NRI International

OUA-IBAR

European Commission

Farming in Tsetse Controlled Area

Environmental Monitoring and Management Component
Project N° 7 ACP RPR 578

Regional and National FITCA database harmonisation

Short term consultancy

From 15-Feb-2002 to 4-Mar-2002

and

from 18-Mar-2002 to 29-Mar-2002

Gilles FOURNIE

Cirad-emvt Report n° : 02-26

April 2002

CIRAD-EMVT
Département d'élevage
et de médecine vétérinaire tropicale
Campus International de Baillarguet
BP. 5035
34398 Montpellier Cedex 5 - France

AUTEUR (s) : Gilles Fournié

ACCES AU DOCUMENT :
Centre de Documentation
du Cirad

ORGANISME AUTEUR : Cirad-Emvt

ACCÈS : Libre

ÉTUDE FINANCÉE PAR : Commission de l'Union Européenne. A travers NR International

RÉFÉRENCE : 7 ACP RP R 578 - Rapport Cirad 02-26

AU PROFIT DE : Programme FITCA de l'OUA-IBAR; volet EMMC

TYPE D'APPROCHE : Mission d'expertise court terme

TITRE : Regional and National FITCA database harmonisation

DATE ET LIEU DE PUBLICATION : Avril 2002, Montpellier

PAYS OU RÉGIONS CONCERNÉS : Kenya ; Ouganda ; Éthiopie

MOTS CLÉS : Base de données ; Suivi environnemental

RÉSUMÉ :

Le Fitca (Farming in Tsetse Controled Areas) est chargé de travailler à l'amélioration des conditions de vie des fermiers dans les zones où la maladie du sommeil est présente. Cette amélioration passe par une lutte contre la mouche tsé-tsé et peut entraîner des modifications de l'environnement.
La mission, menée en parallèle avec des missions de terrain, avait pour but d'étudier l'uniformisation des bases de données disponibles dans les centres Fitca et de proposer la structure d'une base de données pour faciliter le suivi des changements environnementaux.
Les bases de données existantes sont peu nombreuses et sont surtout des fichiers de travail (Ms-Excel ou Ms-Access) contenant des données très spécifiques. Le rapport suggère que la gestion de ces bases soit laissée à la discrétion des centres qui les utilisent, une harmonisation ne semblant pas se justifier pour ce type de données.
Pour la base d'aide au suivi environnemental, un ensemble de données a été défini en collaboration avec les chercheurs de terrain. Cet ensemble de données a été organisé sous la forme d'un modèle relationnel et constitue un noyau d'informations pour le suivi environnemental qui pourra être complété par la suite.
Enfin, le rapport propose l'embauche de spécialistes de la conception et de la gestion de bases de données qui seront chargés de concevoir, de construire et surtout de faire vivre les bases de données constituées au sein du projet.

SUMMARY

1 - Context

1.1. FITCA and EMMC

The FITCA project is Farming In Tsetse Controlled Area. It is a regional program for Kenya, Uganda, Ethiopia and Tanzania. As part of the FITCA project, the Environmental Monitoring and Management Component (EMMC) is in charge of environmental aspects of the fight against tsetse flies.

1.2. Terms of reference for the Short Term Consultancy

Recall of the terms of reference for this short term consultancy: "Regional and National FITCA database harmonisation".

Objective of the consultancy

The Kenyan and Ugandan national FITCA project have already started to investigate their respective areas through different field surveys e.g. animal census, flies density, location and parasitaemia prevalence household surveys…

Accorded to the local situation, the importance given to the different items collected could be highly different from one country to the others. The main advantage of a regional project is to facilitate joined activities throughout the national or regional boundaries. Monitoring and management tools must be enough accurate and flexible to report the major results and changes induced by the control of tsetse flies in the national project areas.

The required consultancy must provide to the regional EMMC component the adequate database tools that will permit a good harmonisation of the different databases developed at the national and regional level.

Methodology

The expert will make an exhaustive inventory of the existing databases developed at the national and regional FITCA project level. He will analyse the structure and links of these different information systems. He will work in close relation with the co-ordinators of the different components of the project. He will propose a comprehensive and well-structured data management system, which could be used by the FITCA community as a whole or on a specific item. According to this list, the STC will start with a field phase and an implementing phase in Nairobi at OAU-ILRI HQ.

Expected results

The expert will provide a comprehensive and evolutive data managenent system including routine software for GIS interpretation.

1.3. Comments

1.3.1. Method

Several missions were made at the same time. They involved scientists from different specialities : veterinary surgeons, sociologists, environmentalists, ecologists, GIS specialists, computer and databases specialists.

In my opinion, the best advantage of this approach is in the mixing of different way of thinking and of different understandings of the problems. But, the planning of the missions didn't make possible to take advantage of this multidisciplinarity as scientists were not all present at the same time and no global brainstorming was planned.

Besides, as for the database part, it was inadequate to do it at a time when scientists hadn't yet defined what data they need to get and to manage.

1.3.2. Terms of reference

The terms of reference are not fully applicable. It is difficult to do an exhaustive inventory of existing databases in a short time mission. It is not possible to design a comprehensive information system when the data details have not yet been defined. Anyway, we have tried to make as useful work as possible.

1. **Exhaustive inventory of FITCA databases at both national and regional levels**...
   It would have been more efficient to send to national FITCA co-ordinators a form to fill in asking for the list and the description of the different files they have produced. We briefly met people and, when computerised data was available, they shown us the files and explained the structure. These people would have had more time to answer to a survey and the information would have been more accurate.
2. **Monitoring and management tools**...
   We will give our opinion about tools used and discuss the use of some other tools.
3. **A comprehensive and well-structured data management system**...
   We will try to structure and organise the different data the scientist will determine useful for the environmental monitoring. We will choose a relational database format to organise these data, as the relational model provides many way to access and combine data. Anyway, this work will only be a draft of the future database as the database model was designed while scientists had not yet finished their work. As said before, the database work comes too soon and we cannot design a comprehensive data management system. Ideally, scientists must have finished working and decided what data they want to use before a database design could begin.
4. **A comprehensive and evolutive data management system including routine software for GIS interpretation**...
   We saw that we cannot meet the "comprehensive" part. As for the "evolutive" part, the use of a relational system should do the trick. But, FITCA will need to hire (or to train) database specialists both for the design of the database and for the maintaining and evolution of this database. We will give the basis of the database scheme but the database designer will have to complete this scheme and put it in operation. As for GIS, as far as I know, ArcInfo is a well known software and is able to access data from a relational database. I don't know more on this subject.

Existing data

As for the FITCA teams we met, only the FITCA Kenyan team possesses computerised data. In Uganda, a survey has been planned but no data is in a computer form. In Ethiopia, the FITCA team has not yet received computers.

The Kenyan team showed us their files. These files are all Excel files; one of them has been converted for use with MS-Access, but does not use any relational structure.

- FITCA(K) baseline survey. This survey is aimed to provide FITCA(K) with some basic information about households and livestocks for people owning zero-grazing units. It is a 11 pages survey with many elements.
- Tsetse traps statistics. For each trap and for every month, this file contains data collected on the ground: number of flies caught, teneral status of flies, etc. Due to the repeated structure for this file (every trap, every month) and the small amount of elementary data, a small relational structure could be used to represent this data in an efficient way.
- Crops. This survey provides FITCA(K) with information about crops. It is a two pages survey with a rather simple structure.

Other surveys are yet to be computerised.

- Use of draught animal technologies.
- Cattle survey

As for the Uganda, the FITCA(U) team told us about an ongoing baseline survey. The questionnaire is 23 pages long, it contains 122 questions many of which will give place to complex answers: for example, the list of treatments applied to animals, "*Please complete the table below describing the fixed assets possessed by the household. (Asset; Years since acquisition or construction ; Type of structure)*" , "*What products do you usually obtain from the forest/uncultivated areas? (Product; Quantity per week/month; Who collects the product?)*", ...

Anyway, as for the FITCA(K) baseline survey, several tables will be needed to structure this data with a relational scheme.

1.4. Database tools and file formats used

In FITCA(K), data is stored in Excel files or MS-Access files.

Most of data is stored in Excel files. These files contain mainly results of surveys and measures on the ground.

Access has also been used for some data. However, the structure of these Access files does not take advantage of relational structures allowed by Access so a great part of the interest of this tool disappears.

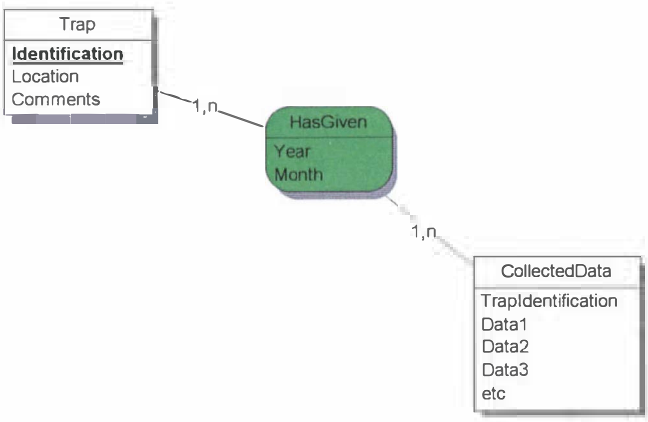
The person in charge of data input and data management is fully operational in Excel but she has not yet acquired necessary skills to step into the management of relational databases.

Yet, it seems really obvious that the use of a Relational Data Base Management System (RDBMS) is indispensable for future database growth. Excel files will not easily allow to make computations other than simple statistical calculations. On the other hand, RDBMS easily allow to compare or group data in many different ways. For example, the files of monthly collected data are constituted from data gathered from all the traps: so, there is a group of data which repeats itself for each trap, the whole repeated for every month. This kind of tri-dimensional data is difficult to represent in an effective way in a spreadsheet, which by essence is bi-dimensional.

1.5. Suggestion

Using a RDBMS instead of Excel files would help in querying the files and extracting information out of them. As these files are used for everyday work, they should be left under the control of the FITCA(K) team. They should be converted from Excel to a relational model. These files are not intended to become a complex information system, so a personal RDBMS like MS-Access could be used.

As an example, monthly collected data from traps could be represented by the following structure:



That allows to use SQL to extract data following various axes: grouped by trap, by date, according to the values of measures, etc.

In the same way, the data which could be shared between different files should become a table in a relational database structure. For example, a farmers table could be referenced to by several other data structures.

Concerning the data management, I would suggest that the data structure should be decided by the local database manager (at FITCA(K) level) under the supervision of the database designer (at FITCA(R) level). [see below the discussion about the organization].

Database for environment changes monitoring

1.6. Method / Decisions made

The scientists of the team, after the studies on the ground and different discussions, established a preliminary list of needed data and the relations between these data. From this list, a relational model of data was built and is shown below.

An important point about this database is the possibility to exchange information between the relational data base and GIS tools. This point should be met by the use of Arcinfo. Through ODBC (Open Data Base Connectivity), the GIS software ArcInfo is able to get data stored in databases: it is for example planned to use ArcInfo to generate information concerning the nearness from a river of points recorded in the database; this information will be calculated by ArcInfo after fetching points from the database and will then be recorded by ArcInfo into the database.

1.7. Data availability

Some of the data is available at the FITCA level (either in computer form, either in a paper form). These data are not the only ones used in the database scheme. FITCA will have to get some data from other institutions.

Interesting data may exist (we heard about population surveys, geographic files, ...) in other national organisms (agricultural services, ministries, ...) but it is not possible to look for these data without knowing precisely what is needed. For this purpose, we suggest a form which could be sent to different services when useful data will have been decided.

☞ **Request**

Object: Request for an access to data

Purpose: For the FITCA project, we need some data that your organism may eventually possess. We appeal to you to know if you have this data, if you accept to share it and what are the access conditions.

Data looked for: *{DataType}*

Purpose: *{DataPurpose}*

☞ **Answer**

Availability for data: *{DataAvailability}*

Description for available data: *{DataDescription}* *{DataAccuracy}*

File format for available data: *{DataFormat}*

Description of data structure: *{DataStructure}*

Where the elements for the request are:

- *{DataType}* is the name of the kind of data looked for. We can foreseen "Geographic data", "Demographic data", ...
- *{DataPurpose}* is the description of the intended use of the data. It should contain a precise description of the elementary data expected.

And the elements for the answer are:

- *{DataAvailability}* specifies the existence or absence of the data and of the access conditions to these data.
- *{DataDescription}* describes the data: source, contents, ...
- *{DataAccuracy}* is used to precise the age of the data, its degree of completeness, ...
- *{DataFormat}* defines the file format where the data are saved in. We can have Excel files, Access databases, Oracle databases, etc.
- *{DataStructure}* is the full description of the contents. It defines the columns of the databases and what their content is. For a relational database, it includes a description of the relationships.

1.8. Database structure

1.8.1. Data Model

This is a conceptual database model based on the brainstorming in Busia with Burkhard Bauer (coordinator for FITCA(K)) and the consultants team. The initial model was designed by Renaud Lancelot and I modified it partly to incorporate data needs for other scientists (agro-pastoralism data, ecological and botanical data).

CONCEPTUAL MODEL

The model shown below tries to put in a relational form a collection of data which could be used to do environmental monitoring. The entities defined here and their attributes come from preliminary discussions with the STC scientific team.

As the information system design process had to be done at the same time scientists where on the field to decide which data will be useful, final decisions about data have not yet been taken. So, the present model will need to be adapted, completed or simplified by the FITCA database team.

Each entity possesses an attribute named "ID". This attribute helps in identifying in an unique way the instance described by this entity. It may be an unique number, automatically assigned by the database system. It may also be a value or a combination of values given by the data manager (for example, ISO country codes can be used). This attribute is not explicitly shown in the following description of the entities.

Entities often have attributes named data1, data2 and data3. These are placeholders that I have used when I had no idea of what data would be useful.

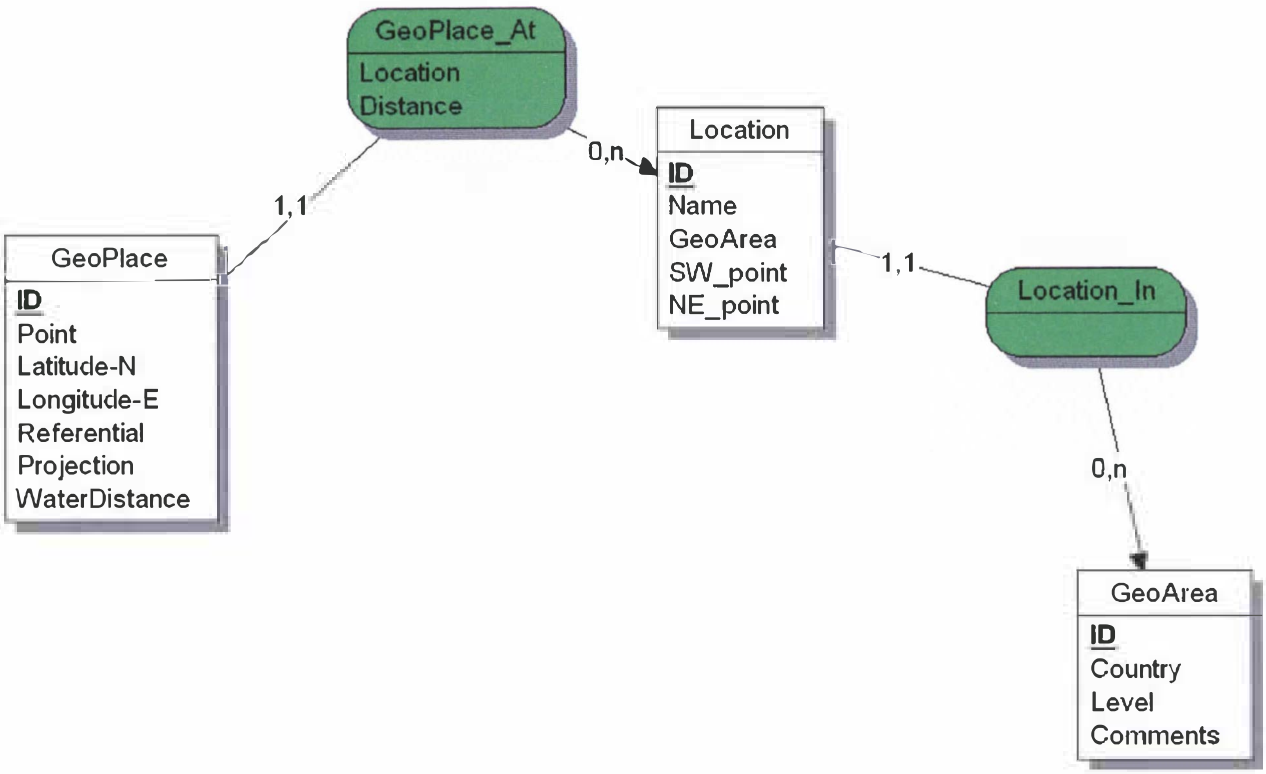
The model is split into sub-models to be easier to read. Obviously, when an entity name appears at different places, it refers to an unique entity.

We will describe in the next sections the different entities and the relations between these entities. Entities have been grouped by "groups" to make the relations more concrete but these groups have no meaning in terms of RDBMS. Besides, entities may be common to several groups; even if they appear several times, they will not be duplicated in the database.

As for the database structure, without being too much technical, the entities will become tables of data (lines of data whose columns will be the attributes; each line will have a unique identification key. This key is an attribute or a combination of attributes) and the relations will become other tables (with each line containing both the attributes of the relation and the keys of the linked entities).

1.9. Location

This sub-model shows the entities used to localise a place (either a point or an area) and the relationships between these entities.



1.9.1. Geoplace

GeoPlace is the main entity for localising objects. Every entity which has to be localised will be associated a GeoPlace. The GeoPlace defines either a point (using GPS co-ordinates) or an area. In the first case, the GeoPlace will have its Point attribute defined (true) and will have GPS co-ordinates; in the latter case, it will be associated with a Location element.

Attributes:

- Point
  - ✓ Point = 1 if GeoPlace identifies a specific point (GPS co-ordinates have to be given).
  - ✓ Point = 0 or NULL if GeoPlace identifies an area (which is further described by a Location).
- Latitude-N
- Longitude-E
- Projection (UTM...)
- Referential
- WaterDistance is meant to give the distance of the GeoPlace from the nearest water source (lake, river, ...). This value should be computed with the help of ArcInfo : ArcInfo will fetch data out the database, compute the distances and store them in the database.

Relations:

- Location... The relation to the Location entity identifies the Location the GeoPlace belongs to. If the GeoPlace can not be assigned to a Location, the relation should points to the closest Location. This relation possesses an attribute :
- Distance
  - ✓ Distance = 0 if the Point is inside Location
  - ✓ Distance measured in km if Point is outside Location. This is to allow to specify Points that are in the vicinity of a village, for example.
  - ✓ This attribute makes possible requests like :
    - "select ... where location=xxx and distance=0" to find Points inside a Location
    - "select ... where location=xxx and distance<10" to find Points inside a Location and Points within 10 kms around
    - "select ... where distance!=0" to find all the Points which are not associated with a Location
    - "select ... where distance>10" to find all the Points far from any Location
  - ✓ **Question** : is this information useful? Is it possible to collect this information?

1.9.2. Location

The Location entity defines the administrative locations where all GeoPlace should be localised. It may be either village names, either district names, either county names, either sub-location names, etc. depending on the GeoArea in which the monitoring is done.

- Name
- Co-ordinates of the enclosing rectangle

Relations:

- GeoArea... The relation to the GeoArea entity identifies the area the Location belongs to.

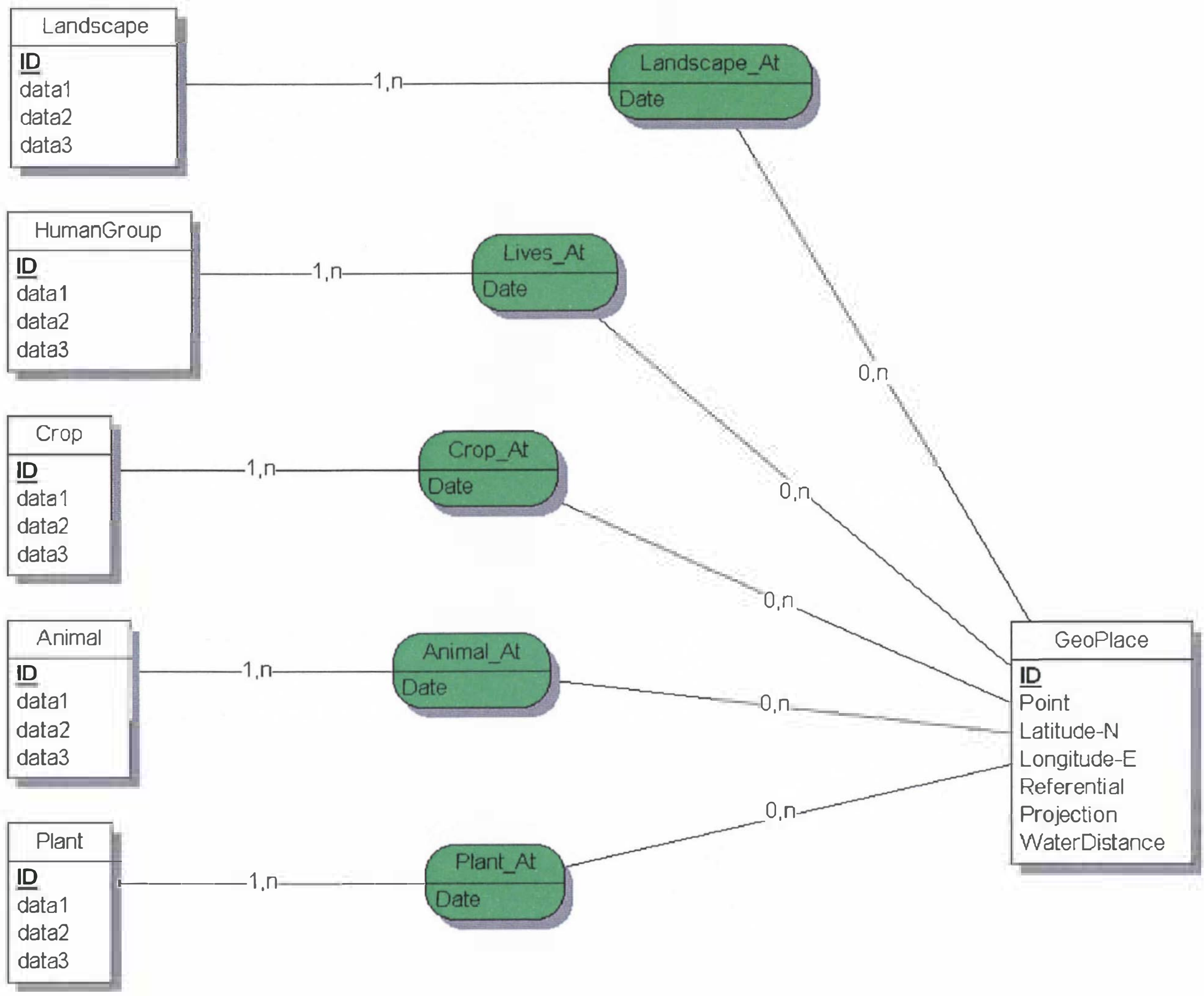
1.9.3. GeoArea

Each Location must be associated with the area under monitoring. This entity indicates the country where the area lies and the administrative level used to identify Location elements.

- Country (et = Ethiopia, ke = Kenya, ug = Uganda, ...)
- Level identifies the level (district, county, village...) for Location. It may be "villages", "districts", "counties", "sub-locations", etc. depending on the area where the monitoring is done. There can only be one level per GeoArea to maintain consistency between data for an area and to make possible detecting environmental changes depending on time.

1.10. Landscape

This sub-model shows the entities used to describe a landscape existing at a specific GeoPlace. A landscape will be described using a set a entities: Landscape describes the relief, the physical landscape; HumanGroup describes the type of the population (farmers, fishermen; ...); Crop describes the different crops; Animal describes the animals; Plant describes the plants.



1.10.1. Landscape

To describe the physical landscape, the useful attributes will need to be defined. We can imagine a few attributes shown below. To be useful, the Landscape entity should describes kinds of landscapes and not specific landscapes. So it will be better, for example, to define a few categories for rain falls (something like : low, medium, high) than to use numerical values. If these values are useful, they should belong to the relation between Landscape and GeoPlace.

- Relief type : plains, mountains, ...
- Hygrometry : rain falls, humidity, ...

Relations:

- GeoPlace… This relation indicates where a specific physical landscape can be found. This relation possesses an attribute :
- Date. The date when the landscape has been described.

1.10.2. Animal

This entity is used to describe all animal species : domestic animals, wild animals and, of course, tsetse flies.

- Species
- Sub-Species
- Domestic / Wild

Relations:

- GeoPlace… This relation indicates where specific animal species can be found. This relation possesses attributes :
- Date. The date when the animals have been found.
- Number. The number of animals found at this place.

1.10.3. Plant

This entity is used to describe all wild plant species. Crops are defined in another entity. **Question**: Should the two entities been merged?

- Species
- Sub-Species

Relations:

- GeoPlace… This relation indicates where specific plants can be found. This relation possesses attributes :
- Date. The date when the plants have been found.
- Number. The number of plants found at this place.
- Surface. The surface occupied by plants.

1.10.4. HumanGroup

This entity is used to describe human groups.

- Type of group: farmers, fishermen, …

Relations:

- GeoPlace… This relation indicates where specific human groups live. This relation possesses an attribute :
- Date. The date when the human group was living at that GeoPlace.

1.10.5. Crop

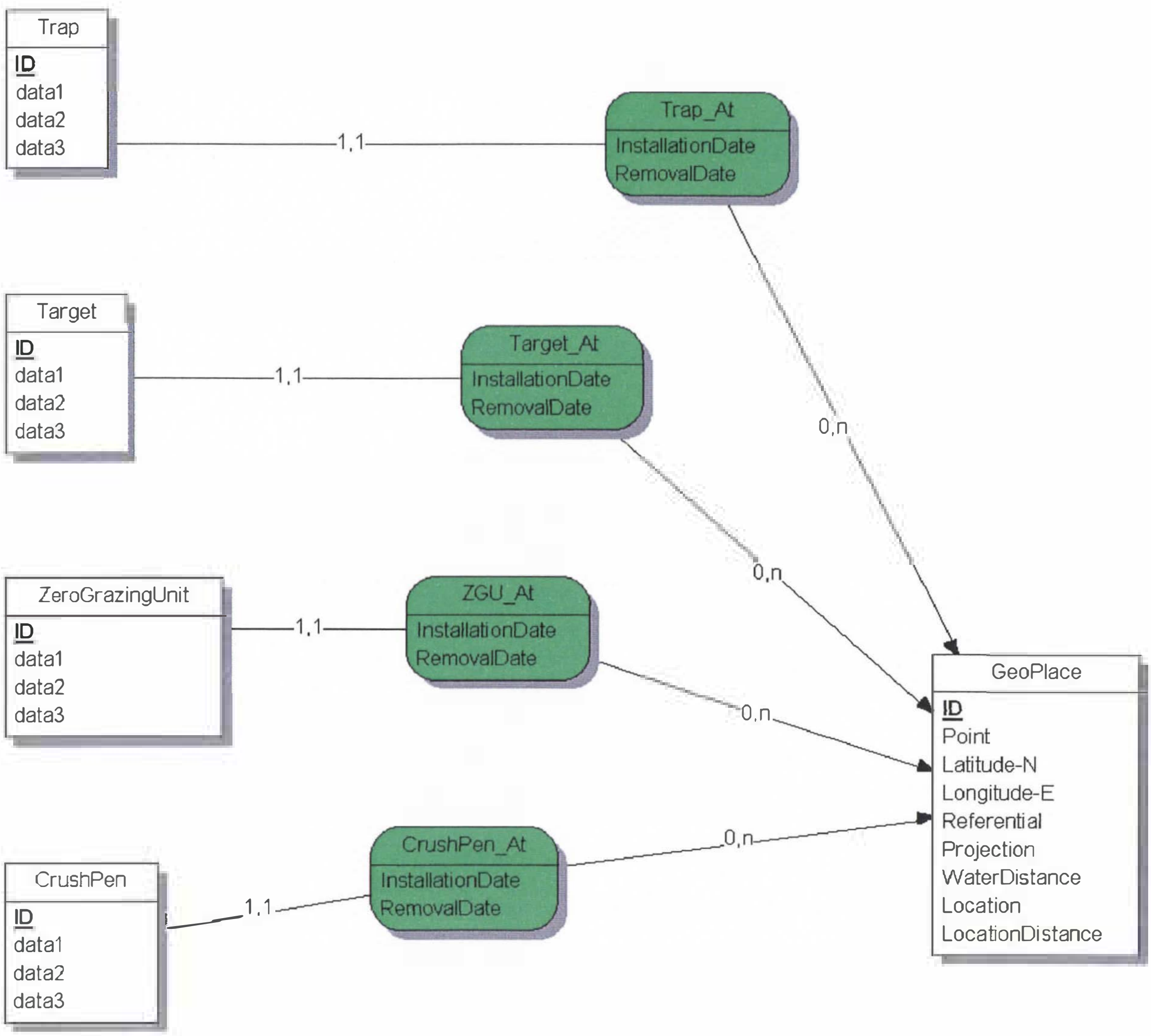
This entity is used to describe crops.

- PlantName

Relations:

- GeoPlace... This relation indicates where specific crops can be found. This relation possesses attributes :
- Date. The date when the crop has been described.
- Production. The month/annual production (a unit will have to be chosen).

1.11. Primary control measures



1.11.1. Trap

This entity is used to describe traps. The main and probably only information needed is an identifier for the trap. We can eventually add a description or comments.

- Description

Relations:

- GeoPlace... This relation indicates where traps have been installed. This relation possesses attributes :
- InstallationDate. The date when the trap has been set up.
- RemovalDate. The date when the trap has been removed.

1.11.2. Target

This entity is used to describe targets. The main and probably only information needed is an identifier for the target. We can eventually add a description or comments.

- Description

Relations:

- GeoPlace... This relation indicates where target have been installed. This relation possesses attributes :
- InstallationDate. The date when the target has been set up.
- RemovalDate. The date when the target has been removed.

1.11.3. ZeroGrazingUnit

This entity is used to describe a zero-grazing unit. We can describe it with its dimensions, the number of animals it contains, the name of the zero-grazing unit owner, etc.

- Owner
- Surface
- Number of Animals

Relations:

- GeoPlace... This relation indicates where zero-grazing units have been installed. This relation possesses attributes :
- Installation Date. The date when the zero-grazing unit has been set up.
- Removal Date. The date when the zero-grazing unit has been removed.

1.11.4. CrushPen

This entity is used to describe a crush-pen. . The main and probably only information needed is an identifier for the crush-pen. We can eventually add the name of its owner.

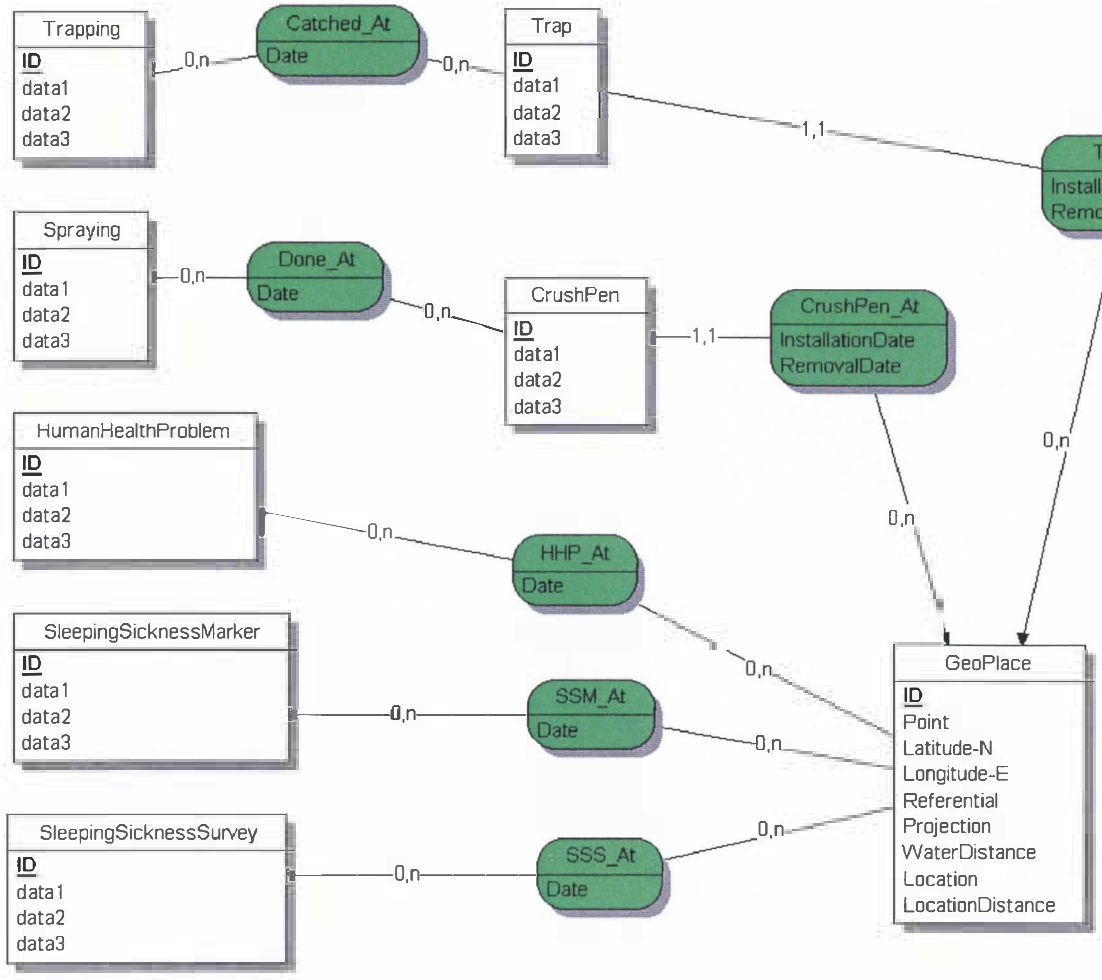
- Description

Relations:

- GeoPlace… This relation indicates where crush-pens traps have been installed. This relation possesses attributes :
- InstallationDate. The date when the crush-pen has been set up.
- RemovalDate. The date when the crush-pen has been removed.

1.12. Collected data

This sub-model shows the entities and the relations used to describe and manage the data collected: from the traps, from the crush-pens, from the health services, etc.



1.12.1. Trapping

This entity is used to describe data collected from the traps. We can imagine attributes for the total number of flies; the number of males, females, teneral flies; the number by species; etc.

- Data1… To be defined…

Relations:

- Catched_At… This relation indicates the trap from which data has been collected. This relation possesses an attribute:
- Date. The date when the data was collected.

1.12.2. Spraying

This entity is used to describe data collected from the crush-pens. We can imagine attributes for the number of animal treated; the quantity of products used; etc.

- Data1… To be defined…

Relations:

- Done_At… This relation indicates the crush-pen from which data has been collected. This relation possesses an attribute:
- Date. The date when the data was collected.

1.12.3. HumanHealthProblem

This entity is used to describe data collected about the sleeping sickness problems.

- Data1… To be defined…

Relations:

- HHP_At… This relation indicates the GeoPlace where the data has been collected. This relation possesses an attribute:
- Date. The date when the data was collected.

1.12.4. SleepingSicknessMarker

This entity is used to describe data collected about sleeping sickness markers. We can have the number of blind dogs; etc.

- Data1… To be defined…

Relations:

- SSM_At… This relation indicates the GeoPlace where the data has been collected. This relation possesses an attribute:
- Date. The date when the data was collected.

1.12.5. SleepingSicknessSurvey

This entity is used to describe data collected from sleeping sickness surveys.

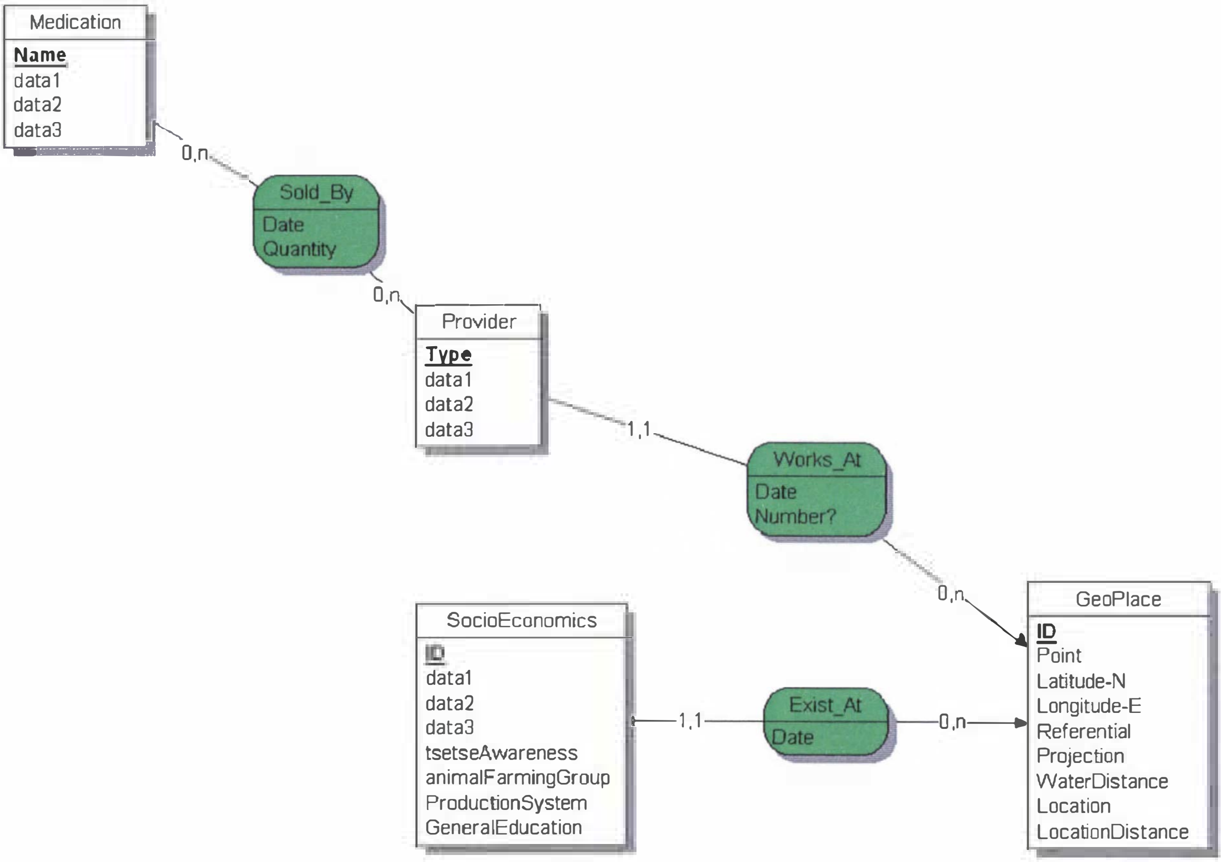
- Data1… To be defined…

Relations:

- SSS_At... This relation indicates the GeoPlace where the data has been collected. This relation possesses an attribute:
- Date. The date when the data was collected.

1.13. Other control measures

This sub-model shows the entities and the relations used to describe control measures other than primary control measures.



1.13.1. Provider

This entity is used to describe different kind of medicine or chemical products providers.

- ProviderType. The list of values could contain "physician", "laboratory", "chemist", etc.

Relations:

- Works_At... This relation indicates the GeoPlace where is this provider type. This relation possesses attributes:
- Date. The date when the data was collected.
- Number. Number of providers of this kind.

1.13.2. Medication

This entity is used to describe different kind of medications.

- Medication. The name of the medication sold…
- Etc.

Relations:

- Sold_By… This relation indicates the Provider type who sold the medication. This relation possesses attributes:
- Date. The date when the data was collected.
- Quantity. Number of doses sold.

1.13.3. SocioEconomics

This entity is used to describe different indicators about the level of information in the population.

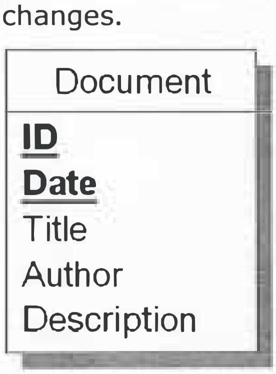
- Education level.
- Tsetse awareness.
- Production System.
- Animal Farming Group. Milk association/co-operative…
- Etc.

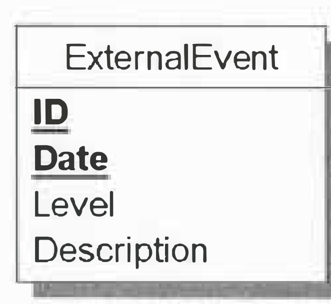
Relations:

- Exist_At… This relation indicates the GeoPlace where the data applies. This relation possesses an attribute:
- Date. The date when the data was collected.

1.14. External informations

These entities are disconnected from the rest of the model. They will contain informations which may be related to the project or which could be pertaining to explain environmental changes.





1.14.1. Documents

This entity refers to books, articles, etc which may be useful to understand the reasons of environmental changes.

- Source. Where the information comes from.
- Description / Abstract
- Type
- Title
- Localisation. Where to find the document (library, ministry, etc.)
- Publication date

1.14.2. ExternalEvent

This entity describes events which may have an impact on the environmental changes. For example, a change of the agricultural policy of the country, or exceptional weather conditions, etc. These entity can only be related to the rest of the model by the date, which allows to correlate events and environmental changes.

- Source. Where the information comes from.
- Description = rules, political events, economical events, international events which could explain environmental changes independently of the control measures against tsetse flies.
- ImpactLevel = Local, National, Regional, International
- AreaName
- Date

1.15. Software

For the database management, we advise against to the use of MS-Access, at least for the central database (the Environment Monitoring database), as MS-Access is more a tool for office automation than for management of large data bases. Nevertheless, MS-Access could be used to manage small local databases and as a tool for the input and the consultation of the base (exports, printings).

Anyway, any professional RDBMS with an ODBC driver could be used. But we will advise the use of either SQL-Server or Oracle: SQL-Server, another product from Microsoft Corporation, is the professional RDBMS from Microsoft. Oracle is the leader in RDBMS.

According to the GIS specialists, ArcInfo is a reference for the geographic treatments and it allows to connect to a relational database using ODBC.

1.16. Organisation around the database

Both for the management of existing databases and for the development and management of the monitoring database, we strongly advise to hire database designer(s) and database manager(s).

We can think of a database designer, for FITCA-Regional in Nairobi, in charge of the design of the databases and of the co-ordination of the works in the different countries. His/Her job will also be to do the conversions from other databases to FITCA databases, to do exports from FITCA databases, to maintain a central database aggregating the data for the monitoring database... Even if FITCA decide to work with MS-Access and not with a professional RDBMS, it is important that the database designer is used to work with RDBMS like Oracle, Sybase, DB2, SQL Server, etc. as this knowledge is very important to design evolutive databases.

And, in each country, a database manager in charge of maintaining local databases and the local monitoring database. The database managers will work with and under the "control" of the database designer. The database managers will be warrant of the input of data (eventually, with the assistance of a secretary) and of the quality of this input.

The database designer will be in charge of checking the overall consistency of the data; he/she will be responsible of making tools to help database managers to maintain good quality of data: input forms with automatic checking against reference tables, scripts or programs to verify consistency of data, etc.

General synthesis, comments and recommendations

1.17. Method

This mission could have been done differently: the database work should have begun when the field scientists had finished working on the ground and had precisely defined what data they need. At that time, a brainstorming between field scientists, FITCA people and computer specialists could have permit to define the information system structure.

So, due to the way the mission was organised, the design of the information system is subject to adaptations. Second, as stated by the STC reference terms, the information system should be flexible; an information system must evolve to adapt to new demands or new environments. These points and the ones given in the next section lead to the hiring by FITCA of a database designer.

1.18. Putting in operation

A special care should be taken about issues concerning people: designing a database will be a useless job if people are not able or have no time to get data from the ground or to input it in the database. In a similar way, a good database design will not be useful if data is not accurate or consistent in the database (for example, if measures are given in different units : sometimes meters, sometimes feet...).

Furthermore, to put the information system in operation, tools will have to be developed (and, later, maintained, adapted, ...). These tools will be :

- Tools to help people insert data into the system : Access entry forms or Delphi/Visual Basic programs. These tools will contain integrity checking functions to help in maintaining a consistent database.
- Tools to control or to help administering the database : extraction of periodic reports, extraction of lists of values, integration with GIS tools, etc.
- Tools to help people query or extract data from the system : Web site with querying forms, Delphi/Visual Basic programs, etc.

And, of course, the feeding of the database will have to be done; either by keyboarding the data, either by importing it from other data sources.

So, I strongly recommend the hiring of one database designer and of database managers (at least one per country). Obviously, people already working for FITCA can also be trained to become efficient in these fields.

For the database team to be efficient, I suggest that the database designer and the database managers are hired at Kenya-Regional level. The database managers will then be put at disposal for the different FITCA national teams. They will work under the authority of the national structures for local data processing but will have to make reports about these works to the database designer. They will fully depend from the database designer for all the common data management.

Ideally, the database designer should be aware of all existing and of new FITCA databases, either they exist at central or local level.

Appendices

1.19. Profile for a database designer

Computer skills

- Windows 95/98/2000
- Perfect knowledge of SQL and SQL-Server. MS-Access appreciated.
- Good practice of Visual BASIC or Delphi
- Good practice of Perl or Python or C language
- Knowledge of Oracle, mySql, ... and free software (Linux) appreciated

Scientific knowledge

- GIS
- Biology

Activities

- The database designer will be in charge of the overall design for the central databases for the FITCA project. He/She will be based in FITCA-Regional in Nairobi.
- He/She will need to design databases (good knowledge of SQL and RDBMS), to develop entry forms or database applications (knowledge of MS-Access and a RAD tool like Delphi or Visual Basic)
- He/She will be responsible to develop and maintain a Web site for FITCA with access to databases (querying, extracting, ...). So, a good experience in designing Web sites, interfacing them with databases and securing the servers will be needed. [If a decision to go Web is made]
- He/She will be in charge of integration of data get from other databases and extractions of data. So, the knowledge of script languages like Perl or Python or of a general purpose language as C will be appreciated.
- Even if the project is based on Microsoft tools (Windows, SQL-Server, ...), the database designer should be aware of the Free Software world: Linux, MySql, PostGreSql, etc.

1.20. Profile for a database manager

Computer skills

- Windows 95/98/2000
- Good practice of SQL and MS-Access
- Knowledge of SQL-Server appreciated

Activities

- The database manager will be in charge of the local databases for the national FITCA projects. He/She will be based in FITCA national places.
- He/She will be responsible of maintaining local data which will be integrated in the central database (knowledge of SQL and MS-Access required). He/She will work on this aspect with and under the control of the database designer.
- He/She will be responsible to design (in co-ordination with the database designer), and maintain local databases, for local use.

Personalities and consultants team

Title	First name	Last name	Institution	Function	Town	Country
Ms.	Adeline	AKINYI ODUOR	FITCA(K)	Data management	Nairobi	Kenya
Dr.	Burkhard	BAUER	FITCA(K)	Tsetse specialist ; FITCA Kenya co-ordinator	Busia	Kenya
Mr.	David	GITAU	FITCA(K)		Busia	Kenya
Dr.	Ambroise	GIDUDU	LIRI	Project Co-Ordinator FITCA Uganda	Entebbe	Uganda
Dr.	Simon Mutiri	KARANJA	FITCA(K) / KETRI	Veterinary doctor	Busia	Kenya
Ms.	Lucy	KINUTHIA	FITCA(K)		Busia	Kenya
Dr.		MERISYA	Busia district	District Veterinary Officer	Busia	Kenya
Dr.	Martin	ODIIT	LIRI	Medical Doctor LIRI/FITCA component	Tororo	Uganda
Dr.	Francis	OLOO	FITCA(K)		Nairobi	Kenya
Ms.	Winfred	OLUBAI	FITCA(K)		Busia	Kenya
Dr.	Charles P.	OTIM	LIRI	Acting Director LIRI	Tororo	Uganda
Dr.		SO OROT	Bondo district	District Veterinary Officer	Busia	Kenya
Dr.	Christopher	LAKER	LIRI/FITCA	Agricultural Economist	Entebbe	Uganda
Dr.	Simon	GOULD	FITCA(U)	Technical Assistant	Epopo	Uganda
Title	**First name**	**Last name**	**Institution**	**Function**	**Town**	**Country**
Ms.	L. Maida	AWORI	OAU/FITCA(R)	GIS specialist and data manager	Nairobi	Kenya
Dr.	Daniel	BOURZAT	OAU/FITCA-EMMC	Regional co-ordinator Environment component	Nairobi	Kenya
Dr.	Stéphane	DE LA ROCQUE	CIRAD / CIRDES	Veterinary doctor	Montpellier	France
Dr.	Gérard	DE WISPELAERE	CIRAD	Geographer, Ecologist, RS and GIS specialist	Montpellier	France
Dr.	Marcel	DJAMA	CIRAD		Montpellier	France
Mr.	Gilles	FOURNIÉ	CIRAD	Computer scientist	Montpellier	France
Mr.	Gilles	KLEITZ	CIRAD		Montpellier	France
Dr.	Renaud	LANCELOT	ISRA / CIRAD	Veterinary doctor	Dakar-Hann	Sénégal
Dr.	Joseph M.	MAITIMA	ILRI / FITCA	Regional Ecologist	Nairobi	Kenya
	Meshack	NYABENGE	ILRI	GIS and RS specialist	Nairobi	Kenya
Dr.	Karl Heinz	POLITZAR	OAU/FITCA(R)	FITCA Regional Co-Ordinator	Nairobi	Kenya

CIRAD-DIST
Unité bibliothèque
Lavalette