Hunt Institute for Botanical Documentation
5th Floor, Hunt Library
Carnegie Mellon University
4909 Frew Street
Pittsburgh, PA 15213-3890
Telephone: 412-268-2434
Email: huntinst@andrew.cmu.edu
Web site: www.huntbotanical.org

Paper prepared for The Comp. Center / Cotton Mill
Izmir, Mar. 25, 1974

Introduction and Background

A. General. The free-field data format which we have adopted is not necessary, in the sense of the computing machine, but is essential to provide the flexibility required of the types of data to be included in each data bank for genetic resources conservation work. There are several data banks necessary to perform the work, and each data bank conforms to the same pattern, but with different amounts of data for each bank.

To review the basic format for each data bank, the following definitions are provided:

(1) item: that object (collection of seeds, or any other) upon which the data are gathered.

Example: an accession (individual collection) is an item. Or, may be any other object upon which to base a set of descriptions, such as the individual pieces of furniture in a furniture store.

(2) Descriptor: a single basis for description of an item, defined over the set of items to be described.

Examples: the day of the week may be a descriptor.
" month " " " "
" year " " " "
" color of the seed may be a descriptor
" height of the plant " " " ", etc., etc.

(3) Descriptor state: for each item, the precise definition of that item, under each descriptor

Example: item = an accession
descriptor = color of seed
descriptor state = red, or brown, or yellow, etc.

The descriptor state is that which is key-punched, and not the descriptor itself.

(4) Data bank: a series of items, of any number, each item associated with a series of descriptors, of any number. The data bank is a logical set of related items. The data bank may be divided into smaller subsets when any data bank becomes too large for convenient manipulation in the computer. As an example of subdividing a data bank,

take the case of the collection data bank. At the beginning, all the collections of all types (species) of plants may be incorporated in a single collection data bank, but as this bank becomes too large, the bank may be subdivided into separate banks for each type (species) of plants that are found in the collection data bank. Thus:

A collection data bank for wheat,
" " " " " rye,
" " " " " barley, etc., etc.

The above provides the basis for input to the TAXIR system, and one which will be employed, not only at the Izmir center, but in every similar center around the world. This facilitates exchange of information between centers.

B. The TAXIR System for information storage and retrieval is a general IR system for any type of data which may be defined as in A. above. It is an integrated set of programs which provides the necessary basic functions required in storage and retrieval: editing, correcting, changing, adding or deleting data, a complete boolean-algebra based querying system, a set of printing options, and a format generator for further data processing (such as statistical analyses, mapping or plotting, or performing other computing functions. The main distinction of TAXIR is an efficient, set-theoretic means of compacting the data and providing addresses for easy retrieval.

The system is not large, in comparison with other general IR systems--it requires between 18 and 20 K 60 bit words of memory (CDC machine) for the program package. Perhaps it may be subdivided, and run as a series of subprograms, but to do so requires very experienced and expert programming.

Reprogramming from a CDC machine to an IBM machine is also an expensive process, even though the system is written in FORTRAN IV. As with most IR systems, TAXIR is machine-dependent, particularly in the input-output functions.

C. <u>Embedding</u> the TAXIR system into an overall documentation system for genetic resources work requires a high level of several skills. The documentation system requires a management function, a data user function, a computer software function, and expert financial (cost analysis) function. We intend to provide this set of skills by employing a team already experienced in all the above-named functions for genetic resources documentation systems. We hope to provide these skills all over the world, but without the experts, nothing can be accomplished. We hope not only to install the systems, but also to train both the documentation experts and the computer programming experts wherever a decision is made to install our documentation system. Clearly, a team approach is required.

D. <u>The global nature</u> of our work indicates that we will need to use a very large central computing facility for the full functioning system. But, by careful systems design, we should be able to provide each center with at least a basic IR capability.

We hope to be able to establish such a function in Izmir. We want to provide the necessary expertise, so that work can be done locally, but it must be understood by all concerned that all the work of documentation cannot be accomplished with present facilities or staff.

We therefore must proceed here with a few steps, all of which provide some useful results, and which fit easily into the larger system without causing any later revisions.

One item that will be needed as an adjunct to any computing system employed for documentation of genetic resources is a tape drive. Magnetic tapes will provide the only means of long term data storage, as well as providing the best means to share data between computing centers.

Programming and production jobs to be performed.

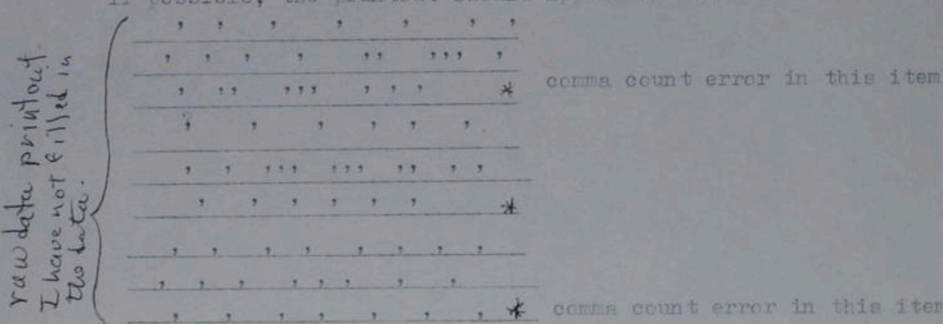
Using the input format already adopted (free-field, with the comma as a symbol for separators between descriptor states, and an asterisk as the end symbol for each item), the next two programming jobs are editing procedures.

1. The first step in editing is to ensure that each item is defined by the same number of descriptors (which, of course, vary between different data banks). We refer to this step as "the comma count". The documentation specialist at the agricultural center must provide the computing center with a description of each data bank as follows:

   1. The approximate number of items included (purely an estimate).
   2. The exact number of descriptors for the data bank. That is, there must be a common number of descriptors for each item in the data bank.
   3. The exact order of the descriptors for the items.
   4. The names of each descriptor.

The computing center must write a program which counts the commas for each item, and prepare a printout which indicates when an error in the input occurs. The printout should associate the error notification with the accession number of the item (or indicate the error on the same printout as the raw data). This latter depends on the character count/line capacity of the printer.

If possible, the printout should appear as follows:



If the comma count error message cannot be easily fit onto the same page of the printout as the raw data, then an alternative would be to make a separate printout, in which the items with incorrect comma counts are listed, preferably giving the accession number of the incorrect item.

This is an effective check on one aspect of the input. With the comma count, the documentalist can easily return to the input data and determine where a descriptor was ommitted, or where a superfluous commas was added during key-punching.

2. The second step in editing can be called the "control vocabulary". What the printout contains is a list of the descriptors and, under each descriptor, all of the <u>discreet</u> descriptor states. The order of the descriptors will be the same as that provided to the computing center by the documentalist. The descriptors should be numbered sequentially.

If a descriptor state for any descriptor has already been listed, there is no need to repeat that descriptor state in the listing. Thus, if the descriptor consists of a list of numbers, there is no need to repeat any number under that descriptor. The same may be done for alphabetic data. (As in the case of the first data bank, q.v., the collector's name is the same one throughout, therefore the printout for descriptor number 9. Collector/s will be but one collector, Pervis Parvaneh.) The only exception, of course, is if there is a misspelling of the name. It is important to record all variations in any descriptor state in the printout, to check all spellings and numerical recordings.

This provides the documentalist with another means to discover input errors and to clean up the input data.

Additionalnotes:

1. There are a few symbols used as indicators in the TAXIR system, and these should not be used to describe any input data. These are:

comma , = a separator between descriptors.
asterisk * = the end of the descriptors for an item.
parentheses () = a boolean combination.

SEMINAR ON DOCUMENTATION
D.J. Rogers
Crop Ecology and Genetic Resources Unit

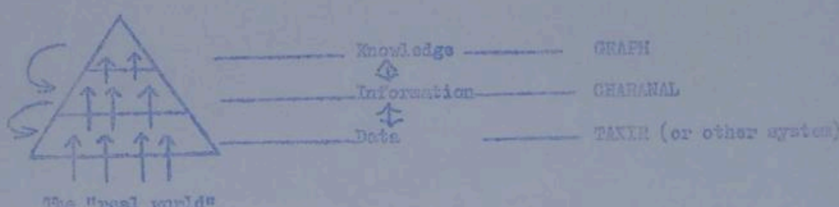
SEMINAR No. 1
4 March 1974

A. Introduction to the Seminar
   1. Background philosophy for "data management".
   2. Some understanding of computers — hardware and software.
   3. User's description of TAXIR, and some related programmes.

B. Why ? — a common basis to begin
   1. Many different systems for very specific purposes.
   2. Very few systems designed for all information needs.
   3. Very few systems are user-oriented.
   4. Need for a system that is general, to be used by people not expert in computers.

C. What is a "system" ?
   1. Generally, an organizational structure to accomplish a series of tasks.
   2. In GRO, all the functions associated with genetic resources.
   3. As a part of documentation, the computer programmes necessary to take care of the data.

D. Embedding documentation into a larger system
   1. The "triangle of knowledge".



B. Basic concepts for our documentation systems

1. Reflection (or model) of methods already known and understood, that is, the biological discipline of taxonomy.
2. The methods employed consider the <u>structure</u> of information, <u>not</u> the <u>substantive content</u> of information.

   Example :

   Words consist of letters (including consonants and vowels), and this is the structure of a word.

3. Example of the taxonomic structure useful in building an IR system.

   a) The binomial scientific terms

Genus - a noun		<u>Triticum</u> - wheat
Species - an adjective		<u>aestivum</u> - a particular kind of wheat.

   The above combination designates very precisely one kind of wheat.

   b) Using the above structural format, in the IR milieu,

Descriptor	- a noun	colour
Descriptor state	- an adjective	red, green, blue, etc.

   c) The Latin binomial is an <u>exclusive</u> system — the two terms allow no overlap with anything else.
   Likewise, a descriptor and associated descriptor state are non-overlapping, or <u>mutually exclusive</u>

   Example :

   Descriptor - colour of stem
   Descriptor states - red
   yellow
   red-yellow (this state used only for those plants exhibiting <u>both</u> red and yellow).

SEMINAR ON DOCUMENTATION
D. J. Rogers
Crop Ecology and Genetic Resources Unit

SEMINAR 2
6 March 1974

A. Review of last Seminar

1. The triangle of knowledge — our interest in the basal portion.
2. Comparative methods — from the model of taxonomy.
3. Generalities on methods of description.

B. Purpose of this Seminar

1. Further description of methods of data description.
2. To guarantee that our methods accomplish the necessary tasks.

C. Continuing discussion on the methods of description

1. The <u>item</u> : the <u>object</u> which forms the basis for description. Ex. : in genetic resources, the accession may be the item ; in other applications, a person specialising in some aspect of genetic resources function, etc.
2. <u>Data</u> may be <u>extrinsic</u> to the item (ex. : collection number), or may be <u>intrinsic</u> to the item (ex. : shape of some structure of the item)

D. The classification of types of descriptors (please see attached example)

1. Purpose of classification of types
   a) To insure coverage of required information for any purpose.
   b) To provide a framework to compare one set of data with any other similar set of data.

E. Structure of Descriptors, in terms of their function.

1. NAME descriptor : usually words — any combination up to 90 letters.
2. CODE descriptor : a form of the NAME descriptor, but shortened to facilitate recording. (ex. : the months of the year, represented by numbers, consecutively, from 1 to 12.
3. ORDER descriptor : Numerical or alphabetical, defined in a "from-to" format, giving maximum range of included data, from smallest to largest.

F. The concepts of a <u>data bank</u>

1. For sake of efficiency, necessary to organize data into logical sets.
2. For sake of costs in data gathering, need to keep within bounds of time, personnel capacity, and budget.
3. For sake of efficient computer use, need to keep sets of data within some limit of size.
4. When any one data bank becomes too large, may be subdivided, again using some logic.

Provisional, not fully developed,
CLASSIFICATION OF DESCRIPTOR TYPES IN GRC WORK

1. "Finding information"

Accession information
- accession numbers
- accession dates
- accession types

Collection information
- names of collectors
- collector's numbers
- collection dates

Nomenclature information
- scientific names
- common names
- numeric designations of individuals

Origin
- country of origin
- state or province
- precise locality
- latitude/longitude
- donor of accession

Storage information
- physical location of accession in store
- conditions of storage
- bag, envelope, box, can, etc.
- storage temperature, etc.

2. Organismic information

Biochemical
physiological
genetic information
- chromosome number
- crosses

Morphological
- plant dimensions
- plant habit
- root characteristics
- stem (or tuber) characteristics
- leaf characteristics
- flower characteristics
- fruit characteristics
- seed characteristics.

3. Pest and disease information

(Resistance to, types of organisms, etc.)

Bacterial
Fungal
Viral
Insect
Nematode
Other pest or disease

4. Environmental information

Ecological
Environmental damage
wind
water
drought
other

5. Rejuvenation or regeneration

Germinability tests
History
dates of tests
where work done
when returned to storage
etc.

6. Use information

Breeder's data
Specific application
food quality
advantageous or detrimental quality
etc.

7. Other categories, as required

SEMINAR ON DOCUMENTATION

D.J. Rogers
Crop Ecology and Genetic Resources Unit
Seminar 3.

8 March 1974

A. Before proceeding to discussion of TAXIR, need a short discussion on computing machines.

1. What they are and what they do - movie from IBM
2. The physical equipment referred to as "hardware"
3. The sets of written instructions which direct the functions of hardware - the programmes - referred to as "software"

B. Basic Concepts of hardware

input - compute - output

1. Input:

Any of several devices by which data and software are presented to the machine, such as keypunch, paper tape, or magnetic tape. Also refers to "that which is put in", data, for example.

2. Computer:

Consists of a Central Processing Unit (CPU) and associated memory units, such as discs or drums.

(a) In the CPU all instructions (programmes) are executed
(b) Size of CPU is a measure of the capacity of the machine - measured usually in two terms: size of internal memory, and speed of computation:
  (i) internal memory of CPU referred to as "core" memory;
  (ii) memory is measured in "bits", "bytes", and "words";
  (iii) a "bit" is a single on-off signal, and one to several bits required to store a single alphabetic or numeric character;

(iv) a "byte" is a term peculiar to IBM machinery, consisting of 8 bits;

(v) a "word" is a set of bits, determined as one unit. Design of hardware determines the length of a word. IBM, for example, has standardized on a 32 bit word, consisting of four 8-bit "bytes"; UNIVAC has a 36 bit word, and Control Data Corporation (CDC) has either 48 or 64 bit words. (This latter is by far the best configuration for our efforts);

(vi) Size of CPU memory varies, depending on design and cost. A small size machine may have 8 000 words, a medium size, 16 – 32 000 words, and the largest computers may well have over 200 000 words. In the jargon of the computing field, 1 000 = "K", so a small machine will have "8K" memory, medium "32K" and large one "200K".

(c) Speed of the operations can approach the speed of light, measured in extremely small fractions of a second, the smallest fraction being the "nanosecond", or 1/1 000 000 of a second;

(d) Associated memory to CPU may be of two general types:

(i) magnetic tape – the tape stores electronic signals much the same as in a tape recorder. While very large quantities of data may be stored on "mag" tape, this is the slowest type from which to recover the data for processing in the CPU;

(ii) disc or drum memory – frequently referred to as "random access" memory, stores incredibly large quantities of data, frequently over 1 000 000 bits. Much faster recovery of data than from tapes.

3. Output:

from computer takes several forms, usually printed, but may be another mag tape, another set of punched cards, or microfilm, or sometimes a display device like a TV tube, referred to as a cathode ray tube (CRT).

Printing machines vary in speed, increasing as the size of the machine increases, up to about 2 000 lines of 128 characters each line, per minute.

Output in the form of graphs or maps is produced on a separate piece of hardware, either a "flat-bed" plotter, or moving paper plotter.

C. Basic Concepts of software

1. The idea of a predetermined method of directions:

(a) In our mind, we have a set of instructions to add, subtract, multiply, order, and follow through some procedure to accomplish some task;

(b) The computing machine has no such "built-in" set of instructions, and must be directed to accomplish each step in the correct order to complete any task. The task can be any type, from solving equations, retrieving data from storage, or printing a computer-drawn art work;

(c) The above indicates that we must use a "stored" programme, to direct the computer;

(d) All functions in the computer are accomplished by answering a series of yes-no type questions, or, in terms of electronic signals, instead of yes-no, on-off. This means of signals to direct the computer is called "machine" language, and the programme looks like a series of random zeros and ones. This is very confusing, and boring, to human beings, and machine language programming is extremely prone to human error.

2. The concept of a "higher" language for software.

(a) Since we understand, and feel more comfortable with, instructions which look more like natural language, computer specialists have developed several programme languages which can be "interpreted" by the machine. These languages have become universally accepted by all computer manufacturers, so that any machine can use these "higher" languages. One such language, of greatest general acceptance, is called by the acronym FORTRAN, for "formula translation";

(b) FORTRAN is the language used for TAXIR, which we will describe in the next seminar.

SEMINAR ON DOCUMENTATION

D.J. Rogers

Crop Ecology and Genetic Resources Unit

Seminar 4
11.3.74

Description of TAXIR:

A. Description of the Program:

1. 2 000 FORTRAN statements (more or less) instruct the computer to perform all necessary functions required in information storage and retrieval.
2. The FORTRAN used is one that is functional on medium to large size computing machines.
3. Each FORTRAN statement causes the machine to perform some necessary activity.
4. The first such statement, for example, essentially turns on the machine - START.
5. From this first statement onward, the TAXIR program places the data bank in storage, and then makes the data available to retrieve (to ask questions).
6. Other FORTRAN statements cause the machine to produce (print) a number of useful outputs.

B. Commands or statements, in certain FORMAT, allows the user to perform functions which are required;

1. The first command made by the user tells the computer to DEFINE ITEMS from cards or tapes; with each item, one defines the descriptors (and their type) which tells the machine which set of descriptors will be included in that data bank - separate data banks can be built from one set of input data cards. In the Introduction Unit, a data bank for accessions for the herbarium, for the storage functions (1) Long term storage; (2) Distribution; (3) Working collection.
2. After the command BUILD DATA BANK, the program automatically lists (prints) all items and the full set of descriptors, and descriptor states, with a list of errors in the data. The list of errors tells the user all the mistakes made by the keypunch operator; i.e. an item with too few descriptors, or an illegally prepared descriptor state.

3. Another command, executed after all corrections are made, causes the computer to provide a CONTROL VOCABULARY (CV). The CV is a list of each descriptor, and under each descriptor, each descriptor state. The CV tells the user all the words he has used in defining each item. It tells the user whether one word has been spelled several different ways (such as country of origin may have been spelled more than one way "TERKEY", "TORKEY" and "TURKEY"; where obviously only "TURKEY" is correct).
4. With the Control Vocabulary, one may also discover that several descriptor states are synonymous, and unnecessary.
5. The next command available is "CORRECTION", to correct the mistakes made in any descriptor, or descriptor state.
6. Still other commands are "ADD MORE DESCRIPTORS/DESCRIPTOR STATES (where more data are required); "DELETE DESCRIPTOR/DELETE ITEM" (where duplication has occured, or some other mistake in the data has been made).

C. Questioning the data banks

1. After all the problems in "B" have been solved, one is ready to use the data bank in any of many different ways:

    a) Complete lists of all items in the bank, with certain descriptors;
    Example: PRINT: items with country of collection, TURKEY"
    The results will be a collection of all types of accessions made in Turkey, including everything the data bank contains from this country.
    b) Precisely defined answers to specific requirements:
    Example: PRINT: PRECISE LOCALITY, ASPECT, LATITUDE and LONGITUDE for items with GENUS SECALE <u>and</u> SPECIES montanum but <u>not</u> SPECIES CEREALE.
    c) Accession No.
    Example: PRINT: ROOM, SHELF no., BOX NUMBER for items with GENUS TRITICUM <u>and</u> SPECIES AESTIVUM <u>and</u> VARIETY .... <u>and</u> LOCALITY AFGHANISTAN <u>or</u> IRAN.

Note use of certain words; <u>and</u>, <u>or</u>, <u>not</u>. These permit formulation of many types of precise questions, giving the user a limitless number of ways to derive the most benefit from his data.

Because of the freedom in formatting data, and because of the flexibility of the TAXIR program, the user is provided with a long-sought tool.

The ability to change, modify, add or delete information is part of the system.

TAXIR has a number of allied programs which can be used, for example:
- Statistical analysis of data
- Plotting and contouring on maps
- Character analysis and clustering program (TAXIMETRIC PROGRAMS)

TAXIR is, therefore, not just a computer program, but rather a number of programs integrated into a system. Learning to use the system takes some time, but the results and benefits far outweigh the costs.

SEMINAR ON DOCUMENTATION

D.J. Rogers

CROP ECOLOGY AND GENETIC RESOURCES UNIT

Seminar 5

13.3.74

A. Review of Seminars

1. The parts of the documentation function

a) Data-oriented, precomputer
b) Computer-oriented
c) Management functions required to integrate these

B. Flow of Work in a) above

1. Determine objectives of data bank(s)

2. Determine the item for each data bank

3. Determine types (classification) of descriptors needed in each data bank

4. Choose the descriptors, and the order needed to collect the data

5. Determine the descriptor types (NAME, CODED, ORDER)

6. Choose means of preparation for computer input

a) Key punch most common
b) Paper tape
c) Mark sensing
d) Mag tape

C. Flow of Work in A.1.b) above

1. Choose the software package needed (determined by objectives)

2. Choose appropriate computing facilities

a) Closed-shop computer center operation
b) Open-shop computer center operations

3. Determine whether appropriately trained personnel in computing center

D. For Both B. and C. above, Determine Budgetary Requirements

1. Factors in budget

a) Is the data bank central to functions of your unit (organization)
b) Personnel trained in data gathering
c) Is there a line item in your budget for computing functions
i) Includes trained programmers, software costs, and computer time
d) If no line item, determine possibility of outside contract
e) Is there a documentation group for your unit (organization) to take care of the whole documentation function. What are their costs?

E. Discussion of Seminar Participant Papers

1. Instructions given not clear - assignment was to:

a) List a set of data banks of interest to the individual or unit
b) Prepare for one data bank a list of descriptors

E. 1. c) Flow for this would have been to follow Steps in B. above

2. Most problems in design of descriptors

a) Definition of descriptors: a single basis for comparison, defined over the full set of the data bank. Further, the descriptor is (by rule) mutually exclusive, and non-overlapping

b) Examples to illustrate a)

i) Dates

The day of the month is a single basis for comparison
The month of the year " " " " " "
The year " " " " " "

Why important? To make a precise query, for example, on flowering times, days more important than months

ii) Names of persons

In almost all languages there is a family designation (in English, the surname) and a designation of the individual (in English, the given name(s). Therefore:
Family name or surname is a single basis for comparison
Given name " " " " " "

Why important? If we are to pinpoint a single individual in a query, must have the capacity to ask for Smith, John, as different from Smith, Robert. This can happen in a long data bank with many names in it

iii) Names of plants

Genus name is a single basis for comparison
Species name is a single basis for comparison
Authority for genus name is a single basis for comparison
Authority for species name is a single basis for comparison

Why?

3. Overlap in requirements between different disciplines

a) Several data banks included requirements for
climatic data
soil data
ecological data

b) Documentation function, if performed correctly, would allow sharing of the expertise among all the various functions of FAO

SEMINAR ON DOCUMENTATION

D.J. Rogers

Crop Ecology and Genetic Resources Unit

Seminar 4
11.3.74

Description of TAXIR:

A. Description of the Program:

1. 2 000 FORTRAN statements (more or less) instruct the computer to perform all necessary functions required in information storage and retrieval.
2. The FORTRAN used is one that is functional on medium to large size computing machines.
3. Each FORTRAN statement causes the machine to perform some necessary activity.
4. The first such statement, for example, essentially turns on the machine – START.
5. From this first statement onward, the TAXIR program places the data bank in storage, and then makes the data available to retrieve (to ask questions).
6. Other FORTRAN statements cause the machine to produce (print) a number of useful outputs.

B. Commands or statements, in certain FORMAT, allows the user to perform functions which are required;

1. The first command made by the user tells the computer to DEFINE ITEMS from cards or tapes; with each item, one defines the descriptors (and their type) which tells the machine which set of descriptors will be included in that data bank – separate data banks can be built from one set of input data cards. In the Introduction Unit, a data bank for accessions for the herbarium, for the storage functions (1) Long term storage; (2) Distribution; (3) Working collection.
2. After the command BUILD DATA BANK, the program automatically lists (prints) all items and the full set of descriptors, and descriptor states, with a list of errors in the data. The list of errors tells the user all the mistakes made by the keypunch operator; i.e. an item with too few descriptors, or an illegally prepared descriptor state.

3. Another command, executed after all corrections are made, causes the computer to provide a CONTROL VOCABULARY (CV). The CV is a list of each descriptor, and under each descriptor, each descriptor state. The CV tells the user all the words he has used in defining each item. It tells the user whether one word has been spelled several different ways (such as country of origin may have been spelled more than one way "TERKEY", "TORKEY" and "TURKEY"; where obviously only "TURKEY" is correct).
4. With the Control Vocabulary, one may also discover that several descriptor states are synonymous, and unnecessary.
5. The next command available is "CORRECTION", to correct the mistakes made in any descriptor, or descriptor state.
6. Still other commands are "ADD MORE DESCRIPTORS/DESCRIPTOR STATES" (where more data are required); "DELETE DESCRIPTOR/DELETE ITEM" (where duplication has occured, or some other mistake in the data has been made).

C. Questioning the data banks

1. After all the problems in "B" have been solved, one is ready to use the data bank in any of many different ways:

   a) Complete lists of all items in the bank, with certain descriptors;
      Example: PRINT: items with country of collection, TURKEY"
      The results will be a collection of all types of accessions made in Turkey, including everything the data bank contains from this country.
   b) Precisely defined answers to specific requirements:
      Example: PRINT: PRECISE LOCALITY, ASPECT, LATITUDE and LONGITUDE for items with GENUS SECALE <u>and</u> SPECIES montanum but <u>not</u> SPECIES CEREALE.
   c) Accession No.
      Example: PRINT: ROOM, SHELF no., BOX NUMBER for items with GENUS TRITICUM <u>and</u> SPECIES AESTIVUM <u>and</u> VARIETY .... <u>and</u> LOCALITY AFGHANISTAN <u>or</u> IRAN.

Note use of certain words; <u>and</u>, <u>or</u>, <u>not</u>. These permit formulation of many types of precise questions, giving the user a limitless number of ways to derive the most benefit from his data.

Because of the freedom in formatting data, and because of the flexibility of the TAXIR program, the user is provided with a long-sought tool.

The ability to change, modify, add or delete information is part of the system.

TAXIR has a number of allied programs which can be used, for example:

- Statistical analysis of data
- Plotting and contouring on maps
- Character analysis and clustering program (TAXIMETRIC PROGRAMS)

TAXIR is, therefore, not just a computer program, but rather a number of programs integrated into a system. Learning to use the system takes some time, but the results and benefits far outweigh the costs.