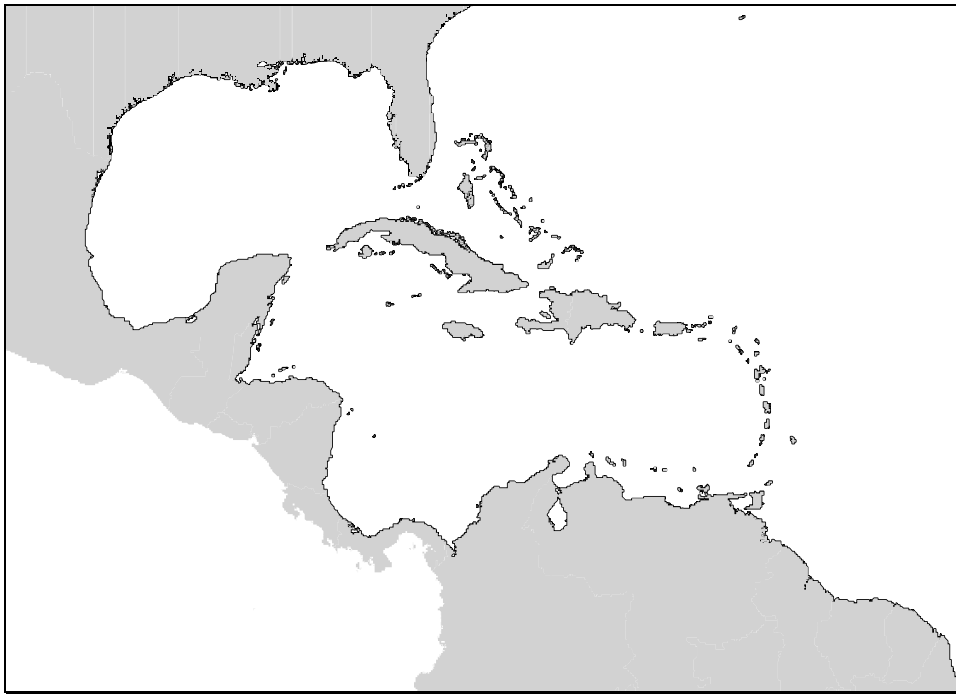**Caribbean Environment Programme**

United Nations Environment Programme

Environmental Data and Information System:
SIMARNA - The Cuban Experience

P

Note**:**

The designations employed and the presentation of the material in this document do not imply the expression of any opinion whatsoever on the part of UNEP concerning the legal status of any State, Territory, city or area, or its authorities, or concerning the delimitation of their frontiers or boundaries. The document contains the views expressed by the authors acting in their individual capacity and may not necessarily reflect the views of UNEP.

TABLE OF CONTENTS

I. BACKGROUND

In late 1987, a multi-disciplinary group began to design what would later become the Environmental Data and Information System (SIMARNA). The System has been in full operation since mid-1989 at the National Commission for the Protection of the Environment and Rational Use of Natural Resources (COMARNA) of Cuba.

The current version of SIMARNA is the result of an exercise which was initially conceived in phases. The first stage comprised of the design and collection of national data and information on matters concerning the environment, including aspects related to the country, its political and administrative division, investments for the protection of the environment, environmental assessment of projects, legal and technical standards, etc. This phase was completed and began its operation in 1988.

The second phase, more complex and of greater scope, took into consideration the design and inclusion of data and information on natural resources, basic sanitary infrastructure, major sources of pollution of island coastal waters, economic utilization of wastes, and other aspects, both on a national as well as provincial scale. It was completed during the first half of 1989, the year in which, simultaneously, the task of collecting national and provincial data was undertaken, using questionnaire that were sent out and returned, mainly through the mail.

Since the beginning of 1990, work has started on the establishment of provincial databases in the respective territories, making good use of the limited capacities that are available in the provinces of the country. In addition, data transmission is being establish through the use of floppy disks and on a smaller scale, modem, until the necessary hardware for the national network is in place.

SIMARNA was conceived to be fundamentally a descriptive database including natural components of the environment and, in addition, to process a significant amount of information on various fields that are related to the management and protection of the environment.

II. OBJECTIVES

To the authorities of the country, SIMARNA makes a steady supply of systematic and comparative data and information on the environment. SIMARNA also prepares regular reports on the state of the environment, and continuously arranges consultations with experts and institutions on various aspects of interest.

III. GENERAL DESCRIPTION OF THE SYSTEM

3.1. Software and Hardware

The DATAEASE programming language, version 4.2, was used mainly because this language is flexible, assimilating dynamic change in database structure, without at the same time impairing stored data.

Among its advantages, in addition, is that of allowing the user to introduce changes in database design, without altering the integrity of the system. This makes operation easy for a wide range of interested persons, starting from those who review and consult the information, and those who obtain the outputs, to those who, as system administrators, require modifications in the database or in its access structure

The operation system MS-DOS was used (version 3.10 or later). The system requires a minimum configuration consisting of:

- IBM compatible micro-computer with 640 K RAM;
- Hard disk with at least 10 MB available;
- Colour video with graphic possibilities;
- Character printer.

DATAEASE is a parametric database programming language. It is based on the relational model, and has great capacity and flexibility in its application.

The capacity of DATAEASE version 4.2 is as follows:

File per database:	2,000
Records per file:	2 billion
Fields per record:	55
Characters per record:	4,000
Active relations:	100
Index files per data entry format:	2,000
Reporting formats per database:	2,000
Screens per data entry format:	16
Field Types:	8
Values per CHOICE field:	99

File extensions are as follows:

DBM:	Data file
DBA:	Form definition file
DBR:	Reports format file
DBI:	Data import format file
DBF:	Data entry format file
Inn:	Form index files, with nn from 01 to 77

3.2. Characteristics

The SIMARNA system is organized around two basic functions:

- Updating (data entry); and
- Retrieval (data output).

The information contained is grouped into five modules: Modules A,B, C, D and E. The modules contain the following information:

Module A: National data and information;
Module B: Quality and protection of various components of environment;
Module C: Provincial information;
Module D: Economic utilization of wastes;
Module E: International Environmental Indicators.

Module D contains information on the economic utilization of wastes, broken down by sectors of the national economy, among which, the agricultural, sugar and food sectors are the most noticeable.

The decision to include this set of data and information in the system, was a direct response to the ever growing importance attained in the fight against the pollution of coastal and inland waters and the atmosphere, by the utilization of residues, generated during economic and social activities. The residues can be utilized economically as energy sources, animal feed, biofertilizers and agricultural irrigation. For developing countries, this is a truly effective option to diminish the negative impact of wastes on natural resources and, at the same time, to increase the recycling and re-utilization of residues generated, whether directly or as secondary raw materials.

Finally, module E contains data and information relating to selected international environmental indicators. This module is programmed in CLIPPER-87.

All data and information handled by SIMARNA original from three main sources:

- Year-books and regular publications from the Cuban State Committee for Statistics;
- Publications and documents prepared by the governing bodies of departments involved in protection, as well as by other agencies of the State's Central Administration, and by scientific research institutes; and
- Publications and documents issued by provincial bodies and organizations.

As a whole, without taking into account Module E, SIMARNA includes:

- - 168 entry forms;
- - 162 relations between formats;
- - 125 reports and other output forms.

3.2.1. Updating of data and information

Data updating follows the arrangements by modules as described above. The organization of data entry formats allows this to be done directly and promptly.

Entry formats of Module A are as follows:

- Territory;
- Climate;
- Major elevations;
- Population by age and sex;
- Number of human settlements;
- Socio-economic indicators;
- Welfare and quality of life;
- Economy, by areas;
- Legal standards;
- Technical standards;
- International treaties;
- Scientific and technical programmes by branches and fundamental research in the environment field;
- International projects;
- Water Resources, comprising of 6 entry forms which contain information on the water balance of the country, assessment of the prospective and current resources, economic and social uses of water resources, data on reservoirs and their main features, as well as on the main rivers, and on the development of annual reservoir capacity;
- Land use;
- Soils;
- Fishery resources;
- Aqueducts, including two entry forms which contain data on the national network and other information;
- Health indicators;
- Communal solid wastes;
- Green areas;
- Liquid wastes and treatment;
- Hazardous wastes;
- Global atmospheric emissions;
- Institutions with responsibilities for the environment.

Module B "Quality and protection of various components of the environment" consists of the following entry formats:

- Inland waters;
- Bays and ports;
- Rain water;
- Impact on human settlements;
- Stratospheric ozone;

- Soils;
- Reforestation and treatment of forests.

Module C "Provincial information" consists of the following entry formats:

- Political and administrative division;
- Investments in existing sources of pollution;
- Investment in sewerage;
- Monitoring of the anti-pollution plan;
- Environmental assessment of micro-locations;
- Environmental assessment of project documents;
- Water resources, consisting of 5 entry forms which contain information on ground and surface water resources in the provinces, both in quantity and by main uses, data on reservoirs, as well as rivers receiving liquid wastes, and impact on the ground water; Coastal resources due to the importance of coastal resources for the country, 4 data and information forms were designed on major bays and ports, estuaries, coastal lakes and beaches, with details on aspects of current use, sources of pollution, major impacts, present quality status, protective and restorative measures, etc:
- Land use;
- Forest resources;
- Protected areas;
- Mangroves;
- Flora;
- Fauna;
- Mineral resources, comprising of 4 entry forms with information on mining sites, areas reserved for mining, quarries and hydrocarbons;
- Major tourism and recreational resources;
- Aqueducts, including 2 entry forms which contain data on potable water treatment plants, etc;
- Sewerage, including 2 entry forms which contain data on sewage treatment plants, etc;
- Health;
- Communal solid residues;
- Green areas;
- Historical environmental events;
- Sources of inland water pollution;
- Air pollution sources;
- Hazardous residues;
- Pre-school education;
- Primary education;
- Secondary education;
- Technical and professional education;
- Teacher education;
- Special education;
- Higher education;
- Teaching personnel;
- Other institutions;

- Dissemination.

On the other hand Module D "Economic Utilization of Wastes" consists of the following entry formats:

- General data on sugar mills;
- Sugar mills with fertirrigation;
- Collection of mudwaters;
- "Gicabu";
- Food production from residues;
- Cane conditioning centre;
- Provincial data on torula factories;
- Residues potential for soils;
- Organic matter applied to soils;
- Worm breeding;
- Local feed;
- Liquid feed;
- Tankage;
- Protein paste;
- In cage (poultry-dung);
- Standing (litter);
- Poultry excreta potential;
- Distilleries;
- Breweries;
- Canning and vegetable industry;
- Dairy industry;
- Washing and recovery of soda;
- Meat industry;
- Flour and confectionery industry;
- National data on trash fish;
- Fish processors;
- Recovery of raw materials;
- Use of bagasse;
- Tannery industry;
- Textile industry;
- Recovered oil.

3.2.2. Data and Information Retrieval

The system's data retrieval is similarly organized by modules, and arranged according to the criteria of groups.

The structure for retrieval in Module A is as follows:

- General information on environmental matters;
- National information on major natural resources;

- Institutions with responsibilities pertaining to the environment.

"General information" contains a total of 13 outputs, comprising investments by agencies, environmental legislation, scientific and technical research on various environmental issues.

"Natural information on the major natural resources" has one output related to the use of water resources.

"Institutions with responsibilities pertaining to the environment" contains 2 outputs with information on the relevant institutions and agencies.

Module B has a total of 7 outputs allowing for the quantitative assessment of the quality of the major components of the environment.

The retrieval structure in Module C is as follows:

- Information on the plan for the economy;
- Information on the evaluation of the investment process;
- Information from provinces on major natural resources;
- Indicators of sanitary infrastructure;
- Historical events pertaining to the environment;
- Major sources of pollution;
- Environmental education, training and awareness.

"Information on the plan for the economy" contains a total of 7 outputs, which comprise the sectors of the plan for the economy, relating to protection of the environment.

" Information on the evaluation of the investment process" contains a total of 12 outputs with information on micro-locations and on environmental assessment.

"Information on natural resources" contains a total of 31 data and information retrieval formats on water, coastal, land and forest resources, flora, fauna, protected areas, mineral resources, and the major tourism and recreational resources.

"Indicators of sanitary infrastructure" comprises a total of 5 outputs on aqueducts, sewer systems, community solid residues and green areas.

"Historical environmental events" comprises 2 outputs with information on major events related to the protection of the environment.

"Major sources of pollution" contains 3 data retrieval outputs on sources of pollution of coastal and inland waters and of the atmosphere, as well as on the installations which generate hazardous residues.

"Education, training and awareness" comprises 10 outputs on pre-school, primary, secondary, higher and other education, each broken down by provinces.

The retrieval structure of Module D "Economic utilization of wastes" is firstly, according to economic sectors and, secondly, to the character of data, which are broken down into national and provincial data. The structure is as follows:

- Sugar sector;
- Agricultural sector;
- Food industry sector;
- Textile and tannery sector;
- Fisheries sector;
- Raw material recovery.

"Sugar sector" comprises 9 data outputs concerning sugar mills, fertirrigation with its effluent of mudwaters, food production, cane conditioning centre, distilleries, "tomba" factories, and others.

"Agricultural sector" includes 12 retrievals on organic matter applied to soils, its potential and current use, worm-breeding, local feed processing, fish silage, liquid feed, tankage and protein pastes, poultry excreta in animal feed and others.

"Food industry sector" contains 5 outputs on distilleries, breweries, canning and vegetable industry, meat and flour/confectionery industries, reflecting current and potential uses of residues as animal feed.

"Textile and tannery industry sector" comprises 2 outputs and the "fisheries sector" another 2 retrievals on trash fish used in animal feed and fish processing.

"Raw material recovery" contains 4 outputs on the use of bagasse, tires and recovered oil, glass and other materials.

"The outputs relating to provinces are similar to what already has been described.

Module E, "International Environmental Indicators" was programmed using the Clipper-87/DBEDIT technique, which allows to add, modify, omit and show items on a screen. It consists of 20 databases, as follows:

CPOBLT: Total population;
CPOBLUR: Urban and total population;
CVIDAR: Life expectancy at birth;
CENF: Main causes of infant mortality in underdeveloped countries;
CMORT: Deaths per 1,000 live births;
CAGUA: Population with potable water service;
CALCANT: Population serviced by sewer systems;
CHIDRAUP: Water resources, for all uses per capita;
CHIDRAUE: Water resources, structure by use;
CTIERRA: Land use;
CEROSION: Erosion;
CFOREST: bForest resources;

CPROTEM: Protected areas;
CFAUNA: Fauna;
CFLORA: Flora;
CCOSTA: Coasts;
CATMOS: Atmosphere;
CRES: Generation of wastes (municipal, industrial, hazardous);
CRESU: Generation of urban wastes;
CFERT: Fertilizers and Pesticides.

This module is structured into a major programme (titled COMARNA), 5 procedures which group related subjects and 10 programmes with one single subject.

The procedures are:

- POBLAC.PRG: Population, comprising:

  Total population;
  Urban and rural population.

- SALUD.PRG: Health, containing:

  Life expectancy at birth;
  Infant mortality;
  Major causes of infant mortality.

- RECHIDCA.PRG: Water resources, comprising:

  Per capita for all uses;
  Structure by uses.

- ATMOSFER.PRG: Atmosphere, comprising:

  SO2 concentration;
  Concentration of particulate matter in suspension.

- RESIDUAL.PRG: Wastes generation, containing:

  Municipal, industrial and hazardous;
  Urban, from selected cities.

Additionally, the programmes on a single subject are:

CANTAGUA.PRG: Drinking water;
CANTALCA.PRG: Sewerage;
USOTIERR.PRG: Land use;
EROSION.PRG: Erosion;

RECFORES.PRG:	Forest resources;
AREAPROT.PRG:	Protected areas;
FAUNA.PRG:	Fauna;
FLORA.PRG:	Flora;
COSTAS.PRG:	Coasts;
FERTILIZ.PRG:	Fertilizers and perticides.

IV.   CURRENT LIMITATIONS OF THE SYSTEM

The present main limitations for the development of the system, both qualitative as well as quantitative, are related to:

Statistical processing of data series contained in the system, through the appropriate programme. To solve this problem, version 4.2 of DATAEASE can work in association with DATAEASE GRAFTALK, to allow the above-mentioned processing to take place with the same database already created, and without the need of exporting. The acquisition of GRAFTALK by the system administrators would be a significant improvement in the quality of operation, and would enhance both the processing and presentation of data and information.

Geographic presentation of data and information. To date, there is neither the possibility nor the material means for the geographic presentation of the contents of the system, whether by political and administrative division, watersheds, regions or otherwise.

The acquisition of the necessary hardware for this (plotting of curves and corresponding software), would contribute to enhance the efficiency of the Cuban Authorities in their management of the protection of the environment.

Installing hardware and establishing electronic transmission capabilities in the Provincial Commissions, would enable the completion of the national network.

Although work is being done on the installation of the provincial database, in centres which are not in the system, the non-availability of own computer hardware is a factor unfavorable to its consolidation and operation.

V.   NETWORKING SYSTEM BETWEEN NATIONAL AND PROVINCIAL DATABASES

5.1. Current situation

During 1988-1990 and up to the present date in 1991, data and information entry into SIMARNA has been based mainly on direct collection of the data originating from three existing basic sources, which are:

- Statistical year-books and regular publications on Cuba by the State Committee for Statistics, the governing body for information in the country;
- Documents and publications by the governing agencies for departments concerned with the protection of the environment, and by other agencies in the Central Administration of the State and by scientific research institutes.
- Documents and publications issued by provincial institutions and entities.

Because data collection is mainly done on a yearly basis, this collection, as well as data entry follow a previously established national and provincial timetable, which is permanently monitored and evaluated by the National Commission.

The data and information flow through both national and provincial databases is shown in figure 1.

Data transmission and communication has been mostly done by mail or handcarrying, after the National Commission issued forms to the sources of the system.

With the planned installation of databases in the provinces, and with existing limited capabilities in mind, data transmission and communication will be done by floppy disks, and on an experimental basis in one province, through modem.

Since the establishment of SIMARNA and commencement of operation, specific databases were created, to cover the particular interests of external users, the request of other users, Thus, the database for the National Water Resources Institute was developed, including all data and information concerning the inland waters of the country and related aspects. In addition, a database on the Economic Utilization of Wastes was installed in the Central Planning Board.

A simplified database, based on SIMARNA, was developed for the office of the President of the National Commission. Following this trend, in 1990 the official statistical sub-system on matters concerning the environment was established under the directorship of the State Committee on Statistics.

5.2. Prospects

In order to become a reality, the national network needs to be provided with the necessary computer hardware, allowing for the system's continued development and, moreover, allowing for increase in the number of current and potential users who need its services for the efficient management of the protection of the environment.

In future, the national network for data and information transfer on the environment, including the National Commission and the Provincial Commissions, should mainly function through modems located in the District Centres. The modem capabilities are yet to be created.

Data flow between the District Centres and associated territories will be done mostly through floppy disks. Map 1 shows what has been described above.

VI. METHODS OF DATA QUALITY CONTROL

As was stated in earlier pages, the data and information sources of the system are the national and provincial monitoring and research institutions, with recognized experience in various environmental and economic fields. Numerous information and data sets, come directly from the Cuban State Committee for Statistics, which is the governing body for information, as well as from other agencies in charge of specific tasks in environmental protection.

Following the procedure established, the national database is fed with data and information issued by institutions at national level, and likewise, the respective provincial databases are loaded with data obtained from the provincial entities. Figure 1 shows this situation.

Data quality control is mainly determined by two factors:

- Authority and specialization of the bodies issuing the data in the field in question;
- "Double checking" of data between the national and provincial entities, which is mostly done on a set of entry forms concerning natural resources.

These institutions are staffed with personnel skilled in monitoring and research. A characteristic feature of the work developed with the system to date is that the responsibility for data entry into the national, as well as provincial databases, is with those members of staff of the National and Provincial Commissions who are specialized in the pertinent field of the environment, and act as controllers and censors of the quality of the information to be entered.

If there are data and information which present some degree of uncertainty, direct links are established between those in charge of the system and the respective issuer, so as to ensure proper clarification.

The National Commission for the Protection of the Environment in Cuba (COMARNA) uses data and information issued by all these institutions in a comprehensive manner, in order to diagnose the state of the national environment, and as a support for decision-making. COMARNA does not fundamental data, nor does it include institutions which undertake such an activity. The National Commission co-ordinates the implementation of national policy on the environment, and it governs the National System of Protection, for which it makes use of SIMARNA. When created, SIMARNA took into consideration a number of favorable characteristics of the country. Among these, the following can be mentioned:

- Economic and social, as well as scientific and technical degrees of development, and the integration of these in the management of the protection of the environment;
- Level of development of the national institutions concerned with monitoring and research, which ensures a high quality of collected data and information;
- Available background information on the importance and priority attached to the protection of the environment, and the national use by the authorities of the natural resources;
- Guarantee that the data and information on the environment will be used for proper decision-making in the management process;

- Existing infrastructure and possibilities of extension;
- Experience, training and specialization to the personnel administering the system.

Finally, the data and information of Module E "International Environmental Indicators", are collected from prestigious publication published by the United Nations System and specialized international institutes.

In conformity with the development plan for SIMARNA over the next few years as conceived by the National Commission, at least 7 District Centres should be created with the necessary computer hardware, so that the National Network can be establish as stated in previous pages. It is anticipated that this Network will be capable of transmitting, data and information contained in the system electronically.

A number of District Centres will have other provinces associated with them, in case these provinces do not coincide with the political and administrative division of the country.

Map 1 shows the proposed structure of the Network.

The Centres will be linked to the National Centre, which already has a basic computing configuration acquired in 1989 through the CEPNET Regional Programme. However, it needs to be strengthened.

The foregoing, together with other specifications forms the bases of the project proposed by Cuba to Regional Co-ordinating Unit (RCU) of UNEP's Caribbean Environment Programme, Jamaica in the latter part of 1989, as part of the regional effort to strengthen and increase national capabilities in the handling of data and information on the environment.

National Centre

District Centres

Associated Territories

Map 1. National Network

VII. REGIONAL AND INTERNATIONAL LINKS

Since the establishment of SIMARNA, within the CEPNET Regional Programme, the National Commission has been considering the feasibility of integration of SIMARNA in the RCU, Jamaica, through electronic data transmission.

Based on the original design of SIMARNA, its administrators developed a prototype system, simplified and adapted to the interests of the region. This would become a part of RCU's

databases, and would be available to the countries participating in CEPNET. This activity could pave the way for the establishment of a Regional Databases, if all the other countries were to undertake similar activities.

Bearing in mind that the CEPNET plans to hold a workshop, for exchanging experiences on this particular aspect, it would seem advisable to request all participating countries to draft proposals for discussing the database content prior to this workshop. This would be followed by the specific design of regional entry formats, and the definition of the system's output.

A matter that will demand regional attention, is the one relating to the software to be used as the database programming language. This is based on the fact that there are countries that have been establishing databases. and software has not been necessarily similar.

The way in which data and information will be transmitted electronically between proposal National Centres and the Regional Centres, will also have to be determined, as well as the feasibility of establishing communication between the National Centres. This will also include the choice of the transmission software, and the budget allocated for the communication.

At present, SIMARNA has no established links with Data Centres in the Caribbean region, nor to any other geographical area.

VIII. OTHER RELEVANT ASPECTS

In accordance with the specific characteristics of the country and the responsibilities of the National Commission, SIMARNA has been designed to include only those sets of data and information planned to be utilized in outputs and specialized consultations. This ensures that the system remains updated and dynamic, as an instrument to assist decision-making, and the preparation of systematic reports on the state of the environment.

Although in the first stages quantitative and descriptive aspects were mainly considered, later significant qualitative aspects were included. Such is the case with entry forms for examples, for sources of pollution of coastal and inland waters and sources of pollution of the atmosphere.

Since 1990, SIMARNA includes Module B "Quality and protection of various components of the environment", which includes a group of qualitative elements concerning the state of waters, atmosphere, soils and forests. Added to this are quantitative indicators concerning the protective measures that are being applied.

During the preparation for the design of a data and information system, to be used by COMARNA, various alternatives were examined. Among where the following:

- Development of a descriptive system, especially oriented to the compilation of general and specific data on areas concerning the environment and the components of these, with mainly quantitative information, but supplemented with indispensable qualitative data;

- Development of a basic data network on the quality of the environment, which would include a set of qualitative indicators obtained from the State on major natural resources (coastal waters, inland waters, atmosphere, soils);
- Compilation and mapping of data and information on the characteristics and state of the environment, by means of a Geographic Information System (GIS).

In accordance with these basic perspectives, which have not wrongly been given priority nor are they contradictory, and taking into consideration the responsibilities of the National Commission for the protection of the environment in Cuba, it was decided that, in its first stage, a descriptive-quantitative system would be designed, and later the network on the quality of the environment would be included. The network began in 1990 and is currently in the phase of data collection.

The linkage of all the results in Geographic Information Systems, would mean the completion of the integral design of the system. This will depend on the possibility of acquiring the necessary computer hardware.

IX. BIBLIOGRAPHY

Technical Project for the design of a data and information database on the environment. 4th version. COMARNA, ACC. January 1987 (typed document).

Users Manual: Data and Information System on the Environment, CEDISAC-COMARNA, ACC. R. Barrera, J. Hernández, J.M. García, D. Fonseca, A. Blanca. December 1987 (typed document).

Applying DATAEASE to a database on the Environment. R. Barrera. Informática 1988. Palacio de las Convenciones. La Habana, Cuba.