta indica (Neem)	2
Baikiaea plurijuga	2
Brachystegia spp	1
Brachystegia, Julbernardia, Burkea, Uapaca, Parinari, Pericopsis, Faurea	1
Burkea africana	1
Dacryodes excelsa	1

Species or groups of species	Number of functions
Canarium schweinfurtherii	2
Celtis mildbraedii	2
Chlorophera excelsa	2
Chrysophyllum albidum	2
Cordia millenii	2
Cynometra alexandri	2
Dipterocarpus scaber	7
Entandrophragma angolense	2
Entandrophragma cylindricum	2
Entandrophragma utile	2
Erythrophleum guineense	2
Erythrophleum suaveolens	2
Eucalyptus deglupta	6
Fagara angolensis	2
Faurea spp	1
Guarea cedrata	2
Holoptelea grandis	2
Julbernardia paniculata	1
Lagerstroemia speciosa	6
Licania ternatensis, Sloanea cambasa etc	1
Litsea spp (Medang)	2
Lovoa brownii	2
Maesopsis eminii	4
Mildbraediodendron excelsum	2
Mitragyna stipulosa	2
Morus lactea	2
Olea welwitschii	2
Parinari caratellifolia	1
Pericopsis angolensis	1
Pinus caribaea	2
Pinus khasya	3
Piptadeniastrum africanum	2
Pterocarpus, Afzelia, Burkea, Diplorhynchus, Diospyros, etc	1
Pterygota mildbraedii	2
Pygeum africacanum	2
Rhizophora racemosa	6
Schrebera arborea	2
Shorea spp	2
Swietenia macrophylla	1
Symphonia gabonensis	2
Tectona grandis	16
Trichilia splendida	2
Trichilia spp	2
Uapaca spp	1

ATROFI-UK is accessible and searchable via the internet. In addition to the catalogue itself there is an enquiry form and there is also a data entry form designed to elicit information about tropical forestry datasets that a reader or his institution may hold. The information required is brief but sufficient to be able to make a judgement about whether the dataset is of interest to ATROFI. A more complete questionnaire tailored to the type of data will then be sent to those holding relevant datasets.

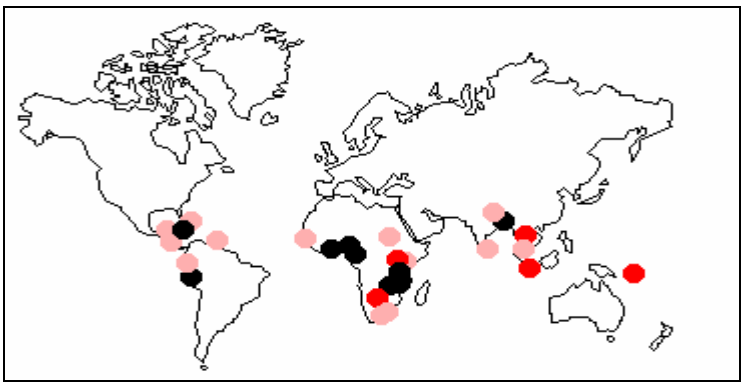


Figure 2. The map shows countries for which there are records in the database. A pink/grey circle indicates 1 record, a red circle indicates 2 and a black circle indicates 3 or more records.

4.2 Volume functions compilation

The database at present contains 142 functions. All functions have been converted into metric units but the database indicates whether the original data was recorded in imperial units. Once all the volume functions have been added to the database, a hard copy compilation will be produced in the form of an OFI Occasional Paper.

4.3 Workshop

A two-day workshop, entitled "Maintaining Forest Data for Future Use", was held at Reading on 30-31 March 2000 to consider archival policy and practice for historic and current tropical forest inventory data. A copy of the workshop programme is given in Annex 6. The speakers covered a wide range of experience in the archiving of data and included representatives from NERC, FC, NRI, Hunting Technical Services, UK National Digital Archive of Datasets and a consultant who worked for FAO on FRA2000. The meeting was attended by 25 people including the speakers; a full list of participants is given in Table 3.

The workshop consisted of two parts: first, presentations on the form and functioning of archival systems and, second, a discussion on archival policy for tropical forest data. In the discussion session several questions were asked. These related to two main areas, firstly what do we do with ATROFI? and secondly how do we prevent the same situation occurring again? The questions as posed to the workshop were:

- Is ATROFI useful and complete enough?
- What facilities can be offered to individuals?
- What are the implications of consultancy company responsibility?
- What are the implications of returning data to country of origin?
- Is archival of this data an international responsibility?
- What is the feasibility of handing the data to the Public Record Office for archival?

- How can we avoid this situation happening again?

A full transcript of the workshop has been produced (Baker, 2000); this is available as a separate publication and will also be available on the project web site. It is reproduced as Annex 8.

Table 3 List of workshop participants

Trevor Abell	Natural Resources Institute
Kevin Ashley	National Digital Archive of Databases
Mark Atkinson	Web Page Consultant
Allesandro Baccini	Consultant to FAO (with experience of FRA2000)
Nell Baker	Oxford Forestry Institute
Graham Bull	Woodland Surveys, Forestry Commission
Eberhart Bruenig	Forestry Consultant
Melvin Cannell	Institute of Terrestrial Ecology
Henry Coleman	Timber Export Development Board, Ghana
Geoff Collett	NERC Environmental Information System, Monks Wood
Ian Dale	Statistical Services Centre, Reading University
Janet Foster	Archive Consultant
John Healey	School of Agriculture and Forest Sciences, University of Wales, Bangor
Brian Kerr	Commonwealth Secretariat
Sandro Leidi	Statistical Services Centre, Reading University
Paul Philips	Institute of Ecology and Resource Management, Edinburgh University
Barbara Pickersgill	Plant Sciences, University of Reading
Andy Roby	Department for International Development
Michael Roper	Association of Commonwealth Archivists and Records
Julie Smith	Oxford Forestry Institute
Paul Smyth	Huntings Technical Services
Roger Stern	Statistical Services Centre, Reading University
Jenny Wong	Forestry Consultant
Howard Wright	Oxford Forestry Institute
Ma Xiangquing	Chinese Researcher based in IERM, Edinburgh University

The workshop came to a number of conclusions. ATROFI was considered to be a useful metadatabase. However, the following improvements could be made to it:

- A global map could be provided as part of the front-end*.
- Related publications and grey literature could be scanned into the database.
- A field could be added stating what the data has been used for in the past as well as what it was collected for.
- Links could be made with GIS systems by providing geo-references for each data set.
- A record of data sets that are known but not yet fully described could be included.

- A field could be added to describe the amount of assessment that is still required to understand and use the data set.
- Fields could be added describing what the long term risk is to the storage of the data and how valuable it is.
- It is important to collect feedback from the users on the usefulness of the database. A short questionnaire making such enquiries should be included on the web page*.

* suggestion has been implemented

The following points summarise the main outcomes of the workshop:

1. Further work is required to collect information on data sets in the UK. In particular the NRI and OFI data holdings still need to be fully described. In addition valuable data is available in MSc and PhD theses on tropical forestry topics held within University libraries (in particular Oxford, Bangor, Edinburgh and Aberdeen). Although this data is relatively secure, summary descriptions of it (including an abstract of the dissertation) in the ATROFI database would render it more accessible.

2. The ATROFI database should be maintained at an appropriate institution (within the UK for the time being). A small grant needs to be obtained to cover the costs of maintaining the database.

3. The database, currently on the web, should be widely publicised to attract further funding, more studies for inclusion and ideas for expansion.

4. The database could be expanded to include information on European holdings of tropical forest data.

5. Many data sets described on the database are not usable in their current form but require significant cleaning and sorting work. It was decided that, in general, this should be the responsibility of potential users. Funding should be sought to undertake this work for particularly valuable data sets.

6. Many data sets are owned by government departments overseas. It is the responsibility of the potential user to seek permission to use data.

7. Once the UK entries are complete, copies of the database should be sent to forestry departments in countries where the data originated. Requests for repatriation will be considered on a case by case basis and copies of data sets will be provided on receipt of funds to cover the costs of creating and sending such copies.

8. Many data sets, although described in the database, are not secure. At present there is no UK strategy for the assessment and archival of tropical forest inventory data and some data is still at high risk of being lost. Such a strategy needs to be developed covering data collected by individual consultants, private companies, universities and government departments.

9. At present DFID projects require consultancy companies to take the responsibility for the maintenance and storage of project data. This raises issues that need to be resolved including copyright ownership, access to data, knowledge of the existence of data, long term security of data, possible use of data to generate unfair competitive advantage and hidden costs. DFID contracts need to specify more clearly the roles and responsibilities relating to data ownership, archival and access.

10. The British Government should take responsibility for the management and safe storage of forest inventory data collected on all overseas projects that it funds. The UK government is responsible to various interests that include: tax payers, overseas governments whom it is assisting and the international community as a whole.

11. DFID needs to secure advice on an appropriate data management policy for both bilateral and research forestry projects including a list of potential criteria for assessing and valuing data. A good starting point for the latter are the criteria used for data inclusion in ATROFI.

12. It is possible that responsibility for the management and storage of third party forestry data more properly lies with an international organisation. This needs to be investigated. FAO, CIFOR, ICRAF and ITTO should be contacted in this regard.

There is a need to develop better policies in the future to store and archive forest inventory data collected overseas by UK organisations and funded by the UK or by other donors. There is a need to ensure overseas capacity to manage and store such data. It is clear that many UK institutions that collect this kind of data have no data archive policy and hence much of the data is being lost.

A Government Department can relinquish the copyright for information that it collects and this is what DFID do for data collected on bilateral projects. DFID retain the copyright for data collected on research projects. At the end of a bilateral project a project assessment is undertaken but no assessment is made of the data collected on the project or of data management.

All Government Departments are now (or will soon be) required to have a functional appraisal policy, this means that they clarify their functions and on this basis they decide what sort of material should be archived.

4.4 Recommendations on archival policy

Arising from the project and the workshop it became apparent that:

- It would be advisable for new projects to include in their negotiations and final agreement something about allowing access to the data produced in the project. This would clarify IPR issues. In bilateral projects, because fresh data can be politically sensitive, confidentiality periods could be specified so that data could be made available for wider use 5-20 years after it was collected.

- Projects should also be required to undertake data management and, in the case of DFID projects, responsibility for the storage and management of data should be negotiated with partner governments. This should be incorporated into the project memorandum and the project reporting requirements.
- DFID should take responsibility for keeping copies of raw field data collected on their projects.
- In the development of its functional appraisal policy DFID should take this into account. Scientists and the public should be consulted in the development of the functional appraisal policy.
- In particular, DFID should start assessing data on electronic format and submitting this to NDAD for storage where appropriate.

This implies that every project, whether research or bilateral, should have a *data management plan* which would form part of the project memorandum. This plan may be minimal in situations where little data will be collected. This plan would need to cover the following:

1. An assessment of the data to be collected during the project in terms of data type, quantity and quality.
2. Arrangements for secure storage of raw data, including field sheets and protocols during the life of the project. This may include any necessary backup procedures.
3. Intended method(s) for archival after the end of the project including retention periods and responsibilities.
4. Details of ownership, copyright and any IPR issues.

This plan should be updated as necessary during the course of a project. It was suggested at the workshop that DFID projects should be required to undertake an annual appraisal of existing project data, describing data sets and providing valuation of the data and an assessment of their archival value. At the end of a project the plan must be finally amended to include the actual details of the archival methods to be employed, future access and responsibilities. Archiving involves a cost and the question of who pays is an important element to be considered. Some portion of the project budget could be set aside for this purpose. One possibility, especially where local archiving may be questionable, is that at the end of a project all data is passed to DFID who could arrange with the Public Record Office for the sensibly organised data-files to be maintained at NDAD (UK National Digital Archive of Datasets) in London.

It was suggested at the workshop that data management capacity and the presence of an adequate data management plan should become one of the criteria for judging project bids. Certainly project performance in this respect should be included in assessment procedures such as ‘Output to Purpose Reviews’.

5. CONTRIBUTION OF OUTPUTS

The project has succeeded in establishing a meta database listing a number of valuable datasets existent in the UK. This is available on the internet and is thus easily accessible. ATROFI-UK will be widely advertised and brought to the notice of potential users in international and national institutions/organisations and individuals. This will be carried out through IUFRO and CIFOR and by publication of notices through existing forestry networks (e.g. the ODI Rural Development Forestry Network, European Tropical Forest Research Network, UK Tropical Forest Forum) and newsletters (e.g. Tropical Forest Update - ITTO, Unasylva - FAO, IUFRO newsletters, Internet bulletin boards such as FOREST INFORMATION UPDATE). Already some positive feedback has occurred.

Access to information is an essential need for sound decision making. This is especially true for the sustainable management of natural resources. This need is widely recognised in most development policies and goals and, indeed, in international conventions and agreements. In setting up ATROFI, the project has provided the means whereby a greater use could be made of data which might otherwise have been lost. This will depend on the actual data being secured. A high proportion of the included datasets resulted from projects or studies that were initially funded by DFID. The project has highlighted the need for a consistent and secure data management policy by DFID to ensure that, at least in the future, valuable datasets, often collected at considerable expense, are documented and archived. Some recommendations on this aspect have been made.

Sustainable management of natural resources has a direct link with the alleviation and elimination of poverty. The third objective of the UK policy for international development is

> *"Protection and better management of the natural and physical environment"*.

Two of the four specific reasons for the DFID Renewable Natural Resources Knowledge/Research Strategy are:

> *"to add to the global store of knowledge aimed at poverty elimination";* and
>
> *"to ensure an adequate supply of appropriate strategic and multi-country knowledge to underpin bilateral programmes in poorer countries"*.

The project outputs have contributed to these objectives.

For the future, there are still some datasets to be identified and included in ATROFI. Collecting the information on potential datasets proved very time consuming; it worked most efficiently when project staff could visit organisations or individuals. It has also been suggested that the tropical forest data contained in forestry student dissertations/theses in UK university libraries should be made more widely available. It must be emphasised, however, that the identification and documentation of datasets is only a first step. There is still the possibility that they may be lost unless they are permanently secured in some form of archive.

Extension of ATROFI into Europe would be of interest. The meta database could be used for holdings of data from European countries with relatively little modification. This would maximise the benefit to be obtained. To assist in the development of a

data management policy for future research and bilateral projects, it could be valuable to carry out in-depth Cost Benefit Analysis for two or three past projects of what is required to make an electronic copy of the data, properly documented, so that they are acceptable for storage by NDAD.

The future maintenance of ATROFI will need to be considered. The system has been set up to require minimum maintenance and if the addition of further records is minimal then the costs for collection of information and data entry will also be low. The domain name is registered for another year and will have to be renewed. An application to FRP for a small annual maintenance grant will be considered.

The project produced the following publications:

LEIDI, A. (1999) Survey of other initiatives for cataloguing, conserving and using such data. Report for R7277. Statistical Services Centre, University of Reading. 24pp. [summary available at <http://www.atrofi-uk.com>] (Internal report)

BAKER, N. (2000) Report of a workshop "Maintaining Forest Data for Future Use", held at University of Reading 30/31 March 2000. Oxford Forestry Institute. 28pp. [available at <http://www.atrofi-uk.com>] (Report of Workshop)

WRIGHT, H.L., ATKINSON, M., BAKER, N., HEALEY, J.R., LEIDI, A., SMITH, J.P. and WONG, J.L.G. (2000) Documentation of UK holdings of growth and yield, inventory and other data from tropical forests. Final Technical Report R7277. Oxford Forestry Institute. 48pp. (Final report)

WRIGHT, H.L. (2000) A compilation of volume functions held in UK institutions. Oxford Forestry Institute, *Occasional Paper No*. ?? (in preparation)

REFERENCES

Adam A.R. (1989) Appraisal of yield regulation in the high forests of Ghana. MSc Forest Management thesis. University of Aberdeen. Unpublished.

Alder D. (1992) Re-analysis of broadleaved inventories 1969-1981. Belize Forest Planning and Management Project, Belmopan. Consultancy report no. 3.

Alder D. (1995) Growth modelling for mixed tropical forests. Tropical forestry papers 30. OFI.

Alder D. and Synnott T.J. (1992) Permanent sample plot techniques for mixed tropical forests. Tropical Forestry Paper 25. OFI.

Anon. (1996) Forest resource accounting. Strategic information for sustainable forest management. IIED & WCMC.

Hall J.B. (1977) Forest types in Nigeria: an analysis of pre-exploitation forest enumeration data. Journal of Ecology 65: 187-199.

Hawthorne W.D. (1995a) FROGGIE Forest Reserves of Ghana: Graphical Information Exhibitor. IUCN, Gland.

Hawthorne W.D. (1995b) Ecological profiles of Ghanaian forest trees. Tropical Forestry Paper 29. OFI.

Hawthorne, W.D., Filer, D.L. and Turnbull, D.J. (1999) Tree data management, mapping, and the development of the TREMA software. *International Forestry Review*, 1, 87-96.

Johns A.G. (1997) Timber production and biodiversity conservation in tropical rain forests. Cambridge University Press.

Lanly, J P (1981) *Manual of Forest Inventory, with special reference to tropical forests.* FAO Forestry Paper 27, 200p.

Philip M.S. (1994) Measuring trees and forests. Second Edition. CABI.

Rogers, D.J. (1998) Forest type/land cover maps for Guyana. Consultant report, DFID Forest Support Project.

van Rompaey R.S.A.R. (1993) Forest gradients in West Africa: a spatial gradient analysis. Doctoral thesis, Wageningen Agricultural University, The Netherlands.

Vanclay J.K. (1994) Modelling forest growth and yield. Applications to mixed tropical forests. CABI, Wallingford.

Vanclay J.K. (1996) Getting the most out of your permanent plot data. Paper presented at the IUFRO Conference 'Growth Studies in Tropical Moist Forests of Africa' 12-21 November, 1996. Forest Research Institute of Ghana, Kumasi, Ghana.

Wong J.L.G. (1993) Preliminary evaluation of the Cross River State reconnaissance inventory. Inventory data processing consultant report to the Cross River State Forestry Project (ODA assisted), Calabar.

Wong J.L.G. (1995) Timber yields from the Forest Reserves of Ghana: An analysis of the implications of sustainable forest management. Planning Branch, Kumasi, Ghana. Unpublished.

Wong J.L.G. (1996) Inventory for management planning : Experience in Ghana. Paper presented at the IUFRO Conference 'Growth Studies in Tropical Moist Forests of Africa' 12-21 November, 1996. Forest Research Institute of Ghana, Kumasi, Ghana.

ANNEX 1

Project Logframe

Narrative Summary	Objectively Verifiable Indicators	Means of Verification	Important Assumptions
Goal			
Tropical moist forests sustainably managed	To be completed by Programme Manager	To be completed by Programme Manager	To be completed by Programme Manager
Purpose			
Appropriate indicators of ecological and economic sustainability developed and promoted	To be completed by Programme Manager	To be completed by Programme Manager	To be completed by Programme Manager
Outputs			
1. A catalogue, in the form of a database, of all UK holdings of static and recurrent inventory data developed and distributed.	1.1 Paper prepared for conference in Costa Rica by January 1999. 1.2 All potential data owners contacted by September 1999. 1.3 Catalogue completed and distributed by end of project.	Presentation/publication of paper. Project reports. Availability of catalogue on web.	Original 'owners' of data located and permission obtained for use of data by third parties Data used by relevant organisations to improve yield regulation, biodiversity conservation.
2. A compilation of volume/biomass functions computed in the UK prepared and distributed .	2.1 Publication prepared and distributed by end of November 1999.	Availability of publication.	
3. Recommendations on the potential for standardising the archival format and system for the identified datasets	3.1 Report prepared by end of project. 3.2 Workshop held by end November 1999.	Report of workshop	
Activities	**Inputs**	**Means of Verification**	**Important Assumptions**

Narrative Summary	Objectively Verifiable Indicators	Means of Verification	Important Assumptions
1.1 Meeting with all collaborators to arrange details of information requirements. 1.2 Prepare initial format for inventory database. 1.3 Prepare list of UK institutions that may hold forestry data; send out questionnaire. 1.4 Prepare master list of countries, dates and types of data holdings from the questionnaire results. 1.5 Correspond with data owners of prioritised data concerning data property rights issues. 1.6 Conduct follow-up visits to UK institutions holding data. 1.7 Consultations to fill information gaps. 1.8 Modify existing database format if necessary. 1.9 Review the usefulness of the datasets in biodiversity assessment. Consult Bangor ERP project and the FRP RBA project 1.10 Prepare paper for conference in CATIE, Costa Rica. 1.11 Prepare and document protocols for accepting new data holdings into inventory database. 1.12 Set up web page for accessing inventory database. Make copies of database on CD-ROM and floppy disks for distribution. Distribute copies. 1.13 Publish notices in relevant newsletters, journals and bulletin boards advertising the existence of the inventory database. 1.14 Prepare report on database activities.	Staff Costs 46433 Overheads 20067 Capital Equipment 2000 Travel and Subsistence: Overseas 2000 UK 4000 Miscellaneous 18550 Total £93,084		
2.1 Prepare a list of all volume datasets at OFI and other organisations (from survey). 2.2 Contact originators of data and establish their interest. 2.3 Extract data and volume functions from files and compile into a publication. 2.4 Distribute hardcopy and make available on WWW.			
3.1 Survey of other initiatives, both national and international, for cataloguing, conservation and use of such data. 3.2 Investigate the potential for the use of a standardised format for the secure archival of such data and the means whereby			

Narrative Summary	Objectively Verifiable Indicators	Means of Verification	Important Assumptions
datasets at risk can be converted into such a form. 3.3 Insert notes on the potential limitations and uses for each inventory in database. Link with Activity 1.9. [3.4 Prepare report of the potential of the archive for various analyses. 3.5 Present results of project at a workshop to be held at the end of the project. Modify final report after workshop.			

ANNEX 2

List of project staff

Dee Aebischer data collection (Wales)
School of Agriculture and Forest Sciences
University of Wales
Bangor, Gwynedd LL57 2UW

Mark Atkinson webpage design and programming
10 The Grove
Wednesfield
Wolverhampton WV11 1RW
tel: email: mark@atkinsonm.demon.co.uk

Tim Baker data collection (Scotland)
Department of Plant and Soil Science
University of Aberdeen
Cruickshank Building,
St Machar Drive,
Aberdeen AB24 3UU
Scotland, U.K.
tel: email: t.r.baker@abdn.ac.uk

Nell Baker location of data sets, data collection, workshop organisation
Oxford Forestry Institute
South Parks Road
Oxford, OX1 3RB
tel: 01865 271038 fax: 01865 275074 email: nell.baker@plants.ox.ac.uk

Robert Burns
Statistical Services Centre
The University of Reading
Harry Pitt Building, Whiteknights Rd.
P. O. Box 240
Reading RG6 6FN
tel: 0118-9316731 fax: 0118-9753169 email: r.w.burn@reading.ac.uk

Ian Dale database implementation
Statistical Services Centre
The University of Reading
Harry Pitt Building, Whiteknights Rd.
P. O. Box 240
Reading RG6 6FN
tel: 0118-9316731 fax: 0118-9753169 email:

John R. Healey data location
School of Agriculture and Forest Sciences
University of Wales
Bangor, Gwynedd LL57 2UW
tel: 01248-382631 fax: 01248-382832 email: j.healey@bangor.ac.uk

Sandro Leidi database and questionnaire design
Statistical Services Centre
The University of Reading
Harry Pitt Building, Whiteknights Rd.
P. O. Box 240
Reading RG6 6FN
tel: 0118-9316731 fax: 0118-9753169 email: a.a.leidi@reading.ac.uk

Julie Smith data collection, webpage implementation
Oxford Forestry Institute
South Parks Road
Oxford, OX1 3RB
tel: 01865 271038 fax: 01865 275074 email: julie.smith@plants.ox.ac.uk

Jenny Wong questionnaire design, database design, data collection, webpage design
School of Agriculture and Forest Sciences
University of Wales
Bangor, Gwynedd LL57 2UW
tel: 01248-355920 fax: 01248-382832 email: JLGWong@cs.com

Howard Wright project leader, volume data sets
Oxford Forestry Institute
South Parks Road
Oxford, OX1 3RB
tel: 01865 271038 fax: 01865 275074 email: howard.wright@plants.ox.ac.uk

ANNEX 3

Preliminary questionnaire - NOTES

This questionnaire is designed to elicit preliminary information concerning the type and scale of potential candidate datasets for the development of tropical inventory data archives in the UK. The questions are designed to provide a basis for the design of a fuller questionnaire to elicit suitable information for inclusion in a tropical inventory meta-database for UK holdings. All information provided in this questionnaire will be treated confidentially and information provided in this questionnaire will not form part of the eventual meta-database.

For each dataset that you hold, please provide the following information on the enclosed form. Thank you!

Data type: Enter the type of inventory in question (see Table below).
Country: Country in which inventory took place
Date: Dates of enumeration
Area: Approximate area covered by inventory either as km² or indicate whether it is local, regional or national in scope
Number of observations: Number of plots or trees enumerated
Data held: Type of data held i.e. raw, derived or as conventional output (see Table below)
Data format: Type of data holding i.e. paper records, punch cards, magnetic tape, floppy disk (8', 5.25' or 3.5') etc.
Availability: Indicate whether it would be possible to include this dataset in the meta-dataset. Please indicate if other permissions are needed before inclusion and give suitable contacts.
Notes: Other relevant information

Table of tropical inventory types of interest to the UK data archives project

Inventory type	Summary	Data type		
		Raw	Derived	Conventional output
Tree volume data	One-off measurements of tree bole dimensions	Diameter and length of bole sections	Volumes of bole sections and sample trees	Volume functions or look-up tables
Plantation yield tables	Repeated measurement of sample trees or stands	Diameter and height of measured trees	Periodic diameter, height and volume increment	Yield functions as graphed yield curves or look-up tables
Natural forest growth and yield	Repeated measurements of tagged tree diameter in permanent sample plots (PSPs)	Diameter of measured trees often with subjective scores for crown position, bole quality etc.	Periodic diameter and volume increment for trees, species and plots	Growth and yield models usually as computer-based applications
Natural forest inventory	One-off tree measurements	Species and diameter of trees often with subjective crown and timber scores	Stocking density, basal area and volume of trees, species and sample plots	Stand tables of species numbers, basal area and volume by size class

Preliminary questionnaire

Name: ______________________________ Institution: _____________________________________ Date:_____________

Data type		Country		Date		Area		On Tropis	
Number of observations		Data held		Data format					
Availability		Notes							
Data type		Country		Date		Area		On Tropis	
Number of observations		Data held		Data format					
Availability		Notes							
Data type		Country		Date		Area		On Tropis	
Number of observations		Data held		Data format					
Availability		Notes							
Data type		Country		Date		Area		On Tropis	
Number of observations		Data held		Data format					
Availability		Notes							
Data type		Country		Date		Area		On Tropis	
Number of observations		Data held		Data format					
Availability		Notes							

ANNEX 4

Natural forest inventory information collection form

Please give measurement units for all measurements indicated.

Study reference

Title / reference	
Principal investigator	
Implementing institution	
Physical location	
Notes	

Informant details

First name		Surname			
Position					
Department		Institution			
Address					
Town		Postcode		Country	
Phone		Fax		E-mail	

Location and scale of inventory

Country		Date	
Scope of inventory*	National / Regional / Local *(circle alternatives)*		
Province / District		Nearest town	
Forest / Reserve			
Vegetation type*	Montane / lowland / mangrove / dryland		
Area covered (ha)		Number of plots	Sampling intensity (%)

Inventory protocol

Field instructions*	Available / Copy enclosed / Published / In standing orders / Unavailable
Reference for field instructions …… ……	
Sample design …… …… …… …… …… ……	

Plot dimensions				
Plot area		Plot shape*	Circular / square / rectangular / strip / T-shaped / other – specify:	
All tree species enumerated?	Yes / no		All tree species identified?	Yes / no
Principal species …… …… …… …… ……				
Size limits				

Parameters recorded

Parameter	Tick	Notes (e.g. recording limits, tally into class intervals etc.)
Presence / absence		
Tree counts		
Diameter /girth		
Basal area		
Height		

List non-tree parameters recorded

Is this part of a multi-purpose resource inventory	Yes / No

Data availability

Type of data held	Tick
Raw	
Derived	
Summary tables	
Is data coded?	
Is key for codes available?	

If data is aggregated, describe the classes used
List parameters coded

Give the approximate quantity and size of storage media:

Media	Size	Quantity - number
Original field sheets	A4, A3, A2, other	
Hardcopy	Wide computer output, A4, other	
Punch cards		
Magnetic tapes	8 in , 12 in, other	
Magnetic disks	8 in, 5¼ in, 3½ in, mixed, other	
Modern computer format	File size in Mb	Software and version used

Can you access computer readable formats?	Yes / No
Do you need assistance to download older computer readable data?	Yes / No
If yes, please provide details of problem	

Data integrity

How complete is the data?*	Complete / Partial / Patchy
Has data been checked for reliability? i.e. cleaned	Yes / No
Please comment on the usefulness or limitation of these data …… …… ……	

Intellectual property rights

Are you the owner of the data?*	Yes / No / Joint owner		
Is a copy of the data available in country?*	Yes / No / Don't know / Only hardcopy		
If you are not the owner, who is?			
First name		Surname	
Position			

ANNEX 5

PSP study information collection form

Please give measurement units for each measurement indicated.

Study reference

Title / reference	
Principal investigator	
Implementing institution	
Physical Location	

Informant details

First name		Surname			
Position					
Department		Institution			
Address					
Town		Postcode		Country	
Phone		Fax		E-mail	

Location and scale of inventory

Country		Date			
Scope of inventory*	National / Regional / Local *(circle alternatives)*				
Province / District		Nearest town			
Forest / Reserve					
Vegetation type*	Montane / lowland / mangrove / dryland				
Is it a plantation?	Yes / no	Has it been logged?	Yes / no		
Area covered (ha)		Number of plots		Sampling intensity (%)	
Type of study*	Plantation / enrichment / natural forest				

Inventory protocol

Field instructions*	Available / Copy enclosed / Published / In standing orders / Unavailable
Reference for field instructions …… ……	

Sample design …… …… …… …… ……					
Treatment(s)					
Plot dimensions					
Plot area		Plot shape*	Circular / square / rectangular / strip / T-shaped / other – specify:		
All tree species enumerated?	Yes / no	All tree species identified?	Yes / no		
Principal species …… …… …… ……					
Size limits (with sampling fractions for sub-samples)					
Study established		Study ended		On-going study	Yes / No
No. of enumerations		Enumeration interval			

Parameters recorded

Parameter	Tick	Notes (e.g. recording limits, tally into class intervals etc.)
Tree co-ordinates		
Diameter / girth		
Height		
Crown position		
Crown status		
Bole form / quality		
Dominance		
Regeneration		
Planting year		
Other		
Non-tree plot characters		
Other additional information, e.g. rainfall		

Data availability

Type of data held	Tick
Raw	
Derived	
Summary tables	
Is data coded?	
Is key for codes available?	

If data is aggregated, describe the classes used
List parameters coded

Give the approximate quantity and size of storage media:

Media	Size	Quantity – number
Original field sheets	A4, A3, A2, other	
Hardcopy	Wide computer output, A4, other	
Punch cards		
Magnetic tapes	8 in , 12 in, other	
Magnetic disks	8 in, 5¼ in, 3½ in, mixed, other	
Modern computer format	File size in Mb	Software and version used

Can you access computer readable formats?	Yes / No
Do you need assistance to download older computer readable data?	Yes / No
If yes, please provide details of problem	

Data integrity

How complete is the data?*	Complete / Partial / Patchy
Has data been checked for reliability? i.e. cleaned	Yes / No
Please comment on the usefulness or limitations of these data …… …… ……	

Intellectual property rights

Are you the owner of the data?*	Yes / No / Joint owner		
Is a copy of the data available in country?*	Yes / No / Don't know / Only hardcopy		
If you are not the owner, who is?			
First name		Surname	
Position			
Department		Institution	
Address			
Town		Postcode	
Country		E-mail	
Phone		Fax	
Are you aware of any other holders of these data? Please give contact details …… …… …… ……			

Consent

Do you consent to the distribution of information contained on this form to the public?	Yes / No
If you do not consent, who should be contacted for permission to release this information …… …… ……..	
Will there be any restrictions on access to the data? (Please provide brief summary) …… …… …… …… ……	

Signature: Date:

Study ID		Associated study ID	

Please list key references that refer to this study or use of the data (include grey literature). When available, you can attach a printout of these.

Author(s)			
Title			
Year		Source	
Published?	Yes / No	Publisher	
Out of print?	Yes / No	ISBN ref number	

Author(s)			
Title			
Year		Source	
Published?	Yes / No	Publisher	
Out of print?	Yes / No	ISBN ref number	

Author(s)			
Title			
Year		Source	
Published?	Yes / No	Publisher	
Out of print?	Yes / No	ISBN ref number	

Author(s)			
Title			
Year		Source	
Published?	Yes / No	Publisher	
Out of print?	Yes / No	ISBN ref number	

ANNEX 6

Volume study information collection form

Please give measurement units for all measurement indicated.

Study reference

Title / reference	
Principal investigator	
Implementing institution	
Physical location	
Notes	

Informant details

First name		Surname			
Position					
Department		Institution			
Address					
Town		Postcode		Country	
Phone		Fax		E-mail	

Location and scale of volume study

Country		Date	
Province / District		Nearest town	
Forest / Reserve			
Vegetation type	Montane / lowland / mangrove / dryland		
Associated studies			

Study protocol *Circle as appropriate*

Enumeration Year		Is raw data available?	Yes / No
Origin / Related study		Method	Standing / Felled
		Used a relascope?	Yes / No
Name of species			

Species ID		Original units of measurement			Imperial / Metric
dbh Range	Minimum			Maximum	
Height Range	Minimum			Maximum	
Published?	Yes / No	Number of functions		Number of trees	
Availability					
File Reference	Notes				

Give the approximate quantity and size of storage media:

Media	Size	Quantity - number
Original field sheets	A4, A3, A2, other	
Hardcopy	Wide computer output, A4, other	
Punch cards		
Magnetic tapes	8 in , 12 in, other	
Magnetic disks	8 in, 5¼ in, 3½ in, mixed, other	
Modern computer format	File size in Mb	Software and version used

Can you access computer readable formats?	Yes / No
Do you need assistance to download older computer readable data?	Yes / No
If yes, please provide details of problem …… …… ……..	

Signature: Date:

Study ID		Associated study ID	

<table>
<tr><td>Equation Name and units</td><td colspan="2"></td><td>Equation ID</td></tr>
<tr><td>Equation detail</td><td colspan="3"></td></tr>
<tr><td>Volume type</td><td colspan="2">Total / utilisable / merchantable ← circle one only →</td><td>Ub / Ob</td></tr>
<tr><td>Number of trees</td><td colspan="3">Notes</td></tr>
</table>

<table>
<tr><td>Equation Name and units</td><td colspan="2"></td><td>Equation ID</td></tr>
<tr><td>Equation detail</td><td colspan="3"></td></tr>
<tr><td>Volume type</td><td colspan="2">Total / utilisable / merchantable ← circle one only →</td><td>Ub / Ob</td></tr>
<tr><td>Number of trees</td><td colspan="3">Notes</td></tr>
</table>

<table>
<tr><td>Equation Name and units</td><td colspan="2"></td><td>Equation ID</td></tr>
<tr><td>Equation detail</td><td colspan="3"></td></tr>
<tr><td>Volume type</td><td colspan="2">Total / utilisable / merchantable ← circle one only →</td><td>Ub / Ob</td></tr>
<tr><td>Number of trees</td><td colspan="3">Notes</td></tr>
</table>

<table>
<tr><td>Equation Name and units</td><td colspan="2"></td><td>Equation ID</td></tr>
<tr><td>Equation detail</td><td colspan="3"></td></tr>
<tr><td>Volume type</td><td colspan="2">Total / utilisable / merchantable ← circle one only →</td><td>Ub / Ob</td></tr>
<tr><td>Number of trees</td><td colspan="3">Notes</td></tr>
</table>

Please list key references that refer to this study or use of the data (include grey literature). When available, you can attach a printout of these.

Author(s)			
Title			
Year		Source	
Published?	Yes / No	Publisher	
Out of print?	Yes / No	ISBN ref number	

Author(s)			
Title			
Year		Source	
Published?	Yes / No	Publisher	
Out of print?	Yes / No	ISBN ref number	

Author(s)			
Title			
Year		Source	
Published?	Yes / No	Publisher	
Out of print?	Yes / No	ISBN ref number	

Author(s)			
Title			
Year		Source	
Published?	Yes / No	Publisher	
Out of print?	Yes / No	ISBN ref number	

ANNEX 7

MAINTAINING FOREST DATA FOR FUTURE USE

A two-day workshop to consider archival policy and practice for historic and current tropical forest inventory data.

WORKSHOP PROGRAMME

Thursday 30 March

10.30 *Registration and coffee*

11.00 – 12.30 *Session 1*

- Presentation of the project — Howard Wright and Jenny Wong
- Demonstration of the project database ATROFI-UK — Alessandro Leidi and Mark Atkinson
- Issues raised and questions

12.30 – 1400 *Lunch*

14.00 – 17.30 *Session 2*

Relevant archival policies/systems in the UK

- What are archives and why do they matter? — Janet Foster – Freelance Archive Consultant
- Dataset Archiving and Data Services at the National Data Repository — Kevin Ashley – UK National Digital Archive of Datasets
- Data archiving in current and future research projects — Roger Stern – Statistical Services Centre, Reading University
- Inventory and Data Retrieval - the view from NRI — Trevor Abell – Natural Resources Institute
- Archiving Census and Woodland Inventory Data — Graham Bull – Woodland Surveys, Forestry Commission
- Maintaining Forest Data for Future Use – A Commercial Perspective — Paul Smyth – Huntings Technical Services
- Volume/Biomass: Georeferenced Forest Volume Data for Tropical Countries — Alessandro Baccini – Independent Consultant FRA2000
- Managing environmental data — Geoffrey Collett – Environmental

Information Centre, Monks Wood

Friday 31 March

0930 – 13.00 *Session 3*

Group work on defining the elements of an archival policy covering such issues as:

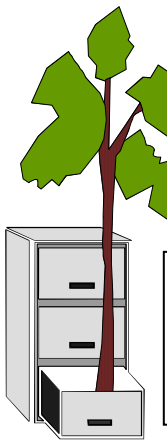
- central archival facility?
- funding mechanisms
- data storage methods/systems
- IPR and repatriation

Final discussions and summing up

13.00 *Lunch and departure*

ANNEX 8

DRAFT REPORT OF WORKSHOP



ATROFI

Archive of Tropical Forest

Maintaining Forest Data for Future Use

University of Reading, 30-31 March 2000

A two-day workshop to consider archival policy and practice for historic and current tropical forest inventory data.

This publication is an output from a research project funded by the United Kingdom Department for International Development (DFID) for the benefit of developing countries. The views expressed are not necessarily the views of DFID.

R7277 Forestry Research Programme

Date of revision: 27 June 2000

Report compiled by: Nell Baker,7 Newman Road, Littlemore Oxford, OX4 3UJ, UK
Tel: +44 1865 777 223, Fax: +44 1865 271 036

TABLE OF CONTENTS

1. Acknowledgements 53
2. Acronyms 54
3. List of participants 55
4. Workshop report 56
   4.1 Introduction/background 56
   4.2 Aims of the workshop 56
   4.3 Structure of meeting 56
   4.4 Papers presented 56
      ATROFI database 56
      Web page 57
      What are archives and why do they matter 58
      Dataset Archiving and Data Services at the National Data Repository 59
      Data management and archiving, the Statistical Services Centre experience 62
      Inventory and Data Retrieval – the view from NRI 63
      Archiving Census and Woodland Inventory Data 66
      Maintaining Forest Data for Future Use – A Commercial Perspective 68
      Volume/Biomass: Geo-referenced Forest Volume Data for Tropical Countries. 70
      Managing environmental data 71
   4.5 Discussion 71
      The ATROFI database 71
      Policy 74
   4.6 Conclusions 77
5. Annexes 79
   Annex 1 Agenda 79
   Annex 2 Database structure 79

1. Acknowledgements

The Forest Research Programme of DFID is thanked for providing funding for the project and for the workshop. The Statistical Services Centre, University of Reading is thanked for providing the venue for the meeting and administrative support. Howard Wright is thanked for chairing the meeting. Sandro Leidi, Lorna Turner and Kelly Watkins are thanked for co-ordinating the meeting at Reading. Nell Baker is thanked for organising the workshop and for reporting on and editing the proceedings.

2. Acronyms

AONB	Area of Outstanding Natural Beauty (UK)
ATROFI	Archive of TROpical Forest Inventory
CAIRS	Computerised Agricultural Information Retrieval System (NRI)
CCT	Computer Compatible Tape
CEH	Centre for Ecology and Hydrology (NERC)
CIFOR	Centre for International Forestry Research (Indonesia)
CLICOM	Climate and Computing (a project)
COPR	Centre for Overseas Pest Research (now part of NRI)
DANI	Department of Agriculture Northern Ireland
DARE	Data Rescue Project (for climatic data)
DAT	Digital Audio Tape
DFID	Department for International Development (UK)
ECN	Environmental Change Network (NERC UK)
EIC	Environmental Information Centre, Monks Wood (NERC)
EU	European Union
FAO	Food and Agriculture Organisation (of the United Nations)
FC	Forestry Commission (UK)
FE	Forest Enterprise (UK)
FIADRS	Forestry Inventory and Analysis Database Retrieval System (US Forest Service, USA)
FRP	Forest Research Programme (DFID UK)
GIS	Geographic Information System
HTS	Hunting Technical Services (UK)
ICRAF	International Centre for Research in Agroforestry (Kenya)
ICRISAT	International Centre for Research in the Semi-Arid Tropics (India)
IPR	Intellectual Property Rights
IRRI	International Rice Research Institute
ISO	International Standards Organisation
ITTO	International Tropical Timber Organisation (Japan)
IUCN	International Union for the Conservation of Nature.
LRDC	Land Resources Development Centre (UK)
NP	National Park
NDAD	UK National Digital Archive of Datasets
NDR	National Data Repository (UK)
NERC	Natural Environment Research Council (UK)
NRI	Natural Resources Institute (UK)
ODA	Overseas Development Administration (now DFID)
OFI	Oxford Forestry Institute (UK)
PRO	Public Records Office of England and Wales
PSP	Permanent Sample Plot
RCHM	Royal Commission on the Historical Monuments
SCDB	Sub-compartment database (UK Forestry Commission)
SSC	Statistical Services Centre (Reading University, UK)
SSSI	Site of Special Scientific Interest (UK)
TPO	Tree Preservation Order
TPI	Tropical Products Institute (now part of NRI)
TRADIS	Tropical Research and Agricultural Development and Information System
TROPIS	Tree growth and permanent plot information system
TSP	Temporary Sample Plot
ULCC	University of London Computer Centre (UK)
UoG	University of Greenwich (UK)
UNEP	United Nations Environment Programme
WGS	Woodland Grant Scheme (UK)

3. List of participants

Trevor Abell	Natural Resources Institute
Kevin Ashley	National Digital Archive of Databases
Mark Atkinson	Web page Consultant
Allesandro Baccini	FRA2000 Consultant
Nell Baker	Oxford Forestry Institute
Graham Bull	Woodland Surveys, Forestry Commission
Eberhart Bruenig	Forestry Consultant
Melvin Cannell	Institute of Terrestrial Ecology, NERC
Henry Coleman	Timber Export Development Board, Ghana
Geoff Collett	Environmental Information System, Monks Wood (NERC)
Ian Dale	Statistical Services Centre, Reading University
Janet Foster	Archive Consultant
John Healey	School of Agricultural and Forest Sciences, Bangor University
Brian Kerr	Commonwealth Secretariat
Sandro Leidi	Statistical Services Centre, Reading University
Paul Phillips	Institute of Ecology and Resource Management, Edinburgh University
Barbara Pickersgill	Plant Sciences, University of Reading
Andy Roby	Department for International Development
Mike Roper	Association of Commonwealth Archivists and Records Managers
Julie Smith	Oxford Forestry Institute
Paul Smyth	Huntings Technical Services
Roger Stern	Statistical Services Centre, Reading University
Jenny Wong	Forestry Consultant
Howard Wright	Oxford Forestry Institute
Ma Xiangquing	Chinese Researcher based in IERM, Edinburgh University

4. Workshop report

4.1 Introduction/background

Current concerns about global warming, loss of global biodiversity and deforestation have thrown into sharp focus the role of tropical forests as a storehouse of carbon and biodiversity and as a potentially sustainable source of timber and other products. However, information about the biomass of tropical forests, their species composition and their productivity, both before and after human impact, is still very fragmented. Such information is needed to prioritise conservation initiatives, as a baseline against which to assess subsequent change, as a means of predicting future changes, as a target for restoration, and as an indicator of best management practices.

Over the years a considerable amount of inventory, growth and yield data from tropical forests have been accumulated in the UK by individuals and various organisations, both private and public. Many of these data are now no longer available in their country of origin. The archiving of these records has often been poor and in many cases has depended on the interest of a single person. There is thus a very real danger that this valuable information may be lost. There is clearly a need to establish a new archiving system to meet modern information requirements; this is the focus of a current DFID Forestry Research Programme project (R7277 Documentation of UK holdings of growth and yield, inventory and other data from tropical forests). On 30 and 31 March 2000 a project workshop was held at the Statistical Services Centre, University of Reading. The following is a record of the proceedings of this workshop.

4.2 Aims of the workshop

The aim of the workshop was twofold. Firstly, to publicise the database that has been created to store information about existing data holdings in the UK (ATROFI-UK). Secondly, to reach a consensus on the main points to be included in a realistic archival policy and practice for such data.

4.3 Structure of meeting

The workshop was chaired by Howard Wright. An introduction to the project was made followed by presentations of the database and the web page. Invited papers were presented indicating existing archive systems in the UK. Discussions were then held covering two main topics. Firstly regarding the future of the database and, secondly, regarding future policy in the UK for tropical forest inventory data archival (see agenda in Annex 1).

4.4 Papers presented

ATROFI database

Sandro Leidi – Statistical Services Centre, Reading University

Prior to designing the database a survey was undertaken of existing databases that might have a similar purpose. Based on this survey, it was decided that the database format should have a very similar structure to TROPIS (Tree Growth and Permanent Plot Information System), using Access as an application to set up the database structure. Fields should be all those included in TROPIS plus additional ones as the information collated by the project comes from different types of inventory, not only Permanent Sample Plot (PSP) inventory. Unlike TROPIS, however, ATROFI does not always contain individual plot information.

The database was named ATROFI-UK as it represents an Archive of TROpical Forest Inventory for the UK. ATROFI-UK, being a meta-database, does not contain raw data but rather a summary of what is in the listed datasets and, most importantly, a contact address of the holder of the raw data. The purpose of a meta-database is to publicise and share the whereabouts and availability of datasets across institutions.

The structure of the database consists of 13 tables: eight main tables containing the information collected in the questionnaires, three junction tables to represent many-to-many relationships and two lookup tables (See Annex 2 for a diagram of the structure of the database).

There are eight main tables in the database, centred around the core **Study** table. This table then links to three different study type tables: Permanent Sample Plots, Natural Forest Inventories and Volume Functions. These subset tables were necessary given that the information requested for each study type differs substantially. The questionnaire layout is also different for each of the three study types.

To preserve referential integrity and so avoid duplicates, a record cannot be entered in the study type subset tables if it does not already exist in the main **Study** table.

The **Person** table contains the names, details and address, when known, of anyone connected with a study. Also the person's role is indicated, whether informant / holder, owner of the data, principal investigator or Intellectual Property Rights contact.

The **Holding** table is separate from the study table, as there may be more than one holding of the same data set within the UK. It is linked to the **Study** table by a many-to-one link via the **Person_Study** junction table. The **Holding** table describes the physical location of records in the UK and their archiving status.

The **Publications** table contains references, whether published or not, that describe any aspect of a study. For example, the protocols used or the methodology in collection of volume data.

The **VolumeFunctions** table contains details of each volume function derived. This table is also a junction table because a Volume study can contain functions for many species and the same species can be studied in many Volume studies.

The **Country** table is a lookup table containing the 235 countries with the standard two-letter abbreviation from the ISO (International Standards Organisation) list. The other Lookup table is the **Species** table, which has over 4,000 botanical names of tropical tree species, including synonyms.

Three junction tables define the many-to-many relationships between pairs of tables:

- **Study_Species** because a study can cover several species and a species may be studied in several studies;
- **Person_Study** because the same person can be involved in several studies (in the same or in a different role) and a study can have many people connected with it;
- **Study_Publication** because the same study can be mentioned in many publications and several studies can be described in the same publication.

The relationships between these tables are shown in Annex 2. Each box represents a table and shows some of the fields. A line linking two boxes represents a relationship. Fields in bold denote the "key fields" that maintain the integrity of the relationships. At each end of the relationship line is a symbol: "1" indicates this table is at the "one" end of a one-to-many relationship; "∞" indicates the table is at the "many" end.

It was decided that access to the database would be primarily online. A rather rigid search system was proposed similar to that in place for the ECN (Environmental Change Network) and FIADRS (Forestry Inventory and Analysis Database Retrieval System of the US Forest Service) database retrieval systems. It is a very efficient method for information retrieval as it forces the user to search within what is held in the database and not to specify a vague query that may yield no result.

Web page
Mark Atkinson - consultant

In December 1999 a dot com domain name was registered for the project and a home page set up. This had the address http://www.atrofi-uk.com. The ATROFI web site shows a summary of the project and its aims. The ATROFI database can also be searched on the web site. The interactive database operates on a series of text files exported from the Access database. These files are searched and displayed using Perl scripts.

Searches can be done separately for each of the three study types or all records can be viewed for each study type. Alternatively, selections can be restricted to the country where the study was carried out, by vegetation type and/or by the name of any person involved in such a study.

The searching is carried out in a number of stages:

1. A first search that gives a listing, for the study type selected, of study title, country and vegetation type. Clicking on a specific study leads to:
2. A summary that characterises the study: its design and protocol, geographical location, implementing institution, scope and year of implementation. Two further options are then given:
3. It is possible to retrieve detailed information about the archival condition of the dataset and, secondly, the contact details of the data holder, to whom questions concerning the release/use of such data should be directed.

The project home page also includes questionnaires to elicit information from interested parties about what type of data they would find useful and what they would use it for. Another questionnaire is provided for users to give details of new data sets.

What are archives and why do they matter

Janet Foster – Freelance Archive Consultant

From our brief discussion this morning it became apparent to me that you had already addressed many of the issues that I intended to cover. For example, you appear to be well aware of the importance of provenance, i.e. where the data came from, who collected it and when. But I will try to cover, from an archivist's point of view, the principle of archival and the way archives and records are managed.

Firstly **what are archives**? And **what is an archivist**? Archival material occurs where a record of activities has to be kept, these may be the decisions and functions of an organisation or of an individual. An accumulation of records from an individual or an organisation or body of individuals that cannot be found anywhere else and is worthy or preservation is an archive.

An archive may also refer to the physical repository; this is where material is processed, stored and where access to the material is provided. Despite long resistance, 'to archive' is now an accepted term and thus the word archive has also become accepted as a verb. In everyday language archive has come to mean anything that one has put away but is not using at the moment. For example, computer files may be put away, the intention here is usually to clear space and not to consider whether the files need to be kept. Looking at the data and making a decision about whether or not it should be kept is also the job of an archivist. In order for data to be archived in the official sense some appraisal of its permanent value has to have occurred.

This appraisal process is called **records management** and happens to records before they are archived. Note that all archives are records but not all records are archived. You may wish to bring this to bear on the data sets that you are looking at keeping.

The appraisal is the most difficult part of the archivist's job. For most records **retention schedules** are made. After a period of retention records are reviewed and a decision in made about whether or not they should be archived. It is first decided how long a record is current and how long it will be semi-current, i.e. its immediate purpose is passed but it may be referred to. A record may be destroyed after this time or it may be considered worthy of archival.

The primary issue in making a decision about archival is **evidential importance**. The archivist attempts to identify those records that will need to be kept to provide evidence of what was done. The second characteristic that is considered is the **information value** of the material, i.e. whether or not a record is of wider interest beyond it original use. These two provide the basis for the appraisal of records. From then one can go on to devise a summary of **criteria for retention**. These points apply equally to paper records and datasets. In assessing records for archival it is paramount that archivists are objective.

An archive should contain records that are rich, concise and limited in quantity. If it is not clear what a record is then it is useless and content is more important than age. However, we tend not to throw away anything prior to 1850.

Archivists give preference to records that provide summarised information. If there is no other use of the data then it is only necessary to keep the summary. But here one must be careful as there are instances where only the summaries were kept because it was felt that the history had been written already and the records were no longer needed and were destroyed. The written history, however, does not contain the evidence and the records should have been kept.

In addition, archivists have to be careful not to concentrate on trendy material, there is a need to be objective and take a wider view of the secondary use that might be made. As it is not possible to predict where research is to go it is hard to make a decision about what might or might not be useful. There are many examples of a particular set of records being used for purposes it was not created for and that one would not have thought of at the time the records were archived.

Archivists also involve themselves in helping the creators of records to decide what might have value in the future. In the University of Essex I have been involved in an effort to search out and save qualitative data from social research. In addition, it is important to consider at the beginning of a project whether the data that it produces will be of a value that merits archival.

As I mentioned before the **provenance** provides the context of the data i.e. its aims, where, when, any sources of existing data and why the particular project was undertaken. The question we seem to be asking here is how can archival consciousness be raised in the organisations that are contracting this work. Your worries seem to be that existing, valuable datasets may not survive.

Dataset Archiving and Data Services at the National Data Repository

Kevin Ashley – UK National Digital Archive of Datasets

SUMMARY

I have been asked to speak to you about how we go about preserving databases in a digital archive. To that end, it will be worth defining what we mean by a few of these terms so that we can be sure we are all speaking about the same thing. Then I will describe the role of the organisations I am involved with in digital preservation, i.e. what we do for the government and what we do for others. I will finish by summarising the essential information and tools that we need in order to do our job. Whoever takes on the task of preserving your database will have similar requirements to ours, i.e. these requirements are not specific to how we go about our job.

WHAT IS AN ARCHIVE?

When those who work in computing talk about 'archives', it's often not easy to be sure exactly what they mean. The term has been used very loosely. Sometimes it refers to any means of storing large amounts of data, such as a tape robot. Often it means some types of data that are more awkward to get at than others, such as tapes in an off-site storage silo. In other cases, it means data that's more than three months old which has been moved automatically to some lower-cost storage medium. The term is rarely used to reflect genuine long-term preservation with some hope of the contents being comprehensible to someone other than its creators, but that is the sense in which I think we are using it today: 'archive' as it is understood by an archivist.

Janet Foster has already given you a good insight into the essential attributes of an archive, and to the role of the archivist. The key aspects I would like to bring out are that the material in an archive usually was not intended for publication: letters and minutes rather than books and pamphlets. The archive takes on the role of preserving and describing the material and making arrangements for access to it. The selection of what ends up in an archive is important, as typically archives seek to preserve material permanently, not for 5 years, or 50 years, but forever. All these points are as true of digital archives as they are of paper, whether we are dealing with databases or documents.

WHAT IS PRESERVATION?

Preservation, however, in a digital archive is different. With paper and parchment our concern is with the physical artefact as well as the information on it. Although we can make copies on microfilm or paper, they are seen to be inferior to the original in some way. Nonetheless, the methods of preserving

paper are well understood. They involve correct environmental conditions, appropriate handling and sometimes chemical treatment of the paper to reduce acid content.

With digital information, our role is somewhat different. There is no concern here about the 'original' entity in any physical sense. A digital copy is identical to its original in any meaningful way. We are concerned about preserving the information content and about ensuring that the information continues to be accessible. This may mean altering it in some way (by changing the storage format) because of changes in technology or software. We need to decide what attributes of the information are immutable and worthy of preservation and which are accidents of technology and may be changed. For instance, the exact storage medium we use is typically not a factor that needs to be preserved. Whether your database is on a CD or a floppy is a matter of convenience, nothing else.

DIGITAL PRESRVATION PROBLEMS

Preservation is hence not as simple as it is in the paper world. The media used are not long lasting, and the methods used to store information change. We cannot examine the contents of a digital archive without an intermediary (hardware and software) to allow us to interpret it. As technology improves, people want information delivered to them in different ways.

We also need to take additional steps to ensure information is not intentionally or accidentally altered. Computer files are usually much easier to change without leaving visible evidence than is the case with paper documents. We can protect against this, but it requires specialist knowledge and techniques to do it. One of the most pressing problems is that context is quickly lost, since it often only exists in the minds of the people who created the information in the first place. Key attributes of the data, that are required in order to interpret it correctly, are often not recorded with the data if it was initially created for use by one person, or one research group.

DIGITAL PRESERVATION ADVANTAGES

On the other hand, digital archives do have advantages. We can copy material easily and cheaply, and the copies are as good as the originals. We can easily provide multiple ways to access the same archival material, suitable for different audiences. We can protect our original material against inadvertent or malicious damage more easily. We can provide access to researchers worldwide, without the need for them to visit us. We can provide very fine-grained control over who can access what parts of the archive, perhaps releasing different fields of a database, or different rows, to different people, without the need for manual methods to sift the material as is necessary with paper. We can automate the checking of our archive against decay, and we can easily represent complex inter-relationships between material in a database. All of these things are either impossible with paper, very difficult or very expensive.

WHAT IS NDR?

So why am I able to talk about this? I am responsible for managing the National Data Repository (NDR) at the University of London Computer Centre (ULCC), which is a single physical repository for large amounts of data from many different clients, along with the expertise to preserve it and provide access to it. We run a number of different services from this, with differing processes for taking material in and releasing it, different costs and different consumers for each service. Our work ranges from providing a simple 'safe-deposit-box', where we store data for an organisation without any knowledge of what the data is, and without providing access to anyone but the owner, through to a full archive service in which we catalogue the material, provide access to the public and provide support for researchers. All of this is operated on a cost-recovery basis.

THE WORK OF NDR

Tasks undertaken by NDR include:

- The conversion of material to standard forms, from proprietary ones.
- Migration to new media or new formats over time.
- The provision of controllable access, and auditable access, so that specific groups can be authorised to get at specific material.
- The provision of search facilities and their integration with other search portals and resource discovery networks (as they are now known).
- The cataloguing and/or indexing of material.
- The gathering of contextual information about material.
- The provision of user support to researchers wanting to access archive material.

I would stress that any or all of these tasks could be undertaken by the original owners of the material, or some other group with an interest in it. We can offer all of this, but we can be involved in only part of it if someone else undertakes to do the rest better or more cost-effectively.

WHAT IS NDAD?

NDAD (the national digital archive of datasets) is the largest single client for the NDR. It is operated for the Public Records Office of England and Wales (the PRO) under a private finance initiative contract, the 'private finance' in this case being the University's own reserves. Our task is to deal with public records which take the form of datasets of one sort or another. The PRO and government departments are responsible for selecting what material is to be preserved (only 2% of paper records are preserved on average, the rest being destroyed at some point in their lifetime). Our role begins once the decision is taken to preserve a specific system: we deal with acquisition, conversion, cataloguing and everything through to provision of public access. We deal both with material that is available for reading by all and with highly confidential material which will not be open to the public for anything up to 100 years from now. Although our primary source material (the databases) is digital, we have to deal with a lot of paper as well, as much of the essential contextual documentation only exists on paper. We are dealing with computer systems old and new, from 1960 to the present day.

Amongst the information we look for is what might be described as micro meta-data: information describing, at the lowest level, the data types of individual fields within a dataset, the descriptions of these fields, any ranges or other restrictions that apply to their values, and so on. In many cases, this information is embedded in the database or application itself. In some cases, particularly where bespoke systems are involved, the answer may lie in the source code of the application.

META DATA

We also look for technical meta-data on a broader level, describing the inter-relationships between the elements of the dataset and the capabilities of the system used to process it. This can include information on what methods could be used to retrieve data (use of key fields, soundex name searches, and so on) and what reporting capabilities existed. Information on provenance and use is also sought, and on policy matters relating to the establishment of the system and what effects its use might have had.

The sources for the metadata are varied. Modern databases may have much of the metadata embedded inside the system, but more often we are looking at supplier and user documentation, internal organisational records, publications, our own specialist knowledge and oral history collected from those who created or used the data.

DIGITAL ARCHIVAL COSTS

I've been asked to say something about the cost factors involved in digital archival work. It is a very complex subject and one I will confess I still don't fully understand myself. However, amongst the factors that certainly have an influence are these:

- Do resources arrive in neat bundles?
- Is metadata attached to the data when it arrives?
- What are the access patterns: frequent or infrequent?
- Is contemporary knowledge of the data available?
- Size of the user base and the number of accessions
- The level of support required for depositors and users

What generally does not influence costs is the raw amount of data. For us that is about 1% of our total operating costs. To model any of these you need to start with a service model of what you are trying to do, and who you are doing it for.

NDAD SKILLS

The skills we have available in NDAD to do our work are varied. We have professional archivists working alongside specialists in databases and their uses. These are backed up by systems specialists who ensure data is kept safe and secure in systems that work 24 hours a day, and user support staff who deal with research queries. Our aim is not simply to preserve a set of numbers, but to document how the information was used, why it was collected, what influence it had and the context in which all of this took place. The information we gather is thus both historical and technical in nature.

As I said earlier, one can conceive of ways in which some of this work is done elsewhere by other groups, perhaps with a greater specialist knowledge about some of the archive's holdings. The

possibilities are endless. Technology is certainly not the constraint any more in dealing with the archiving of computer-based material. The barriers facing us are more often organisational and motivational, and funding, as always, can also be a concern.

Data management and archiving, the Statistical Services Centre experience

Roger Stern – Statistical Services Centre (SSC), Reading University

INTRODUCTION

The problems of data archival in the forestry sector are also experienced in other areas. Examples that come to mind include:

- climatic data archival (The CLICOM project that has been running since 1983),
- agricultural research data (ICRISAT (International Centre for Research in the Semi Arid Tropics) and DFID work)
- agroforestry data (joint work with ICRAF (International Centre for Research In Agroforestry))

DATA MANAGEMENT

In considering data management there are three main topics that need to be covered:

- Inventory of data that are available. This is being covered by the current project creating a database of the meta-data.
- Rescue of the historic data. This is not being covered by the current project but it is important. Current projects undertaking this kind of work include the DARE project for climatic data looking at the Meteorological office holdings. This is not an easy task.
- Management and archiving of current and future project data. This is the main subject of the work at SSC and is the second objective of this workshop.

There are two separate problems here. Firstly, how can the data be managed and archived for efficient work on the project and secondly, who owns the data and what rights do individuals on (project) teams have to the data. Both of these issues are important but it is possible to examine them and attend to them separately. Too often technical discussion is sidetracked by ownership issues. Here, at SSC, we only consider the technical issues, as the solutions are neutral to ownership.

The SSC are the statistical advisors for DFID natural resources projects and they produce good practice guides for data management as part of their proactive work. These guides cover mainly:

- Design of research studies
- Data management (and archiving)
- Analysis
- Presentation

They are all available as free booklets and are also available on the web in html, pdf and help file formats. The most popular guides and best received tend to be the data management guides. Two of which were produced in 1999 entitled:

- ‘Data management guidelines for experimental projects’ and
- ‘Project data archiving – lessons from a case study’.

Two more have just been reviewed and will be available from April 2000 entitled:

- ‘Date entry using a spreadsheet’ and
- ‘Excel for statistics – tips and warnings’.

A further guide to be reviewed in 2000 is:

- ‘Data management – when should you use a database system – Microsoft Access’. (All these guides are available for reading online or downloading at: <http://www.reading.ac.uk/ssc/dfid/booklets.html>)

At the SSC our main aim is to support statistical work but poor data management is the main stumbling block. So in the past two years much of our work has been concerned with support on data management, our involvement in the ATROFI project is an example.

In our experience projects tend to take the view that they can manage data as they go and then leave a copy of everything behind for archival. This is not ideal as it adds extra work in describing and sorting out which data needs to be permanently archived and does not benefit the project team. We propose that new projects should concentrate on establishing a good data management strategy and not just

become sidetracked by archiving. If data are considered generally and are well managed then the project should proceed more smoothly and the archiving work reduced.

The solutions that we suggest for data management are:

- Standard database software can be used for project data but there is little guidance on the use of database software for scientific applications (e.g. the 'moving to access' booklet).
- Special software can be used instead. For example the Epi-info software is very good and very popular. The new version should be released soon and it is free.
- Projects could also consider the types of data they might collect. They can then ask whether a common database structure can be defined and constructed for them that copes with data from different studies. This will help with single analyses and also makes it easier to combine data from different studies within the same project.

The latter option was considered by ICRAF and ICRISAT in the mid 1990s. The work still continues at ICRAF and has led to the Logbook software, which is promising for the future.

CONCLUSIONS
In conclusion it is clear that for all projects constructive work on archiving is only possible when we recognise that there is a problem and that a good solution is difficult and time consuming (as demonstrated by this project). For new projects archiving should be obligatory, as it is important. However, it is less important than good data management during the project and we need to recognise that this is a difficult task.

Inventory and Data Retrieval – the view from NRI

Trevor Abell – Natural Resources Management Division, Natural Resources Institute (NRI)
Andrew Larkin – Librarian, Natural Resources Institute

BACKGROUND
NRI came into existence in 1988, following the move to Chatham and the amalgamation of the Land Resources Development Centre (LRDC), the Tropical Forest Products Institute and the Centre for Overseas Pest Research. LRDC, located close to the Directorate of Overseas Surveys at Tolworth (from which it was initially formed), was involved with natural resource assessment including forest inventory and it is therefore the work of that organisation that is particularly relevant to this meeting when looking at historical forest inventory data.

During the period that LRDC operated, forest inventories were carried out in: -

Ethiopia, Ghana, Gambia, Nigeria, Tanzania, Sudan, Belize (British Honduras), Bangladesh, Jamaica, Fiji, Indonesia, New Hebrides, St Helena and Solomon Islands.

After the formation of NRI, inventories involving our staff have been carried out in:

Belize, Ghana, Guyana, Somalia, Kenya, Malawi and Indonesia.

Additional studies could also be mentioned in other regions where forest mapping and resource assessment has been carried out, though not necessarily including inventory.

During the LRDC era and for the early years of NRI, up to 1996, the organisation was a fully government organisation that was administratively part of the Overseas Development Administration (ODA); in fact, for a few years, we had the rather awkward name of Overseas Development Natural Resources Institute. Work was undertaken with funding from ODA with occasionally financial involvement of other multi-lateral organisations such as the World Bank and FAO (Food and Agriculture Organisation of the UN).

DATA HANDLING
The forest inventories undertaken could be either the product of specific forestry investigations completed as a short to medium term exercise or the component of a long-term multi-sectoral land resource study. In all types of investigation there was some degree of involvement of local

organisations – normally representatives of the forest department in relation to the work of forest inventories. Fieldwork was usually undertaken as a combined exercise between the local forestry staff and the LRD/NRI foresters.

With respect to data analysis and report preparation, there was not a uniform approach, but generally with the longer-term regional studies arrangements would have been made for data to be analysed in country. For the shorter, more localised studies, data analysis and report writing was undertaken often within the UK.

In terms of handling and storage of inventory records this means that in some instances the full field records were brought back to HQ, whereas, in other cases, all the original field records were left in-country together with many of the summary sheets and computer records. Only the final reports and associated annexes might be returned to the UK. Fortunately, in several instances much of the work in terms of volume equation derivation and construction of stand tables was often undertaken with the assistance of OFI (Oxford Forestry Institute), so Oxford has become a repository for at least some of this inventory data.

THE LIBRARY AND RECORDS

Following the transfer of NRI from the civil service to the University of Greenwich (UoG), a decision had to be made on what should happen to the extensive library (the largest single library of tropical natural resources in Europe) and all the other records. Funding support for the library gradually ceased from DFID and the library is being incorporated into the Medway Campus library of the University. The former components of NRI brought their own collections of forestry material, although the LRDC had by far the largest amount. For the most part, the three collections are still housed separately at Medway, but there is an on-going programme to transfer material onto the UoG catalogue. At the moment, the interest is the post -1980 material and it is unlikely that the older archive material will ever be computerised and searchable on one single database

All the LRDC collection (i.e. reports, papers, journal articles and books) has been computerised and available on TRADIS (Tropical Research and Agricultural Information System), a searchable database mounted on the CAIRS system (Computerised Agricultural Information Retrieval System). This database is no longer networked but can be interrogated on a standalone server. This database was compiled only from material held in-house. Field records were never itemised and included on the database – unless they had been converted in some way into a recognised report annex and could be entered as such. There are around 5,000 documents on the TRADIS system with a forestry flavour. The TPI (Tropical Products Institute)/COPR (Centre for Overseas Pest Research) material was placed on a similar in-house database called TRAIS and has around 4,000 items of a forestry nature. The older pre-1979 material can be accessed from traditional card catalogues and this goes back to the 1890s. A separate technical card index again covering the interests of TPI since its formation has around 25,000 references for “forestry” – the majority relate to individual tree species.

At present there is no specific policy for the forestry collection. The NRI collection is managed as a whole with no bias to particular areas. There is an obligation to maintain this collection for future reference and there is considered to be no risk of withdrawal or disposal of the material. TRADIS is being transferred into the UoG system, but there is still a considerable backlog in this work.

There was great pressure on storage space and initially many of the field records were despatched to a central university store with inevitably some disposal of that material considered to be of no further interest. Fortunately, it seems that much of the former LRDC material was saved and was finally properly archived and put in the care of the Public Records Office at Hayes, Middlesex. A total of 300 feet of shelf space was despatched to Hayes in 1999 – this includes all disciplines not just forestry. All correspondence files were considered to Government property and have been separately stored in the Public Records Office.

Reports and analyses have been produced from much of this forest data, but making further use of the raw field data in some other new way, although theoretically possible, in practice will give considerable problems unless the researcher can identify adequate notes to provide clear guidance on inventory design and any coding of records that has been used. All the inventories that have been undertaken since the transfer out of government ownership have been conducted as part of long-term studies and field records will be left in country.

In essence we have:

- A comprehensive library database system – but not a single one-stop search tool for all the collections.
- A continuing programme to merge all the older collections into the UoG system, but this is unlikely to be achieved for the oldest material – at least in the foreseeable future.
- Inventory records and field cards are no longer on site but have been transferred to Hayes. Presently, no overall database for the field records.
- Collections of field cards are known to be patchy and dependent on the level of partnership with in-country organisations at the time the inventories were being conducted.

THE FUTURE

Ideally for any future inventory work, it would be sensible to ensure that a distillation of the key facts are submitted to a web-based database – perhaps under the control of FAO. Data summarised in the form already being undertaken by ATROFI-UK would seem to be ideal: -

Location, forest area, forest type, main species, type of inventory, sampling percentage, parameters recorded, type of data storage and the volume regression equations calculated.

In addition, investigators should be asked to supply stand tables, species lists and overall volume estimations, specifying whether this is for commercial or total volume

FOREST INVENTORY REPORTS – NATURAL FOREST

BOTSWANA

Forest Inventory and Management in the Baikea Forest of North-west Botswana. P. Henry, Project Report 44.

ETHIOPIA

Southwest Ethiopia Forest Inventory Project : an inventory of Magada Forest. D. Chaffey, Project Report 28, 1978 revised 1980.

Southwest Ethiopia Forest Inventory Project: an inventory of Munessa and Shashemane. D. Chaffey, Project Report 29, 1978 revised 1980.

Southwest Ethiopia Forest Inventory Project: an inventory of Tiro Forest. D. Chaffey, Project Report 30, 1978 revised 1980.

KENYA

Kenya's Indigenous Forests –Status, Management and Conservation. Editor Peter Wass, ODA – IUCN 1995, Over view of the KIFCOM project.
Plus Separate forest inventory reports 1-15, Hugh Blackett 1994.

MALAWI

A Report on the Inventory of Dzalanyama Forest Reserve. Unpublished report (T. Abell), Forest Planning Unit, Lilongwe, 1995.

An Inventory of Miombo Woodland in the Lower Shire Valley. T. Abell, 1993.

TANZANIA

Tabora Rural Development Programme, Wood land Ecology Reconnaissance survey, R. Lawton Project Report 76, 1979.

SUDAN

Report on the Forest Development Prospects in the Upper Kinyeti and Ngairigi Basins, Imatong Central FR. Jenkin, Howard, Thomas, Abell and Deane, Land Resource Study 28, 1977.

SOMALIA

Charcoal in Somalia: A Woodfuel Inventory in the Bay Region of Somalia. Neil Bird and Gill Shepherd, Reference B009, NRI, 1989.

GAMBIA
Inventory of the Mangroves above the Proposed Gambia River Barrage at Yelitenda, Gambia. M. Johnson, Project Report 54, 1978.

NIGERIA
Land Resources of Central Nigeria. W. Howard, Land Resource Report 9, 1976.

BRITISH HONDURAS/BELIZE
Inventory of the Coastal Plain of British Honduras. M. S. Johnson and D. Chaffey, Land Resource Study 15, 1974.

Inventory of the Chiquibul Forest, BH. Johnson and Chaffey, Land Resource Study 14, 1974.

Inventory of the Mountain Pine Ridge, BH. Johnson and Chaffey, 1972.

A forest Inventory of Part of the Mountain Pine Ridge, Belize. *LAND RESOURCE STUDY 13, 1973.*

Belize Mountain Pine ridge Forestry Project. J. Sandom, Project Report 209, 1987.

Sustaining the Yield: *Improved Timber Harvesting Practices in Belize 1992-98*. Neil Bird, 1998.

JAMAICA
An Inventory of the Carib Pine Forest in Central and Eastern Jamaica. M Johnson, D. Alder, M. Jefferson, Project Report 81, 1981.

BANGLADESH
A Forest Inventory of the Sunderbans , Bangladesh (3 volumes). D. Chaffey, D. Miller, C. Myers, J. Sandom, Project Report 140, 1985.

HAITI
Method of Inventory - Pine Forests in Haiti. M. Berry and K. Musgrave, Technical Report 2, 1977.

GRENADA
Forest Inventory in Grenada. M. Johnson, Project Report 169, 1985.

FIJI
Fiji Forest Inventory Vol. 1- Environment and Forest Types, Vol. 2 –Catchment Groups of VitiLevu and Kanavu Vol. 3 – Catchment Group of Vanua Levua. M. Berry and W. Howard, Land Resource Study 12, 1973.

NEW HEBRIDES
New Hebrides Codominium, Erromango Forest Inventory Project. M. Johnson, 1968.

INDONESIA
The Forest Resource: Site Investigations along the Trans-Sumatra Highway. Project Record 40, T M Abell, 1979.

SOLOMON ISLANDS
Field Sampling of Camposperm Forest Santa Isabel, BSIP. T. Rees LRD Miscellaneous Report, 1964.

Archiving Census and Woodland Inventory Data
Graham Bull – Woodland Surveys, Forestry Commission

BACKGROUND
The Forestry Commission is responsible for administering grants to Private Woodland Owners in England, Wales and Scotland. Forestry in Northern Ireland is covered by DANI (Department of Agriculture Northern Ireland).

DATA HOLDINGS
Information retained by the Forestry Commission of relevance to this workshop is as follows

Woodland Grant Scheme
The Woodland Grant Scheme (WGS) database is continually being updated as new schemes are applied for. Old data is not deleted and regular backups are made but there is no intention of archiving data on the WGS. The database can be used to make queries about active grants for a particular date.

Constraint mapping
Coloured paper maps at 1:25000 and 1:50000 have been retained showing the earlier grant schemes and constraints e.g. SSSI's, AONB's, NP's, TPO's etc. Digital data snapshots have been saved on Optical Disk or DAT (Digital Audio Tape) in a fire proof safe and off site.

Subcompartment database
Forest Enterprise (FE) manages the State owned forests, as from April 2000 will take six monthly snapshots of subcompartments and this data is stored on the subcompartment database (SCDB). Woodland Surveys have archived 1995, 1996, 1998, and current 1999 subcompartment data on SCDB.

Stock maps
The FE started saving digital versions of stock maps in 1997 but very few old editions have been kept.

WOODLAND SURVEYS HISTORY
The Forestry Commission (FC) was formed after World War 1 in September 1919. Since then there have been five main censuses or woodland surveys where purpose has been to provide information about woodlands in the UK at the regional and national level. These surveys are summarised in table 1 below.

Table 1. UK Woodland Surveys

Year	Woodlands covered	Plot size	Method of plot location
1924	FC and other	0.8 ha	Questionnaire
1930 1938	FC and other	2.0 ha	Sampling
1947	FC and Other	2.0 ha	Complete plus small woods
		0.4 ha	Sampling
1965	Other	0.4 ha	Sampling
1980	Other (except dedicated and approved)	0.25 ha	Sampling plus non woodland trees
1994			Pilot for NIWT to present day live data

1924 Census
Copies of report are available in the research library at Alice Holt. This includes a copy of the questionnaire used within the report but no map data is available.

1930 and 1938 Census
The 1930 census was a survey based on the 1924 questionnaire and copies of the report are available but there are very few documented details of the data that was used. The 1937 survey was never fully completed due to the outbreak of World War 2. A map is provided of areas (in Scotland) that were surveyed. The survey data is available for counties and an incomplete set of maps is also available.

1947-1949 and 1951 Census
The 1947 census was a complete survey of all woods over 5 acres including small woods and hedgerows surveyed in 1951. Reports are available for England Wales and Scotland respectively. Photographic copies of the 1:10560 maps and main published reports have been retained as well as an incomplete set of record maps. The latter provide details of woodland type surveyed. Including: Coniferous high forest, Mixed high forest, Broadleaved high forest, Coppice with standards, Coppice, Scrub, Devastated, Felled, Lost, Thorn colonisation.

The 1951 hedgerow survey was an assessment of Hedgerow volume carried out on sample woods between 1 and 5 acres. 441 maps were selected and a part of each sheet was assessed for small woods. All data and maps related to this survey are stored in PRO.

1965-1967 Census
Only one national report was published and copies are available in the FC. Unpublished county summaries are stored in county offices. FC management boundaries have gone through various changes over the years, and variations are contained within the annual reports, note that the English conservancies are due to change again on 1 April 2000.

1980 Census
The 1980 census produced county, country and national reports which are all available. Reports were also produced for the conservancies of the day i.e. North, South, East and West Scotland etc. All

reports have been saved to microfiche and microfilm. Field maps at 1:10560 and 1:10000 have been distributed throughout the FC.

The forest research library at Alice Holt has a collection of ground photographs taken for this survey. Aerial Photo's by county have been donated to the Royal Commission on the Historical Monuments (RCHM) of Scotland, RCHM of England and the Air Photo Unit Welsh Office, Cardiff. The FC combined information from aerial photographs and Ordinance Survey green plate (purchased on film) to mark up 1:50000 maps with FC dedicated and approved woodland. These were then scanned by Laser Scan (a Cambridge based company) using Laser Track (also used for banknotes) and the data is saved on 9" magnetic tapes.

Current National Inventory of Woodland and Trees
A pilot inventory was started in the Grampian Region Several changes were made to the original design, i.e. moved to sampling 100 Km tiles and to surveying all woodland regardless of ownership rather than splitting woodland types between FC and Other. The main woodland survey covers all areas of woodland greater than or equal to 2ha. In addition a survey of small woods and trees is included covering small woods of 0.10 – 1.99 ha, groups of trees, linear features and single trees.

The main woodland survey made use of 1:25000 photography (LCS88 in Scotland and in England and Wales the FC either purchased existing cover or commissioned contracts to obtain photography). This photography is being distributed to conservancies for further use once field data has been validated.

The Grampian report and data is now lodged with the PRO via ULCC with a 30-year restriction on public access to the data. An Oracle database of sample woods and 1 ha squares is to be archived. In addition digital map data (Vector and Raster) are all archived off site on CD-ROM and DAT tape.

County reports in England andWales, regional reports in Scotland, country reports and Great Britain reports are to be produced. Reports will be available on the FC web site. MS Office versions of reports are to be archived as well as data capture field manuals, field notes, field maps, statistical programmes, and VAX/VMS text files.

DATA USE
The following is an example of how the data is used to monitor and compare change.

The 1980 census revealed that there were 241,000 ha of woodland (>0.25 ha) representing an 11.6% land cover. The 1998 Forestry facts and Figures report showed that there was 247,000 ha of woodland (11.9% cover). The 1998 inventory of woodland indicated that there was 271,000 ha of woodland over 0.2ha (13.1% cover) and 286,000 ha of woodland over 0.25 ha (13.8% cover).

DISSEMINATION
The Woodland Surveys site (www.forestry.gov.uk) is currently under construction within the Forest Research site. It will include links to Local Authorities and other related sites. The FC has a policy to provide information on woodlands, to monitor change and make comparisons with the earlier records.

Maintaining Forest Data for Future Use – A Commercial Perspective
Paul Smyth – Huntings Technical Services

BACKGROUND
Since the 1950s, Hunting Technical Services Limited (HTS) has provided technical assistance in rural development initiatives. In this role, over 1300 studies and projects have been completed in more than 130 countries. This work has included many dedicated forestry sector assignments. All HTS projects generate data of one kind or another. These data are most commonly presented in text and tabular form as technical reports and their appendices. Many studies also involve the making of maps, generally produced through air-photo or satellite image interpretation, supported by ground survey. In recent years, map production has been achieved through digital cartography.

DATA COLLECTION AND DATA ARCHIVAL
Archiving of all this data is an issue that undergoes frequent consideration by HTS. All project reports are kept indefinitely, either in the Company library, or at a storage facility off-site. Working project documentation is retained in our stores for a period of five years, after which there is an active microfiche programme. By and large, copies of all maps either used or produced by the company are kept for internal use and for reference. Similarly, photographic negatives that allow the reproduction of

satellite image mapsheets have been archived. The storage of digital satellite data and digital maps is more problematic. Air-conditioned storage conditions are required, and digital media and data formats can become obsolete. Additionally, the integrity of data held on these media cannot be guaranteed, as there is no procedure for regularly checking readability.

Occasionally a data archive constitutes a deliverable project item, with the intention being to hold the data archive in the recipient country. In such cases, there has been relatively little additional effort involved in creating a second archive for storage at HTS head office.

While we are aware that the forest data accumulated by HTS may have a value that extends well beyond the time frame of the particular study, it is becoming increasingly difficult for us to ensure the long-term maintenance of an archive of our project data.

At HTS we are involved in fairly broad-based natural resources consultancy. In many projects, this involves map making, often using remote sensing data and GIS (Geographic Information System) analysis. There is a general rule of thumb that all data collected on a client's behalf by us remains the property of that client.

The forestry studies I have been involved in myself have been national reconnaissance level mapping projects. In such projects, data have been collected at known-location Temporary Sample Plots (TSPs) in order to characterise the vegetation classes that we have been mapping. Vegetation data is collected at these TSPs and this includes height, density (canopy cover), trees per plot, stems per plot, stem diameter and bole length, if appropriate.

Data collected in the TSPs were used to characterise the vegetation classes that we were mapping, and to assist in the interpretation of areas that we were not able to visit during the field data collection phase of the project. Sample TSP data were fair copied and included in an 'Interpretation Manual' that was a project deliverable (Yemen). I do not know what has happened to the completed field sheets for this project. As the Interpretation Manual was compiled in the UK, I suspect that these have been filed and will be stored at an archival facility off-site.

For Tanzania, all raw field notes were filed and deposited with the Client (the Institute of Resources Assessment, University of Dar es Salaam, Tanzania).

I am not sure, but I would think that field data and working (non deliverable) reports are usually kept; these would be boxed up at project end and sent to the head office where they would eventually end up on the shelves of archive agencies. The (remote sensing and GIS) data sets that we are involved with tend to be larger than those you have been talking about and there is a severe competition for appropriate storage space. We are currently in the process of deciding which digital data to dispose of, given that there is a transfer of use of our climate controlled room, from image-processing lab and tape store to a home for our networking routing equipment.

Project reports are kept indefinitely either in our office or off site. Working documents are retained for five years after which there is an active micro-fiching system. Copies of all the maps made or used are kept in the office for internal use and reference. In the past have we kept them as aperture cards and as film positives.

Digital data archiving is much more of a problem as it tends to require expensive conditions such as air-conditioning. In addition, regular maintenance is needed such as winding on and rewinding Computer Compatible Tapes (CCTs). In the past, we used old imaging systems, data from which are no longer readable without effort on our part to translate the format to something more modern.

CCT data of value has been transferred to new media; there is a certain amount of effort involved in doing this and there is not necessarily any commercial reward for this. We also cannot guarantee the readability of the data. Part of the problem is that propriety formats of the data may not be readable by current systems. There is a cost associated with getting such data into new formats and onto modern media.

For example, some GIS datasets or map composition files contains pointers to other files in other sub directories, and when files are moved the pointers do not work any more. Care must be taken in the archiving of these kinds of files.

Some of the projects have specific instructions on archiving in the project document. For example, in mapping the woodlands of Tanzania, we were required to provide the maps in a medium that they could read and we had to keep a duplicate copy. But such specification is often not the case and there is always the risk that data is may be inadvertently lost or destroyed in country. So our working practices apply to material that is filed at our head office but not to material or data that is left in country. If the data does not come back to head office then the only possible access is via the reports that have been written.

There have been several conscientious individuals in the company who have in the past taken all the old data and kept it in their houses if they feel it is threatened by restructuring.

However, as shelf space is limiting, our archiving policy, and health and safety regulations may limit the physical amount of data that we are able to store.

Volume/Biomass: Geo-referenced Forest Volume Data for Tropical Countries

Alessandro Baccini – Independent Consultant for FAO FRA2000

The FAO FRA 2000 project (Forest Resource Assessment) aims to collect information on forest cover worldwide in order to review the state of the world's forests and to track changes. The project has been active for three years and is now beginning to produce results.

The FAO have a database that is similar to ATROFI but is more basic as it only includes data that is useful for the FRA2000 project, that is spatial volume data. The database includes two tables, one for general information and another providing further details that are added if the data is considered to be of use. When selecting data that might be of use it is necessary for the data to be accompanied by a map and this we found became one of the main constraints on data set selection. If the data cannot be geo-referenced then it cannot be used to determine spatial volume. A second criteria for selection was that all species should have been inventoried as we are interested in total volume. Likewise, it was important that the minimum diameter taken in the study was less than 40cm. Finally we had to be certain that the data was reliable.

This database exists in dbf format. The database was converted to Access in order to produce a report of the work that has been done and soon the data will be moved to an oracle format for maintenance purposes. The meta data is not stored separately from the data itself but the data is linked to a GIS. This makes it possible to cross link data with other information such as climatic data. The data is fed into a model produce spatial forest volume information. The team has also developed a land use index from which one will be able to derive national level statistics.

This briefly explains the data needs of FRA 2000 and demonstrates an important use for the information collected on ATROFI. ATROFI, however, does not explicitly include spatial information and this would be of great value as the demand for data that can be used in spatial modelling is steadily increasing.

The method of data storage and analysis use by FRA 2000 has the advantage that both new and old data. In the recent past there has been a tendency to collect less forestry inventory data and instead to concentrate more on land use mapping. This means that the old inventory data becomes more important and valuable. The older data provides a better indication of the potential carrying capacity or yield as more recent inventory data comes from forests that have already been logged

The report of this work is available from the FAO. In addition if you want a copy of the forestry paper 134 then we can also make it available.

The best data is at the plot level but we only had this for a few countries. Considering the scale of the work, the fact that it was global and we were using 1km square resolution we considered almost

everything we could collate. I did not include forest inventory summaries for areas bigger than 3 million ha. But this also depended on the countries.

Managing environmental data

Geoffrey Collett – Environmental Information Centre, Monks Wood, Natural Environment Research Council (NERC)

The Centre for Ecology and Hydrology (CEH) has recently been formed from the old Institutes of Terrestrial and Freshwater Ecology, The Institute of Hydrology and the Institute of Virology and Environmental Microbiology. The Environmental Information Centre (EIC), at CEH Monks Wood, is the Designated Data Centre with responsibility for managing those NERC Terrestrial and freshwater datasets not catered for by the National Water Archive. EIC is also responsible for developing data strategy policies and procedures.

EIC has a variety of data sources from large thematic programmes, through science programmes and NERC grants to voluntary recorder schemes. Data management within CEH also occurs within data centres that have specific responsibilities such as the Environmental Change Network and the Biological Records Centre.

We produce data management plans for the thematic programmes. These cover Intellectual Property Rights (IPR), copyright, Principal Investigator responsibilities, metadata, quality assurance, recommended data formats and media, stewardship *etc*. Each project has to prepare a data management plan. It has been suggested that the last instalment of grants should not be paid if the data management plan has not been prepared! It is intended to implement these data management policies across CEH.

EIC's role includes ensuring adequate storage, stewardship, validation and dissemination of data. We also supply advice on quality assurance. When data is passed on for storage we require that data is deposited with adequate documentation.

We are also seeking out datasets for recovery and NERC has provided funding until recently for data archaeology. This activity often has an unexpected cost in the need to enlist senior staff to investigate the data from an expert point of view.

Discovery Metadata catalogues are important for data dissemination giving basic information to enquirers about the who, what, when and where of the dataset. EIC has a metadata search engine on the web that looks at an underlying Oracle structure which is intended to contain records on all the CEH data sets.

In handling data requests we keep in mind that the Environmental Information Act says that environmental data has to be publicly available (but not that it has to be freely available). The EIC offers a data sales and licensing service for CEH data products. Our policy is to make data that NERC owns, freely available to bona fide academic researchers.

We also provide guidance on data management to our various research arms.

It is our brief to encourage public interest and awareness of environmental data. Recently we went live on the web with a phenology data set that should could catch public interest. It allows user to update a database of phenological events such as the first flowering of snowdrops. Other examples of widely used products are the Countryside Information System, the Countryside 2000 dataset and the Landcover map.

4.5 Discussion

The ATROFI database

IS ATROFI USEFUL AND COMPLETE ENOUGH?

The ATROFI database, in its present form, was considered useful by all present. The following suggestions were considered appropriate for future development of the database:

- A global map could be provided as part of the front-end that would permit immediate assessment of the geographical range of datasets included.
- It was suggested that it would be useful to include information on the existence of stand tables, species lists or volume estimates that the dataset had been used to produce. It was agreed that the references to literature could provide this information as adding this to the database would be time consuming. It is intended that grey literature should be scanned into the database.
- It would be useful to have a field in the database stating what the data has been used for in the past as well as what it was collected for. The latter already exists but explicit objectives are not included.
- Although time consuming it would be useful to link the data with GIS systems by providing geo references for each data set.
- The database should include records of data sets that we are aware exist but that we have not yet been able to describe in full. Users could be asked to provide any additional information they might have on these data sets. This is particularly important for data sets we are worried about losing.
- It may be useful to add a field to the database that provides information on the amount of assessment that is still required to understand the data set.
- In addition a field could be added describing what the long-term risk is to the storage of the data and how valuable it is.
- It is important to collect feedback from the users on the usefulness of the database. A short questionnaire making such enquiries should be included on the web page. It is important to publicise the database a lot to stimulate interest.

FURTHER WORK ON ARCHIVAL

The project did not complete the collection of information on data sets held within the UK. Most of the relevant individuals have been contacted but work is still required on the data at Oxford and the NRI data held in Hayes. The latter has not yet been seen. In addition there may still be relevant data stored in geography and/or zoology departments that we do not know about. Although many of these have been contacted, when working across disciplines it sometimes takes several enquiries to different people within a department before data sets are discovered that fit our criteria. In some cases there may be only one person holding this kind of data in a department. Other sources of data that have not yet been looked at include management plans and MSc and doctoral theses.

It was agreed that we should concentrate first on data that is at risk of being destroyed or getting lost. In addition, with regard to data stored in Universities or other institutions, the people who know about the data and who are key informants on the data are fast disappearing. Theses and other publications stored in libraries are relatively safe and the data should already be adequately described, filing cabinets full of data, however, are less secure and require input from more informants to understand the data. It was suggested that institutions should be asked to keep data for the next five years to ensure that no further valuable data is lost before it can be described and assessed for archival. Alternatively, it was suggested that there should be an intermediary holding place where data can await assessment before a decision is made about its value and archival.

There is a need to continue with the information collection. The amount of work required to collate information about a data set can be extensive and depends significantly on whether or not key informants are available. It was suggested that the time required to describe a data set (OFI and NRI) should be tested using one or two data sets. In some cases significant work may be required to 'rescue' data sets and this needs also to be assessed. In addition, the process can be assisted, if the web page is publicised and a questionnaire is provided requesting further information about data sets already in the database. Once this has been completed the next logical step is to start including information on tropical forest inventory data held in Europe.

VALUE OF DATA

As well as describing this data the problem of finding an appropriate place to store it is also of concern. In order to make a case for safe storage, we need to be able to determine and describe how important these data are. The value of this kind of data needs to be assessed and some data sets will be more valuable than others. Some criteria need to be developed to value data sets.

In general the data is considered valuable because:

- Collecting forest inventory data is expensive and the cost of storing it pales in comparison. In most cases the data cannot be re-created.
- Some data loses its value as it gets older but the value of forest data increases with time, reaching a peak at about 100 years (depending on the rotation or the life expectancy of the trees).
- The best data is that which provides quantitative measurements at the plot level. Plot summaries and forest summaries are, however, also of value.
- The number of re-measurements and hence the time span increases the value of the data.
- If plot locations are adequately reported then re-measurement is possible. Thus the potential value of the data is increased.
- Much of the older data provides the only information left about the natural 'climax' vegetation of an area.
- Older data may also be the only source of data on potential volume for a site as, since the 1940s, most of these areas have been logged at least once. Older inventories tend to report higher volumes.
- The potential uses of this data need to be listed.
- Ground truthing for remote sensing data.
- Modelling environmental and climate change.

STRATEGY FOR MAINTENANCE OF DATABASE

The database needs to be maintained after the end of the project. The database should be maintained at an institution with a small continuing grant provided. The question is where should the database be held and who should be approached to fund it?

In considering the location for the database the need to survive institutional changes should be considered. The possibility that FRP might be able to provide funding for further maintenance was considered. The inclusion of more data sets in the database would increase its value, enabling it to attract funding for future maintenance. The cost of maintaining the database needs to be investigated; note that the cost of creating the data far exceeds the cost of preserving it.

Funding could be sought from the EU (European Union) for expansion of the database to cover data sets stored in Europe.

The database should be put onto a CD-ROM and needs to be publicised widely, overseas as well as in the UK. This publicity should provoke feedback and queries about data sets and increase potential interest from funding agencies.

CONTACTING DATA OWNERS – ACCESS TO DATASETS

Where data ownership is unclear the project had originally intended to approach source countries to ask about permission to use the data. However, it was decided that instead source countries would be provided with a copy of ATROFI, informing them of the existence of particular data. It would then be up to them to request copies of the data. Potential users of data would be required to contact the source countries for permission to use the data.

This system was chosen because firstly it was felt that it would be difficult for the project to locate officials senior enough to grant blanket permission for data use. Secondly, officials are unlikely to give blanket permission but might be able to give permission on a case by case basis, i.e. for certain uses. Thirdly, and consequently, if approached for blanket permission for data use, officials are more likely to place a total ban on use of the data.

We are also aware that some data sets held in the UK are sensitive and are sometimes being held unofficially. These need to remain confidential at present but should be kept securely as in 20 or so years they will still be of high value, they may no longer be sensitive and they may be the only copy left.

If an originator country decides that the data should not be stored in the UK or in an international organisation, negotiations on ownership will have to be opened.

Policy

EXISTING SITUATION

Data collected by the Land Resources Development Centre, now amalgamated into NRI is currently awaiting assessment for archival value by the PRO. It has been placed in a set of boxes stored in Hayes. This data is not secure and is not yet documented or described. The project needs to make sure that no decision will be made on this data until it has been assessed by the project. DFID is the Government Department that is currently placed to make a decision about this data. If the PRO is made aware that this data is important but DFID decide that the data does not merit archival, the PRO will present it to other organisations for assessment. The PRO needs to be informed that this data is potentially of value. If the data was selected for archival then, although the material contains digital data, it is all on paper and therefore would not be transferred to diskette and stored on NDAD by the PRO but instead would be stored as paper documents.

The Centre for Ecology and Hydrology store data from environmental research in the UK. They are involved in some overseas projects but this is not necessarily stored on the central database but rather on the databases at the individual research stations. There is no formal attempt to co-ordinate with European organisations. Open access to data is considered important although older scientists tend to be more concerned with ownership. Those data sets that are expensive to maintain are to be reviewed regularly. For much data it is cheaper to store it than to review it.

Through project reporting requirements the CEH ensure that data is submitted for storage. At the start of a project one is required to fill in a form that includes issues such as accounting, data management and health and safety. For example, there is a box on this form that one ticks to say that a data management plan has been completed. So the approach to ensuring compliance is similar to that for Health and Safety. Data storage, as long as it is something that occurs through the lifetime of a project, is not very expensive. Adding value, on the contrary can be more expensive, for example, improving the ease of access to databases or providing publicity.

ARCHIVAL PROCESSES

Information stored on electronic media requires a technology for its retrieval and technologies tend to become obsolete. Therefore, the PRO also stores a paper copy of the information if available, preferably on microfiche.

Initially NDAD devoted it efforts to storing valuable digital information that was at risk of being lost. They then included information that might be of immediate interest to the public and only opened to the public two years ago once a critical mass of such information had been stored.

NDAD store digital data that is available on electronic media. Paper records or reports may be scanned in but only if they relate to digital data stored on electronic media.

It is very important to describe data well so that it can be used in the future. The PRO do not undertake to do this kind of work and would not, for example, take on the job of describing the NRI data stored at Hayes, this would be considered DFID's job. Neither would the PRO convert digital data on paper to electronic format. Yet the PRO are keen to help people to get funding for this kind of work.

On the other hand, much of the work of NDAD has been on providing provenance information for data sets. They are able to do this as they have a less data to deal with than the PRO. In addition, the data they receive has not come through an archival system in a department but rather from individuals within departments who have collected or have been working on the data. They encourage the older people to look through the old data as they are more likely to understand documentation of their own era than a present day clerk. Combining an older informant with someone who has technical qualifications and an understanding of current needs is a good way to do this work.

NDAD have also worked on data sets, i.e. cleaning them *etc*. and making them usable. The ECN, on the other hand will store the data and some information about it but will then require users to invest in the full recovery of the data. Most users are willing to do this. This does mean, however, that not all data sets are equally accessible and some of the less accessible data may never get used.

FACILITIES THAT CAN BE OFFERED TO INDIVIDUALS

The facilities for data storage that individuals should be able to utilise were discussed. Here the individuals referred to include retiring foresters, retiring forest researchers or consultants who have completed a certain piece of work. It was noted that the problem is more acute for the past decade than for earlier periods as information is now stored on diskettes and consultants are less tied to institutions that might store their data. There is a need for a system where people can deposit their data. There are no existing facilities of this type at present and data is lost or left in a box somewhere without adequate description.

THE IMPLICATIONS OF CONSULTANCY COMPANY RESPONSIBILITY.

Consultancy companies such as Huntings Technical Services have large collections of data, some of which has been collected on DFID funded projects. Huntings appear to act relatively responsibly with regard to data management and storage yet certain issues of concern were identified with this system.

There are two main implications of a system where the consultancy companies who have collected the data are made responsible for the storage and maintenance of data. These are firstly, secure storage of the data and secondly access to and ownership of the data.

Companies are not the most secure depositories for data as, although they may manage data responsibly they tend not to build in measures for data storage if the company closes. And companies do close down. Archival companies and government departments now exist that can store and retrieve data permanently in a secure manner. Both are able to assure confidentiality of data stored if required.

The ownership and access rights to the data are not entirely clear. It seems that in most DFID projects the data is owned by the company but can be used or distributed by DFID upon request. The company is required to look after that data and ensure that it is available if needed. In some cases companies could claim a certain degree of IPR as they have added value to the data collected. This is not a common problem as DFID contracts tend not to limit ownership to consultancy companies and companies working in development are aware that a more open co-operative mode of operation is needed to achieve good results. Of greater concern, however, is limited public knowledge of the existence of data. Companies who have collected data, catalogue and store that data and know about it. This puts them at an advantage when bidding for projects that might build on or require data collected in the last few decades. Thus, one could argue that by opting for a system where companies are made responsible for data storage could result in a situation where certain companies have unfair advantage over others in a bidding process that is supposed to be fair.

Companies are required to cover the cost of data storage under their overheads. As the stock of data stored by a company increases the cost of maintaining it increases. If these companies do not gain a competitive advantage by keeping this data they are unfairly disadvantaged as they have to either bear or charge higher overheads on project bids. Clearly this needs to be carefully considered in any DFID information management or IPR policy.

THE IMPLICATIONS OF RETURNING DATA TO COUNTRY OF ORIGIN

The implications of relinquishing all responsibility for data storage and instead destroying data or returning data to the country of origin were discussed. This, in some senses, would comply with DFID's present policy on data storage for bilateral projects. Where their interest does not extend beyond completed reports. But:

- The data is considered too valuable to destroy.
- Returning data to countries of origin would be very expensive. It was agreed that this could be undertaken if the country of origin would pay for the data compilation and transport costs.
- Those present felt that data should be stored safely in the UK. Firstly, because much of the data is of value to the country of origin, to the UK and to the world as a whole. Secondly, many of the countries of origin do not have the facilities or the capacity to store this data safely. And thirdly much of the data was collected using UK government funds and it can be argued that the government has an obligation to its taxpayers to store such valuable data for future use. The government is and should be concerned that the investment it has made is good and that data is available for developing countries to use.

In some cases countries may not wish the UK to hold a copy of their data and in this case data should be either returned or retained with assurance that it will remain confidential. This needs to be negotiated with the countries in question.

IS ARCHIVAL OF THIS DATA AN INTERNATIONAL RESPONSIBILITY?
As much of this data is of international importance is there an international responsibility for its archival? This is a possibility and organisations that might be approached in this regard include: CIFOR (Centre for International Forestry Research), FAO and ITTO (International Tropical Timber Organisation). Perhaps a facility could be provided for countries to request safe archival of their data? There may be precedents for such and arrangement in other disciplines. In meteorology there is no central data storage unit but there are international groupings for data storage and there is co-ordination with regard to data use and protocols.

FAO do not consider themselves responsible for data storage but as they are now responsible for the FRA2000 project they may be interested in storing and maintaining data that is of value for this purpose.

ITTO is a potential source of funding as they are interested in using this kind of data.

CIFOR has a policy that all information should be publicly available and therefore it may not be possible for confidential data to be stored with them.

UNEP (United Nations Environment Council) is planning to undertake a biomass project, they might also be interested in setting up a data storage system or in assisting with the funding for such a facility.

The IRRI (International Rice Research Institute) encourages researchers to deposit data with them, their system of data management should be investigated.

ICRAF are advanced in this area and are presently involving themselves in a data management exercise. Reading University is working with them to develop their existing data management software into a more widely usable tool.

It was concluded that the possibility of an international organisation holding responsibility for ATROFI should be carefully considered.

THE FEASIBILITY OF HANDING THE DATA TO THE PRO FOR ARCHIVAL
The feasibility of storing this data with the PRO or NDAD was discussed.

NDAD would only be able to take data that had been collected through a Government Department. Therefore a portion of the data we are concerned about would not be eligible. Other data could be stored on NDAD if funding were provided for the initial 'provenancing' and digitising. The cost of storage after that would be low.

The ECN would be able to store the data described on ATROFI but funding would have to be found to describe and transfer the data to electronic format.

It was noted that it is also important to consider the fact that data storage is permanent once the PRO has been requested to undertake it. It would not be possible to get rid of the data at a later date and therefore this is an option that is only appropriate for very valuable information. As the value of forest data tends to increase with time, this was not felt to be a serious drawback.

Government Departments can relinquish their ownership of data. A drawback of insisting that data should belong to a government department and therefore that it is eligible for consideration for storage via the PRO also gives government departments the right to destroy data if they decide it is of low value. To circumvent this, petitions can be made to government departments or to the PRO expressing the concern that certain data is of value.

ARCHIVAL POLICIES IN FUTURE
There is a need to develop better policies in the future to store and archive forest inventory data collected overseas by UK organisations and funded by the UK or by other donors. There is a need to ensure overseas capacity to manage and store such data. It is clear that many UK institutions that collect this kind of data have no data archive policy and hence much of the data is being lost.

A Government Department can relinquish the copyright for information that it collects and this is what DFID do for data collected on bilateral projects. DFID retain the copyright for data collected on research projects. At the end of a bilateral project a project assessment is undertaken but no assessment is made of the data collected on the project or of data management. It is necessary to find out if DFID's project completion report requirements say anything about this.

All Government Departments are now (or will soon be) required to have a functional appraisal policy, this means that they clarify their functions and on this basis they decide what sort of material should be archived.

- It would be advisable for new projects to include in their negotiations and final agreement something about allowing access to the data produced in the project. This would clarify IPR issues. In Bilateral projects, because fresh data can be politically sensitive, confidentiality periods could be specified. So that data could be made available for wider use 5-20 years after it was collected.
- Projects should also be required to undertake data management and, in the case of DFID projects, responsibility for the storage and management of data should be negotiated with partner governments. This should be incorporated into the reporting requirements.
- Those present at the meeting felt that DFID should take responsibility for keeping copies of raw field data collected on their projects.
- In the development of its functional appraisal policy DFID should take this into account. Scientists and the public should be consulted in the development of the functional appraisal policy.
- In particular, DFID should start assessing data on electronic format and submitting this to NDAD for storage where appropriate.

The DFID representative asked the project if they might be able to submit a list of criteria for the valuation and appraisal of digital data for archival. The project was also requested to provide advice on policy in general.

To start with we could build on our existing criteria for selecting data for inclusion in ATROFI.

It was suggested that all raw forestry data collected in the field should be kept, as this would not represent an excessive amount of data. If other sectors were included in this policy then the quantity of data might be unmanageable.

It was suggested that DFID projects should be required to undertake an annual appraisal of existing project data, describing data sets and providing valuation of the data and an assessment of their archival value. DFID should be requested to put this into practice.

It was suggested that data management capacity should also become a criteria for judging project bids.

4.6 Conclusions

ATROFI is a useful meta database. The following improvements could be made to it:

- A global map could be provided as part of the front-end.
- Related publications and grey literature could be scanned into the database.
- A field could be added stating what the data has been used for in the past as well as what it was collected for.
- Links could be made with GIS systems by providing geo-references for each data set.
- A record of data sets that are known but not yet fully described could be included.
- A field could be added to describe the amount of assessment that is still required to understand and used the data set.
- Fields could be added describing what the long-term risk is to the storage of the data and how valuable it is.
- It is important to collect feedback from the users on the usefulness of the database. A short questionnaire making such enquiries should be included on the web page.

Further work is required to collect information on data sets in the UK. In particular the NRI and OFI data holdings still need to be fully described. In addition valuable data is available in MSc and PhD theses on tropical forestry topics held within University libraries (in particular Oxford, Bangor, Edinburgh and Aberdeen). Although this data is relatively secure, summary descriptions of it (including an abstract of the dissertation) in the ATROFI database would render it more accessible.

The ATROFI database should be maintained at an appropriate institution (within the UK for the time being). A small grant needs to be obtained to cover the costs of maintaining the database.

The database, currently on the web, should be widely publicised to attract further funding, more studies for inclusion and ideas for expansion.

The database could be expanded to include information on European holdings of tropical forest data.

Many data sets described on the database are not usable in their current form but require significant cleaning and sorting work. It was decided that, in general, this should be the responsibility of potential users. Funding should be sought to undertake this work for particularly valuable data sets.

Many data sets are owned by government departments overseas. It is the responsibility of the potential user to seek permission to use data.

Once the UK entries are complete copies of the database should be sent to forestry departments in countries where the data originated. Requests for repatriation will be considered on a case by case basis and copies of data sets will be provided on receipt of funds to cover the costs of creating and send such copies.

Many data sets, although described in the database, are not secure. At present there is no UK strategy for the assessment and archival of tropical forest inventory data and some data is still at high risk of being lost. Such a strategy needs to be developed covering data collected by individual consultants, private companies, universities and government departments.

At present DFID projects require consultancy companies to take the responsibility for the maintenance and storage of project data. This raises issues that need to be resolved including copyright ownership, access to data, knowledge of the existence of data, long term security of data, possible use of data to generate unfair competitive advantage and hidden costs. DFID contracts need to specify more clearly the roles and responsibilities relating to data ownership, archival and access.

The British Government should take responsibility for the management and safe storage of forest inventory data collected on all overseas projects that it funds. Interests that the UK government is responsible to, include: tax payers, overseas governments whom it is assisting and the international community as a whole.

DFID needs to secure advice on an appropriate data management policy for both research and bilateral forestry projects including a list of potential criteria for assessing and valuing data. A good starting point for the latter are the criteria used for data inclusion in ATROFI.

It is possible that responsibility for the management and storage of third party data more properly lies with an international organisation. This needs to be investigated. FAO, CIFOR, ICRAF and ITTO should be contacted in this regard.

5. Annexes

Annex 1 Agenda

See Annex 7 of main report

Annex 2 Database structure

See Figure 2 of main text