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

The Hunt Institute is committed to making its collections accessible for research. We are pleased to offer this digitized item.

Usage guidelines
We have provided this low-resolution, digitized version for research purposes. To inquire about publishing any images from this item, please contact the Institute.

Statement on harmful and offensive content
The Hunt Institute Archives contains hundreds of thousands of pages of historical content, writing and images, created by thousands of individuals connected to the botanical sciences. Due to the wide range of time and social context in which these materials were created, some of the collections contain material that reflect outdated, biased, offensive and possibly violent views, opinions and actions. The Hunt Institute for Botanical Documentation does not endorse the views expressed in these materials, which are inconsistent with our dedication to creating an inclusive, accessible and anti-discriminatory research environment. Archival records are historical documents, and the Hunt Institute keeps such records unaltered to maintain their integrity and to foster accountability for the actions and views of the collections' creators.

Many of the historical collections in the Hunt Institute Archives contain personal correspondence, notes, recollections and opinions, which may contain language, ideas or stereotypes that are offensive or harmful to others. These collections are maintained as records of the individuals involved and do not reflect the views or values of the Hunt Institute for Botanical Documentation or those of Carnegie Mellon University.

About the Institute
The Hunt Institute for Botanical Documentation, a research division of Carnegie Mellon University, specializes in the history of botany and all aspects of plant science and serves the international scientific community through research and documentation. To this end, the Institute acquires and maintains authoritative collections of books, plant images, manuscripts, portraits and data files, and provides publications and other modes of information service. The Institute meets the reference needs of botanists, biologists, historians, conservationists, librarians, bibliographers and the public at large, especially those concerned with any aspect of the North American flora.

Hunt Institute was dedicated in 1961 as the Rachel McMasters Miller Hunt Botanical Library, an international center for bibliographical research and service in the interests of botany and horticulture, as well as a center for the study of all aspects of the history of the plant sciences. By 1971 the Library's activities had so diversified that the name was changed to Hunt Institute for Botanical Documentation. Growth in collections and research projects led to the establishment of four programmatic departments: Archives, Art, Bibliography and the Library.

GENETIC RESOURCES

COMMUNICATION INFORMATION DOCUMENTATION SYSTEMS

GR/CIDS

A Working Description of an Interdisciplinary Analysis of Communications in Genetic Resources

Genetic Resources and Bionomic Systems Program
University of Colorado
Boulder, Colorado 80302
Cable: GENRESYS

Paper delivered by:
Date:
Place:

INTRODUCTION

The extensive ongoing work within the international community of genetic resources is extremely valuable in the efforts to solve the world's food problems. From diverse fields (such as plant breeding, agronomy, botany, economics, management science, and biometrics) scientists are endeavoring to make significant breakthroughs in the effort to provide an adequate global food supply, often working under difficult conditions. Intimately imbedded in these efforts is a set of problems, tasks, and activities dealing specifically with genetic resources conservation that cut across and overlap into the functional activities of these working scientists. For example, the importance of getting new germplasm from centers of genetic diversity is quite well known to the breeder.

Because of the importance of plant genetic resources conservation, the Consultative Group for International Agricultural Research (CGIAR) formed the International Board for Plant Genetic Resources (IBPGR) in November, 1973 to promote an international network of genetic resources activities. This global network would have the goal of providing an efficient mechanism through which information and physical material maintained at any genetic resources center would be exchanged and shared by any other center or authorized individual user. To realize a global network, the implicit genetic resource work functions currently in operation must be identified and understood. Also, an operating system must be designed and

implemented to satisfy the needs of individual scientists.

During the first year of operations the IBPGR initiated several projects. One project concerns an investigation of current methods of documenting, processing, and communicating genetic resources information. The objective of this investigation is to define and better understand the information problems and requirements associated with genetic resources activities so that the establishment of the global network may be facilitated. The IBPGR is funding an interdisciplinary research project at the University of Colorado through the Food and Agriculture Organization (FAO) of the United Nations to carry out this investigation of Genetic Resources Communication Information Documentation Systems (GR/CIDS).

As a starting point for this investigation the members of the Genetic Resources project have begun to work with individuals in accordance with their specific problems and the data requirements of the centers they represent. Knowing that decisions are made and problems are defined at these centers, the members of the Genetic Resources CIDS project are now working closely with staff scientists at several locations including Centro Internacional de Mejoramiento de Maiz y Trigo (CIMMYT), Centro Internacional de la Papa (CIP), Instituto Colombiano Agropecuario (ICA), and International Institute of Tropical Agriculture (IITA). Work is also being initiated at other centers. By closely working with centers' staff the members of the project expect to successfully fulfill the requirements of those to whom the project is geared.

The approach taken in making the investigation is not necessarily a traditional descriptive orientation. The main objective

is not only to identify the problems associated with genetic resources work and merely document them. The identification of problems is only one objective within a set of objectives that yield a communication network between scientists in the field. Other objectives within the set deal with directing the flow of data to the point where it subsequently can be transmitted or exchanged in response to requests as well as dealing with how new data can be recorded and used. To accomplish the network, it is necessary to work with real data drawn from active and base collections in different parts of the world. Further, computerized methods of information management will be tested at centers having the appropriate facilities. An effort will be made to facilitate the exchange of data analysis procedures that are used by individual scientists at different centers to promote feedback on analytical techniques.

One of the first steps necessary in the investigation, has been to take the organizational structure of genetic resources work and begin to understand the data and information problems that currently exist. The structure of Genetic Resources work needs to be identified. And the real data problems within this structure actually must be utilized as the foundation for the work.

Genetic Resources Functions

Five major functions have been identified in genetic resources work. These are exploration/collection, storage and maintenance, improvement (e.g. plant breeding), introduction, and utilization. While each of these functions has its own particular information problems, a systematic perspective reveals how each is linked to the other via data requirements as well as through the control and

measurement responsibilities of monitoring and evaluation. Figure 1 illustrates the structure of these five functions as they are connected through the data information transmission and feedback flows.

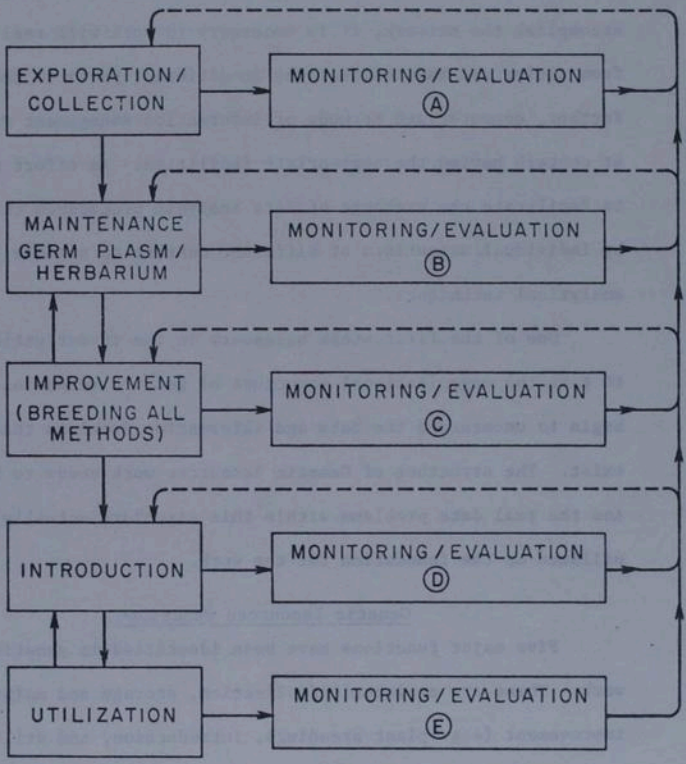

Figure 1. Information/material flows

From workshops that have been held as part of the GR/CIDS Project, it has been possible for scientists (breeders, statisticians, collectors) and administrators (research directors) to describe their own problems within each of these functions. These discussions have produced a more detailed breakdown which better defines and classifies the informational components of the various activities. These are described as follows:

Exploration/Collection

Planning, scheduling and executing collections with limited resources.

Data capture on genetic resources collection and operational taxonomic classification of genetic resources collection.

Documentation of collection information and its integration into a center's operating structure for immediate retrieval.

Diversity mapping for a more effective genotypic-geographic coverage.

Storage and Maintenance

Germplasm storage and initial testing procedures which include rejuvenation, multiplication, and access.

Documentation requirements of plant characteristics, history, field notebooks and literature references.

Herbarium specimen storage, documentation requirements such as annotation validation and access.

Improvement

Breeding objectives based on field inputs.

Field trial screening.

Progeny testing analyses.

Recording of and access to pedigree history.

Analytical methods for complex genotypic-environmental interactions.

Utilization

Expected effect and reaction to new genetic resources material on farming systems in target area.

Actual and expected effect on socio-economic conditions in extended target areas.

WORKING WITH REAL DATA PROBLEMS

The methods being used to investigate the network formation are organized into three categories. Each of the three is being applied within the five functional areas of genetic resources in close co-operation with members of participating centers' staff and on real data. These methods have been classified into the categories of systems analysis and design, development, and transfer. Although these functions are closely related and even overlap for several activities, there are separate functions with each.

Systems Analysis and Design

Using the tools of systems analysis, the members of the GR/CIDS Project have been and will be able to continue to discover the specific details of problems within the genetic resources structure. The analysis itself is simple in theory yet when applied can be a rigorous methodology. For such an analysis the following types of questions should be answered:

1. What part of the genetic resources system is being observed?
2. How does it function?
3. How does it link to other elements of the system?
4. How can its structure be altered to perform better?
5. How can the internal feedback mechanisms be made more effective so the system is more self-corrective?
6. What types of monitoring and evaluation techniques can be utilized for measuring progress?

This type of analysis provides a descriptive understanding from which an evaluation can be made of the operation of the network. The network is visualized as consisting of genetic resource operating units.

These units are in fact, genetic resources workers, operating individually or at centers which have a genetic resource component.

The first step in generating a descriptive understanding of the system is to determine what is known. This is done as follows: Pertinent literature is reviewed. Conversations are held with individual genetic resource workers. They are then asked to provide in writing their description of the network. Finally, there is a working through of a real data problem. From this information, comes an initial model which will continuously improve as it evolves. (See Figure 2)

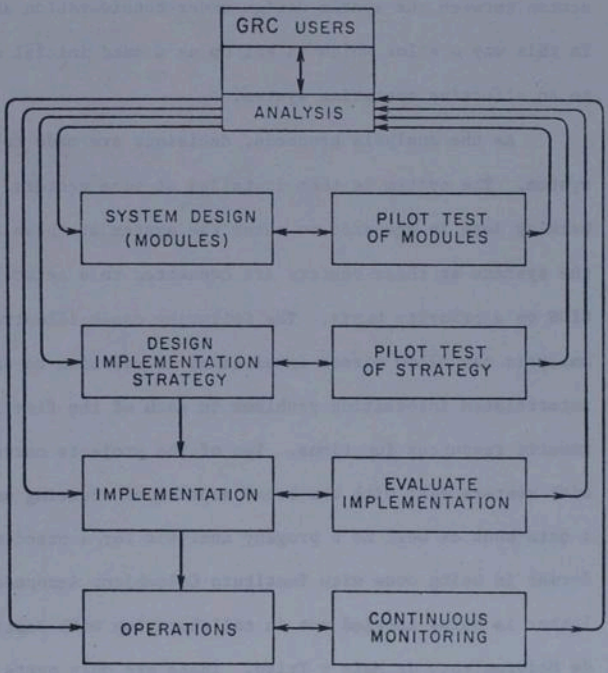

Figure 2. Flow diagram summarizing the process of the systems approach

The initial network provides a means of determining the types of system changes desired by the users collectively. The Genetic Resource workers and the GR/CIDS Project members are together responsible for developing a system to meet effectively the requirements of the users.

To test the system design a pilot project is set up. This determines whether the design changes meet the predicted necessary requirements. The system's design is continuously improved through testing the most suitable techniques and using them to the required capacity. Within the comparative testing there is a dynamic interaction between the system design under consideration and the user. In this way a Pilot which is set up as a good initial system expands to an effective operating system.

As the analysis proceeds, decisions are made to implement the system. The system is then installed at some centers. Current and backlog data is prepared to enter the system at these centers. And the systems at these centers are connected to a network of global CIDS on a priority basis. The following cases illustrate how systems analysis can allow direct involvement in addition to uncovering the interrelated information problems in each of the five identified genetic resources functions. Two of the projects currently underway with centers deal with the broad aspects of creating and evaluating a data bank as well as a progeny analysis for a breeding program. The former is being done with Instituto Colombiano Agropecuario while the latter is being carried out in collaboration with Centro Internacional de Mejoramiento de Maiz y Trigo. These are only parts of larger projects at both centers.

The ICA project specifically deals with maize collections in South America (especially the Andean variation). Data on these collections resides in three locations. Some is on machine readable computer tape. Most is in the journals of the leaders of the breeding program. And perhaps the most critical is in the minds of scientists and other field personnel. The immediate task is to work with the scientists to establish procedures for extracting this data and to organize it into computerized form. When this is completed, the data can be effectively evaluated, analyzed and updated. The resulting banks can then provide easily accessible information which is useful in breeding programs and future collections planning and implementation. The maize progeny testing at CIMMYT exhibits the type of multi-site international project in which the specific analysis is directed by the breeders. Interaction with the GR/CIDS Project takes place through the development and installation of an integral mini computer assisted system of CIMMYT. A combined team from CIMMYT and the GR/CIDS Project have already worked out a schedule to develop the data/information flow, computer methods, test the results, and evaluate the information based on selection criteria. The entire analysis is to be completed by the end of this year so that the international breeding program can effectively use the data for the next round of tests.

DEVELOPMENT

An important consideration within the systems approach is the investigation of computer assisted methods as they apply to genetic resources work. Preliminary results of pilot projects have indicated that the requirements for a computerized information system for genetic

resources work can be met by EXIS (Executive Information System).* EXIS is a set of software programs that can virtually operate on any large scale computer. It was designed to meet center specific problems as well as to function as a probing tool. By July, 1975, the programs will be operable on minicomputers. The process employed in implementing EXIS provides great flexibility for the user in working with the system.

The design of the computer programs assists the user in structuring working data banks, extracting subsets of the entire data bank, and in formulating the output in any manner specified to present information to others. Thus the system places the burden on the effective use of the computer as opposed to demanding highly specialized programming skills from the user. Specifically what the system allows is as follows. The user can input data in any quantity and form as long as the user carefully, clearly and completely describes the data. A query system is provided that assists the user in determining and extracting subsets of information. The user has simple access to additional computer programs for further processsing of information (statistical, clustering, mapping, graph drawing, simulation). The user is provided with the capability of designing output in standard forms, such as catalogs.

An important consideration is the system's anticipation of changes in user requirements and demands and its ability to locate or

---

*EXIS includes the EXIR Program. EXIR is the current name for the updated TAXIR computer program. TAXIR is the acronym for TAXonomic Information Retrieval, a general purpose, computer assisted information system developed at the Taximetrics Laboratory, Department of EPO Biology, University of Colorado, David J. Rogers, Director. The development of the initial system was supported by a grant from the National Science Foundation, Office of Scientific Information Services GN656.

develop more effective techniques to be applied specifically to genetic resources work. Changes suggested by breeders and scientists working with genetic resources are being introduced into the work program of the development group. Such modifications include linking with the programs of the Statistical Analysis System (SAS) developed at North Carolina State University. Ease of installation and implementation in genetic resources centers has been another important consideration. The first installations have already been implemented in Mexico and Colombia.

The ultimate success of the system depends on its integration into the real data needs of the centers staff. In this manner genetic resources workers will be able to work with their own methods in creating, documenting, accessing and exchanging information on real germplasm banks.

TRANSFER

The transfer function is designed to facilitate the analysis and development aspects of the GR/CIDS Project. One of its main roles is in the installation and implementation of computer assisted methods in those centers where the members of the GR/CIDS Project are working with scientists and breeders. Installation and implementation consist of more than the introduction of software packages. Included are short range orientation and training programs in which members of the GR/CIDS Project work directly with technicians and others at the centers, at workshops and at various sites. The preparation of proper documentation and materials for such workshops and working sessions is one of the aspects of the transfer function.

Another primary role is related to the preparation of data for those centers without the necessary computer capabilities. The GR/CIDS Project is able to offer data service bureau facilities to aid in the preliminary steps of data preparation and data analysis. These are crucial to the development of the network of CIDS. Data preparation assistance includes the translation of data into machine readable form and then into useable data banks so that breeders and scientists can analyze the information available to them.

Funds are currently being used for several pilot projects. These include the training of technicians and others in the methods of data preparation and the development of more effective computer analyses than are currently functioning at certain centers. All pilot projects work with data from national centers or international centers with gene banks of primary importance. The transfer function is able to facilitate the acquisition of data for the GR/CIDS Project members so that the investigation of such information can be implemented.

Various preliminary analyses are being done on data banks by the GR/CIDS Project members to facilitate the investigation of information that breeders and scientists require. Such analyses at present include descriptive statistical methods, associative analyses and more complex cluster analyses including measurements based on information theory. Standard data base analysis and maintenance procedures are being defined throughout the network which will facilitate the communication of information regarding the data to the users of such data banks. Shown below is an example of the banks presently under analysis.

DATA BANKS

CIAT (Mexico
CIATPHS
8656 Items
36 Descriptors
Phaseolus germplasm collection

CIMMYT (Mexico)
CIMMYT Catalog
10,000 Items
200 Descriptors
Maize germplasm collection

ICA (Colombia)
ICAZEA
3653 Items
42 Descriptors
ICA maize germplasm bank

INIA (Mexico)
INIAZ
96 Items
52 Descriptors
INIA Zea mays germplasm bank

PCIM (Peru)
PCIMZEA
6000 Items
64 Descriptors
Maize germplasm bank

IZMIR (Turkey)
IZMIR
2316 Items
43 Descriptors
Major section of Izmir
Regional germplasm bank

USDA-NSSL (United States)
200 Files of Representative Collections
100,000 Items
40 Descriptors

CONTINUING WORK

A set of coordinating activities have come from the workshops and the pilot projects. Individual scientists at several centers have agreed to participate and contribute to the GR/CIDS Project by serving on one or more groups.

The Data/Information Group is concentrating on the requirements for establishing data banks. The work includes all facets of information flow from the exploration/collection phase to the utilization phase of the genetic resources function.

The Network Investigation Group aids in the investigation of the information network. Procedures are being developed for the exchange of data between scientists working at centers throughout the world.

The Quantitative Models Group is organizing the types of analyses currently in use and also those needed in evaluating, testing and measuring the data in each of the five genetic resources functions.

As the information grows in each of these areas, it will contribute to the investigation of the network concept. This work and the knowledge gained from the computer programs installations and pilot projects will provide more information on the specific problems of the working scientists.

In summary the GR/CIDS Project offers a method for investigation and evaluation so that the need and means for an international communication network in Plant Genetic Resources will evolve. In no way are all the answers provided. What is offered are the necessary means for setting up the probe and for continuing from there.