OCCASION

This publication has been made available to the public on the occasion of the 50th anniversary of the United Nations Industrial Development Organisation.

DISCLAIMER

This document has been produced without formal United Nations editing. 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 the Secretariat of the United Nations Industrial Development Organization (UNIDO) concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries, or its economic system or degree of development. Designations such as "developed", "industrialized" and "developing" are intended for statistical convenience and do not necessarily express a judgment about the stage reached by a particular country or area in the development process. Mention of firm names or commercial products does not constitute an endorsement by UNIDO.

FAIR USE POLICY

Any part of this publication may be quoted and referenced for educational and research purposes without additional permission from UNIDO. However, those who make use of quoting and referencing this publication are requested to follow the Fair Use Policy of giving due credit to UNIDO.

CONTACT

Please contact publications@unido.org for further information concerning UNIDO publications.

For more information about UNIDO, please visit us at www.unido.org

UNITED NATIONS INDUSTRIAL DEVELOPMENT ORGANIZATION
Vienna International Centre, P.O. Box 300, 1400 Vienna, Austria
Tel: (+43-1) 26026-0 · www.unido.org · unido@unido.org

UNITED NATIONS INDUSTRIAL DEVELOPMENT ORGANIZATION

UNIDO/IS. /Rev.1 January 15, 1988

by Mieczyslav MURASZKIEWICZ

16761

AUTOMATED SYSTEM OF THE INTIB DIRECTORIES (ATHOS)

Industrial and Technological
Information Section

Vienna, January 1988

TABLE OF CONTENTS

I. INTRODUCTION

The data required for carrying out industrial activities embrace a large spectrum of information, which is stored in different specialized information systems containing specific data on technology, economics, low, etc. Proliferation of various information services and databases is observed, and it is more and more difficult to locate the right service and/or to acquire the adequate information without the assistance of referral systems.

Thus, the role of reference/referral information is growing up. This kind of information, which basically does not provide documents or facts but only refers the user to the appropriate source of information, is of increasing interest to the users who may range from a small-scale enterpreneur, through an engineer, manager or vendor to a decision-maker at the governmental level. One can notice that, presently, the general strategy of information activity is changing from acquisition and proprietary collection approach to accessibility strategy.

The kernel of the accessibility strategy is a system of directories containing data on information sources. This system allows, *inter alia*, to:

(i) Facilitate adequate access to the relevant information;

(ii) Provide the users with an efficient tool for data acquisition;

(iii) Make more widely known and to promote the full use of existing industrial and technological information facilities in both developed and developing countries;

(iv) Facilitate establishment of information policy including linkages at the national, regional and international level.

The Industrial and Technological Information Bank (INTIB), which operates under the auspices of UNIDO, has developed a number of directories (for the list of directories see Annex 1). These directories are used at UNIDO/INTIB HQ to support its Industrial Inquiry Service and to coordinate and promote information activities in developing countries. They are also employed by different information centers, especially by the INTIB focal points in the developing countries, for their specific purposes.

At present, the production of the INTIB directories and related documents including collection, storage and processing of data is done mostly automatically. Needless to say that this process is tremendously time- and labour-consuming, susceptible to errors and can hardly provide the consistency between different directories. Almost the same is true for the usage of the hard copy directories available now. Therefore, the need to automate the process of building and exploitation of the INTIB directories has emerged.

The purpose of this document is to specify the AuTomated system of tHe INTIB directOrieS (ATHOS) which is planned to be implemented in Industrial Information Section of UNIDO. It is assumed that ATHOS will be a user friendly system with an augmented man-machine interface including features typical for expert systems. The intelligence embeded in ATHOS is aimed at increasing the user's comfort and to provide greater effectivness of the system.

An outline of this document is as follows. In Chapter II the characteristic of ATHOS is provided. Special stress is put on the purpose, scope and coverage of the system. Analysis of users, their categorization and their information needs are also presented. Next, the design criteria which determine the process of ATHOS designing and implementation are discussed. Identification of input/output information flow, specification of the ATHOS functions and proposals concerned with the ATHOS architecture complete this chapter.

Chapter III provides the technical specification of ATHOS. Main procedures which can be invoked by a user when interacting with ATHOS are described. Requirements on the retrieval language are presented. Tools offered by ATHOS to the database administrator are listed. It should be stressed that the decision has been made that the detailed design of menu and input/output layouts are left to the ingenuity and creativness of the team implementing ATHOS.

However, specification of data fields and input/output formats were done by the INTIB staffer and given in Annex 4.

Chapter IV is dedicated to the ATHOS database administrator. Some general remarks concerning the administrator are formulated at the beginning of this chapter. After that, responsabilities and tasks which have to be undertaken by the database administrator are determined.

Chapter V contains proposals of hardware and software for implementing ATHOS . Additionally, the implementation schedule and suggestions as to the further work are given.

Bibliographic references and four annexes end the document. The first of them contain a list of directories edited by INTIB, the second one gives 20 industrial sectors covered by the UNIDO/INTIB activities, the third annex compares two software packages viz. dBASE III+ and Mini/Micro CDS ISIS. The last annex - prepared by the INTIB staff memeber - specifies data fields, input/output flow and some tables.

2. CHARACTERISTIC OF ATHOS

This Chapter is devoted to general presentation of the AuTomated system of tHe INTIB directOrieS (ATHOS). The purpose, scope and coverage of the system are discussed. Next, the users and their needs are to be described. Functions of ATHOS will be presented and general information flow from input to output will be outlined. Given these data the architecture of ATHOS will be proposed.

2.1 Purpose of ATHOS

INTIB has been widely known for many years as a supplier of valuable information pertaining to the technology transfer. However, as the volume of data has been increasing the effective servicing of the INTIB users becomes more and more complicated and labour-consuming. The same can be said about maintaining the consistency and integrity of the information available.

In order to enhance the efficiency and quality of the Industrial Inquiry Service offered by INTIB and upgrade the ability to coordinate and promote industrial activities in developing countries it seems to be necessary to modernize the existing INTIB information system. Automation of the INTIB directories is an immediate goal within the process of modernization.

ATHOS is chiefly aimed at automation of the INTIB HQ functions dealing with running the system of directories, however, its output is assumed to be disseminated and used throughout the world INTIB infrastructure, in particular, in the INTIB focal points of various categories. The purposes of ATHOS are as follows:

(i) To facilitate and discipline data acquisition;

(ii) To secure the input data verification, validation and integration into the existing INTIB information resources;

(iii) To guarantee fast and error-free data processing and compilation;

(iv) To provide effective storage of data;

(v) To protect data;

(vi) To reduce the redundancy of data;

(vii) To increase the consistency of data;

(viii) To enable the generation of various documents (reports, statistics, etc.) on the user's request on different media (paper, magnetic carriers, etc.) - thus, speeding up the operations of the Industrial Inquiry Service;

(ix) To facilitate and accelerate information retrieval from the INTIB directories;

(x) To provide fast publishing of the INTIB directories at professional level;

(xi) To enable the transfer of ATHOS data to the UNIDO information systems;

(xii) To enable the co-operation and data exchange between INTIB HQ and the INTIB focal points in the developed and developing countries.

2.2. Scope and Coverage

ATHOS is a referral information system, i.e. it deals rather with the references to sources of information than with sources themselves. Nevertheless, a significant amount of factual data is stored and processed within ATHOS. Bibliographic information may rarely occur in ATHOS.

The scope of ATHOS is determined by the UNIDO statutory activities carried out in 20 industrial and related branches (cf. Annex 2). At present, the ATHOS coverage is defined by the existing INTIB directories and similar documents which are elaborated by the INTIB information officers. A list of the directories and other material is given in Annex 1.

The ATHOS coverage is not fixed definitely. It is expected to be extended as new information sources will be identified and the derived data from them will be incorporated to the system.

2.3. Users and Users' Needs

Identification and categorization of users and their needs is a sine qua non condition of appropriate system design. The review of actual users of the INTIB directories shows that two main categories of users can be distinguished, viz.

UNIDO internal users

Within this category one can identify three subcategories of users having different needs, namely:

(i) Decision-makers and managers; members of this subcategory are chiefly interested in aggregated and synthetized information, especially of statistical nature;

(ii) Industrial operations officers; they basically look for partners who could provide them with adequate information and expertise;

(iii) INTIB staffers; this subcategory again can be divided into two groups, viz.

- Industrial Inquiry Service staffers who work as intermediary information officers between ATHOS and terminal users, therefore, perform typical tasks like indexing of new documents, query formulation, retrieval, dissemination, etc. These officers are particularly interested in such system facilities which would allow to combine answers to different queries and to present the output material in a desired form;
- editors who are responsible for compiling and producing the INTIB directories and related material in a hard copy form.

External users

Two subcategories of users can be identified within this category, viz.

(i) Information center and/or INTIB focal points staffers; their needs are similar to those which were specified above for the Industrial Inquiry Service staffers at UNIDO HQ;

(ii) End users; this group is highly heterogeneous with the information needs ranging from very specific, often factual data to deeply processed and packaged information.

2.4. Design Criteria

The purposes, scope and coverage of ATHOS and the results of its users' analysis, as well as, general principles of complex information systems designing, and discussions with the INTIB information officers determine the following list of design criteria.

(i) ATHOS should be designed as a user friendly information system. This means that the user is not supposed to be familiar with programming and other computer science techniques. Moreover, it is assumed that the system will create an ergonomic working environment to provide the user, as much as feasible at the current stage of computer technology development, with intellectual comfort.

(ii) It follows from the preceding paragraph that ATHOS is not conceived as an ordinary information system which relies only on specialized access mechanisms like commands, index terms, query forms, etc. It should be more sophisticated and more intelligent. It is suggested that the expert system approach be adopted to design the user interface. The intelligent interface could be especially helpful when gaining access to ATHOS information sources and formulating orders dealing with the ATHOS facilities.

(iii) As the ATHOS coverage and lifetime are not *a priori* confined, it is reasonable to assume the openness of the system. The meaning of openness is here twofold. First, data updating and extention of the existing facilities are allowed; second, adding new functions and databases must also be accepted, if required. Technically, the openness of the system can be reached by means of modularity of its structure. Therefore, it is recommended that ATHOS be designed as a modular system composed of interrelated blocks executing specific functions and a set of databases storing the subject-oriented data.

(iv) The ATHOS user-friendliness assumption (paragraph i) gives rise to a dialog language based on the concept of hierarchically organized menus. The list of functions displayed on the screen should allow access to individual programs without knowing program and/or file names. This language should also be equipped with an extensive system of HELP features in order to assist, if requested, the user at any stage of his/her co-operation with ATHOS. Another vital point as to the ATHOS dialog language is that it actually should be multilingual. The ideal is that new languages could be easily incorporated into the ATHOS dialog language.

(v) Functionally ATHOS should be designed so as to:
- facilitate input and output of data, including automatic filling in of certain fields;
- provide semantic (based on thesaurus) and syntactic data control for the indicated data fields. All the errors made, when entering data, should be reported automatically with no chance to proceed without correcting;
- protect the data stored;
- decrease the redundancy of data;
- maintain the consistency and integrity of data;
- produce all the output as professional quality items rather then working material. This requirement is mainly addressed to the function of editing directories and similar documents. They have to be elaborated under such tools like desktop publishing packages.

(vi) Basically ATHOS is conceived to be operated on IBM PC AT compatible microcomputers. However, it is expected that ATHOS information sources will be sometimes transferred to some UNIDO mainframe information systems. Moreover, a possibility to adopt ATHOS for networking should be taken into account.

(vii) It is strongly recommended that ATHOS be implemented on the basis of a ready-to-use software package rather than coded from scratch. The reasons to do so are as follows:

- to reduce the implementation time;
- to save funds;
- to enable the ATHOS database administrator to easily redefine the existing databases and/or define the new ones. One should not neglect the fact that the database administrator is not supposed to be a professional programmer with broad knowledge about implementing databases, so the tools given to his disposal have to be rather simple.

Undoubtedly, this package has to have a programming facility (sort of data manipulation language) allowing an application programmer to define specific procedures and to link them to the main body of the package.

As a comment to the above list of design criteria let us say that the future owner of any information system tends, what is perfectly natural, to automatise as many aspects of information activities as feasible. On the other hand good engineering practice proves that because of economic and pragmatic constraints only 70 - 80 percenteges of users' desiderata are actually implemented in the information system.

2.5. INPUT/OUTPUT INFORMATION FLOW

External, i.e. input and output, data flows are presented in this Section.

INPUT

Four kinds of input information to ATHOS can be distinguished.

(i) Raw data; information dealing with directories and other documents stored in ATHOS belong to this group. Update information is also considered as a part of raw data. It is assumed that some semantic and syntactic data control along with data validation is provided by ATHOS;

(ii) Users' queries; process of query formulation is of great importance and very often difficult for non-experienced (casual) users, therefore, ATHOS should contain expert system facilities which could facilitate the establishment of appropriate queries;

(iii) Commands; they are used to have executed the choosen ATHOS functions like printing a dictionary, generating a standard report, etc.;

(iv) Instructions from the databse administrator, e.g. renaming, reshaping databases, defining new computing procedures, reorganization of the system. etc.

A keyboard, floppy disk drivers and streamers are the input devices for entering information to ATHOS. In the future, a CD-ROM device can be used as inputing device (for large chunks of data non-varying with time).

Output

Four groups of output from ATHOS can be identified.

(i) Answers to the users queries; the relevant data can be given in different formats and on different media (screen, paper, diskette, etc.);

(ii) Documents yielded by commands; various directories, reports, statistics, etc. belong to this group;

(iii) Data extracted from the ATHOS databases to be exported to the INTIB focal points and to the UNIDO mainframe information systems;

(iv) System reports; they are generated for the database administrator to notify him about the status of various facilities of ATHOS, e.g. content of system tables, dictionaries, memory maps, etc.

2.6. Functions of ATHOS

The functions that are performed in ATHOS can be divided in the following groups.

(i) Collecting the sources of information.

(ii) Intellectual preocessing of the sources resulting in extracting data for directories or looking for additional data on identified institutons.

(iii) Data input.

(iv) Collecting user enquires.

(v) Query formulation.

(vi) Consulting and assisting the user and database administrator during their work with ATHOS.

(vii) Information retrieval.

(viii) Printing answers.

(ix) Editing directories.

(x) Data export.

(xi) Generating internal reports.

(xii) Creating new databases.

(xiii) Maintenance, integrity and consistency checking of databases.

Some of these functions are interrelated, i.e. it may be necessary or useful to utilize the results of some functions while performing some other functions. The links between the functions are depicted in Fig.1. Now, the ATHOS functions will be discussed in detail.

COLLECTING THE SOURCES OF INFORMATION

There are two categories of the sources of information that are used to feed the system, viz.

- country reports, industrial surveys, conference and exhibition materials, etc.
- input forms (questionnaires) used for collecting the detailed data on institution activities. For the sake of integrity, some of the existing forms will have to be modified in concordance with the definitions adopted in Chapter III. The existing directories fall also into this category but a warning is necessary: some data contained in them may be obsolete.

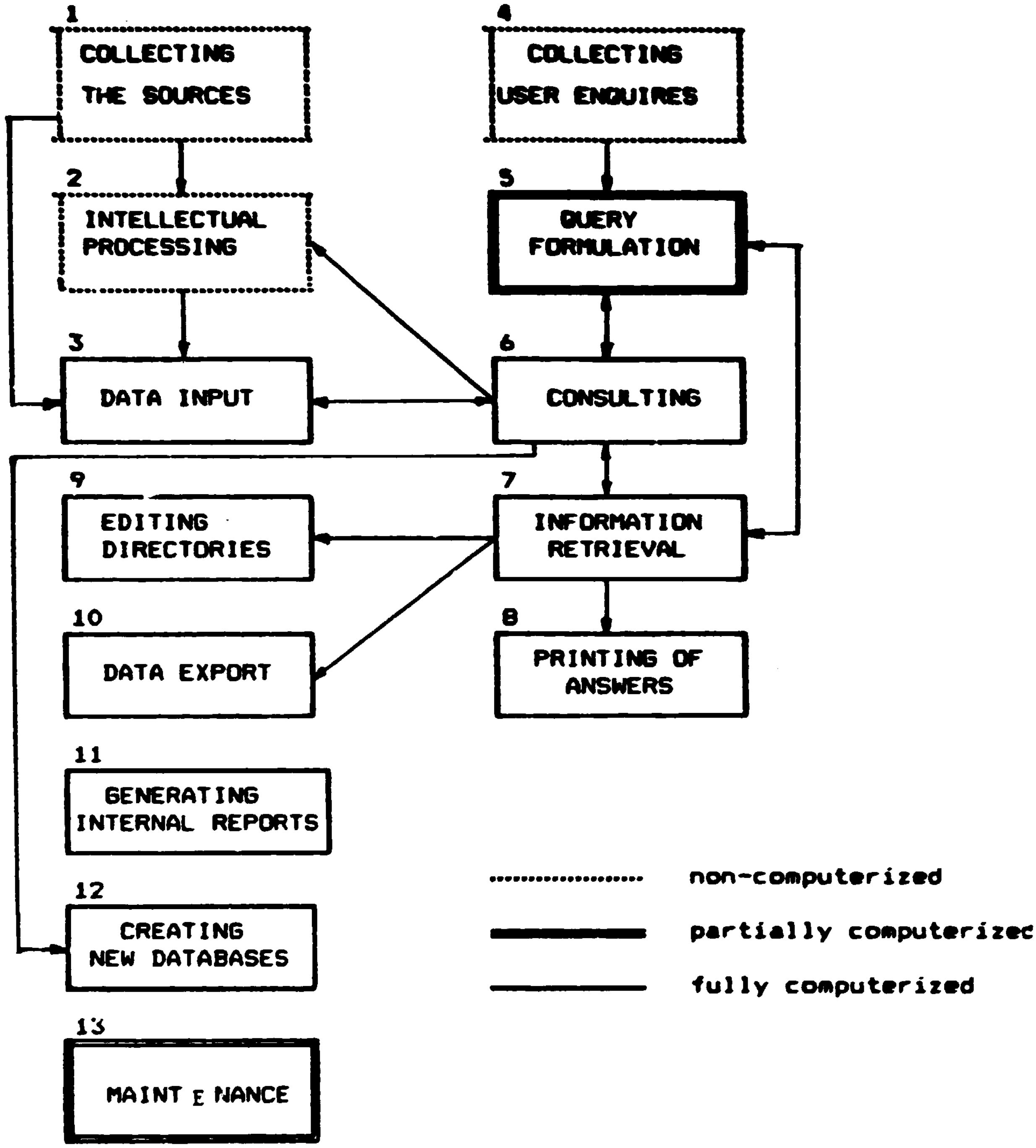

Fig.1 Outline of the ATHOS functions

INTELLECTUAL PROCESSING OF SOURCES

The sources available (cf. function i) must be analyzed intellectually to extract the information on activities, institutions, industrial potential, etc. To facilitate this task, it is recommended that the function (vi) - consulting - be used in order to get the information concerning the activities and institutions already registrated in ATHOS, localization of data items, terminology use, etc. The data extracted from the sources may be incomplete or considered unreliable. Such being the case, the institution in question should be again consulted with the aid of an appropriate input form. One can predict that some indexing of objects on which the information is to be stored in ATHOS will be necessary. The function is terminated with the decision which data (if any) are to be entered into ATHOS and which still have to be verified.

DATA INPUT

The input of data can be divided functionally into data entry (i.e. creating a new record in some of the databases) and update (i.e. changing the value of some data item or adding a new item to the already exusting record). In both cases, the following

facilities should be provided:

- help subfunction (this and next subfunctions are not shown in Fig.1) which can be activated when in doubt about filling in the particular data item;
- data validation subfunction which is invoked automatically for data items whose content must be compatible with the preselected class of data;
- status assessment subfunction that is invoked automatically at the end of filling in the record; the status information reflects the completeness of data, verification level, links with other databases, etc.

To ensure the high quality of the data stored in ATHOS, it is recommended that the function (vi) - consulting - be used for getting expertise on terminology, general database layout, possible links between records, etc.

COLLECTING THE USERS' ENQUIRIES

It is expected that the users' enquires will be supplied in a free format natural language form, therefore, they will need conversion to the menu-driven query system. Nevertheless, after some period of ATHOS exploitation it may be justified to distribute some standard query forms.

QUERY FORMULATION

The user enquiry is a source for the query formulation. It is recommended that the HELP subfunction and consulting function be used to get to know how to establish a query from a syntactic standpoint and a semantic point of view.

CONSULTING

The consulting function provides access to information stored in the Expert Database, Thesaurus, Dictionaries, Tables, System Resources Used and Available, etc. This information should enable one to answer for instance the following questions concerned with:

- scope and coverage of all the databases;
- list of actually available databases (note that some of databases may not be loaded because of memory limitation);
- list of actually system supported facilities, e.g. sorting modes, display and print formats;
- size of databases;
- contents of the relevant part of the thesaurus (globally supported synonyms);
- contents of the relevant part of a database dictionary (locally supported synonyms);
- stopword lists for each database;
- coding tables; etc.

INFORMATION RETRIEVAL

The retrieval functions must, *inter alia*:

- retrieve records pertaining to queries (expressed for instance in a language based on Boolean Algebra);
- allow to mask the ending of a term;
- inform on the number of relevant records;
- retrieve records with synonimic terms automatically;
- provide for interactive search which allows, among others, the immediate modification of query;
- display retrieved documents according to the preselected display format.

PRINTING ANSWERS

The results of a search can be sorted and next printed according to a pre-selected sort-and-print format. Other kinds of documents (drafts of directories, reports, statistics, system status reports, correspondence with users, etc.) generated by means of ATHOS can be printed according to a pre-defined format.

EDITING DIRECTORIES

This is a vital function of ATHOS . It is assumed that technical quality of edited directories will be high. Nowadays, the desktop publishing packages running on powerful microcomputers equipped with high-resolution monitors are used to carry out editing activities. This should be the case for ATHOS.

DATA EXPORT

Any part of each ATHOS database can be exported in the ISO 2709 format to a diskette(s) or to a mainframe by means of direct or telecommunication connections.

GENERATING INTERNAL REPORTS

The reports needed by the ATHOS database administrator (cf. Section 4.2) are produced by this function.

CREATING NEW DATABASES

Creating a new databse (corresponding to a newly introduced directory) is a task for the database administrator. When doing that he must be extremely careful to preserve the compatibility and consistency with already existing databases. Another thing is that he should use as much as possible the existing software modules. The consulting function of ATHOS (cf function vi) can provide suitable information regarding this task.

MAINTENANCE OF THE DATABASE; INTEGRITY AND CONSISTENCY CHECKING

The main tasks concerned with this function are as follows:

- detecting and correcting inconsistencies after each input/update of data;
- setting aside records with non-verified data items after some time, e.g. one year;
- setting aside records with low completeness level lasting for a long time, e.g. one year;
- periodical upgrading the thesaurus basing on the contents of database dictionaries, e.g. every two years.

2.7. ARCHITECTURE OF ATHOS

The basis for any information system architectural considerations is determined by the so-called conceptual model, usually provided by the future system owner. In the case of ATHOS this model was defined by the INTIB staffers. The model is presented, as a rule, in a form of natural language assertions which are then converted into a more formal data model that should be amenable to computerization.

There are three formal fundamental data models which can be used to represent objects and relationships occuring between them. These models are: network, hierarchical and relational ones.

The analysis done in the previous chapters leds to the conclusion that any of the three models could be used to implement ATHOS. A thorough study of ATHOS in terms of the network model proved that the relationships between objects in different directories were scarce. Therefore, it is fairly justified to consider each directory as a substantive database.

Looking at ATHOS from practical point of view one can see several constraints caused by hardware, software, economic factors and system reliability. The analysis of the ATHOS architecture in terms of these constraints also points towards the separate databases organization.

Nevertheless, to gain the integrity of ATHOS and to ensure its step-by-step implementation all the databases must be controlled by the same software. A heterogeneous software solution, though technically feasible, is absolutely not justified by the structuries of directories. Moreover, as ATHOS is automatized from scratch, there is no need to assure communication with other databases exploited under incompatible software.

The proposed architecture of ATHOS is given in Fig.2. One can distinguish five levels in this figure. Now, they are to be presented.

At the first level three categories of objects which co-operate with ATHOS are shown, viz.

- database administrator (cf. Chapter IV);
- users (cf. Section 2.3);
- other information systems (IS), including the INTIB focal points and the UNIDO mainframe information systems, which are supplied with data from ATHOS.

At the level 2 there are two objects. The first one is a man-machine interface which links the ATHOS environment with the ATHOS information resources and processing facilities. In fact, the interface is a dialog language that was characterized in Section 2.4, p.(iv). It should be noted that the role played by the interface is of special importance not only because of the fact

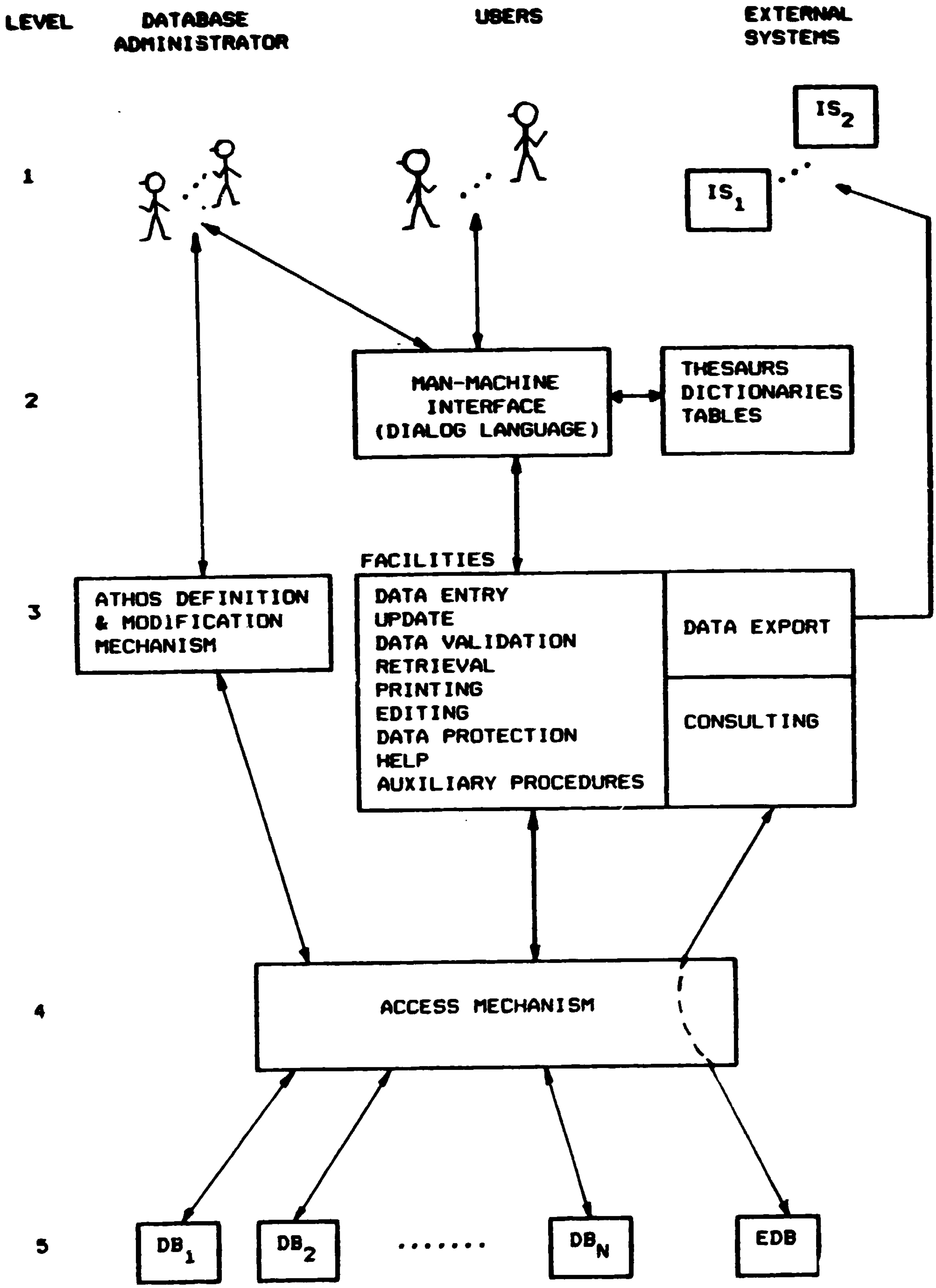

Fig.2. ATHOS Architecture

that it is an intermediary between ATHOS and its environment but also since the user tends to identify the interface quality with the quality of the system as a whole. Therefore, special attention should be paid to the designing and implementing the dialog language.

The next objects at the level 2 are a thesaurus and an array of tables and dictionaries. They are used for facilitating, controlling and validation of input and update. Needless to explain how essential for the quality and usefulness of ATHOS are the contents of the thesaurus, dictionaries and tables.

The following objects belong to the level 3:

- a set of the ATHOS facilities which implement the ATHOS functions defined in Section 2.6. Moreover, auxiliary procedures (e.g. generation of a dictionary, notyfing the user about the number of documents in a database, etc.) have been added to these facilities. The ATHOS facilities are presented in detail in Chapters III and IV;
- the ATHOS definition and modification mechanism which is accesible for the database administrator only (for the role and duties of the database administrator see Chapter IV). Within this module all the tools to define and redefine the database are available. This module also contains a manipulation language which allows to code advanced and sophisticated procedures operating on the data stored in ATHOS, and to implement newcomming functions of ATHOS.

The only object at the level 4 is an access mechanism which should provide efficient access to the ATHOS databases. The design criterion (vii) from Section 2.4 says that a ready-to-use software package is recommended for implementing ATHOS. As the access mechanism constitutes an integral part of the package, so the structure and functioning of this mechanism depends on the package rather than on the ATHOS implementator.

Two kinds of objects can be found at the level 5:

- the ATHOS proper databases ($DB_1, \dots, DB_n$) which contain the INTIB directories and other documents (cf Annex 1);
- The Expert Database (EDB) which contains information on the $DB_1, \dots, DB_n$ databases. EDB is mainly used by a consulting facility to assist the user in formulating queries and commands to ATHOS. This is reflected by a special link joining EDB, via the access mechanism, with the consulting facility in Fig.2.

The linkages depicted in Fig.2 represent relationships and flow of data/commands within ATHOS. In the light of so far discussions the meaning of the linkages seems to be self-explanatory

III. TECHNICAL SPECIFICATION OF ATHOS

This chapter contains detailed information on components of ATHOS. Descriptions of procedures available for the ATHOS users are given. Requirements vis-a-vis the retrieval language are provided only; technical specification of such a language is not needed since the philosohphy of the retrieval language, as a rule, cannot be changed within the software package choosen to implement the system. The last section in this chapter specifies the tools useful for the database administrator. Annex 4 contains the specification of all the fields occuring in directories to be handled. Also, notes on Input/Output design, as well as some tables, are presented in the Annex.

3.1 PROCEDURES

The ATHOS functions and facilities were discussed in Sections 2.6 and 2.7, respectively. Now, let us take more technical look at those of them which are available to the ATHOS users. Within ATHOS the functions take a form of procedures. Basically, there are two kinds of procedures, viz. these ones which already exist as a part of the software package to be used, and those that must be written by the team who will implement ATHOS. The procedures below are presented in a semi-formal PASCAL-like language in order to express in precise manner their purposes.

It is assumed that the mechanisme allowing to define and drive menus is available. The procedures presented below are supposed to be invoked from m nus. However, the detailed layout of the menus and their hierarchy is not proposed. This is left to the ATHOS implementation team. Practice shows that freedom in that respect yields better results than imposing rigid patterns.

DATA ENTRY AND UPDATE

Data entering has to be guided by an input data form. Tasks handled by the data entry procedure are as follows:

- inputing data from a keyboard;

- controlling and validation of data through calling the data validation procedure;
- automatic filling in the indicated fields; the tables mapping comming data into new data are assumed to be available, e.g. the table which maps country names into country codes;
- make a report on the data entered;
- updating the inverted file (if any);

More formally, the procedure looks as follows:

```
procedure DATA ENTRY;
  comment semantical_vailidflag, syntactic_validflag and
          creat_flag are supposed to be set up;
  call PROTECTION procedure;
  if prot_level ≠ "data entry" then leave the procedure;
  while field is available do
     begin
     get a field content from a keyboard;
     if semantical_validflag = on or syntactical_validflag = on or
        creatflag = on then call VALIDATION procedure;
     end;
  while field is available do
     if autoflag = on then call AUTOMATIC filling in procedure;
  call REPORT procedure;
  call UPDATE inverted file procedure;
end DATA ENTRY.

procedure AUTOMATIC;
  get a calculation PATTERN from a table for the specified field;
  calculate the value of this field according to this PATTERN;
  assigne this value to the field
end AUTOMATIC.
```

DATA VALIDATION AND STATUS ASSESSMENT

This procedure includes the following:

- controlling the data; three levels of control data should be provided, namely,
    - lexical (based on the thesaurus and/or dictionaries) and syntactic control,
    - syntactic control,
    - no control;
- creating new complex data from atomic data and putting them to the databases.
- gathering information on data entering and updating.

```
procedure VALIDATION;
  comment semantical_vailidflag, syntactic_validflag and
          field_creatflag are supposed to be set up;
  if semantical_validflag = on then call SEMANTIC procedure;
  if syntactical_validflag = on then call SYNTAX procedure;
  if field_creatflag = on then call AUTOMATIC procedure;
  set up a table for REPORT procedure
end VALIDATION.

procedure SEMANTIC;
  while a word from a field is available do
     begin
     get a word;
     consult a thesaurus (or dictionary);
     set up a consultflag;
     if consultflag = on then call WARNING & CORRECTING procedure;
     end
end SEMANTIC.
```

```
procedure SYNTACTIC;
  while field is available do
     begin
     call PARSER;
     if parserflag = on then  call WARNING & CORRECTING procedure;
     end
end SYNTACTIC.

procedure CALCULATE;
  get a calculation PATTERN from a table for the specified field;
  calculate the value of this field according to this PATTERN;
  assigne this value to the field
end CALCULATE.
```

Sorting and Printing

Usually sorting precedes printing. The sorting procedure has to assure:

- sorting according more than one key;
- short time of ordering.

The printing procedure has to be able to:

- print hits of retrieval according to the pre-defined format(s);
- print a whole database according to the pre-defined format(s);
- print the reports generated by ATHOS in the pre-defined format(s).

A symmetric procedure to printing is a displaying one. It has to perform the same tasks but on the screen, rather than on paper.

```
procedure SORT/PRINT;
  comment sortflag, kind_of_printflag, kind_of_displayflag, hit,
        whole and report flags are supposed to be available;

  if printflag = on then
     begin
     if sortflag = on then call SORT procedure;
     get print format according to kind_of_printflag;
     if kind_of_printflag = "hit" then call PRINT_HIT procedure;
     if kind_of_printflag = "whole" then call PRINT_WHOLE;
     if kind_of_printflag = "report" then call PRINT_REPORT procedure
     end;

  if displayflag = on then
     begin
     if sortflag = on then call SORT procedure;
     get display format according to kind_of_displayflag;
     if kind_of_displayflag = "hit" then call DISPLAY_HIT
                                              procedure;
     if kind_of_displayflag = "whole" then call DISPLAY_WHOLE;
     if kind_of_displayflag = "report" then call DISPLAY_REPORT
                                                 procedure

     end;
end SORT/PRINT.
```

EDITING

Editing directories and other related documents handled by ATHOS is to be performed by means of desktop publishing facility (DTP). The data formatted, according to the predefined format, originated from ATHOS can be processed under DTP.

```
procedure EDITING;
  call DTP procedure
end EDITING.
```

DATA PROTECTION

The data protection procedure controls access at four levels:
- the entrence level; access to ATHOS;
- the database administrator level; access to all software and data resources of ATHOS
- the input/update/output level; access to data resources only;
- the browsing/retrieval level; access to browsing and searching functions of ATHOS only.

```
procedure PROTECTION;
  for i = 1,2,3 do
  begin
    get password from a keyboard;
    if password = "OK" then
       begin
       establish appropriate access;
       set up protection_level;
       leave the procedure
       end
  end
end PROTECTION.
```

DATA EXPORT AND BACKUP

This procedure allows to load down into the diskette(s) a part or whole the database(s) in the ISO 2709 format. Backuping under a pre-defined format is included in this procedure, as well.

```
procedure EXPORT/BACKUP;
  comment exportflag and backupflag are supposed to be available;
  if exportflag = on then
     begin
     convert file to ISO 2709 format;
     load file to a diskette(s)
     end;
  if backupflag = on then load file on a diskette(s)
end EXPORT/BACKUP.
```

CONSULTING

The consulting procedure :

- provides access to information stored in the Expert Database, thesaurus, dictionaries, tables, etc.
- answers to questions put by users on various aspects of the ATHOS resources;
- helps formulate a query;

```
procedure EXPERT;
comment values of ATHOS_structure, resources and query flags are
        supposed to be available;
  call EXPERT COMMUNICATION INTERFACE procedure;
  get access to EXPERT DATABASE;
  if ATHOS_structure = on then call STRUCTURE procedure;
  if resources = on then call RESOURCES procedure;
  if query = on then call QUERY ASSISTANT procedure
end EXPERT.
```

HELP

The HELP procedure serves all the functions performed by ATHOS. HELP has to be available wherever and whenever. It is left to the team implementing ATHOS to prepare the help messages.

```
procedure HELP;
  comment global_menuflag, global_input_formflag, .....,
          function_menuflag, field_input_formflag, ...
          are supposed to be available;
  if global_menuflag then display the menu help-message;
  if global_input_formflag then display the format help-message;
  ...
  if function_menuflag then display the function help-message;
  if field_input_formflag then display the field help-message
end HELP.
```

3.2. Outline of the Retrieval Language

The retrieval language is a part of the man-machine interface (dialog language) mentioned in Section 2.7, Fig.2. This language has to fulfill the following postulates:

- to exhibit large expressive power; in fact the language should be informationally complete what means that any data item stored in the database can be reached by a query expressed in this language;
- to allow formulation of "narrow" (precise) queries in order to reduce the level of information noise;
- to be easy and user friendly;
- to provide necessary statistics on the retrieval done.

There are several types of information-retrieval languages. Many of them are based on Boolean Algebra which provides a convenient and natural way of expressing information needs. The retrieval language proposed for ATHOS belongs also to this category. Below the syntax and semantics of the ATHOS retrieval language are sketched.

For the sake of clarity and to avoid exhausting terminological discussions the term "access point" is introduced. An access point is any searchable element defined for a given database. It can be a descriptor, key-word, key-phrase, etc. The access point is assumed to represent the class of the records from the database that contain this access point in their bodies.

Beside the precise access point defined above, the ATHOS retrieval language should allow to use a right-truncated access points. This technique, referred to as root searching or right truncation, allows to search on leading sequences of characters. The system will automatically preform a logical OR operation between all access points having the specified root.

Precise access points and right-truncated access points are not the only operands which may occur in a query. Another operand, called an ANY term, can also be used when formulating the query. An ANY term is a collective term standing for a pre-defined set (cluster) of access points. Whenever the ANY term is used in the query, the system automatically execute a logical OR operation on all access points of the cluster associated with that ANY term.

The boolean operators are necessary to build queries. It is assumed that at least three basic operators, viz. AND, OR and NOT are available within the ATHOS retrieval language. Additionally, field level and proximity operators which are more restrictive types of the AND operator are supposed to be available within the ATHOS retrieval language. Such operators are especially useful for natural language processing.

A boolean expression is a sequence of operands, operators and parentheses. The syntax of the expression is as in normal logic. Semantics of the expression can be viewed as a set of actual records relevant to it. For instance, a query (which is a boolean expression)

A * B + C

refers to thoses records which contain the A and B operand (both of them must occur) or the operand C.

Further details on the syntax and semantics of the retrieval language are needless at this moment because ATHOS, as it says the criterion (vii), is to be implemented by means of a ready software package which usually contains a pre-defined language for retrieval. Modifications or development of such a language are time-consuming.

It is assumed that the ATHOS retrieval facility will be able to inform the user on the number of records relevant to each operand and each subexpression.

It follows from Section 2.6 that the user can call for assistance when establishing a query. Of special value would be lexical support comming from a thesaurs or a dictionary. But not only semantic aids are necessary; it seems that, in particular for casual users, syntactic support of the kind offered by the Query By Example language, could be useful. Another thing is to have the facility allowing to build the final query step by step. Breaking the query into elements allows to check at each step the number of record retreived, but also to display at the screen the records themselves, so as to verify at any time the logical validity of the search formulation in terms of relevance of the records retrieved.

3.5. Database Administrator's Tools

Two kinds of tools are to be available for the database administrator, viz.

- tools offered by the package to be used for implementing ATHOS;
- some tools developed during implementation of ATHOS.

The standard tools of the first category are as follows:

- a language to define the logical structure of the database including definition of all the fields;
- an editor to design input worksheets;
- an editor to design output formats. Three types of formats may be defined: for a screen, for a printer and for a desktop publishing package;
- a language to define selection techniques and rules of building inverted files and dictionaries;
- an editor to design menus and a facility to link them together;
- a data manipulation language.

Of the tools to be established by the team who will implement ATHOS, the database administrator should have access to the following:

- an array of procedures for data validation;

- procedures for status assessment (cf. Section 2.6, description of the data input function);
- data protection procedures;
- procedures mentioned in points (i) and (ii) in Section 5.3 (Implementation Schedule);
- tables containing the messages used by the HELP facility.

IV. DATABASE ADMINISTRATION

This chapter is dedicated to a databse administrator to whom so many references have been made so far. The relationship of the database administrator with ATHOS has been already presented in Section 2.6, Fig.2.

4.2. GENERAL REMARKS

The database administrator is granted responsability for matters that deal with the databases as a whole. He controls the overall structure of the data and is responsible for safekeeping of information stored in the database. Controlling the data does not mean that the database administrator knows the content of records, however, he has to know their structures and relationships between them.

In reality the database administrator may be a department or group rather than one man. A complete understandig of the database, its organization, its economics, its design criteria, and the requirements of its many users is too much for one man.

In the database literature a distinction between the so-called application programmer and the database administrator is made. The application programmer is a professional who writes and maintains programs written in the data manipulation language. These programs implement specific tasks which cannot be performed by means of the facilities available within the software package used. The stored, compiled versions of these programs can be invoked by commands as many times as needed. Typically, the application programmer is treated as an advanced user who is familiar with sophisticated techniques offered by the database management system.

In this document, however, the application programmer is considered as a member of the database administrator team. The reason to do so is the nature of ATHOS (cf. its purposes and functions described in Sections 2.1 and 2. , respectively) and the organization of the Industrial Information Section which is to be its owner.

4.2. TASKS OF THE DATABASE ADMINISTRATOR

The ATHOS database administration responsabilities are the following:

(i) Participation in the creation of the original description of the database structure and the way that structure is reflected by the files to the physical database.

(ii) The granting to the various users of authorization to access the databse or parts of it.

(iii) Description of new databases and modification of the existing database descriptions.

(iv) Introducing slight extentions of the ATHOS processing facilities. Defining and implementing new functions cannot be excluded either. In this case the database administrator acts as an application programmer.

(v) Generating status reports dealing with the exploitation of ATHOS.

(vi) Supervising and participation in inputing the data to ATHOS, data export and dissemination of the ATHOS software among the INTIB focal points.

(vii) Making backup copies of the database and reparing damage to the database due to hardware or software failures or misuse.

(viii) Upgrading database quality, in particular, during the database lifetime the database administrator should participate in all the enterprises which are aimed at diminishing the information noise generated by ATHOS.

Additionally, the database administrator should be ready to assist all the ATHOS users, if requested, and to take care of the system development and its promotion.

As far as the tools for defining new databases and procedures (items iii and iv) are concerned it is recommended, for the former, that the language based on menus and tables to be fill in be used, and, for the latter, the language of PASCAL or C type.

V. CONCLUSIONS

This chapter is devoted to technical and organizational issues and tasks to be undertaken in order to implement ATHOS. In particular proposals dealing with hardware and software selection as well as sugesstions concerning implementation schedule are presented. As it follows from previous considerations, ATHOS has been conceived as an open system and it can be extended according to the emerging future needs. This is why, this chapter is to be concluded with a sketchy development plan of ATHOS.

5.1 Proposal of Hardware

The choice of the class of hardware for running ATHOS was practically made in the design criterion vi, Section 2.4 where the IBM PC AT compatible microcomputer was recommended. This was done as a result of discussions with the INTIB staffers. It is quite clear that in the current stage of the INTIB development where the usage of such microcomputers is a commonplace this choice is plausible.

Before entering upon any details concerning hardware configuration let us evaluate the needs for storage of data. The total volume of the data in question can be estimated according to the following formula:

$$V = NDB * ANR * ALR$$

where: V - total number of ATHOS databases
NDB - number of databases
ANR - average number of records in a database
ALR - average length of a record (in characters)

At present, given the existing ATHOS directories (cf. Annex 1) and assessment made by the INTIB staffers there is:

NDB = 20
ANR = 1000
ALR = 2000
so V = $40*10^6$

It is expected that the volume of the ATHOS database within 5 years will grow up by a factor of 2. It should be stressed that the above estimation includes the data stored in databases only, however, it does not take into account the derived data like inverted files, speeding up catalogues, dictionnaires, etc. It is well-known from database administrators' experience that the volume of derived data files ranges from 30 to 60 percentages of the original database volume. So the presented needs, for 60% overhead, can be estimated as 40+24=64 Mbytes, and after 5 years as 128 MBytes.

The volume of the EDB database, thesaurus and other dictionnaires and tables (cf. Section 2.7) can be neglected vis-a-vis the volume of the ATHOS databases.

Since the memory requirement for ATHOS is relatively high, the hard disk to be employed must be of large capacity (more than 128 MB) and short access time. In a case such a hard disk is not available the databases must be split by the database administrator so that the relevant part of them be stored on the hard disk.

Below is a specification of hardware configuration for ATHOS:

(i) IBM PC AT compatible microcomputer with:
- 12 MHz clock
- at least 3 MB RAM
- at least 1 hard disk of no less than 128 MB
- 1 floppy disk drive (5.25") of 1.2 MB
- 1 floppy disk drive (5.25") of 360 KB or 1 drive (3.5") of 0.7 or 1.4MB

(ii) 1 keyboard
(iii) 1 high resolution monitor with an appropriate control card
(iv) 1 mouse
(v) 1 matrix dot printer
(vi) 1 laser printer
(vii) 1 streamer

Justification of items iii, iv and vi is that professional quality of editorial work (desktop publishing approach) is required from ATHOS .

5.2. Proposal of Software

At present, there are many database software packages for IBM PC AT compatible microcomputers available on the market. The most popular of them is dBASE III+ by Ashton-Tate. Other examples are PARADOX, R: SYSTEM V which belong to the same class as dBASE III+. On the other hand there exist several information-retrieval packages of which Mini/Micro CDS ISIS 2.0 by UNESCO is the most advanced as far as the scope of the facilities available is concerned.

Comparison of selected features of dBASE III+ and Mini/Micro CDS ISIS 2.0. was done in Annex 3. The comparison leds to the conclusion that for the ATHOS functions implementation the Mini/Micro CDS ISIS 2.0 package is more suitable than dBASE III+ and similar relational database management packages.

Many desktop publishing (DTP) packages are available on the market, e.g. Ventura Publisher, PageMaker/PC, Frontpage, Pagemater, AdvantTex, Rim System. After discussions with the INTIB staffer who is supposed to work with DTP the XEROX VENTURA PUBLISHER ver.1.1 which can run on IBM PC AT compatible microcomputers is propsed.

5.3. Implementation Schedule

The standard software package which is to be used for implementing ATHOS (Minin/Micro CDS ISIS was suggested in that respect in Section 5.2) will be called PACKAGE later on. It is proposed to implement ATHOS in eight stages:

(i) Preparation of software tools for semi-automatic defining

- logical structure of a directory (field definitions)

- input worksheets
- sorting patterns
- output formats
- retrieval patterns (specification of searchable fields)
- help feature

These tools are to be used not only for implementation; they will be added to the ATHOS Definition & Modification Mechanism (cf. Section 2.7, Fig.2, level 3) being a part of the PACKAGE, and will be given to the database administrator's disposal.

(ii) Preparation of software tools for semantic control

- handling the thesaurus
- handling the dictionaries
- handling the tables

(III) Implementation of the shell of the consulting module including

- user communication interface
- assistance for query formulation
- access to the Expert Database (EDB, cf. Section 2.7, Fig.2, level 4 and 5)

(iv) Implementation of the ATHOS facilities (cf. Section 2.7, Fig.2, level 3) through

- selection and activation of some facilities available within the PACKAGE, e.g. data entry, update, retrieval, printing, etc.
- implementation of the rest of facilities, e.g. data validation, editing, data protection, etc.

(v) Implementation of all the databases to be handled by ATHOS viz. DB_1, $,DB_n$ and EDB (cf. Section 2.7, Fig.2, level).

(vi) Creation of an experimental databases.

(vii) Testing, evaluation and refinement.

(viii) Installation of ATHOS at UNIDO/INTIB HQ.

Fig.3 presents the schedule of the ATHOS implementation. Manpower needed to implement ATHOS can be estimated as 10 man-months.

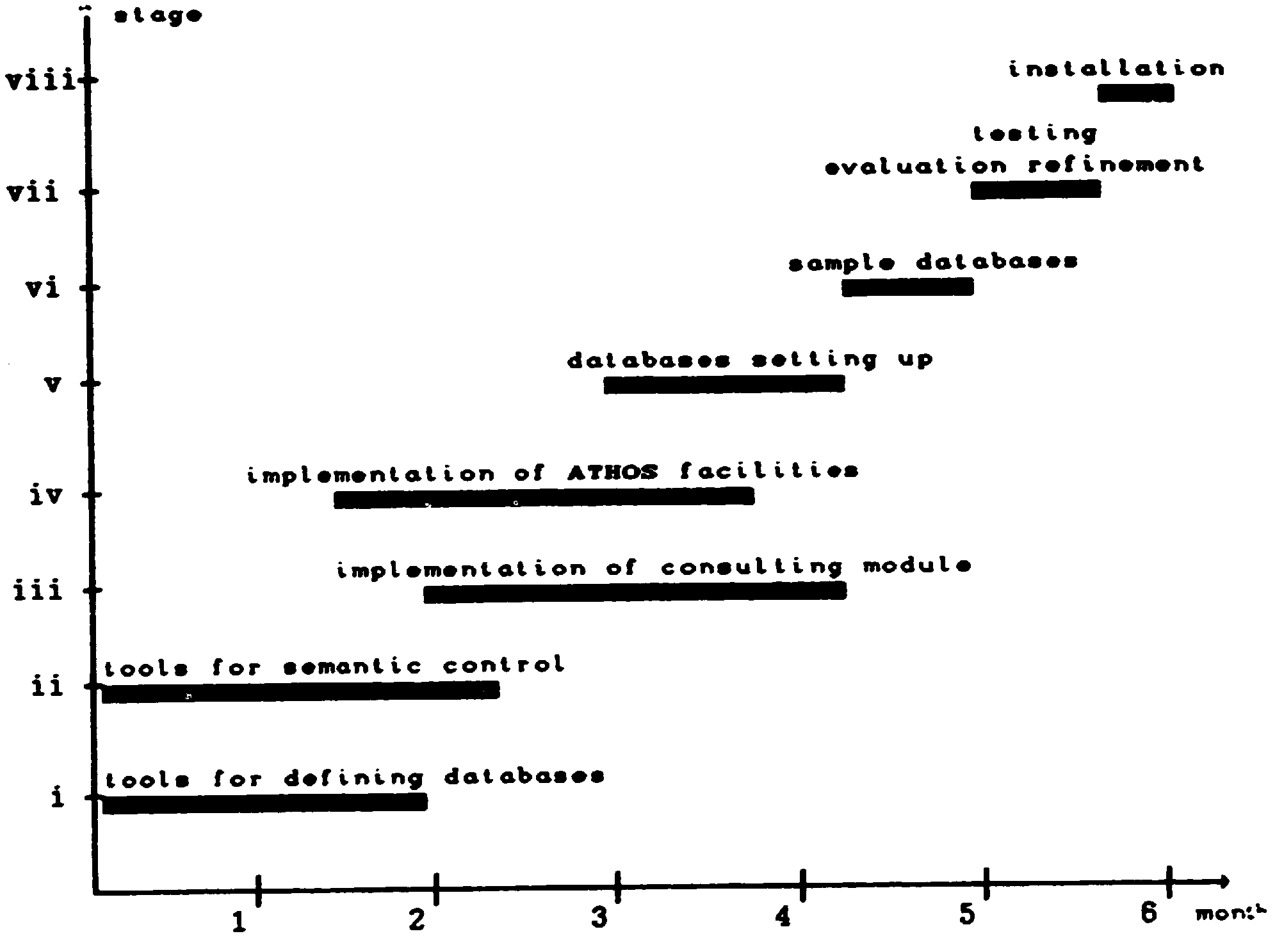

Fig.3. Schedule of the ATHOS Implementation

5.4. Further Work

A list of further work which might be undertaken to enhance and to improve ATHOS is proposed below.

(i) Refinement and optimization of

- logical structure of the ATHOS databases
- procedures implementing the ATHOS functions

(ii) Extention of the consulting module towards communication in a semi-natural language.

(iii) Establishment of conversion modules to other information systems.

(iv) Adapting ATHOS to use it in a network.

(v) Establishment of the thesaurus (multilingual ?) and/or the classification system.

(vi) Designing and implementating facilities to handle multilingual databases.

VI. REFERENCES

1. Bańkowski J., Wysocki A., Guidelines for the Establishment or Redesign of Industrial and Technological Information Service System, Including Selection of Software and Hardware, UNIDO/IS.597, Jan.15, 1986.

2. CDS ISIS (Mini-micro version), UNESCO, Paris, 1985

3. CDS ISIS (Mini-micro version), Preliminary Guide to Version 2, UNESCO, Paris, 1986.

4. PC's Handbook for 16 Bit PC User's, Microway Enterprises Co. Ltd.

5. Programming with dBASE III+, Ashton-Tate, 1985.

6. Nolan R.J., Informatics for Industrial Development, UNIDO/IS. 415, Nov.25, 1983.

7. Wysocki A., UNIDO Industrial Information Medium Term Programme, UNIDO, Jan., 1986.

8. Role of INTIB. Round Table Discussion of an Advisory Group of INTIB Users, UNIDO/ID/WG, Sept.20, 1985.

ANNEX 2

LIST OF INDUSTRIAL SECTORS TO BE DELT WITH BY INTIB

1. Food processing with special emphasis on vegetable oils and fats
2. Fertilizer
3. Leather and leather products
4. Capital goods with special emphasis on energy related equipment and technology
5. Petrochemicals
6. Building materials
7. Fishery industry
8. Industrial manpower training
9. Agricultural machinery
10. Non-ferrus metals
11. Iron and steel
12. Pharmaceuticals
13. Industrial financing
14. Wocd and wood products
15. Textile and wearing apparel
16. Electronics
17. Agro-industry
18. Low-cost transport
19. Pulp and paper
20. Energy: New and renewable sources of energy;
    Non-conventional sources of energy;
    Energy for rural requirements

ANNEX 3

COMPARISON OF DDBASE III+ AND MINI/MICRO CDS ISIS

The dBASE III+ package is a general relational database management system. It implements the idea of relational model of the real world. Therefore, it is especially useful if the structure of data to be stored in a database can be organized as a set of relations. It has to be stressed that the process of the relational model buiding, though partially algorithmic due to Codd's normalization procedure, is relatively difficult.

<u>Technical specification of dBASE III+</u>

Number of records	- 1 billion maximum
Number of bytes	- 2 billion maximum
Record size	- 4000 bytes in .dbf file 512 KB in .dbt file
Fields	- 128 maximum
Character fields	- 254 bytes maximum
Data fields	- 8 bytes
Logical fields	- 1 byte
Memo fields	- 5000 bytes maximum or the capacity of the word processor used
Numeric fields	- 19 bytes maximum

15 open files of all types.
10 open database files. A database file counts as two files if memo fields are used.
Seven open index files per active database file.
One open format file per active database file.

CDS ISIS is a generalized Information Storage and Retrieval system designed specifically for the management of structured non-numerical databases. CDS ISIS can handle unlimited number of databases, however, it is impossible using its standard mechanism to define relationships between the databases.

Technical specification of Mini/Micro CDS ISIS

Unlimited number of databases.
Number of records in one database - 16,000,000

Records of unlimited variable length.
Each record may contain up to 100 types of fields of maximum 1650 characters length.
Each field may be arbitrary split into subfields and repeatible structures.

Length of access point in the Inverted file - 30 characters.
Total length of sort keys up to 256 characters.

Table 1 compares the selected features of dBASE III and Mini/Micro CDS ISIS.

TABLE 1

	dBASE III+	Mini/Micro CDS ISIS
whole package user friendliness	good	good
database administrator friendliness	good	good
information retrieval friendliness	low	good
response time for a query	long	short
compatibility with other UN information systems	low	Micro ISIS has been developing by UNESCO and widely used in UN. It is logically compatible with the mainframe CDS ISIS broadly used in UN, including UNIDO
suitability for implementing information systems	average	high
price	$ 385	free of charge
quality of documentation	good	good

ANNEX 4

FIELD DEFINITIONS, INPUT/OPUTPUT FLOW AND TABLES