MTA SZTAKI Közlemények 34/1986 pp. 53-78

A DISTRIBUTED DATABASE MANAGEMENT SYSTEM IN TERRITORIAL PLANNING

P. Dipotet[2], A. Benczur[3]

INTRODUCTION

Programs concerning the territorial planning and the long range development of economic regions are extremely needed, but only seldom carried out in developing countries. In Cuba, conditions have been created[1] in order to successfully face those important Programs. Regional planning is concerned with the collection, filtering, storage, processing, protection and dissemination of quite a lot of different kind of data (geographical, social, economical, etc.) spatially distributed. In other words, a very efficient and complex information system is needed, helping the experts in the regional planning environment, to make good decisions in the adequate time.
The former information system belongs then to a territorial-multi-site organisation,[1] with decentralised operations, but requiring intersite information for integrated control at the higher level.
The need of a decentralised management with local control and at the same time cooperating at a higher level for greater good, makes distributed databases very desirable for this application. With cheaper ad more

1 DIPOTET P.-81. Elaboración de un Sistema Hombre-Maquina para la Proyección del Desarrollo de Nuevos Territorios. Habana. IMACC-ACC. T-31904

2. IMACC-ACC Habana, Cuba
3. ELTE University. Budapest

reliable hardware and communication links, this technology is a real alternative to centralised processing. In our case, without big installed mainframes and with good quality micro and mini-computers available, the distributed approach has been adopted in our project.
In this paper we present the preliminary ideas of the distributed database system we are designing for a long range regional planning Program.

PROBLEM STATEMENT

When we analyze the regional planning problems[1].We observe there are external factors, that we must take as compulsory, and internal factors, particular for the Region, bounding, to some extent the development of the Region.
The external factor , policy for long range development of the country, give us:

- global requirements in products, raw materials and services from the Region;
- the external resources to be allocated in the Region during the planning period;
- indicators for social and institutional infrastructure to be developed in the Region.

The main internal factors are: natural and human resources; population requirements; tradition, experiences and material resources accumulated in the Region.

The economic development depends on the efficient management of external and internal reources in the sense of satisfying actual and forecasted requirements.

In order to have a clear picture of the former factors and, in general, of the Region current situation and trends, a retrospective analysis of its economic development was performed. The former analisys allowed us to get more precision about the objectives, tasks and functions of the Program.

Main Objectives of the Program

To elaborate alternatives for the long range socio-econimical development of the region allowing to satisfy, as much as possible, the needs, motivations and aspirations of the local population; and to maximize the regional contribution to the global social product of the country

To maximize the satisfaction of people needs, motivations and aspirations.

To develop the food and agricultural production needed to satisfy regional and national requirements.

To develop the assortment of industrial products, according to regional and national requirements and local possibilities.

To develop the social and productive infrastructure satisfying the regional development requirements.

To look for the balanced rate and proportions of the regional development

PRINCIPAL TASKS OF THE PROGRAM

To study the regional way of life and to caracterize the factors and indicators of the social development of the region.	Social Models	System for simulating the long range development alternatives of a developing region, and also to supervise and control the performences of selected plans.
To develop a model to prognosticate the population growth in the region.		
To develop models simulating alternatives for the long range agricultural and food production in the region		
To develop models simulating alternatives for the long range planning of the industrial sector of the region.		
To develop models and systems simulating alternatives for the long range planning of the building sector of the region.	Economical Models and Systems	
To develop models and systems simulating alternatives for the long range development of the regional infrastructure.		
To develop systems updating the economical, material and resources balances and communicating to given people new information and information changes.	Balance and Information Systems	

Some Indicators of the Socio-Economic Development of the Region

In order to know and to follow the development of the region and to have references allowing us decision making, some indicators are next given. These indicators give us quantitative information which can be used in economical planning, supervision and control processes.

One known indicator is the relation between fixed assets and the labor. It measures the intensity of economical investments in the region.

We denote it $F_1 = f_1 \frac{f.a.}{w_r}$, where f.a.= fixed assets in $ (residual value)
w_r = number of workers.
f_1 = coeffient of propertionality

Indicators of economic efficiency are:

Productivity = $f_2 \frac{IP}{w_r} = F_2$, IP = Internal Production (S)
w_r = Number of workers

Cost per dollar $= F_3 = f_3 \frac{C}{P}$ C = total investment (S)
P = total output production (S)

Expected time to recuperate investments $= F_4 = f_4 \frac{I}{P}$
I = Investment (S)
P = production per year (S/Year)

Indicator of way of life (Related to the national media):

$$F_5 = f_5 \frac{E_r \cdot S_r \cdot O_r \cdot F_{2r} \cdot N_r \cdot m_r \cdot 1/F_{3r}}{E_n \cdot S_n \cdot O_n \cdot F_{2n} \cdot N_n \cdot m_n \cdot 1/F_{3n}}$$

where E = Expected life

S = mean salary

O = Ocupation coefficient = $\frac{w_r}{h}$, h=number of in habitents

N = cultural (school) level

m = migration coefficient = $1+g\,\frac{n}{h}$

n = number of inmigrants

g = coefficient (100 in our case)

$()_r, ()_n$ = regional, national mean value respectively

Former indicators are simplified ones. In developing countries these indicators are frequently affected by technological, skilfulness and tradition, organizational, etc. One of the main points in the organizational factor is the quality of the actual information system in our geographically distributed and complex environment.

In our developing region, a dynamic approach in the development of the Program[2], was applied (Figure 1). This approach allowed us to gradually improve the available information about de region, and to use also the new information in decision making.

With the improvement of the information base and, therefore, the new knowledge about the object and its environment, it is now possible to develop an information system for planning the long range development of the region. In other words, it is needed to perform the simulation of long range investments planning projects for the region, and this work must be based on some kind of man-machine system, helping the "man" (responsible for decision making) to evaluate the consequences of different plan alternatives and to control the deviations of actual plans (Fig. 2.)

The proposed information system must be geographically distributed with one center (regional node) and 14 sub-regional nodes). Each site should be capable of operating independently. In our case, each node has strong relation with the center, and weak relations with the other nodes. Most informational activity originates from pre-defined transactions that are executed repeatedly using the same statements.

2 Dipotet P. Models in the socio-econimic development of a new agricultural region.
Közlemények 27/1982. Budapest. MTA SZTAKI

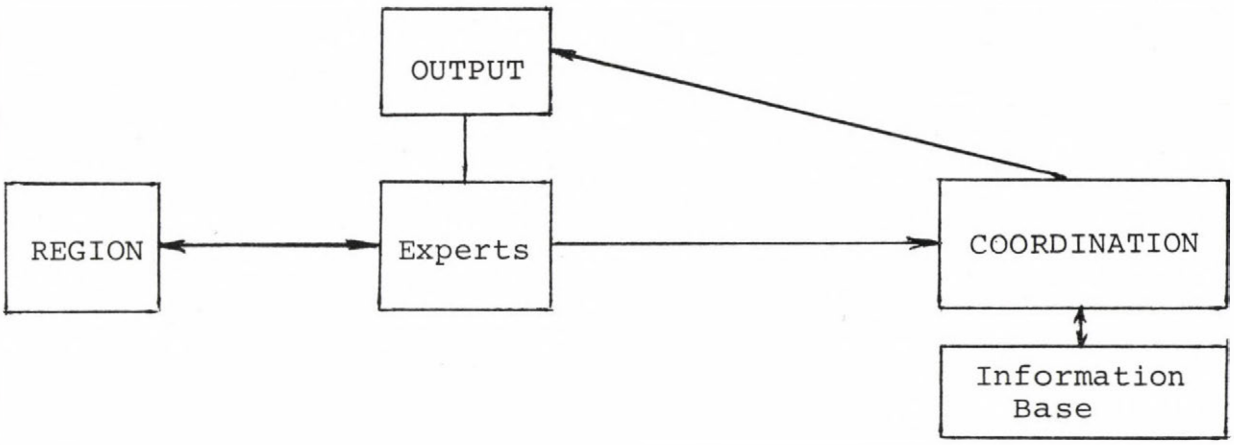


Figure 1. Preliminary system to study the Region

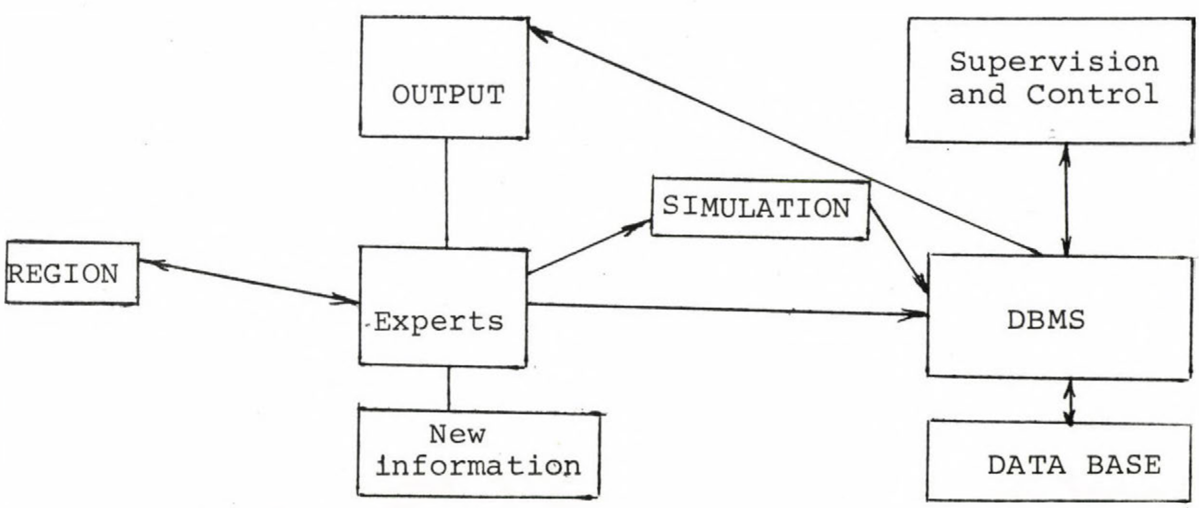


Figure 2. System for long range planning of a developing Region

In our case we require a distributed data base management system for the following reasons:[3]

- Economic. Large databases exceed the storage capacity and processing power of a single machine. In the former three years the cost of micro-machines have been lowered to be competitive with the large mainframes on which DBMS systems have traditionally relied. For several reasons, the cost of communication links between computers has also been reduced.

- Organizational. Departments expected to use DBMSs are both geographically and organizationally distributed but require integration due to new requirements and organizational changes. DBMSs must therefore be distributed, in order to model effectively the territorial organization that they serve. They must provide to each component a degree of local responsibility for, and control over, the data that it "owns".

- Technical. Performance and availability are enhanced when data can be situated close to the majority of its users, rather than in a distant center. The impact of the failure of one machine - or the link to it - is reduced, particularly when redundant copies of the same data are maintained on other machines.

- Reliability. Local data processing and site autonomy in general, allow us to have an up-to-date information, reducing what might be called "information float", which is not desirable for territorial gevernment decision making.

[3] B.G. Lindsay. Query Processing in R*. North Holland. 1984.

PROBLEM ANALYSIS

The proposed distributed database is purpose-built and a good control can be exercised on the nature of each node and their contents. We may recognise two types of users, the global user processing the data of the distributed database under the control of some kind of Distributed Database Management System (DDBMS) and local or nodal user (figure 3) processing the data of a particular node under the control of the Nodal Database Management System /NDBMS).

Due to the presence of global and nodal DBMS we have two levels of control. The global control is centralised. All global processing are controlled by a central computer (a node act as the center) onto which all global transactions must be chanelled. The NDBMS controls the activies of each node, but these nodes are fully autonomous having its exclusive local users who need not be aware of the existence of the DDN (Figure 4). In this case some of the data captured by a node is contributed to the DDB and, of course, some of the data of the DDB is contributed to the node (Figure 5).

Next we present some definitions used in this paper.
A fragment (a set of records) is the elementary object of a DDB.
A logical fragment is a part of a logical Nodal Database.
Physical Nodal Databases consist of physical fragments
A logical Node is the owner of a Logical Nodal Database
A Physical Node is responsible for the Physical Nodal Database
A site is responsible for a physical Node.
A Node is a set of physical Nodal Databases.

Figure 3 represents the generalised architecture of the proposed homogeneous and centralised DDB.

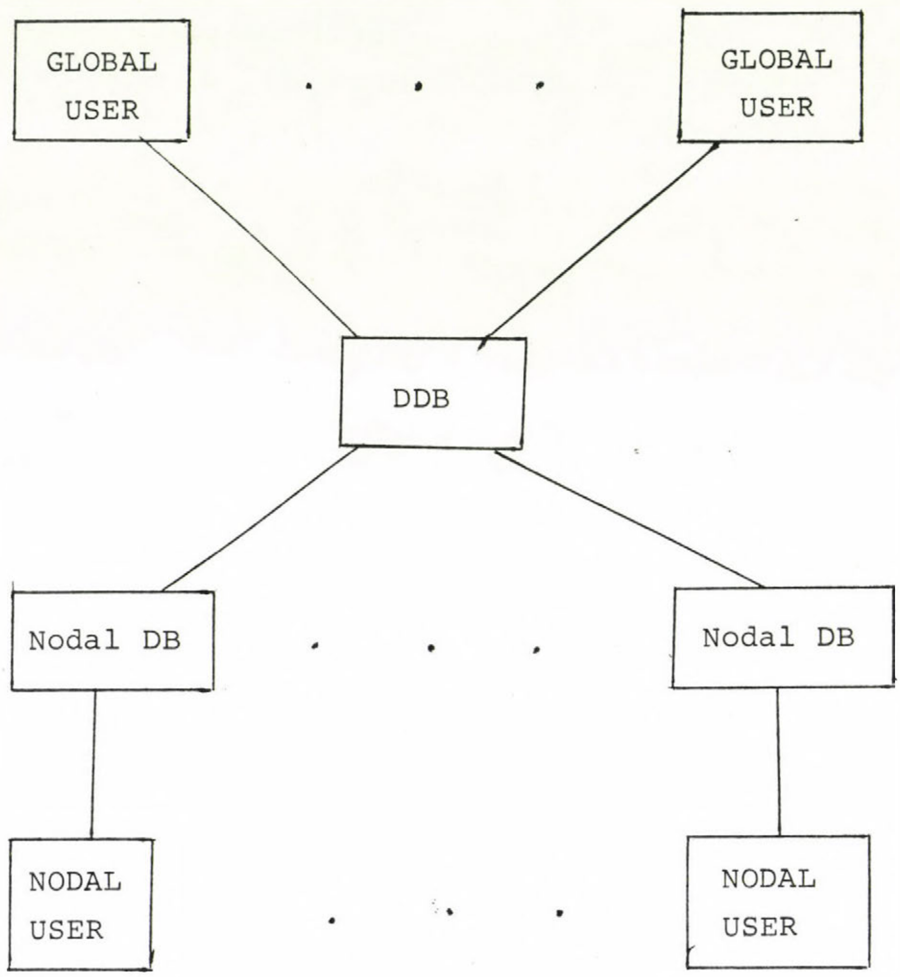


Figure 3
The distributed data base

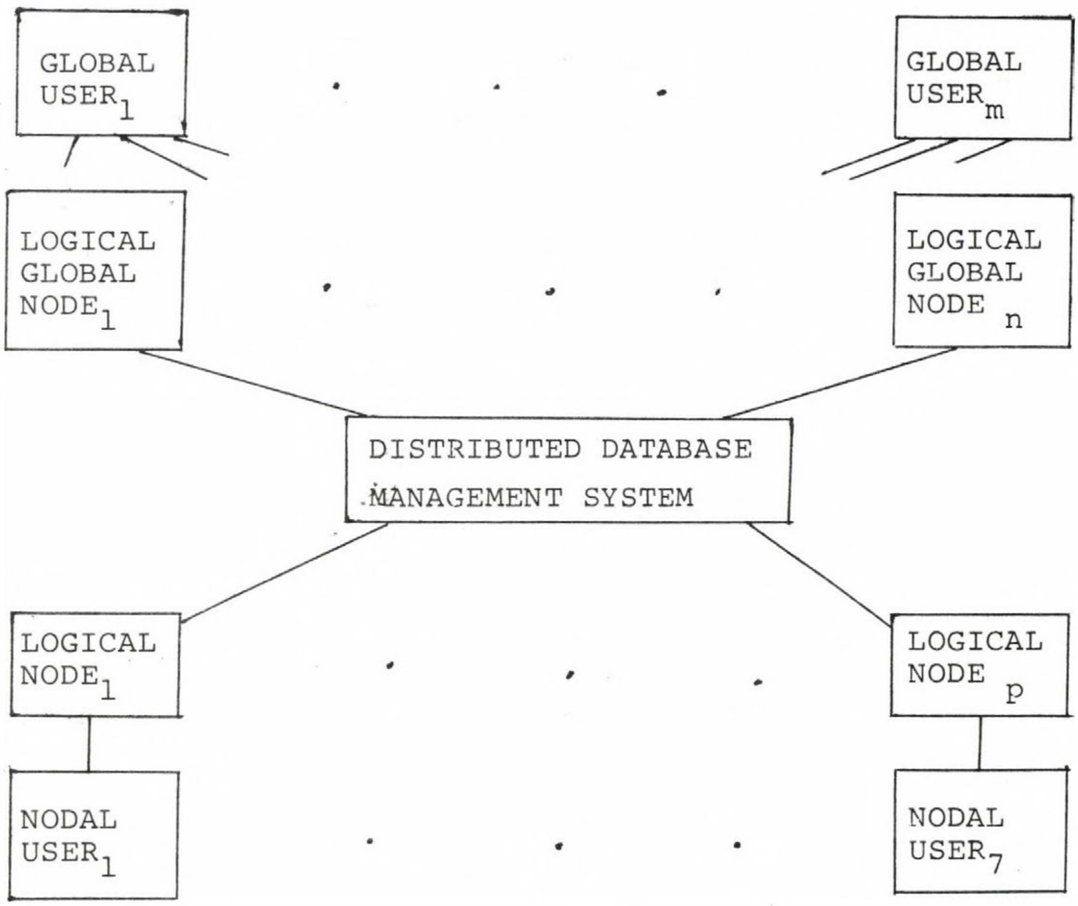


Figure 4
LOGICAL DDB

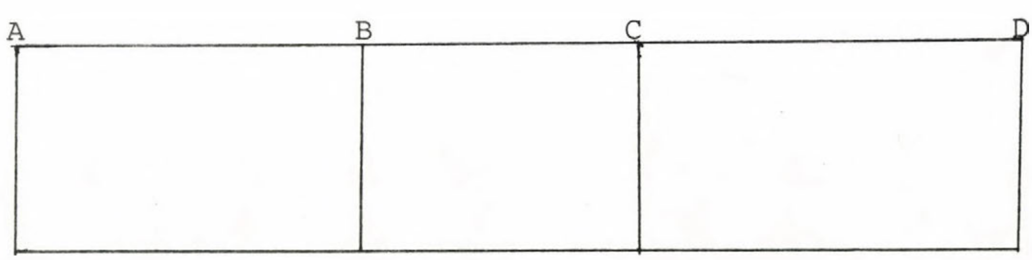


Figure 5 Logical Nodal Database

Nodal user is responsible for the collection of information A-C. The Logical Node contributes information B-C to the DDB. Information C-D is received from the DDB.

The proposed DDB is fully homogeneous allowing only identical data models at each node and having as well identical machines.
Ideally a global user should be able to formulate his transactions treating the DDB as a single database, without having to specify where the data of his interest reside (data transparency). In our case in a Logical Global Database it is possible to have the same Logical data than in a set of Logical Nodal Databases.

In order to improve processing efficiency in the DDB, a statistical analysis of users data requirements was performed.[4] This analysis allowed us to fragment data and to distribute and replicate them "optimally" over selected nodes to improve the overall performance of the DDB.

The processing of global transactions requires at least one user language that is acceptable in all nodes. When a user transaction is submitted, the DDBMS have to decompose it down to subtransactions optimally - one subtransaction for each node, taking the replicated data into account. The aim is to reduce the total cost of the execution, including the cost of communications.

In our particular case, privacy is a very sensitive issue, and safeguards for its protection is very important. We are implementing some protection facilities in all logical nodes.

[4] Dipotet P., et al. 1985. "Introduction of Mathematical Modelling and Computer Techniques in Territorial Planning." TEchnical Report. IMACC-ACC Habana.

Constraints are specified in the nodal data model /and also imposed by user programs/ in order to assure semantic integrity. The protection of the database against errors caused by the simultaneous update of related data by two or more user programs (interprocess integrity, also known as concurrency or internal consistency) is tackled by the NDBMS. The protection of the DDB as a whole,against the potential inconsistency that can be caused by the update of replicated data or related data is faced in our case (centralised control) using a locking strategy.

In our conditions, the need for database recovery is of paramount importance. It arises due to[5]: user program failure; physical storage failure (in parts or whole); processor failure; DBMS failure; Communication link failure. Up to now communication link failure is a serious problem for our proposed DDB. We are providing adequate (and expensives) backup and recovery facilities to enhance the reliability of the proposed DDB. Next future we will use special (dedicated) lines improving the quality and reliability of the communication links.

In this purpose-built DDB, the global requirements have priorities respecting local requirements, dictating completely which data should be in which node.

[5] Rien Van de R.P.; Litwin W: "Distributed Data Sharing Systems"
North-Holland, 1982.

In our case a single command is called a transaction (query or update). Transactions[6] are the units of atomic interaction between the DDB and the external world.

The DDB consists of relations (logical). Each relation is partitioned into subrelations called logical fragments which are the unit of data distribution, meaning that each may be stored at any of several sites in the system. Logical fragments are defined and the assignment of fragments to sites is made when the database is designed. A stored copy of a logical fragment is called a physical fragment.

User transactions are unaware of data distribution or redundancy. They reference relations, not fragments. It is DDB responsibility to translate from relations to logical fragments, and then to select the physical fragments to access in processing any given transaction.

[6]Bernstein P.A. et all.: Introduction to a System for Distributed Databases ACM Transactions on Database Systems. Vol. 5, No 1, March 1980

In our proposed DDB there are two types of transaction, global and nodal transactions respectively. The system of global and nodal logical databases gives a distribution of the logical fragments.

A transaction must specify the local or global database wanting to access, but only global transactions can have access to logical Global Databases.

Then, our DDB is a collection of three elements: Logical Databases, Physical Databases and Communication Network.

The Logical Databases plan and control the distributed execution of transactions, performing the following tasks:

- fragmentation, the LDB translates queries on relations into queries on logical fragments and decides which instances of stored fragment to access;
- Concurrency control, the LDB synchronizes the transaction with all other active transactions in the system.

The Physical Databases store all data managed by the DDB. PDP respond to commands from LDB /Read, Move, Manipulate, write/.

The Communication Network interconnects LDB and PDB (Global and Nodal) providing the following services:

- guaranteed delivery, allowing messages to be delivered in all circunstances;
- transaction control, guarenteing update sequenses and validation;
- site monitoring, to keep track of which sites have failed and to inform sites impacted by failures;
- network clock, a virtual clock kept approximately synchronized at all sites.

SELECTED EXAMPLE

Our application exhibits two important characteristies: First, the activity requires an integrated database. That is the activity entails access to a single pool of information by multiple persons and dependencies (global users). And second, the users of the information and ist sources (Nodes)are distributed geographically. Then, centralized control and Global Databases are needed to ensure operation in accordance wich the Regional overall policy and goals; and decentralized processing in Nodal Databases is required, for reasons of performance, reliability and flexibility of function. By meeting both former goals (centralized control and decentralized processing) in one system, the DDBMS offers very good benefits to our Program as it is possible to see in the following example:

Within a Region, it is needed to develop databases according to different functions (labor, transport, finance, health care, education, inventory, etc.) and the higher (regional or global) level must have complete access to the lower (sub-regional or nodal) levels. Nevertheless, subregions need to develop their respective plans, and require data from the higher level and from other subregions (Figure 5). In Figure 6 we present the most important actual logical data records (forming the basis of the respective logical Nodal Database) related to people (Persons) living or working in subregions. Data item identity may be used as a common (numeric) key for records shown in Figure 6.

It is apparent that the relational data model approach can be successfully applied, operating with the former four databases to get new information about the Personal (in our case Logical Nodal) Databases shown in Figure 6. The higher (regional) level contributes (Figure 5) data about national and regional educational, health care, migration and housing possibilities, normatives, etc.

Personal record (in Personal and Housing Departments)

IDENTITY	ADDRESS DATA	FAMILY DATA

Labor record (in Labor Department)

IDENTITY	ADDRESS	LABOR

Health Care record (in Health Care Department)

IDENTITY	ADDRESS	HEALTH CARE

Public Education record (in Public Education Department)

IDENTITY	ADDRESS	PUBLIC EDUCATION

Figure 6.

Logical Data Records related to persons in the Region

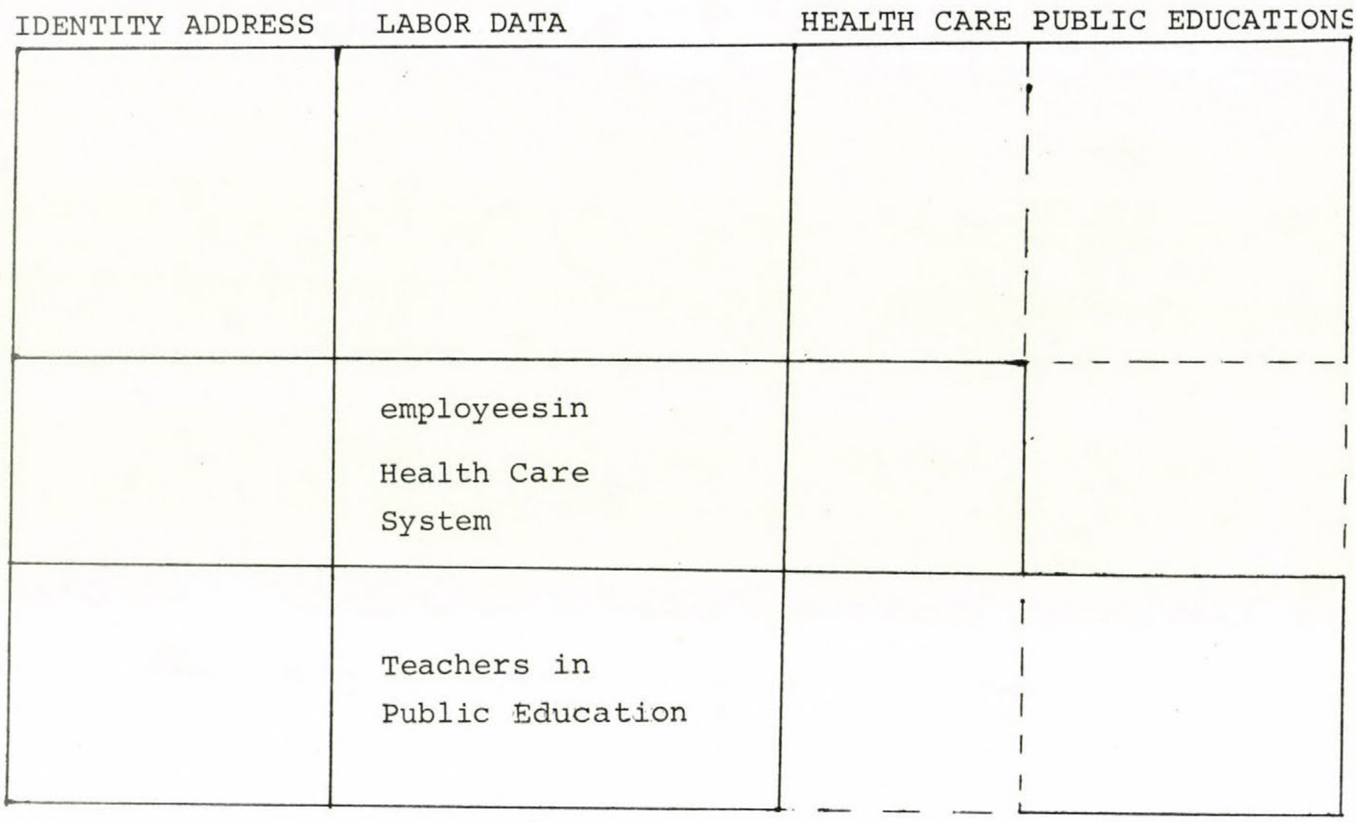
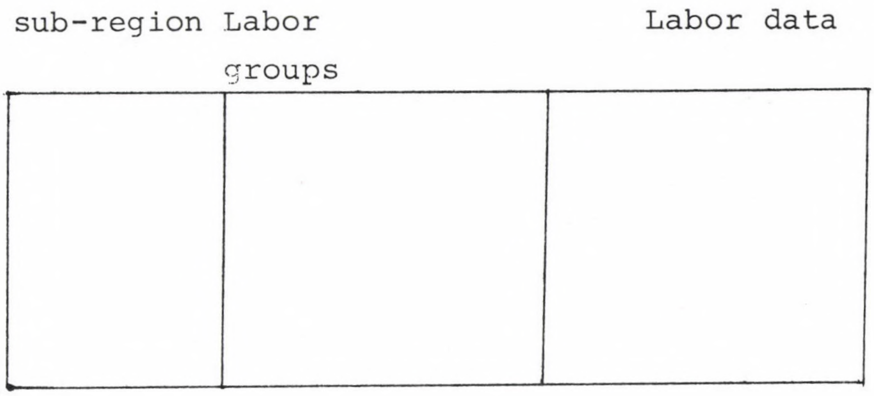


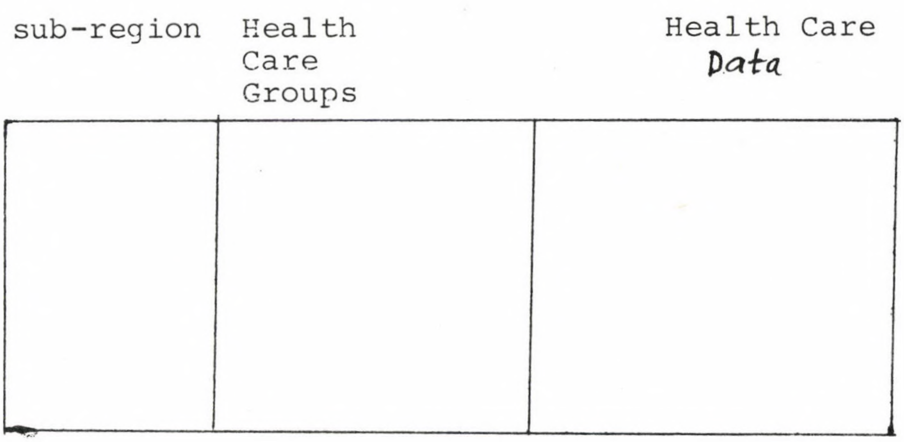
Figure 7

Personal (Logical Nodal) Database

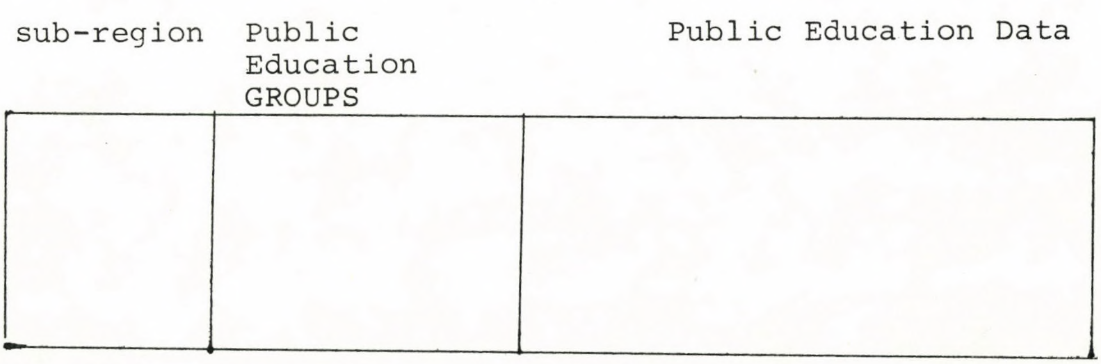
For example Health Care Logical Nodal Database consists of identity, address and health care data for each person living or working in the sub-region, and Labor data concerning employees in Health Care System of the sub-region.



Logical Global Database (LABOR LGDB)



HEALTH CARE LGDB



Public Education LGDB

Figure 8. Logical Global Databases

In the Personal LNDB, of course, using union and projection (relational operations) it is possible to derive logical fragments aswering query requirements on database relations

In Figure 8. some Logical Global Databases related with personal data, are shown. An important task for global users is to have statistical data, concerning their respective interests about LNDB. In This case, data in LGDBs are aggregated data contributed from LNDBs. Access to both Global and Nodal Databases, is controlled in sites by database administrators and also by secret passwords in the DDBMS.

It is convenient to maintain directories (for Global users) containing identity locations (identity key vs LNDB where they occur) and usage statistics. It is possible to use directories as ordinary (global) user data. Directories speed up the information retrieval process and, knowing the LGDB content, allow global users to have a complete picture of available information in the DDB.

CONCLUSION

We have discussed several aspects of the design of a DDBMS within a Program for long range planning for a development region. Constraints imposed to the system, and previous knowledge and evaluation of main global and nodal users requirements; allow us to purpose a centralised control, decentralized processing, homogeneous DDB. In our case, the relational data model is an appropriate one satisfying users and system requirements.

Next future we will present a report describing also other aspects of the proposed DDB we didn't face in this paper

Egy osztott adatbázis-kezelő rendszer területi tervezés céljára

P. Dipotet, A. Benczur

Összefoglaló

A szerzők a Kubában megvalósitandó területi tervezési rendszer számára szánt osztott adatbázis-kezelő rendszer tervezetét ismertetik.

Разделенная управляющая система базы данных в территориальном планировании

П. Дипотет, А. Бенцур

Р е з ю м е

Авторы дискутируют несколько аспектов осуществления разделенной управляющей системы базы данных для целей долгосрочного регионального планирования на Кубе.