International Livestock Research Institute

DATA MANAGEMENT POLICY

Introduction

Valuable data are generated from ILRI research activities. The important task of managing data starts from the design of a project and continues through the life span of a project and beyond. Good data management involves checking and verifying data and maintaining the information that describes the context in which the data was collected. Good data management also involves data organization and manipulation for analysis and ability to carry out comparisons with other data sets, making the data available to others and archiving for future use. Several recent occurrences have shown that much of the data at ILRI is not well managed particularly data that has been properly archived. Data archiving must be part of an overall data management plan.

There are a number of changes both within and outside ILRI that mean that we need to approach data management issues in a systematic manner in order to enhance our ability to deliver. For example:

- The CGIAR ICT-KM strategy contains a project on building a strategic alliance across research groups for common data standards and exchange. One notable activity will involve syndicating data to many partner website for easy access and developing common standards/descriptors in several critical areas.
- Scientists have a growing need to be able to integrate data from many different studies.
- Increasing collaboration requires sharing of data within and outside ILRI.
- Outputs are no longer just peer reviewed papers (with built-in quality assurance) but a diverse set of publications and recommendations.
- Emphasis on monitoring and impact needs long- term studies with a requirement to look after data over many years.
- There is a trend towards data sets generated with public funds being considered public property, with donors requiring projects to produce public data archives.
- There is trend towards the raw data used in any science publication being made publicly available.

Currently at ILRI much of the data are not archived in a form that makes them accessible for further analysis or to meet any of the above-mentioned external requirements. Therefore, there is a real need to develop a comprehensive policy within ILRI for adequately managing research data generated from experimental, survey and case study scenarios.

Archiving starts at the beginning of a project

The need to archive data needs to be taken into consideration right from the beginning of any project. There is a need to use electronic methods for storing the data, ensure that standards set by the institute, the CGIAR, donors or collaborators are met. There is also a need to use standard software for the type of data, being aware of how the data will be presented externally from the beginning, ensure the protocol and all documentation that is required to correctly interpret the data and to view it in the correct context is available and kept with the data from the beginning.

Who is responsible for data archiving?

The Principal Investigator in charge of each experiment/survey/case study project is responsible for ensuring that his/her data are stored with relevant, detailed documentation in such a form as to allow retrieval of all data by any body unfamiliar with the research. The investigator's supervisor is responsible for ensuring this is done correctly. Data archiving should be part of the clearance process when a scientist leaves the institute. Theme Directors should ensure that for each project a data management plan is clearly defined and followed throughout the life of the project.

Where should the archived data be stored?

Copies of the data plus documentations should be kept in at least three places:

- In the unit that generated the data.
- In an overall archive at ILRI-Kenya on an intranet server (for the electronic format).
- At ILRI-Ethiopia

Ideally, the archive should be in an electronic format. At Headquarters, the overall archive should be maintained by IT services on an Intranet server (for electronic archive) and by the Library (for hard copies).

Who authorizes access to the data archive (when not made public)?

Internal access

The Theme Director and the Principal Investigator will have full access rights to the data. Scientists involved in the project should have access to the data related to their project. The DG, the DDG-R will have full access rights to the overall archive.
In the IT services, only the IT Manager should have the full access rights to the overall archive.

The Theme Director in consultation with the Principal Investigator and the DDG-R should authorize access to the archived information related to the study led by Principal Investigator. At the project level, the Principal investigator should authorize access to the data to other ILRI staff not directly involved in the project.

External access

The DDG-R in consultation with the Theme Director should authorize access to the archived information when it is not made public to non-ILRI staff. Intellectual Property Rights of ILRI and its collaborators should be addressed before authorizing data access.

When should access to the archived data be made public?

Access to archived data is a very sensitive issue. For short and long term experiments public access (via internet for example) of archived data should be 3 years after the completion of the research. The results of the research should be made public at any time. Public access of the data should also take into account the Donors requirements.

What about sensitive research areas such as vaccine development?

Patents and copy right issues should enter into account to properly address this issue. Confidentiality is another element to consider. The archive of such data should remain at the level of the unit until the official release of the discovery/major findings. This archive could then be transferred to the overall archive.

What about GIS Databases?

ILRI GIS Team is part of a "Spatial Database Clearing House" in which the team contributes a small part of their data for map viewing using Internet Map Services. Currently, the Kenya database (metadata & GIS files) could be accessed through the Internet. GIS Databases are in the "Clearing house" metadata format. Any GIS database to be developed at ILRI should follow this format for compatibility.

What information should be archived?

The information to archive will consist of the documentation and the computer files.

The documentation of each data set should consist of:

- Final concept note or project proposal
- Research Protocol including details of sampling methodology of treatments and experimental/survey design, scientists involved, institutions, collaborators, funding agencies
- Details of all variables recorded, including variable definitions, units of measurements and coding scheme and information on how missing values are being treated. Details of derived variables including calculation formulas. The units of measurement should be expressed in the International System.
- Results of laboratory analysis of samples
- Reports, maps, publications, references.
- Index of all computer files stored in the archive (with a brief description)
- Documented computer programs used for manipulating data
- Instructions for the sequence of running programs to create intermediate data files
- Completion date, duration and budget.

- Results of the research.

Computer files should include:

- Master files (one) of the raw data – they should be in a format accessible by one of ILRI's standard software. We should make sure that the software used for data entry is available and accessible to ILRI. If not, it should be part of the archive.
- Cleaning and validation rules should be indicated.
- Program files should have internal documentation as to allow others follow the logic of data manipulation.

One file (the index file) should be created and should contain the listing of all the files for the experiment/survey together with a brief description of each file and any relevant instructions on the use/sequence of computer programmes.

When does the archiving process start?

Data Management of which data archiving is a component should start at the beginning of the project. It is a time consuming process and needs to be planned in advance.

What about past studies that are not archived in the way suggested above?

Archiving of past experiments will have value if they can be properly documented. This may not always be the case for some projects. We strongly recommend that data and documentation of past studies be archived, if completeness of the relevant documentation and related files can be ensured.

What about updates/corrections of the archive?

It should be the responsibility of the Principal Investigator to initiate the process through his/her Theme Director in collaboration with IT Services or Information Services. The Principal Investigator should make sure that the updated information is included in all the archives at all locations.

Software for data archiving

We should capitalize on existing web-based software when selecting software for data and text archiving. The system for retrieval of archived data should allow easy retrieval of partial or full information related to a specific experiment/survey/project. IT services, Information/Library services should assist in the selection of the appropriate software. The design of the Database for the archive is a separate issue and should not be included in a policy document.

Intellectual Property Rights issues

Data management should also consider Intellectual Property Rights related issues as ILRI exercises them over the publication and or dissemination of all research data. The issues to consider are copyright ownership, liability and third party copyright.

Data ownership

Data collected on ILRI funded projects and executed by ILRI are property of ILRI. Data collected through collaborative research projects are jointly owned by ILRI and collaborators. If ILRI policy on data management is different from that of collaborators then the Memorandum of Agreement (MOU) or the Letter of Agreement (LA) should clearly state which policy should prevail.

Copyright Notice and Disclaimer of Liability

When the data is made available to the public, there should be an appropriate copyright notice that indicates that ILRI and collaborators are the copyright holder, and allow certain uses of the data. If someone relied on the data--after it has been made available to the public--and got unfavorable results then ILRI is not liable.

Third party copyright

If the data does incorporate data acquired from others, then we need to (a) have the appropriate permission when making the overall database available to the public (b) acknowledge the copyright or other IP owner.