

CSNT/89/08

ECONOMIC COMMISSION FOR LATIN AMERICA AND THE CARIBBEAN
Subregional Headquarters for the Caribbean

CARIBBEAN DEVELOPMENT AND CO-OPERATION COMMITTEE

INTERNATIONAL DEVELOPMENT RESEARCH CENTRE

COLLOQUIUM ON STATISTICS AND THE NEW TECHNOLOGIES

Holiday Inn, Port of Spain

3 - 5 October, 1989

**THE ESTABLISHMENT OF INTEGRATED NATIONAL
AND REGIONAL STATISTICAL AND BIBLIOGRAPHICAL
INFORMATION DATABASES AND REMOTE ACCESS TO THEM**

Prepared by

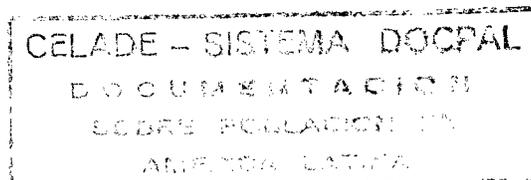
Wilma Primus

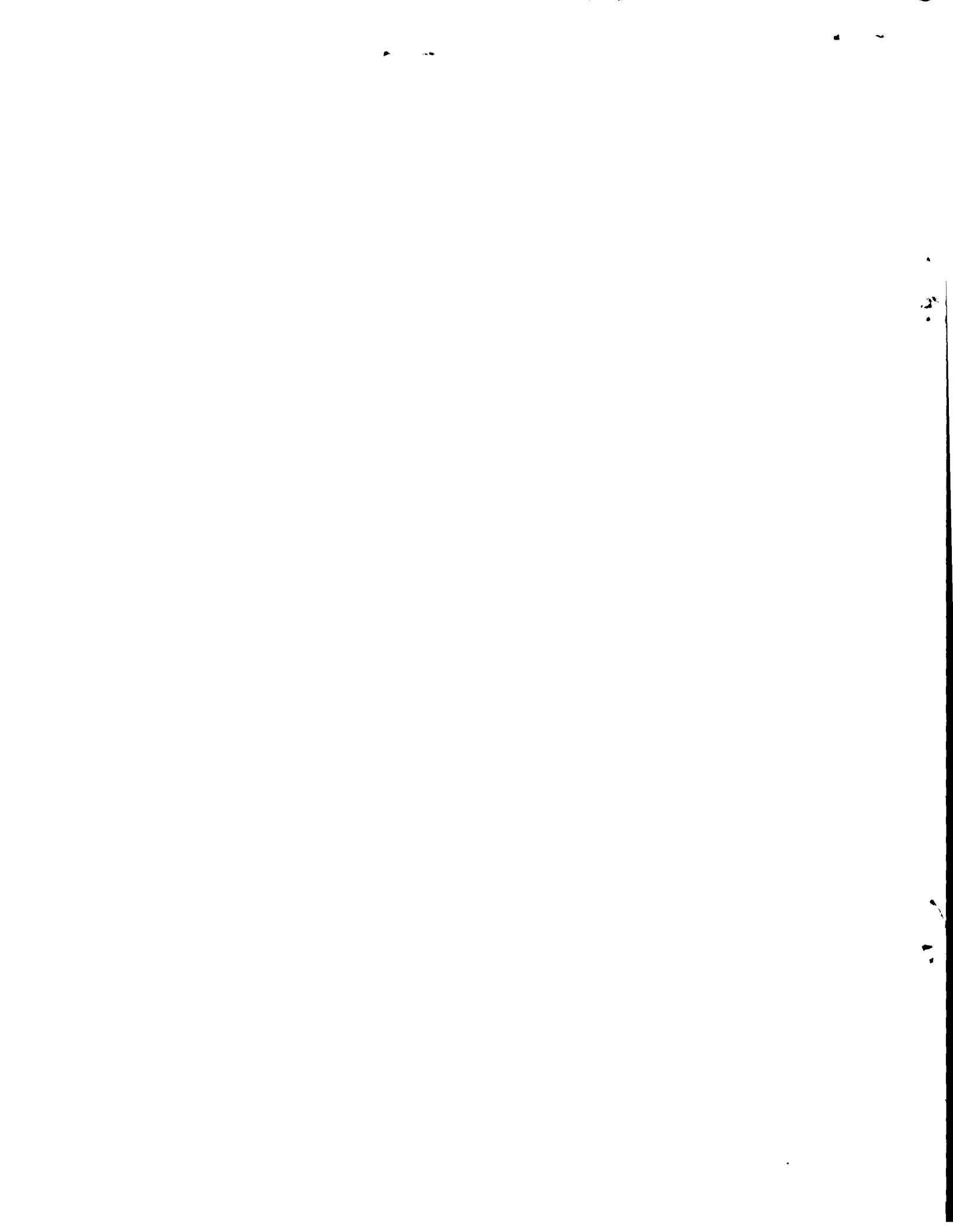
Project Co-ordinator

Caribbean Documentation Centre

UNECLAC

TRINIDAD AND TOBAGO





THE ESTABLISHMENT OF INTEGRATED NATIONAL AND REGIONAL STATISTICAL AND BIBLIOGRAPHIC INFORMATION DATABASES AND REMOTE ACCESS TO THEM

Introduction

Libraries play an important role in facilitating organized information flows and in achieving coordinated bibliographic control of documentation. Without this control it will be difficult to identify or locate information on any given topic contained in the published literature. For many years bibliographic access has been through the use of conventional card catalogues and numerous volumes of published bibliographies, indexes, abstracts and directories. The exponential growth in publications taxed the ability of librarians to store and retrieve relevant information in a timely fashion. Fortunately, the increasing power, flexibility and availability of computers and rapid changes in communications technology enabled impressive progress in the development of computer-based bibliographic files which can be searched interactively from remote terminals. This facilitated resource sharing and networking and today several such networks are being developed in the region. There is considerable interest by information specialists and others in the region to extend these computer based information handling techniques to include statistical data but the difficulties experienced in using statistical sources have been a deterrent. What seems to be needed is the development of a comprehensive concept of data collection and supply at the national and regional levels and the identification and formulation of operational follow-up activities for the establishment of databases of frequently requested data sets.

Bibliographic Databases

In the Caribbean, the establishment of national and regional databases is the result of a co-ordinated effort, based on structured sectoral or mission-oriented systems and networks designed to provide identified groups of users with timely access to information to meet specific needs. Each network comprises information units, coordinated in each country by a national focal point and at the regional level by a regional institution. Input to the system and services to the users are decentralized, while overall planning, training and maintenance of the regional databases are the responsibility of the regional coordinating centre. The national focal points identify, collect and analyze relevant national literature. This activity may be shared with other participating libraries but the focal points retain the overall responsibility for coordinating the documentation efforts in order to ensure complete coverage of the national literature. Many of the libraries use microcomputers to create the national databases and to provide the national input on diskettes to the regional coordinating centres, which then merge the bibliographic records and maintain a centralized database. In this way skills and capabilities are gradually being developed.

These databases permit rapid search of thousands of documents on the basis of carefully selected sets of terms which attempt to match the interest of the user with the contents of the documents. The result of such a search is a concise listing (sometimes with abstracts) of potentially relevant documents which can then be

evaluated. Online bibliographic search systems have proven to be very useful.

Nevertheless, today's economists, market researchers, planners, consultants, etc. are involved in complex development situations involving different disciplines and sectors. Access to the traditional primary and secondary literature only partially satisfy their needs and a higher value is being placed on rapid access to quantitative data from a wide range of sources. These users need to have the data broken down in various ways to enable them to get better analyses and to get more dimensions into the analyses. In addition many of these users operate in a technologically changed environment and are eager to manipulate data in machine readable form. They see the need for integrated statistical databases which bring together basic statistics from many different fields and data collection programmes and recompile them for many different purposes.

Integrated Statistical Databases

The development of integrated statistical databases should be based on priority needs and should be constructed with some well specified set of uses in view.

A pre-condition to the development of fully integrated statistics is a consistent **conceptual framework**. In the absence of a comprehensive framework, consistency would be the main instrument for the integration of data - consistency of concepts, definitions and classifications; consistency in reporting units and weighting patterns, in standards of documentation. Other considerations are flexibility to accommodate unforeseen uses, political sensibilities, maintenance of confidentiality, a clear definition of the scope of the database and sources of input, and identification of prospective use. Standardized definitions and classifications are important and where these have been developed and widely accepted, they should be used as a matter of course.

Proper Documentation is critical and every database should be accompanied by a data description that will permit someone unfamiliar with the file to use it. This is a time consuming task which must be included in all budgeting. Documentation is required for text files and data files. The text files contain descriptive information on data availability and sources. They help the user to become familiar with the database structure and sources of data and are essential for interpretation of the data files - which contain the structured statistical data.

Where input is decentralized, an input manual should be prepared, describing the required data elements and input forms and instruction for software installation. Training is also a critical element in decentralized systems.

Some institutions in the region, including libraries, are attempting to meet user demands through the creation of statistical databases. The energy and trade information databases follow the concept of networking initiated by the bibliographic systems.

The Caribbean Energy Information System (CEIS)

CEIS is a cooperative network in which libraries in twelve Caribbean countries and four regional institutions share and exchange information in support of Caribbean energy activities. The system is coordinated by a regional focal point (the Scientific Research Centre of Jamaica) and stores, retrieves and disseminates bibliographic, quantitative and referral data in the energy sector.

System Design - A feasibility study was undertaken by an energy expert, a scientist and a librarian to identify potential user requirements and priorities. Subsequently, potential network institutions were named, agreements signed, and a meeting held to determine the type of system needed. A systems consultant was then contracted to design the system and to identify hardware and software requirements.

Hardware and Software:- At the national level, microcomputers, Lotus 123 and CDS/ISIS are used to create the national databases and at the regional coordinating level, an HP minicomputer, SPSS-X and Minisis store, analyse and retrieve energy data relating to the twelve participating Caribbean countries.

In terms of quantitative data, five modules have been designed for the petroleum database and six for alternative energy. The petroleum modules are as follows:

1. Supply - stores supply data on all energy sources including potential resources, proven source inventories of New and Renewable resources, production data (from oil wells, etc.), fuel imports, exports, and ocean losses.
2. Transformation - stores transformation data on all energy sources including electricity generation, installed capacity, refinery inputs, outputs and losses.
3. Consumption - stores consumption data by sector and resources (including prices).
4. Economic and financial - stores data on the cost associated with extraction, processing and consumption as well as general economic data, energy import costs, public or private debt, balance of payments, consumer price index.
5. Demographic/technical/social - stores technical data including urban/rural population, employment, per capita income; transportation sector - passenger vehicles, freight transport, etc.

The Alternative Energy Modules, which will become operational shortly, will store data relating to wind, solar, hydropower, biomass, geothermal energy, and energy conservation.

The Referral Database stores data on energy research projects, energy expertise, and products and services available on energy in the region. The printed output is a **Directory of Energy Research in the Caribbean**.

The **Bibliographic Databases** provide access to published and unpublished energy information available in the region and to material held in the International Development Research Centre Energy Research Group database.

Considerable documentation has been prepared on the databases. There is a very detailed input manual which deals with data collection sources, database file description specifications, description of data elements and the actual input sheets for each module. Instructions are also provided on how to install and use the worksheets in Lotus 123, print the input forms and update the worksheets.

Workshops and in-house training have helped participants in the use of the spreadsheet software and in reformatting the data.

Each national focal point collects the data in the formats in which they were issued from the respective sources (energy, economic and financial, or sectoral) and, where necessary, reformats the data before entry on the CEIS input sheets. Data are submitted on a half-yearly basis and run from 1985.

A **Caribbean Energy Report** is published and disseminated and the data can be used to establish models and provide the basis for planning and forecasting.

Another system, in the area of trade data, is being developed by the **CARICOM Secretariat**. This database includes country product data, volume traded and value, data for exporting and importing country, exports by products, and trade balances. **DBASE III** is used, with a window to Lotus 123.

ECLAC's Demography Unit is also creating a data bank of country and regional tables providing data on population, births, deaths, migration, marital status, health, and education. The data are drawn from published sources. **ECLAC's Statistical Unit** has also developed a databank which stores and retrieves time series of major economic and social indicators for Caribbean countries. Lotus 123 is used in both databases.

The **Agricultural Information System of the Association for Caribbean Transformation (ACT)** is a statistical database consisting of five modules usable by decision makers from policy level to production and marketing strategies. The database provides weekly average wholesale prices for 42 domestic food crops for nine Caricom countries; production cost data on these crops and technical coefficients; import/export levels and direction of trade, and retail agricultural input prices. Lotus 123 and **DBase III** are used to store and analyse the data on a microcomputer.

Despite these responses to specific demands, what seems to be needed is a coordinated national/regional approach to collecting, processing and storing socio-economic, natural resources and other types of frequently required data sets which can be accessed in accessible machine-readable form. There is a need for compromise between the requirement for depth and detail and the requirement for broad coverage, and for guidelines for harmonization of the data and for the creation of databases. There seems to be a need to examine more closely the search languages of numeric

databases, which are influenced by the specific database, the subject, and the search elements. The search languages of bibliographic databases follow the same principle of boolean logic and keyword searching.

A regional effort would aim to collect, store, measure and analyse Caribbean economic, social and other activities and the region's relationships with the rest of the world. Such a database could include

- General statistics (general data on short-term economic trends, on population and employment, industry, agriculture, prices, services, transport, finance and general accounts)
- National Accounts, finance and balance of payments
- Agriculture, forestry and fisheries
- External trade
- Miscellaneous data, e.g. governments' expenditure on R&D

The data should be available online and perhaps in spreadsheet form.

Access to Bibliographic and Statistical Databases

Computerized data dissemination is developing along two parallel paths - distribution of databases in some tangible storage medium (diskettes and tapes are used to exchange data by existing systems) and distribution through online interactive networks.

At present ECLAC's host computer provides remote online access to three regional bibliographic databases covering socio-economic, patents and agriculture, and to two international databases - LABORDOC and INFOTERRA (produced by ILO and UNEP respectively). ACT's Agricultural Information System is also online for remote access. Despite the fact that the infrastructure for computer-based communication is in place and electronic messaging services and host computer facilities are being offered by some telecommunications authorities, there are limitations to remote access to both services.

A growing use of the telecommunications facilities is foreseen and a trend to decentralization of remote access, which would enable micro to micro linking and access by small retrieval systems at the specialist user's workplace to larger national and regional on-line systems.

The Role of Libraries in the Dissemination Process

Existing bibliographic and data services are underutilized. Information specialists can help by creating directory type referral and inventory mechanisms to direct users

to relevant sources. These referral devices could be in printed and machine-readable form, with regular updates. On-line directories of regional and international data sources could be made available for access.

There is a need for greater awareness among librarians of available data sources and data handling techniques. Statisticians and database producers can play a more positive role in ensuring that librarians, as intermediaries, are trained in the use of these databases, since they are in an excellent position to promote and market them. They should also draw on their experiences and overall knowledge in file design, bibliographic database creation, user manuals and promotional activities. Database producers should also prepare user friendly online promotional material on the content and use of their products.

Conclusion

The computerization of data collections will continue to increase to meet demand. A structured and coordinated approach is critical for the region, especially since the costs associated with the creation and use of databanks are high. International and regional cooperation should be pursued and an exchange of existing experiences in the building of databases should be encouraged.