1320105

KELLY, KATHLEEN

DESIGN OF A COMPREHENSIVE DATA BASE ON DESERT PLANTS

THE UNIVERSITY OF ARIZONA M.L.ARCH 1982

University
Microfilms
International 300 N. Zeeb Road, Ann Arbor, MI 48106

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark __√__.

1. Glossy photographs or pages _____
2. Colored illustrations, paper or print _____
3. Photographs with dark background _____
4. Illustrations are poor copy _____
5. Pages with black marks, not original copy _____
6. Print shows through as there is text on both sides of page _____
7. Indistinct, broken or small print on several pages __√__
8. Print exceeds margin requirements _____
9. Tightly bound copy with print lost in spine _____
10. Computer printout pages with indistinct print _____
11. Page(s) __74__ lacking when material received, and not available from school or author.
12. Page(s) _________ seem to be missing in numbering only as text follows.
13. Two pages numbered _________. Text follows.
14. Curling and wrinkled pages _____
15. Other ___________________________________

DESIGN OF A COMPREHENSIVE
DATA BASE ON DESERT PLANTS

by

Kathleen Kelly

A Thesis Submitted to the Faculty of the

LANDSCAPE RESOURCES DIVISION

SCHOOL OF RENEWABLE NATURAL RESOURCES

In Partial Fulfillment of the Requirements
For the Degree of

MASTER OF LANDSCAPE ARCHITECTURE

In the Graduate College

THE UNIVERSITY OF ARIZONA

1 9 8 2

STATEMENT BY AUTHOR

This thesis has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgement of their source is made. Requests for permission for extended quotations from or reproduction of this manuscript in whole or part may be granted by the copyright holder.

SIGNED: Kathleen Kelly

APPROVAL BY THESIS COMMITTEE

This thesis has been approved on the date shown below:

Warren D. Jones — 12/8/82
Warren D. Jones
Professor of Landscape Architecture
Date

Wallace H. Fuller — 12-8-82
Wallace H. Fuller
Professor of Soils, Water & Engineering
Date

Michael M. McCarthy — 12/8/82
Michael M. McCarthy
Professor of Landscape Architecture
Date

ACKNOWLEDGEMENTS

The two people who have been most helpful and influential in the work described here are not members of my Graduate Committee because they have no official affiliation with the University of Arizona. They are Peter Z. Ingerman, who appreciated my science fiction day dream approach to computer applications well enough to encourage some much needed discipline and to teach me much of what I know about computers today; and Marianna G. Holland, who has consistently and unflinchingly provided me with helpful criticism and perspective.

The official members of my Graduate Committee, Warren D.Jones, Wallace H. Fuller and Michael M.McCarthy, have had enough faith in me to insist that I go as far as necessary to make them understand the concept and motivation behind this work.

Thanks are also due to Rod Replogle, U.S.Forest Service, for introducing me to video discs, and to Katherine Alfone for reviewing a draft of this work.

TABLE OF CONTENTS

ABSTRACT

This work describes the motivation and ideal criteria for a comprehensive, interdisciplinary, computerized catalogue of desert plants, tentatively named PLANT. Recent attempts to develop plant data bases with some similarities to PLANT are recounted and the hypothesis that such a broad system is technologically feasible is satisfied, although the still exploratory field of Very Large Distributed Data Bases is judged likely to be the most effective way to establish PLANT.

The main body of the research described here is the compilation of a lexicon of the data elements to be included in PLANT and their functional specifications for the system. (Completed sections of the lexicon are included in the Appendix.) The lexicon, as developed, is judged to be extremely unwieldy in its drive to comprehensiveness and complete data independence. Further, its compilation raised significant questions about the classification and synthesis of data about plants from many different disciplines, which will be addressed in the next stage of research.

CHAPTER 1

INTRODUCTION

Work on the plant information system described in the following pages, and tentatively named PLANT, first began in 1977. It was motivated by the enormous difficulty of finding and organizing for future use information on plants which were native to or suitable for use in hot deserts in the Middle East, where the author was working as an environmental planner.

Statement of the Problem

Environmental planning acts as a bridge between the work of scientists such as ecologists, geologists, soil scientists and so forth, and the work of professional designers--planners, landscape architects, engineers and architects. Therefore, it deals with plants both as elements of and indicators of a pre-existing environment, and as tools, raw materials and ornaments in the managed environment which will be imposed on the earlier one.

As long as one trains and practices environmental planning in a fairly restricted geographic area, the fact of having to find and integrate so many different kinds of information on plants is not a big problem. Even when a question does come up under these circumstances, the planner is usually in touch with local experts who can supplement her (or his) own training and experience.

In the humid, temperate regions of the earth, people have studied plants and their relationships with both the natural and human environment for thousands of years. This information has been digested over and over again in a variety of forms and is generally available to any interested person. In arid regions, this has not been the case, especially in those regions which are being settled for the first time, or which have had a long tradition of nomadism. In many of these cases, there is not even a good list of the plants which grow there or which have become adapted. To the degree that studies of plants' relationships with their environments have taken place at all, they have not been extensive and are often locked up in ministry filing cabinets (Kelly & Schnadelbach 1976). Information about plants in deserts has certainly not been digested and integrated into widely available forms in the majority of cases, and the data required for good environmental planning, if it exists at all, is often only in the specialized languages and journals of many different disciplines (Holland 1979a).

When one is charged with assessing the current environment on a site slated for development, as well as the possibility for establishing some agriculture, industry, windbreaks and planning some sort of regional flood control, and compiling a list of suitable ornamental plants, the search for information ranges all the way from the original manuscripts of early travelers (often in Latin), and the lore of gardeners, to journals in range management, forestry, soil science, geology, hydrology and to individuals in these specialties.

Desert environments are not so resilient nor accomodating as humid, temperate environments are in the face of human development. They are difficult to deal with in the first place and are quick to react to mistakes. In the face of burgeoning development of human settlements in deserts around the world, a great deal more scientific inquiry into aspects of the desert environment, including plants, is taking place now than ever before. However, there is as yet no central depository of information on desert plants (Kelly & Schnadelbach 1976) to compare with those for humid, temperate (and even to a certain degree, tropical) plants. Nor has the digestion of information about desert plants, the integration of data from all the interested disciplines that we accept without question for temperate plants been accomplished.

Hypothesis

It is proposed that the vast strides recently made in the development, collection, storage and manipulation of information can be brought to bear to serve environmental and natural sciences and disciplines in a way which has never before been technically feasible, to facilitate human development of arid regions by providing a central depository of information on all aspects of desert plants, designed in such a way as to permit easy flow of information about them from one discipline to another.

Criteria

Ideally, the means for solution of the problem as stated, and of proving the Hypothesis, would have the following characteristics:

1. Comprehensiveness Across Disciplines - Every discipline with any interest in plants would be represented, including, for instance, landscape architecture, wildlife biology, horticulture, agronomy, plant pathology and genetics, soil science, ecology, ethnobotany, and others.
2. Comprehensiveness Within Disciplines - Within each discipline, all information concerning plants would be incorporated in full. Thus, for instance, this would not be a data base on soils, but it would include all the kinds of data available on the relations between soil and plants, such as the need for and tolerance of specific elements, of elements in combination, soil texture, and necessary or helpful microorganisms, etc.
3. Comprehensiveness Globally - The solution would have not geographic limitations whatever, neither with regard to information included in the system, nor access to data in the system.
4. Independence of Data - The data would be stored in such a way that their use was not limited by any structure imposed by the discipline from which they came, nor by anticipated uses of the data. This is a requirement of the ideal data base management system (See Chapter 3). It is unlikely that it will be achieved; nonetheless, it is an important criterion

against which to measure the relative succes of the solution. This criterion also requires that data not be aggregated. Thus, rather than just indicating that a plant is a potential allergen, the data base should also be able to specify what factors enter into such an estimation.

5. Ease of Access - In order to meet these criteria, the solution developed must be easy to use--technically, financially and geographically. A person's ability to use the system cannot be limited by the fact that she (or he) was trained as a range manager rather than as a landscape architect or computer programmer. The system should not be expensive to use, nor should anyone be denied access because of her (or his) geographic location.

6. Accomodation of Hard and Soft Information - The system would provide for the inclusion of both precise, verifiable information, as well as imprecise, general or as yet unverifiable information, thus accomodating both the need for and the availability of both kinds of information, in such a way that they cannot be confused.

7. Interdisciplinary Correlation - The system would permit easy correlation of information across disciplines, especially with regard to newly developed information.

8. Updating, Correction, Refinement, Etc. - The system would make these operations as painless as possible, as the system would be meaningless without them on a continuing basis.

9. Arid Land Plants - The system would be designed specifically

for use with plants which are native to or suitable for use in arid and desert regions, because that is where the need is greatest, and because of the author's background and needs. However, the work should be done in such a way that as much as possible of it can be applied to other groups of plants, or so that other groups of plants can be eventually integrated into this system.

10. Illustrations - The system would include illustrations of the plants and their parts and of any other aspects of information in the system that would assist in its effectiveness.

Possible Uses of PLANT

An important part of planning the proposed information system will consist of establishing the standard sorts of reports and lists and correlations to be built into the system. This phase of planning will require close coordination with specialists from many different fields. However, it may be instructive to see a very preliminary, illustrative list of the things the system might be expected to do:

1. Assemble lists of appropriate plants for specific land uses in specific environmental conditions, such as revegetation of eroded areas, mine wastes, construction areas, etc.
2. Select plants for experiments in horticulture, forestry, etc.
3. Maintain a catalogue of the holdings of botanical and herbarium collections, and of seed banks, providing information on where such plants can be obtained or observed.
4. Provide a *de facto* key to desert plants. (Even in the

one case in which a scholarly, systematic flora has been developed to dat outside of the United States, Flora Palaestina (Zohary 1966), no key has yet been developed.)

5. Help determine what the widespread presence of a plant on a site might mean in the ecological workings of the place.

6. Explore possible connections between plants, and plant families, in plant pathology for the purpose of controlling pests and diseases.

7. Help select plants for harvesting or conversion of toxic wastes and toxic soil constituents.

8. Identify exogenous species that might be useful for some specific purpose, by the use of ecological analogues.

9. Evaluate the likely ecology of imports vis a vis allergies, pests, diseases, tolerance of local conditions, etc.

10. Aid in vegetational aspects of all kinds of land use change, through ecological analogues, cultural and ecological information, etc.

11. Aid in the design and management of wildlife habitats.

12. Aid in identification of plants which may provide important new medicines, dyes, etc.

Approach

It has been assumed from the beginning that a computer based information system would be necessary to deal with the problem stated above. Beyond that, an adequate response to the hypothesis of this work and to the stated criteria required three basic tasks. The first was an investigation into work that had already been done in establishing interdisciplinary, computerized data bases on plants. This investigation was singularly unfruitful until 1979, when a personal contact who knew of this work provided a two page description of AVIS (See Chapter 2.). One important system, developed by USDA was not even "discovered" until after the first draft of this work had been completed. There seems to be no way of knowing for sure that the investigation of others' work in this field is complete.

The second task has been an exploration of some of the technical requirements of the information system that would satisfy the hypothesis and the criteria for its proof. This has involved inquiry into the nature and progress of data base management systems (DBMS's), and recent advances in computer hardware, as well as video disc technology. These are all rapidly advancing fields, and advances in them tend to be made by commercial enterprises. They are rarely discussed in scholarly publications at a level which is useful to the development of specific applications. Consequently, this portion of the inquiry relied heavily on discussions with acknowledged experts in these areas, and on the more informal and user oriented sorts of periodicals which have become characteristic of these fields.

The third task began as a simple list of what should be in the data base. Such a list is an essential part of the data dictionary required for a DBMS. Although making machines do new tricks is always fun, the assembly of this lexicon, as it is known, tedious in and of itself, became the most interesting part of this work because of the other questions it raised. These are far reaching and cannot be done justice here, but a suggestion for exploring them is one of the major conclusions of this work.

Common to all three tasks has been the principle of trying to determine the characteristics of the ideal system, without regard to financial, temporal or political considerations. The reason behind this very conscious decision is that with any complex problem, there is generally a range of possible solutions to it rather than just one. The range of possible solutions in a case like this will vary in their cost, difficulty and usefulness. It is not possible to weigh the various trade-offs that are represented in such a range of possible solutions without knowing the characteristics of the full range. A search for a solution which eliminates possibilities at the "top" of the range because of assumptions about cost, for instance, makes the job easier, but may restrict it unnecessarily. Knowledge of the characteristics of the ideal solution, on the other hand, permits a well informed paring down to a feasible solution at the top of the range of possible solutions.

CHAPTER 2

PREVIOUS WORK ON COMPUTERIZED PLANT DATA BASES

For a variety of reason, information on work others are doing on computerized plant data bases is very difficult to come by. When the people involved in such work do publish in scholarly journals, which is rarely, they most often write on aspects of the data itself, without much discussion of their use of computer sciences. Computer searches of their titles and abstracts, therefore, seldom index such articles under such key words as "data base" or "computers" or information system." As a result of the same circumstances, when one does find out that a certain system exists, most often the details of the computer and information science side of the system are not available in the literature and must be gleaned from interviews and correspondence with the people involved.

In a few cases, geographic and technical access to the system were so restricted as to lead to a suspicion that information was being hoarded, a despicable but common practice, which it is one of PLANT's primary goals to overcome. Wide knowledge of the existence of these systems tends to be similarly restricted.

As a result, to the degree that what follows is at all complete in covering previous work in this area, it is due to serendipity, and not the extensive scholarly attempts to research the subject fully.

David Pitt (1982) has just completed an excellent survey of microcomputer applications in landscape architecture, in the course of which he was able to gather information on several very interesting plant selection programs, a few of which share some of the criteria of PLANT. Although none of the systems Pitt surveyed is on the scale of the system envisioned here, his article will be a much needed contribution to the literature of a sort which has been previously available.

Of the systems about which it was possible to find some information prior to Pitt's survey, most are data bases. One is a data base management system specifically designed to deal with plants. Of the data bases, only one is believed to be healthy and expanding. The others have all either been disbanded, stagnated in their original demonstration form, or are about to be disbanded.

This in itself is a dismal finding and the author scanned Index Medicus (an index to publications in medicine and related fields) for the last two years to see if this phenomenon was a function of the subject matter or of some other factor, such as inadequate technology. Index Medicus revealed extensive activity in both the compilation of data bases and in data base management with regard to medical care and to human physiologic systems.

It is not within the scope of this work to explore why there is such a striking difference between medicine's adoption of the latest advances in computer and information sciences, and the apparent lag in adoption of these advances in landscape architecture and related fields. Nonetheless, one cannot help but speculate as to whether the reasons have to do with national priorities which may have affected funding for such efforts, resistance on the part of the landscape architectural profession to the use of computers, a generally low priority on systematizing environmental information, or some other reason.

A discussion of the five systems about which some detailed information was available follows:

Plant Science Data Center

This system is little more than a catalogue of the collections of botanical gardens in the United States (Holland 1979a). It is useful only as an inventory technique for the participating institutions and to find where a certain species may be observed, occasionally where seeds or cuttings may be obtained. It is a valuable data base, but does not incorporate a DBMS. It is the one system that appears to be definitely extant at this writing.

PIN

PIN was established early in the 1970's, primarily for assisting in land reclamation, as an activity of the Western Energy and Land Use Team of the Department of Interior's Fish and Wildlife Service, now headed by Philip Dittberner. The data base in PIN originally covered the plant species native to Montana, Wyoming and Colorado. Later Utah and North Dakota species were added. It included taxonomic, geographic, biologic and ecologic information on 4000 species with more detailed information on 600 species which were thought to be especially important to land reclamation following mining activities.

PIN included a DBMS which included regular updating and some correlation of the factors listed in the data base. The capacity for remote terminals was built into the system, however, only one remote terminal was ever installed. The system was used quite a lot, according to Dittberner (1982), but is expected to cease operation as of October 1, 1982 as a result of budget cuts. If this happens, the PIN files will be put on hard copy so that manual searches will still be possible. However, no provision has been made for updating the files while they are on paper.

AVIS

The Arid Vegetation Information System (AVIS) was developed by Marianna Holland (1979a, b, c) at the University of Arizona and the Boyce Thompson Arboretum. The original goals for AVIS are the most similar to the goals for PLANT of any system discussed here. AVIS was specifically designed to overcome the very dispersed nature of information on desert plants and to enable plant information to cross disciplines easily and promptly. It called for technical and financial ease of access, convenient updating and expansion and was to have been international in scope.

The criteria Holland established for AVIS also included cost effectiveness and the ability to implement the system on a pilot basis within a relatively short period of time while funding was available. This had a profound effect on the system's ability to meet its original criteria. Holland took the position that once AVIS had been implemented on a pilot basis, it would be possible to expand the catalogue over time to include not only more plants than the original 50 Arizona natives, but also more scope. It is not clear that such expanded scope would ever have been as broad as PLANT's, but it would certainly have overcome the limitations that now restrict access to AVIS and its ability to be used for work outside the United States.

The most basic differences between PLANT and the original plans for AVIS, on the basis of several long conversations with Holland, are philosphical. For instance, taxonomic information was specifically excluded from AVIS "because the scientists dealing with this information are very aware of its location,...and this information is not often needed by other users." (Holland 1979a,p.13) That would be considered an unacceptable limitation for PLANT because of the criteria for comprehensiveness within and across disciplines, because the use of PLANT as a _de facto_ key, which would require taxonomic information, has already been planned, and because taxonomists ought to have access to other perspectives on plants as well as _vice versa_.

In general, Holland notes, "Information types were included or excluded on the basis of their need by a broad range of users...." (op.cit.p.13) It is indeed critical to the success of AVIS, or any other interdisciplinary system, that people from all relevant disciplines be consulted as to what information they would like to have in the system and what they would like to get out of the system. The difficulty is in trying to predict the ways information may be used by someone outside one of those disciplines at some time in the future. The fact is that it's not predictable, and PLANT is being designed to facilitate cross fertilization among disciplines, even if it is not possible to say right now that some specific instance of it will take place.

Finally, AVIS, like PIN and AEGIS (see below) to a certain degree, was designed in part on the basis of the kinds of information on desert plants which were readily available. PLANT, on the other hand, is designed much like a library for which most of the books have not only not yet been bought, but not even written. In the case of desert plants, this analogy is very close to the truth as they have not had nearly the amount of study that temperate plants have had over the centuries.

As has been noted above, these apparent limitations on AVIS were not a function of limited vision on the part of its author, but of the decision to try to implement the system on a pilot basis. (Therein lies a major limitation of PLANT.) However, as of this writing, no support for expansion of AVIS beyond its demonstration stage has been found

AEGIS *

AEGIS was begun in approximately 1972 under the direction of James Duke at the Economic Botany Laboratory of the USDA. On the basis of discussions with Dr. Duke, Allen Atchley, who currently serves on his staff, and Weldon Lodwick, who was for several years their consultant on computer aspects of the system, the original comception of AEGIS appears to have been similar to that for PLANT in the sense that it was to be a sink for information from many different fields about plants. The original impetus for the system was the desire of the US government to assist foreign governments in finding substitute crops for the opiates which the US had requested them to stop producing. For this purpose, files were accumulated on ecological tolerances and cultural practices for more than 1000 agricultural plants. These files included 30 ecological variables and 18 variables concerned with yield, as well as an unknown number dealing with cultural practices.Other files came into the possession of the AEGIS staff, including a file on forestry with 40 additional variables ranging from cultural practices to caloric content, and a folk medicine file with some 88,000 entries.

AEGIS was housed on a mainframe computer and was specifically international in scope, but access to the system was always through

---

* This discussion is based almost entirely on conversations with Drs. Duke, Atchley and Lodwick, according to whom there is no published information on the computer or information science aspects of AEGIS.

Dr. Duke's staff. Furthermore, it appears that AEGIS was never much more than a depository of information. A search of the files to answer a specific question was analogous to literally sorting through a giant chart. AEGIS did not incorporate a data base management system and, as a result, searching, using the accumulated information for any useful or imaginative purpose was extremely time consuming and expensive, both in computer and person hours. In discussing how expensive the system was to run, Dr.Atchley said that it cost $50,000 a year to "store the information on-line," a cost which could have been reduced by the use of a data base management system and a combination of batch and interactive (on-line) processing. These preliminary conclusions were confirmed by Dr. Lodwick.

The importance of AEGIS in this context has several aspects. For one thing, it is a great comfort to find a group of people with so much training and experience in scientific pursuits who agree on the desirability of such a comprehensive data base on plants. Second, the experience with AEGIS is a powerful confirmation of the vital necessity of using a DBMS in PLANT, and of selecting the characteristics of that DBMS very carefully. Third, it is very exciting to know that these vast files of basic information exist, and their existence has prompted more serious consideration of the possibility of implementing PLANT as a distributed data base. (See Chapter 4.)

However, AEGIS also highlights a problem which is not dealt with directly in this work, but which keeps cropping up again and again, and which will have an important bearing on the eventual feasibility of PLANT or any similar system. Access to AEGIS was only through the staff of the Economic Botany Laboratory. When questioned about this practice, Dr. Atchley expressed reservations about other people's ability to "interpret" the results of any given search adequately. This very secretive approach to information was one of the early motivations for undertaking the development of PLANT, and may in the end be the greatest impediment to its implementation in any form.

AEGIS has been nearly totally dismantled because of budget cuts (Duke 1981). Some of its files have been sent to other agencies and organizations which have the funds to keep them up to date.

CAPS

In contrast to the other systems discussed here, CAPS (Computer-Assisted Plant Selection) was designed solely for use by landscape architects in selecting plants. However, by stretching the Apple II microcomputer on which CAPS is housed to its limits, the CAPS system has met several of the criteria for a data base management system, especially with regard to easy access. Almost anyone who can find the power switch can use CAPS at a first sitting to draw up a list of plants which will meet given criteria. With only a little experience, CAPS can be adapted to include other plants than those in the original data base and, to some extent, other plant characteristics. In addition, CAPS is completely transportable as both the data base management system and the data base exist on an independent medium, discettes, which can be used with any Apple II anyplace in the world.

CAPS also, in excellent cooperation with the Apple II's operating system, does a good job of data bookkeeping of the sort that is required to update the data base, generate new reports and so forth. As a further aide to wide accessibility, it permits the user to interact directly with the computer in English or English-based BASIC. (Johnson & Harsh 1981)

CHAPTER 3

MANAGEMENT OF THE DATA BASE

In the broadest sense, any body of data is a data base, and every data base is managed by some system, even if that system consists of nothing more than just leaving it alone. The card catalogue of a library is a data base (as is the body of information in the books represented by the catalogue entries). It is managed by a fairly sophisticated system which incorporates the three key characteristics which define a data base management system (DBMS):

- Data Independence
- Data Bookkeeping
- Accessibility

Data Independence

Each book in the library is represented in the card catalogue by at least two entries, one for the author of the book and one for the title of the book. In most cases, books are also listed under a particular subject thought by the librarians to be of interest to users of the library. But it is not necessary for a person who wants to find a certain book to divine which subject it is indexed by, because the basic data (the author and title) are also listed independently in the catalogue. These pieces of information have not been aggregated in a way which prevents the user of the catalogue from having direct access to them.

In contrast, books are not listed in the catalogue by publisher or date of publication. This information is available in the catalogue, but only if the user already knows the author or title of the book.

Although some people, when exploring a subject, go directly to the subject index of a card catalogue and may go no further, many others are grateful to have direct access to the basic data in the catalogue; that is, they are grateful for the data independence expressed in the design of the catalogue. This data independence gives them the freedom to develop bibliographies, or indices, which address their interests precisely, and to engage in scholarly detective work which can make use of, but is not dependent on, the librarians' notions as to which subjects to index and which books fall into those categories.

As these people have discovered, the use of indices is often a time saver, but has some disadvantages. The first of these is the selection of which things to index by, a choice which usually must be made in order to avoid the very inefficient chore of indexing by every subject or characteristic. Does one index by the things which are seldom indexed elsewhere, the things one is least likely to remember offhand, or by the things likely to be requested most often? Any one of these decisions represents trade-offs. Something is lost in each case that may be reflected later in the way the data base is applied to actual problems.

Use of a selective set of indices is limiting in this sense, although it is usually the most efficient way to proceed. However, a set of indices which aggregates data and stands in replacement for the basic data it refers to literally loses a large amount of information and totally violates the concept of data independence. Instead the indices must complement the basic data.

Data Bookkeeping

Apart from selecting the books to be purchased for the library, the most critical function of a librarian is data bookkeeping: creating and maintaining the card catalogue. New books must be assigned a "number" (which encodes information about the subject and author of the book), and entries must be made in the card catalogue bearing that number under the book's title, author and subject, if an index to that subject exists. When books are removed from the library's collection, or reassigned to a special collection or different subject, all the cards referring to that book must be located and changed to reflect the new location, classification or the book's removal from the library's shelves In order to do this, there must be a complete list of the subjects in the subject file (an index to the indices) and a list of all the subjects under which each book is listed. These lists in turn must be kept scrupulously up to date as part of the very basic task of insuring that all the data on each card in the catalogue is correct.

All these chores collectively are known as data bookkeeping. They tend to be tedious and repetitious for the most part, but are absolutely vital to the usefulness of any body of data. A data base which has no provision for these tasks cannot be said to have a data base management system in the modern sense.

Accessibility

Accessibility in data base management systems refers to aspects of the ways users interact with the systems. First and foremost among these is the requirement that a user need not be a specialist in the system itself in order to use the system. For instance, one need not be a professional librarian in order to use the card catalogue. It takes only a few minutes orientation to enable a new user to look a book up in the catalogue and derive enough information from the entry there to go and find the book. Of course with more training and with experience, a person can stretch the catalogue to its limits deriving all sorts of information from it, often more information that the librarians themselves are aware is contained in the catalogue.

A second aspect of accessibility in data base management systems forbids exclusivity on the basis of the content of the data base (as opposed to its management system). A person with a job as a clerk in the geology department is not restricted to the parts of the catalogue that refer to clerking and geology, no matter how defined. That person may just as easily use the catalogue to investigate political science or evolution or the works of Doris Lessing (and to find important connections among them that the original indexers of this hypothetical collection never dreamed of).

Ideally, every person should be able to have access to every data base. A person studying at the University of Arizona should, ideally, have access to the data bases at the University of Moscow, or at General Motors. Clearly this is not the case and a long list of reasons for limiting access to data has grown, sometimes it appears even faster than the body of data generally has grown. This aspect of accessibility remains, nonetheless, an ideal toward which data base management systems aim.

History of Computerized Data Base Management Systems

From the point of view of a user or data manager (librarian) of a computerized DBMS, new information is very easy to accomodate in the system, because the bookkeeping of adding new information every place it belongs, deleting obsolete information wherever it exists, etc. is handled by the DBMS. A computerized DBMS incorporates programs to produce the sorts of reports or lists users are likely to need from the data base, but should a new kind of report be needed, the DBMS can produce it fairly easily with the appropriate instructions (programs) because the data are stored in a way which is independent of the reports the system was originally envisioned to produce. It is not necessary, therefore, to change the way the data are stored to produce the new report or list. Furthermore, by using a computer in combination with a DBMS, it is possible to include and provide access a great deal more information, and more detailed information, than is possible otherwise.

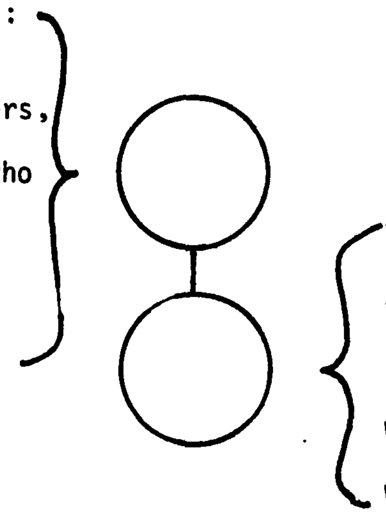
In fact, computerized data base management systems make so much sense logically that is difficult to appreciate why they are new, why files, whether manual or machine based, have not always been arranged with the data independence required to permit many new combinations or data and to avoid extensive repetition of data. Therefore, it is worthwhile to recap the main elements in the history of computer sciences from which DBMS's evolved.

Computers were originally invented to compute, not to store and process large amounts of data. They were built with a small amount of random access memory (RAM), what used to be called "core" memory, within the machine, sufficient only to carry on computations and the programs by which they were accomplished. (The original UNIVAC had only 1000 words of RAM.)

Access in RAM is truly random. The time is takes to get to any location in RAM is completely independent of the last location to which the machine was directed. And programming for early computers was oriented toward using this small amount of RAM as efficiently as possible, which meant, among other things, for as short a time as possible. Given the small amount of RAM and the relative slowness of early computers, infinitesimal fractions of seconds took on very large proportions, and the programmers (the only people in actual communication with the computer) found ways in all aspects of their work to shave millionths of seconds off each operation.

A model of the computer environment at that time would have had just two main parts:

The users and programmers, basically synonymous, who were in intimate communication with the computer; and



The computer, in which any data being used were physically stored, as well as the programs which operated on them.

In time, the use of computers for processing large amounts of data became a high priority, and advances were made in both the speed with which the machines operated and the way they operated. Aside from the greater speed, the most significant of these advances for DBMS's was the development of the concept of virtual memory and denser memory devices.

Virtual memory refers to using the still relatively small amount of RAM on a time-shared basis. The computer's operating system keeps track of the places in RAM that are not in active use or permanently committed (by a user or the operating system itself), and assigns these places to data or instructions as required. When these places in RAM (identified by "address spaces") are no longer in use, the computer's operating system frees them up for other uses by copying whatever is there to some auxiliary storage outside of RAM. Peter Ingerman (1981) calls this development in computer systems an illustration of the Law of Conservation of Agony. The incredible agony of keeping track of what space is available for input of material from an auxiliary medium such as a tape or disc, and moving it out of the way when it is no longer needed, has not been diminished, but only shifted from the programmer to the computer's internal operating system.

While data are in RAM, they can be processed very rapidly and randomly. However, while data are on an auxiliary medium this is not the case. This is particularly true of magnetic tape which was the most widely used auxiliary memory until fairly recently. Data on magnetic tape are stored in series. One cannot get to a record in the middle of the tape without going through one half the tape. More precisely, it takes almost no time at all to go from one record on a tape to the one physically next to it, but the length of time it takes to get from one randomly selected record to another is a function of their respective locations. As a result, even thought RAM could be used more effectively in this scheme, time was still an important factor, with respect to getting to the appropriate record and overcoming the inherent clumsiness of most serially arranged files in gaining access to information out of sequence.

At first, programs for computer systems using virtual memory and auxiliary memory just performed such tasks as determining when one had arrived at the right record, identifying the contents of the record and determining whether or not it was in the right sequence. Actual manipulation of the files on magnetic tape was a horrendous job. Consequently, although much work went on to ease it, data bookkeeping tasks such as updating were not easy to perform, and the idea of rearranging the data to produce new types of reports was frowned on to say the least.

Another major step in the evolution of DBMS's was the development of the computer language COBOL, which was specifically designed so that its use was not dependent on the use of any particular make or model of computer. As a result, for the first time, programmers could build on each other's work in ways and with an ease that were never possible while they were segregated by the use of different systems. They started using COBOL to structure and manipulate files, and developed indexing schemes to get access to the files faster.

Programmers were still, however, the only people who were able to communicate effectively with computers, and that was generally their full time job. Programmers controlled the way records were stored internally in the computer and, to a very great degree, the way that information was available externally as well.

Nonetheless, the development of COBOL, even though it eased communication among programmers without yet permitting the eventual users of the data to enter the dialogue, was important in two respects. First, it was the first hint of an element of a computer based system that dwelt neither in an individual computer nor in an individual computer programmer. Second, by facillitating communication among programmers, a lot of note swapping was able to take place, very informally at first, but gradually in a very structured way through various associations and journals.

This was one clear case in which there was no such thing as too many cooks. As a result of the note swapping, there became available a library of ways to program for various data processing tasks and it became possible to pay more attention to the needs of users (the people who used the statistics and reports produced by the computer systems, but did not generally have direct access to the machines) without using the computers inefficiently.

Meanwhile, a third important step in the evolution of DBMS's took place. That was the development of an improved rotating memory with which access time to a specific record was still a function of the distance between it and the last record read, but could be no more than some maximum, very small period of time--the time it takes a disc, for instance, to make one complete revolution. This is not random access memory as we find internally in a computer, but it is so fast that the difference between access time in RAM and the access time with the use of an improved rotating memory device is insignificant for most purposes.

Recapping, the use of RAM was made more efficient by not relying on it for the permanent storage of data or programs. The substitution of an improved rotating memory for the inherently serial memory of magnetic tape as an auxiliary storage medium made it possible to put data into RAM much more selectively and much faster (out of sequence), while a library of programs for doing data bookkeeping was being built up.

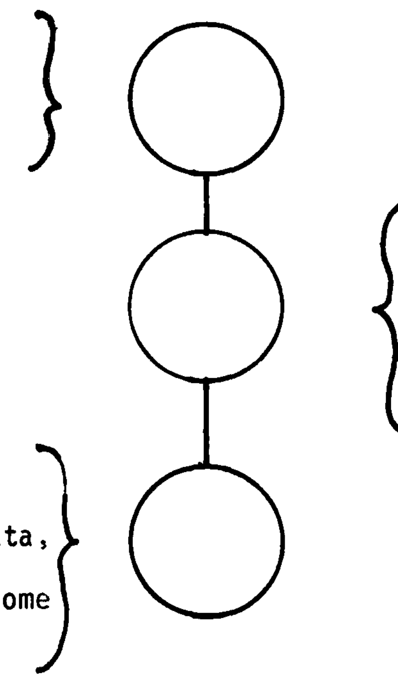
Simultaneously, there have been nearly continuous advances in computer hardware which have made it less expensive to buy and operate computers. As a result of that, and the growing influence of computers in our lives, more and more people became familiar enough with computers to imagine new applications for them without necessarily being able to program them. (The influence of science fiction in this regard cannot be discounted.)

It was inevitable that these developments would stimulate the imagination of both computer programmers, from the point of view of making the systems do new tricks, and others, who never got closer to a computer than filling out an IBM card, from the point of view of making their work easier and better. One result of all this is modern computerized data base management systems.

The heart of a DBMS is a body of computer programs designed to permit communication between people who have data or need reports, and a computer which neither knows nor cares about either. The DBMS checks that new data are legal (e.g. "purple" is not legal as a description of tree height), puts data in the form required for physical storage in the computer, manipulates them in whatever ways are necessary for producing reports, and so forth. Having done this bookeeping, the DBMS can then report what data are available and produce reports on the contents of the data base on demand, retranslating the data from the terms in which they are stored to the terms required by the user.

The definition of a modern data base management system requires that it be what is called "user-friendly" or a "friendly system" in order to meet the criterion of accessibility. Therefore, the two element model of the past is replaced by a model of the computer environment with three elements:

The users, who may be managers, researchers, librarians, etc.;



The DBMS which is composed of people, computer programs, and machinery; and

The computer and the physical storage of data, whether in RAM or in some auxiliary device.

In this model, users who are not programmers are acknowledged as part of the system, with the DBMS acting as an intermediary between them and the physical storage of data. The computer does what it does best, and the people who use the system do what they do best. Neither the computer nor the people who use it have to contort their "intelligence" to be able to work together, nor do they have to become fluent in each other's language.

It is important to recognize that people are also an important part of a DBMS. They are responsible for keeping the system in good working order, controlling updating, writing new programs when they are necessary, and helping the users getting the maximum potential out of the system. Some of the recent, commercially available DBMS's include such powerful and comprehensive programs that the users themselves can take on many if not all of these functions with relatively little training. Therefore, a DBMS allows many more people to use the same data base effectively, whether through a person in the system or directly. And the data base can be much broader and more detailed than would have been possible without such a system.

The Data Budget

We have been so indoctrinated by experience and training over the years to be frugal in our use of computer time and memory, and consequently so structured in the way we think about data, that it is difficult now to think about large amounts of data in what might be called a spendthrift way, even at the expense of scientific interest and flexibility. As a result, many data bases established over the last few years have not taken full advantage of the advances in data base management systems. Some function like giant card sorters, in effect, building a new index each time a question is asked. Most impose structures on the data base left over from the days when saving a millionth of a second was more cost effective than data independence.

But a few moments reflection on the possibilities that open up when one tries to think about data like a millionaire will show that they are endless and potentially very important. For instance, it is no longer necessary to automatically eliminate data that we know will be used very infrequently, or even data that we cannot predict a specific use for.

While it is still only reasonable to get the best possible idea of the way a body of data will be used and the sorts of combinations of the data that will be required before the system is established, there can be a very large number of these, and that number can be added to, relatively easily, after the system is established--if it has the characteristics of a true data base management system, especially data independence.

Of course, even a millionaire has a budget of some sort, a large one perhaps, but a budget nonetheless. And the same is true of the possibilities for data bases with DBMS's. The budget is just very large.

CHAPTER 4

ENVIRONMENT OF THE SYSTEM

Many of the criteria for PLANT must be addressed in the mechanical and spatial arrangements through which the system will be implemented. These are known collectively as the environment of the system.

Hardware

The most basic element of a system's environment is the kind of computer hardware in which the system is housed. When work on PLANT first began, there were three basic, and generally mutually exclusive, choices in that regard:

- Small systems specifically designed for home and small business use. The TRS-80 and early Apples are typical of this group, known as microcomputers here. *

- Medium sized systems, such as the Cromemco, which are designed to function as dedicated computers for business and research, and which now also provide an interface with mainframe computers. These are known here as minicomputers. *

- Mainframe computers, the biggest both in size and capacity, and the direct, if distant descendant of the early UNIVAC and its ilk.

---

* The terms "minicomputer" and "microcomputer" are not well standardized. As used in this work, "mini" will always refer to the medium size computer and "micro" to the small, so-called personal computer, in accordance with the usual order of these prefixes.

The three types of computers are distinguished by increasing random access memory (RAM) and by their ability to use higher capacity auxiliary memory (in the order listed), and by the operating systems they use. In general, minicomputers and mainframes both use an operating system known as CP/M, while microcomputers use a variety of others depending on the manufacturer and model.

These distinctions have blurred considerably in the last two years. A few microcomputers have been developed with much more RAM and capacity to use auxiliary memory than ever before. At least one can be adapted to interface effectively with larger computers using CP/M. Nearly all good, recent microcomputers can communicate with others of their kind by telephone lines, but are still very limited in the amount of user traffic they can carry at one time.

The minicomputers now available can perform much of what it took a mainframe to do ten years ago, and most can be expanded beyond their basic configuration to serve several users at one time, even if they require the use of different languages and programs.

Mainframe computers, on the other hand, are being configured to reach down to smaller users through time sharing. These very large computers still, however, hold a monopoly on the ability to serve a very large number of users requiring many different functions at one time.

Another distinction between mainframe and small computers which used to prevail was that mainframes were used, for the most part, indirectly in batch mode, while the smaller computers could all be used interactively. (The difference between batch and interactive modes is analogous to the difference between writing a letter and receiving a written reply, on the one hand, and having a telephone conversation, on the other.) AVIS was housed on a mainframe and was set up for batch processing exclusively. Batch processing limits access to a data base by requiring either detailed knowledge of the system under which the data base is stored or an access to an intermediary who has such knowledge, or both. Holland considered the possibility of using a combination of batch and interactive modes for access to AVIS (1979a, p. 40) to ease access, but at that time, interactive processing was relatively expensive. In the time since then, the difference in cost between batch and interactive processing on mainframe computers equipped to handle both has been eliminated for many purposes (largely because of increases in the cost of paper, but also because of improvements in the efficiency of interactive systems and their interfaces with batch systems). Even cost conscious organizations such as the University of Arizona Computer Center will soon be switching to a hybrid batch/interactive mode for many uses previously restricted to batch mode.

CAPS operates only in the interactive mode characteristic of the Apple II, and of microcomputers generally. Although CAPS appears to have stretched the Apple II to very near its limits, performing an array of data management chores few thought possible on such a small computer, anyone who obtains the relatively inexpensive discs holding the CAPS programs and data base and can get access to an Apple II can have access to CAPS. Unfortunately, an Apple I will not do, nor a TRS-80, as those computers have different operating systems.

CAPS would also be limited in its ability to be expanded to a system with the scope of AVIS or PLANT because of the limitations of the Apple II in processing very large amounts of data. The number of data management chores and the number of plants envisioned by either AVIS or PLANT demand a greater capacity than any microcomputer now on the market. But the criteria for technical and geographic ease of access to PLANT require a system in which direct interaction with the computer is possible.

The ideal hardware environment for PLANT will take full advantage of the best that each kind of computer has to offer, as well as of the blurring of distinctions between them. At this writing, the best choice for housing the data base itself and the DBMS is probably a mainframe computer. This is so primarily because of the mainframe's capacity to accomodate many users of the system at one time. It would also permit the use of a commercially available DBMS on a time share basis (see Software following). However, the cost of storage of the data base on a shared mainframe could present a problem after the system is well established.

Access to PLANT would ideally be through any sort of terminal that can communicate with a mainframe. A "smart" terminal, such as a minicomputer would be used for establishing the data base and possibly the DBMS or adaptations to it. All material to be transmitted to the main system would be assembled and edited on a minicomputer in the interactive mode and then sent to the mainframe in the batch mode. Such a hybrid batch/interactive system avoids the costs, delays and complexity of interacting with the whole system while merely performing data bookkeeping tasks, as was the case with AEGIS.

The actual users of PLANT may reach it through smart terminals or "dumb" terminals, which can read the material in computer files, but very little else. (The GEAC terminals in the University of Arizona Libraries are dumb terminals.) Dumb terminals are very inexpensive relative to minicomputers, and therefore easy for small organizations and individuals to buy, or for large organizations to buy many of, thus increasing the ease of access to PLANT.

Access

There should be two means of access to PLANT: by subscription and by individual request. In the first case, an individual, company or other organization would subscribe, just as to a magazine, to what may be called a "room" of the library of information available in the data base. Each subscriber would have a unique code, possibly but not necessarily a number, which would be keyed to the body of data the subscriber is authorized to use (by virtue of her (or his) needs and an appropriate fee). The subscriber code would also indicate to the system whether or not she (or he) was authorized to change or add to the body of information.

Thus, a landscape architect who specializes in deserts would specify that in advance and have unlimited access to the room of data she (or he) expected to need for that purpose. A landscape architect who performs environmental assessments as well as suggesting plants for various sorts of landscapes would require a different, perhaps larger, part of the data for her (or his) work. A taxonomist searching for the identification of an unfamiliar species, or testing a new scheme for naming some troublesome family of plants would require access to yet another partition of the data base. There would probably be some overlap among all three partitions.

A large organization incorporating many different disciplines, such as the College of Agriculture, would require access to a very large body of data, possibly both as a collection of small rooms and as one very large room. This sort of subscriber would probably have more than one access code to permit the most cost effective use of the system and to permit some users to add information to the data base.

It is likely that subscribers will occassionally come up with questions that cannot be answered within their subscription rooms, and that others who are not subscribers will also want to consult PLANT from time to time. Therefore, provision will also be made to fulfill these individual requests through people who are specialists in PLANT. (These special requests will also serve as a source of feedback on needs the subscription service is not meeting.)

It is hoped that this configuration for access to PLANT will insure that everyone from a school child to the world's botanical gardens will have as good access to PLANT as they have to their telecommunication systems and postal service. It also helps to insure that the only obstacle to use of the system by individuals and organizations around the world is language as, except for local plant names, the system will speak in English.

Software

Ways to obtain the required DBMS software fall into three broad, overlapping categories. The quickest to implement and the least expensive in the short term is to use a commercially available DBMS on a time sharing basis. Most such systems have been developed for business applications, a few for museum and library collections, and several of them are very good for these purposes. They are, for the most part, hierarchical in structure. That is, the data are organized along mutually exclusive "branches" of a "tree" which has as it's "owner" or "trunk" some unique identifier, such as a plant name.

A hierarchical system is not ideal for PLANT and would require extensive manipulation to avoid imposing an inappropriate structure on the data base. A good time sharing DBMS has the advantage, however, that advances in both hardware and software must be incorporated into the system promptly and with minimum disruption to time sharing customers for competitive reasons. In a field of such rapid change, this would be a considerable advantage.

The second broad choice is the outright purchase of an already developed set of programs. These are available in a wide variety of types, including network and relational DBMS's, which attempt to overcome some of the rigid structure of hierarchical ones. However, they are very expensive at the outset. In addition, a buyer is not so well assured that other buyers' reactions to the system will ever be incorporated as improvements. These systems also vary widely in the training, debugging and service available after purchase.

Neither a time shared DBMS nor a purchased one would suit the purposes of PLANT precisely, and some local adaptation would be required in either case. The degree to which these available systems can be changed or adapted varies considerably and would have to be one of the major consideration in selecting a system from either category.

The third choice of DBMS's avoids this problem, but at great cost, by developing a whole new system specifically for PLANT. Much though it may sound like it, this would not require reinventing the wheel. It is, in fact, the most attractive alternative in the respect that the fewest compromises in the arrangement of the data would have to be made. To illustrate, the use of an already existing DBMS would be analogous to trying to turn an extremely attractive dormitory building into luxury apartments, possible, but difficult and full of compromises which might endanger the original purpose of the enterprise. Whereas, a brand new DBMS designed specifically for PLANT would not have the advantages analogous to an old, established landscape, and a structure which, if not perfectly suitable, had finished settling years ago, and proven itself sound.

The choice as to how to obtain a DBMS has not been made yet. When the choice is made, it will depend on several factors, among them:

- The funding available for implementation, and its timing
- Further experience with and new developments in commercially available systems; and
- The nature of the structure of the data in PLANT.

A Distributed Data Base

The latest thinking on data base management systems (Delobel and Litwin 1980, IEEE 1980) is an attempt to accomodate the many different sorts of environments that data bases can be found in by permitting interaction among systems on different hardware and in different structures. Ideally a distributed data base management system (DDBMS) can use information from many different sources and distribute it to many different users in forms and through hardware as varied as in the source data bases.

For example, in the case of PLANT, the basic data in the system might be stored in many different computers at such places as the USDA research facilities in Beltsville, Maryland, and universities and botanical gardens around the world. It is likely that these portions of the data base would be organized by subject in some cases (e.g. the folk medicine file of AEGIS) and by geographic area in others (e.g. the five state PIN system).

As in the original concept for PLANT, most users would have access to some partition of this very large data base determined by their needs. The DDBMS would perform the same functions as in a centrally located DBMS, but these would probably be somewhat more complex in order to take proper account of the different characteristics of the various data bases being used.

Recent thinking and experimentation with DDBMS's (Polk and Byrd 1981) indicates that they are likely to be much more efficient and cost effective than centrally located data bases. Furthermore, they can take explicit account of the different requirements various disciplines have for their data, while simultaneously rendering it available to other disciplines. The concept is the data base equivalent of having one's cake and eating it too, and would overcome many of the theoretical difficulties encountered in assembling the lexicon (See Chapter 6).

Were this approach to be adopted for PLANT, there would be some shift in emphasis of future work. Although establishing a solid understanding of the structure of comprehensive information on plants would still be a basic task, work on the lexicon would cease in favor of inventorying the computer based information now available or being assembled, inventorying the sorts of information that are likely to be useful in various disciplines and beginning the design of the DDBMS to serve them.

The reason for this change in approach is that a DDBMS is by definition never finished. Just as a DBMS is specifically designed (mostly by virtue of data independence) to enable the production of new reports and the incorporation of new and new kinds of information, a DDBMS is specifically designed to be able to incorporate the addition of whole new systems to the original, as well as new sorts of manipulation of the material in the data base.

It is judged by Polk and Byrd (1981) that experimental testing of DDBMS's themselves will not be effective. Perhaps this is so because of the high cost of creating artificially the complexity required for a proper test, or alternatively, the disruption to current users of non-artificial data bases. However, much theoretical and conceptual investigation into the design of DDBMS's has taken place. The French National Institute for Research on Information and Automation (INRIA) has undertaken a project known as SIRIUS to research and actually implement a totally distributed data base (Delobel and Litwin 1980). Scientists in Germany are also active in exploring the theoretical requirements of DDBMS's (Delobel and Litwin 1980).

There appears to be much less interest in DDBMS's in the United States, however, Polk and Byrd (1981) report that C.J.Date at IBM has begun work on the possibility of a so-called higher language (a metalanguage) which could be incorporated into any of a number of standard computer languages for the purpose of allowing data bases with different structures to be used in a single DDBMS.

The principle disadvantage of a DDBMS is political, in that institutions are often unwilling to permit wide access to their data, one of the very problems that PLANT is designed to overcome.

CHAPTER 5

ILLUSTRATIONS: VIDEO DISCS

In order to meet the general criteria for PLANT, illustrations of the plants in the system must have the following characteristics:

- They must illustrate as many of the data elements as possible. Thus the visual record should include illustrations of all significant plant parts, the plant in various stages of maturity, the plant in its native habitat and in a managed landscape, in all its uses, what it looks like when afflicted with the pests and diseases to which it is susceptible, what it looks like when harvested, micrographs of specially adapted cells, diagrams of genetic material and so forth;
- They must be easy to duplicate and transmit;
- They must be durable;
- They must be easily accessible; and
- They should minimize the requirement for librarian type functions (by people).

Clearly, no physical collection of photographs and drawings can meet these criteria, and resolution of this problem has been one of the most serious problems affecting the technical feasibility PLANT. Within the last three years, however, the technology of video discs has been developed to a point at which archival storage of optical information on discs is now a thriving and growing industry.

There are fundamental differences in production and in reading between LASER read video discs and conventional video tape, which make LASER read video discs ideal for archival (that is, permanent and not requiring updating) storage of information, particularly visual information. *

Both record electronic impulses representing FM waves, the same frequency modulated waves by which all television and a lot of radio broadcasts are transmitted. In the case of video tape, the impulses are recorded directly onto magnetic tape and read as electronic impulses directly off the tape. Anything which interferes with this direct reading (such as a protective coating on the tape) to keep it from wearing out reduces the performance of the tape. LASER read video discs, however, are physically imprinted with tiny pits, and covered with a reflective and then a protective coating. To read the disc, a LASER is flashed onto the surface and reflected directly back to its source where the surface is smooth, but refracted where the surface is pitted. The pattern in which the LASER light is

---

* The following description of the production of LASER read video discs is of the patented process developed by Phillips N.V., and used under license by DiscoVision, a joint venture of RCA and IBM. There are other processes under development (e.g.Drexler 1981) for specialized processes, but the Phillips process is in commercial production and is the one used for all but one of the applications noted below. This description is based on Mount 1981 and Lauzzana 1981.

refracted correlates directly with the pattern of the tiny pits, and therefore with thepattern of the electronic impulses which they represent. The LASER reader interprets the pattern of refraction as electronic impulses and these are converted back into FM signals.

By virtue of reading the FM signals indirectly, rather than directly as is the case with tape, video discs can achieve a permanence which is estimated to be longer by an order of magnitude than film, but which has not been tested fully yet, because of the newness of the technology. This permanence is one of the things that renders video discs so ideal for archival storage of any material, but particularly for slides and drawings which are known to deteriorate with use and storage over so short a period as ten years.

The second important difference between video tape and video discs stem from the way information is arranged on them and is directly analogous to the equivalent difference in magnetic tape and discs for use as auxiliary memory devices for computers. A tape is a continuous medium which must be read, or at least scanned, in sequence. An item at the end of the tape cannot be read without scanning the whole tape. Tape is an ideal medium for reading things in large bunches and in alphabetic or numeric order. But it is very cumbersome for reading individual items out of sequence, as might happen if one wanted to compare the mature habit of Acacia and Zizyphus trees.

In contrast, all of the information stored on a video disc is available at one time. A LASER read video disc rotates at 1800 revolutions per minute and therefore the time it takes to move from any one piece of information to another ranges from a maximum of 2.5 seconds to a minimum of 1/30 of a second.

The high speed at which LASER read video discs constitutes a third vital difference between them and video tape with regard to archival storage. An image which is read and displayed 30 times each second is interpreted by our eyes as constant; this capability of video discs is call "freeze-frame." With freeze-frame capability, we can look at a single image for as long as we like, hours if necessary, without damaging the medium on which the image is stored since it is read indirectly rather than directly.

The latest television technology permits very high resolution of images and excellent manipulation of color, so that we can be assured that, to the degree the original images--slides, drawings, etc.--used to imprint the disc were true, the image read off the disc will be true as well, without distortion in line or color.

Discs provide a very dense storage medium, currently 54,000 images per side, and the latest equipment can search both sides of a disc simultaneously. As a result, even with ten images per species, a single disc could accomodate information on 10,800 plants in a space the size of an ordinary long playing record.

The cost of committing visual material to a master disc is very low and going down. (Master video discs share the permanence of copies, unlike the master discs now used for long playing records.) The real cost in producing a video disc of plant images is in acquiring the images, putting them on continuous film (Cooper 1981), and producing an absolutely accurate catalogue of what is on the disc and where (by the sequence number identifying each frame). That cost is very high, as it requires a great deal of manual labor and expert judgement, and any skimping on it would destroy the utility of making the disc in the first place.

This technology has not been fully exploited in establishing permanent archives of visual information, but a number of demonstration discs have been produced. The Boston Museum of Fine Arts (Sorkow 1981) put 2,000 slides from their collection on a disc. The Jet Propulsion Lab has committed a large volume of engineering drawings to disc. Sears Roebuck is experimenting with the use of video discs as catalogues. The University of Nebraska produced a disc for the American College of Hematology, and Skaggs is now producing a disc to help train their personnel in identifying different kinds of produce (Mount 1981).

Finally, there is the question of the interface between the DBMS and the illustrations that accompany it. Optical information is different in kind from digital information and also takes up a great deal of space relative to digital information. For that reason, although video disc systems which permit easy access to both kinds of information on a single disc are now being developed, it seems unlikely that it will be efficient in the near future to use such a system for a very large data base such as that envisioned for PLANT. Rather, they will probably operate independently, with the file on each plant including complete notations as to what illustrations are available and where they are located in the disc library.

CHAPTER 6

THE LEXICON

An essential part of every DBMS, and of every large data base is a data dictionary (Date 1981, Lomax 1977) which provides functional specifications for the data base. It lists and defines the elements of data, the structure of the data, the means by which they can be correlated and listed, how the data are updated, etc. The earliest, most time consuming and most thought provoking part of the work discussed here has been the assembly of the list and functional definitions of the elements of the data base for PLANT, a critical part of its data dictionary. The list of data elements and their functional (for PLANT) definitions are known here as the lexicon.

Every discipline has its own language, its own classification system, names for things and syntax, or rules by which information may be organized. The criteria for PLANT demand that the languages of many disciplines be transcended in a certain sense. In cannot be a straightforward translation across disciplines, for that is not always possible. Nor do we mean to violate the names and syntax which have been developed so carefully over the years. Rather, a way must be found to enable the disciplines to speak with each other through this system.

The lexicon provides first of all a list of all the elements of the data base, much like an inventory or catalogue. Therefore, its first requirement is that it be complete, actually list every item of information that the data base will accomodate. It is not necessary that each item of information be known to exist, or that a specific use for each item of information can be envisioned at this stage. It is important to consider the full range of possible information, so that when we come upon an interesting or useful body of information, we know how to fit it into the overall scheme, have a place to put it, and so that we can find it later when we do have a use for it. And it is important to have an easily accessible place to put new information as it is developed, since one of the major criteria by which PLANT will be judged is its ability to shorten the length of time it normally takes newly developed information to cross disciplines. There is no sure test for whether the lexicon is complete at any stage. It is only possible to ask onself and colleagues whether there is anything else that should be on the list.

Second, the lexicon requires a precise, functional definition for each data element, what may be termed the data specifications. The functional definition consists of two parts, the range of possible values that the element can take on (Lomax 1977) and the number of possible values that each element can take on.

The range of possible values can be thought of as all the permissible answers to a question about the data element. It may consist of no more than "yes," "no," and "it depends," as for instance to a question about a plant's having roots in the soil. The possible values in the case of average leaf length at maturity may consist of "all integers from 1-999 in millimeters." In some cases the range of possible values will consist of a list of terms. For instance, the range for "Common designation" will have a finite list of terms including such things as "tree," "shrub," "perennial herb," and so forth.

The possible values are very repetitious and will therefore often refer to reference files which hold all the possible values for a number of data elements. Questions about color will arise many times with regard to different plant parts. Accordingly, there will be a comprehensive reference file on color which will be used in each and every case in which a color is called for. The color reference list, in turn, will be defined by a widely available color key (probably one of the Munsell guides).

In many cases, the possible vales of a data element are not well defined nor even universally agreed upon. The names of the fragrances that parts of plants give off provide a goo example. In these cases, we can do little more than list the referents of most fragrances, such as "lemon, sandalwood, rose," etc. But there is a list of possible values, and someone who tries to enter "my backyard after a summer rain" or "a new house" will get an error message, and be referred to the list of possible values.

The second part of the functional definition for each data element is the number of values it is permitted to have in any one case--its occurrence. Most of the data elements are permitted to have no value, signifying that the information is not available. Some can have only one value, and others can have more than one.

The colors of many flowers will, for instance, have more than one value. Finally, a few data elements such as "Scientific name" must have one and only one value.

The functional definitions insure that the possible values for each data element will be the same, of the same character and in the same units, without regard to which discipline the information comes from or which is requesting the information. Leaf length provided in Angstroms or light years must be converted to millimeters, no matter what. The combination of the DBMS and the lexicon will also insure that anyone who tries to tell PLANT that a leaf's length at maturity is purple will get an error message.

The ideal DBMS requires complete independence of all data so that each item of information can be used for any purpose. In other words, the data cannot be aggregated or structured in any way which might prevent its being used for any purpose. Very often, the easiest way to satisfy this requirement is to break the data down into very small pieces, and this has been done to the maximum degree possible in developing the portion of the lexicon completed so far.

However, breaking data down into very small pieces may require a degree of precision that is unavailable (Dittberner 1982) or inappropriate. For some purposes, it's not even very useful. For instance, whether a tree grows fast or slowly is often an important consideration in deciding whether or not to use it in a specific landscape. As used by landscape architects, these terms are very broad and are generally consistent only within each individual's use of the terms. It is not clear that defining these terms in centimeters growth per week in each of several carefully defined stages of growth, under equally carefully defined conditions would enable landscape architects to use those terms more consistently. Horticulturists, agronomists and plant physiologists might be interest interested in such precise information, and so, if it is available PLANT ought to accomodate it. However, the more general terms serve an important function, in spite of their imprecision, and therefore the lexicon includes them, as well as more precise indicators of growth rate.

A conscious effort was made in assembling the lexicon to break information down in ways which would permit use of this work eventually for all plants, not just desert plants. The only useful technique discovered so far for insuring that no term in the lexicon unintentionally excludes a plant to which it should apply has been to look for exceptions to the definition, and the dissection of data, established. But even within the class of desert natives, this test proved unsettling many times.

Knowing whether a plant is a tree or a shrub or an annual or herbaceous is generally very useful. But how does one define a tree? Any definition which includes woodiness, or an outer cambium layer, leaves out palms, many of which are clearly trees. One of the commonnly identified characteristics of trees is a canopy which sits at some interval above the ground. But the branches of many conifers reach right down to the ground. So the existence of a canopy separated from ground level cannot be used in the definition. In fact, any definition which requires branches at all will leave out many important trees.

At some point, one finds oneself asking the question, "What is 'treeness'?" And the only answer which seems to satisfy is that many things which are rooted in the ground, have an upright habit independently of any outside support, and one or a few main trunks are often called trees, and plants which don't have these characteristics are very seldom called trees.

This kind of question came up many times and attempts to deal with them yielded surprising results. Among these, was finding that whether or not a plant had roots in the ground and establishing the manner of a plant's emergence from the ground turned out to provide very useful partitions in the data about plants. These are also aspects of plants which are common to all, regardless of their natural habitat, and they convey some information which is useful for other purposes in addition to providing broad, natural classifications. Accordingly, they are given considerable attention in the lexicon.

Except for paring the list of data elements down, work has been completed (and included here in the Appendix) on the portion of the lexicon describing the plant in the following areas:

1. Roots - Designed to distinguish between deep and shallow rooting plants, to highlight unusual root arrangements, and to catalogue shoot/root ratios by both weight and linear dimensions.

2. Soil Penetration - Originally a description of tree trunks, but now includes information on underground trunks and other less typical interfaces with the soil as well.

3. Habit - In the main, describes the plant's silhouette, but also includes the degree of shade provided, if any, by the plant out of leaf.

4. Foliage - Descriptions of both individual leaves and the foliage in general, when young, mature and dormant. Also includes shade provided by the plant when in leaf.

5. Flower - The attempt to preserve correct botanical usage while still providing descriptions which will be useful to non-botanists is most difficult here. It is, so far, the longest of the description sections as a result.

6. Fruit - In addition to a straightforward description of of the fruit, this section includes information on whether the fruit is edible by humans, birds or animals, poisonous, etc.

7. Miscellaneous - Includes information on fragrance, growth rate, size at various points in the life cycle of the plant, spontaneous vegetative reproduction, and a space for a few very imprecise comments such as "fantastic plant."

Work has begun on several other sections of the lexicon, including Ecology & Culture, Uses, and Sources & References, and the section on Names has been completed (included in the Appendix). The portions of the lexicon dealing with various specialties such as range management and plant genetics, for instance, will have to be assembled by small teams which would include specialists in those fields as well as people with a full understanding of the requirements and goals of PLANT in order for the information from those specialties to serve both their own disciplines and others'.

The completed portion of the lexicon has received some excellent criticism from Holland which was echoed in a recent communication from Dittberner (1982). As it now stands, the lexicon asks for a level of detail of description which may never be available in some areas. In other areas, it calls for a level of detail which may be available for a large number of plants, but for which there is no immediately obvious use. The problem in many of these cases is in trying to predict ahead of time which elements of information are going to be needed and which are not. Given the aim to determine the characteristics of the ideal system for PLANT, there is great reluctance to eliminate data elements from the lexicon without better information on which ones are safe to eliminate.

On the other hand, Dittberner relayed that in the course of setting up PIN there was a similar experience. The team working on PIN's equivalent of this lexicon wanted to provide for as much precision and quantification of information about plants as possible. Their finding was that the data which would provide the desired quantification was simply not available and so they opted for qualitative terms in most cases. Dittberner also reported (1982) that Dr. Richard Hodder at Montana State University has continued to pursue this desire for more precision and quantification in plant description, but has not yet published his findings.

This criticism, experience, and the next stage of work on PLANT, which involves trying to determine the structure of such comprehensive information on plants, should provide the needed clues for paring the lexicon down to its most useful proportions.

CHAPTER 7

CONCLUSIONS

This work has had two themes: first a search for the ideal assemblage of technologies necessary to bring PLANT into existance in accordance with the criteria stated above; and, second, compiling a list, the lexicon, of what should be in the data base.

There is no doubt that the technology exists to create the comprehensive system discussed here. The computers (large and small), the DBMS's and video discs required by the system are all commercially successful facts, presenting no impediment to development of the system. On the contrary, recent work on distributed DBMS's for very large data bases indicates an even greater likelihood of technical feasibility by virtue of eliminating the need to centralize the collection and input of information into the system.

Aside from the fact that the one portion of the lexicon completed so far is extremely unwieldy in its drive toward comprehensiveness and complete data independence, the conclusions to be drawn from that part of the work are not so straightforward. Just as one would be unwilling to consciously eliminate a certain species from an ecosystem without having some good estimate of the consequences of such an act, the author has never felt that the conventional wisdom on which of the items in the lexicon can be eliminated with impunity is backed up with enough understanding of the relationships among those various pieces of data.

One of the most striking observations made in the course of work on the lexicon was that so many of the standard and time honored ways of classifying plants either break down under careful scrutiny or don't convey any useful information in this context. (In narrower cases, for instance within each discipline, this is not necessarily the case.) It seems significant that most people would agree that a palm has more in common with an elm tree than with a perennial grass, and that some others would categorically deny that a monocotyledonous, herbaceous perennial could have anything in common with a dicotyledonous, woody, branched plant. Educating people in the taxonomic niceties of palms and elms will not change the fact that many palms have "treeness" and that they perform the same functions as "true" trees all the time. So, to the degree that any structure imposed on the data base in PLANT does not permit us to see the ways that a palm is like an elm on the one hand, and also like a perennial grass on the other, it is literally hiding information from us.

This is a very obvious example, merely a cartoon of the problem in general, and one that has often been dealt with, primarily by introducing additional redundancy into a system. But it is completely analogous to the subtler difficulties that arise when imposing any sort of structure on a body of data.

The data base in PLANT could conceivably be thought of as a giant chart, with the plants designating the rows and the data elements heading the columns. This is essentially no structure and any useful manipulation of the information in the data base is inordinately difficult and time consuming (as well as expensive) as a result, especially given the very large size of the matrix.

A variety of ways to make it easier to examine and manipulate large bodies of data are incorporated in the programs in data base management systems. But these are structures too, even when they are temporary, and have just as much capacity to hide information as to reveal it by our choice of DBMS, which data elements to examine and which plants definitely do or do not have those data elements. The devices incorporated in DBMS's will be essential tools in the development of PLANT, but only after careful consideration has been given to what structure we would like them to express.

Ideally, we would impose no structure at all, but rather articulate whatever structure is inherent in the information which will compose PLANT's data base. Information about plants, like information on most natural systems, is complex, overlapping, at time redundant, at other times full of apparent anomaly and uniqueness. Each such body of data is naturally connected to myriad others which share these messy characteristics. The vast majority of systems for establishing some structure for information aim specifically to simplify it in some way, to make it more manageable. This is as true of traditional taxonomy as it is of statistical methods and data base structures devised for purposes outside the natural sciences.

The intuitive conclusion, as a result of work on both the lexicon and on data base management systems, is that this "messiness" in the data on plants and other natural systems is an indication that these data are different in kind from the data of physics and chemistry which have given rise to most of the data structures and the statistical techniques now available. It may also indicate that the data should not be simplified until we have a better understanding of their complexity.

It is proposed that by accepting the "messiness" of the data, carefully and rigorously, we may be able to detect a structure actually inherent in that data. We may find indicators that can be used with confidence for simplification of the lexicon, retaining only the items which pertain exactly to the characteristics of natural systems, especially those of a relatively small scale at which even sophisticated statistical techniques have little or no real scientific value, or in aggregations of small systems where the same is true (Renard et al. 1974).

A methodology call Q-analysis, developed within the last fifteen years by R.H.Atkin (1974) and using the principles of set theory and algebraic topology, holds out the promise of being a system which can deal with the complexity and need for synthesis in natural systems. A key aspect of Q-analysis is the requirement to establish and define carefully the sets of information (similar to the column headings on the matrix described above) and the hierarchy inherent in these sets. This aspect of Q-analysis alone may turn out to be its greatest contribution to investigations in the natural sciences, as we are seldom as careful as we should be in this respect.

However, with this vital preliminary step accomplished (which incidentally can only be done by human beings), the methodology goes on to apply precise mathematical techniques to examine the connectivity in a body of data organized by this hierarchy of sets. Many of the applications of Q-analysis to date have been trivial (e.g.Freeman 1980) and have not, therefore, yielded any information which could not have been gotten by some more orthodox means. Work on more complex data, however, has brought to light information that was literally lost in the simplification and aggregation of data required for most standard statistical techniques. (Gould 1981).

If the structural language of Q-analysis does indeed provide the tools for a better understanding of natural systems, the implications for our ability to manage, protect and teach about the environment are enormous. Within the context of PLANT, such tools would immediately make the possibility of a distributed data base much more attractive and realistic than any scheme tried so far. But like any new, potentially profound, idea, Q-analysis must be tested. An adequate test for this purpose would require a body of data that is not overwhelming in its overall size, but as rich and complex as is usual for natural systems. It should also be a body of data for which the conclusions of Q-analysis can be readily compared with the conclusions of other sorts of analysis, whether statistical or judgemental. Such a test will be the next step in this work.

In short, what began as a manual file on plants for the author's own use has turned into an inquiry into the very basic nature of information on natural systems. It is hoped that this is merely a large loop in a spiral which will eventually lead back to bringing PLANT into reality.

APPENDIX A: GLOSSARY

AEGIS : a plant data base developed by the USDA Economic Botany Laboratory. See page 16.

Batch mode, batch processing : indirect communication with a computer. See pages 35 and 36.

CAPS : Computer Assisted Plant Selection. A program developed by students at Ohio State. See page 18.

DBMS : Data Base Management System. See Chapter 3.

DDBMS : Distributed Data Base Management System, used in this work especially with reference to very large data bases. See pages 42 through 44.

Dumb terminal : a computer terminal with very limited capacity to interact with a larger system and little or no inherent capacity as a computer. See page 38.

Interactive mode, interactive processing : direct communication with a computer. See pages 35 and 36.

PIN : Plant Information Network. A data base and DBMS developed by the US Department of Interior. See pages 12 and 13.

RAM : Random Access Memory. See page 25.

Remote terminal : a computer terminal located at some distance from the computer itself, usually by telephone lines.

Smart terminal : a computer terminal which has some inherent computer capabilities, as well as the capacity to operate interactively with a larger system.

APPENDIX B : THE LEXICON

The lexicon is a list of the elements of data originally slated to be included in the PLANT data base, along with the functional specifications for each data element. The lexicon was compiled with the objectives of including the smallest useful unit of information, accomodating both quantitative and qualitative information, and permitting its use by all interested disciplines.

The column labeled "Data Element" lists the pieces of information in the lexicon. "Possible Values" gives form that information may take. "Occurence" indicates whether there can be no information on that data element, only one piece of information, or more than one. "Priority of Access" ranges from I to III, "I" having the highest priority.

Only the "Names" and "Description" portions of the lexicon are included here. A description of the other sections originally planned and of the lexicon in general may be found in Chapter 6.

	Data Elements	Possible Values	Occurrence	Prior-ity
	Scientific name	All botanical names; all with genus & species; some with subspecies, cultivar etc. up to four parts Some repeating repeating elements	1	I
	English common name	All common names in English; most contain more than one word Many repeating elements	0-1+	I
	Obsolete scientific names & synonyms	All obsolete botanical names, all with genus & species; some with subspecies, etc. up to four parts Many repeating elements	0-1+	III
	Spanish common name	All Spanish common names Tilda and accent	0-1+	III
	Arabic common names	All common names in Arabic; must be associated with region in which used; transliterated Many repetitions and repeating elements	0-1+	III
	Hebrew common names	All common names in Hebrew; transliterated Probably repeating elements	0-1+	III
	Farsi common names	All common names in Farsi; may need to associate with region in which used; transliterated Many repetitions and repeating elements	0-1+	III
	Common designation	Tree Shrub Vine Annual Perennial Herb Grass	0-1+	I

Continued next page

NAMES

	Data Element	Possible Values	Occurrence	Prior-ity
	Common designation	Palm Epiphyte		
		Others may be added		
	Common names in other languages	All common names except in English, Spanish, Hebrew, Arabic and Farsi; must be associated with geographic area; transliterated when appropriate	0-1+	III
		A holding area in preparation for new name files		
	Nursery & other industrial names	All scientific, all English common names and others	0-1+	III
		In most cases, will refer to a name listed elsewhere		
	Family	All botanical families	0-1	II
	Former family	All botanical families	0-1+	III
	Tribe	All tribes in Compositae, Graminae and Leguminosae	0-1	

	Data Element	Possible Values	Occurrence	Prior-ity
	Existence of Roots	None Soil roots only Soil & aerial roots Soil & prop aerial roots Soil & aerial/soil roots Aerial roots only	1	III
	Extent of root system	Extensive Not extensive	0-1	II
	Primary roots, morphology	Single tap root More than one tap root Branched tap root Fibrous roots, coarse Fibrous roots, fine Fibrous roots, medium Tap and fibrous roots Storage organ= fleshy Fascicled= Cluster of approx. equal sized roots	0-1	II
	Secondary roots	Single tap root More than one tap root Branched tap root Fibrous roots, coarse Fibrous roots, fine Fibrous roots, medium Storage organ Adventitious roots	0-1	III
	Primary direction of root growth relative to ground plane	Parallel Perpendicular Angular Parallel & Perpendicular Does not relate to ground plane	0-1	III
	Root depth relative to height of plant	Deep Shallow Intermediate	0-1	II
	Root extent relative to width of plant	To drip line Beyond drip line Not as far as drip line	0-1	II

	Data Element	Possible Values	Occurrence	Prior -ity
	Root/shoot ratio by weight	All integers from 1-999:All integers from 1-999	0-1	III
	Root/shoot ratio by volume	All integers from 1-999:All integers from 1-999	0-1	III
	Location of root attachment(primary)	At soil line Below soil line Above soil line	0-1	III
	If root attach ment above or below soil line, at what distance	All integers from 1-99 in centimeters	0-1	III
	Root exposure	No primary soil roots naturally exposed Some primary soil roots almost always exposed Many " " " " " "	0-1	II
	Changes with culture	Aerial roots with high humidity only Deep watering discourages shallow roots Roots more extensive with drought Roots more extensive with more water than naturally Shallow watering encourages shallow roots	0-1+	II

	Data Element	Possible Values	Occurrence	Prior-ity
	Primary shoot, nature	Herbaceous Woody Suffrutescent = Woody at base only	0-1	
	Primary shoot, name	Culm Trunk Stem Foliage Single leaf More than one leaf Inflorescence	0-1	
	Organ emerging from soil	Primary shoot Bulb Corm Rhizome Tuber Crown Root(s)	0-1	
	Diameter of primary shoot when mature at soil line	All integers from 1-999 in centimeters	0-1	
	Diameter of primary shoot when mature 30cm above soil line	All integers from 1-999 in cm	0-1	
	Diameter of primary shoot when mature 1m above soil line	All integers from 1-999 in cm	0-1	
	Diameter of primary shoot when mature "at breast height"	All integers from 1-999 in cm	0-1	
	Color of epidermis of primary shoot when immature	All listed colors	0-1	
	Color of epidermis of primary shoot when mature	All listed colors	0-1	

	Data Element	Possible Values	Occurrence	Prior-ity
	Texture of epidermis of primary shoot when immature	Regular vertical ridges=Same number in any cross section Irregular vertical ridges (Includes furrows) Smooth Peeling in regular patches Peeling in irregular patches Obvious lenticels=Lens shaped spots Thorny=Spiney Corky Tomentose Sclerophytic Rough	0-1+	
	Texture of epidermis of primary shoot when mature	Regular vertical ridges=Same number in any cross section Irregular vertical ridges (Includes furrows) Smooth Peeling in regular patches Peeling in irregular patches Obvious lenticels= Lens shaped spots Thorny=Spiney Corky Tomentose Sclerophytic Rough	0-1+	
	Character of primary shoot in youth	Straight and upright Straight & horizontal Leaning Curved Gnarled Mallee Thicket Climbing	0-1+	
	Character of mature primary shoot	Straight & upright Straight & horizontal Leaning Curved Gnarled Mallee Thicket Climbing	0-1+	

	Data Element	Possible Values	Occurrence	Prior-ity
	Visible changes in primary shoot over length	None Bulges at base Bulges in middle Bulges near top Tapers toward top Tapers toward bottom	0-1	
	Water storage in primary shoot	No Yes, but not visible Yes, visible bulge	0-1	
	General cross section of primary shoot	Round Quadrangular Round w/ ridges/furrows Quadrangular w/ ridges/furrows	0-1	
	Primary shoot jointed	Yes No	0-1	
	Primary shoot has internodes	No Yes, hollow Yes, solid	0-1	
	Primary shoot changes w/culture	More than oneprimary shoot w/training Only one primary shoot w/training Naturally one primary shoot Naturaly more than one primary shoot More branching w/more water More branching w/less water Taller w/more water First branch higher w/more water More branching w/cold First branch lower w/cold	0-1+	

	Data Element	Possible Values	Occurrence	Prior-ity
	Height of first branch above soil after ten years, naturally	All integers from 1-999 cm At base of primary shoot	0-1	III
	Normal maximum height of first branch above soil	All integers from 1-999cm At base of primary shoot	0-1	III
	Main branches, those attavhed directly to primary/ shoot are usually	Horizontal Horizontal w/ascending tips Horizontal w/descending tips Upright Arching/Fountain shaped Pendulous/More than arching	0-1	II
	Overall appearance of branched portion of plant w/o foliage in first five years	Columnar Pyramidal Upright/Not as much as columnar or pyramidal Candelabra Globose Oval Flattopped/Umbrella shaped Vase shaped/ Weeping Mallee	0-1	I
	Overall appearance of branched portion of plant w/o foliage in maturity	Columnar Pyramidal Upright/Not as much as columnar or pyramidal Candelabra Globose Oval Flattopped/Umbrella shaped Vase shaped Weeping Mallee	0-1	I
	Sunlight at noon, equidistant from center of plant and dripline is reduced (when plant is not in leaf)	75%+ 50-75% 25-50% 25%- Evergreen	0-1	III

	Data Element	Possible Values	Occurrence	Prior-ity
	Leaf buds, position	Terminal Axillary Lateral Adventitious	0-1	
	Leaf buds, texture	Scaly Hairy Resinous Naked	0-1+	

	Data Element	Possible Values	Occurrence	Prior -ity
	Leaves appear like	True leaves Needles Appears leafless, even in growing season	0-1+	I
	Leaves are	True leaves Needles Scales arranged to resemble leaves Vestigial Absent	0-1	I
	True leaves are	Simple Compound Lobed Two lobed Three lobed Divided } Increasing distance toward Finely divided } midrib, lacinate is to the Lacinate } midrib.	0-1	I
	Leaflets, lobes, divisions are arranged	Palmately Pinnately Bipinnately Tripinnately	0-1	I
	Leaf attachment	Petiole Sessile Perfoliate Sessile rosette Sheath, open Sheath, closed	0-1	III
	Leaf arrangement	Opposite Alternate Whorled	0-1	II
	Leaflet arrangement	Opposite Alternate Whorled	0-1	III
	Leaflets, number	Single number or range, all integers from 1-99	0-1	

FOLIAGE

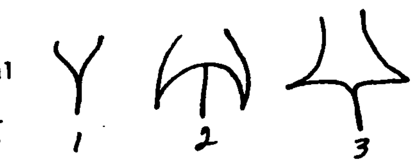
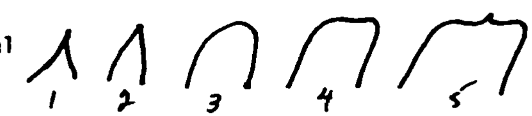
	Data Element	Possible Values	Occurrence	Prior-ity
	Leaf shape	Fan Heart=Cordate Sword/Lance Needlelike Ovate=Petiole at broad end Orbicular=Round Variable Obovate=Petiole at narrow end Linear Oblong	0-1+	I
	Leaflet shape	Fan Sword/lance Needlelike Ovate Orbicular Variable	0-1	III
	Leaf margin	Smooth=entire Undulate=Wavy in three dimensions Wavy in two dimensions Crenate=round teeth=Scalloped Dentate=teeth pointing out Serrate=teeth pointing to apex of leaf	0-1	II
	Leaflet margin	Smooth=entire Undulate=Wavy in three dimensions Wavy in two dimensions Crenate=round teeth Dentate=teeth pointing out Serrate=teeth pointing to apex of leaf	0-1	III
	Leaf size at maturity, w/o petiole Width	All integers from 1-9999mm Range	0-1	II
	Leaf size at maturity, w/o petiole Length	All integers from 1-9999mm Range	0-1	II

FOLIAGE

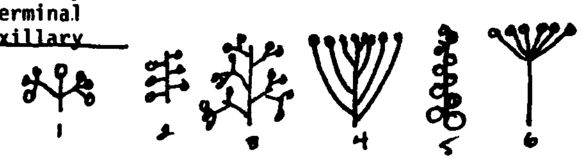
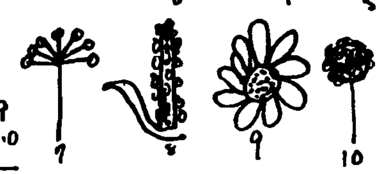
	Data Element	Possible Values	Occurrence	Prior-ity
	Length of petiole	None All integers from 1-999cm Range	0-1	III
	Diameter of petiole away from leaf	None All integers from 1-999mm	0-1	III
	Length of leaf w/petiole at maturity	All integers from 1-1-9999mm Range	0-1	II
	Leaves stipulate	Yes No	0-1	III
	Petiole color when young	All listed colors	0-1	III
	Petiole color when mature	All listed colors	0-1	III
	Veining prominent	Yes No	0-1	II
	Pattern of veining	Palmate Parallel Two parallel veins Three parallel veins Pinnate Reticulate Single midrib	0-1	III
	Color of veining	Similar to leaf color All listed colors	0-1	III
	Texture of upper leaf surface, young	Tomentose=Wooly Glossy Dull Pubescent=Fine, soft, short hairs Glabrous=Smooth, no hairs Glaucous=Waxy Villous=Long, silky, straight hairs Hirsute=Moderately stiff hairs Hispid=Stiff, bristly hairs	0-1	III

	Data Element	Possible Values	Occurrence	Prior-ity
	Texture of mature upper leaf surface	Tomentose Glossy Dull	0-1	I
	Texture of young lower leaf surface	Tomentose Glossy Dull	0-1	III
	Texture of mature lower leaf surface	Tomentose Glossy Dull	0-1	III
	Stomates	Both sides of leaf Lower side only Variable	0-1	III
	Average number of stomates/cm^{2}lower	All integers from 1-999 (no units	0-1	III
	Average number of stomates/cm^{2}upper	All integers from 1-999 (no units)	0-1	III
	Leaf color when young	All listed colors All listed colors w/spots of all listed colors All listed colors w/stripes of all listed colors All listed colors w/margins of all listed colors	0-1+	I
	Leaf color when mature	All listed colors All listed colors w/spots of all listed colors All listed colors w/stripes of all listed colors All listed colors w/margins of all listed colors	0-1+	I
	Texture of individual leaves, young	Fine Medium Coarse	0-1	III

	Data Element	Possible Values	Occurrence	Prior-ity
	Texture of individual mature leaves	Fine Medium Coarse	0-1	III
	Character of individual young leaves	Stiff Arching Leathery=Coriaceous Succulent Flexible Add to both lists: Parchment-like=chartaceous Hyaline=Iransparaent or transluscent Scarious=Thin & dry, but not green or transparent	0-1	III
	Character of individual mature leaves	Stiff Arching Leathery=Coriaceous Succulent Flexible	0-1	I
	Appearance of leaves in mass	Open, fine Open, Medium Open, coarse Dense, fine Dense, medium Dense, coarse	0-1	I
	Life cycle of leaves	Evergreen Evergreen w/some leaf drop and new leaves all year Drought deciduous Cold deciduous Nearly evergreen except annual preparation for new leaves Deciduous	0-1+	I
	Nearly evergreen, drop leaves in	All months	0-1	I
	Nearly evergree, new leaves in	All months	0-1	I
	In leaf under normal conditions	All months to all months	0-1	I

	Data Element	Possible Values	Occurrence	Prior-ity
	Individual leaves are	Opaque Transluscent	0-1	III
	Sunlight at noon, equidistant from center of plant and dripline is reduced, when plant is in leaf	75%+ 50-75% 25-75% 25%-	0-1	I
	Leaf base	Asymmetrical Cuneate 1 Sagittate 2 Hastate 3  	0-1	
	Leaf apex	Asymmetrical Acuminate 1 Acute 2 Obtuse 3 Emarginate 4 Mucronate 5  	0-1	
	3-D leaf morphology	Involute=Rolled or folded on itself Convolute=Coiled on the longitudinal axis Plicate	0-1	
	Grass leaves auriculate	Yes No	0-1	
	Grass leaves ligulate	Yes No	0-1	

FOLIAGE

	Data Element	Possible Values	Occurrence	Prior -ity
	Flowers/inflorescence	No true flowers True flowers, easily identifiable as such True flowers, not easily identifiable as such	0-1	II
	Most conspicuous aspect of inflorescence (to people)	Individual flower Clusters of flower parts Total inflorescence Petals Stamens Pistils Sepals Bracts Pedicel Fruit Buds	0-1+	II
	Major categories of inflorescence	Solitary, terminal Solitary, axillary Multiple, terminal Multiple, axillary	0-1	II
	Multiple flower arrangements	Cyme 1 Raceme 2 Panicle 3 Corymb 4 Spike 5 Umbel, flat 6 Umbel, round 7 Spadix 8 Head, daisy-like 9 Head, clover-like 10 Catkin=Ament  	0-1	II
	Individual flower shapes	Bell-shaped/Lilly of the Valley Orchid-like Tubular/Nicotiana Trumpet shaped/Morning Glory Saucer shaped/Poppy Pea Flower Star-like/		

	Data Element	Possible Values	Occurrence	Prior -ity
cont'd	Individual flower shapes cont'd	Cup shaped Daffodil shaped Iris-like Two lipped Assymetrical other than the above Symmetrical other than the above		
	Petals	None Completely united Completely separated United at the base, separated at outer edge	0-1	III
	Number of true petals/flower	All integersfrom 0-99 Double flower, many	0-1	III
	Color of petals	All listed colors	0-1+	I
	Number of bracts	All integers from 0-99	0-1	II
	Color of bracts	All listed colors	0-1+	II
	Number of sepals	All integers from 0-99	0-1	II
	Color of sepals	All listed colors	0-1+	II
	Apparent color of total inflorescence	All listed colors	0-1+	I
	General size of individual flowers	Tiny Small Medium Large Huge	1-0	I
	Diameter of individual flowers, avg.	All integers from 1-999mm	1-0	II
	Length to pedicel of individual flrs	All integers from 1-999mm	1=0	II

FLOWERS

	Data Element	Possible Values	Occurrence	Prior-ity
	General size of inflorescence	Tiny Small Medium Large Huge	0-1	I
	Diameter of inflorescence	All integers from 1-999mm	0-1	II
	Length of inflorescence	All integers from 1-999mm	0-1	II
	Length of pedicel	All integers from 1-999mm	0-1	III
	Habit of inflorescence	Upright Arching Pendant	0-1	I
	Flowers on single inflorescene are	Dense Open Spray=Orchids	0-1	I
	Second most conspicuous aspect of inflorescence (to People)	Individual flower Clusters of flowers Total inflorescence Petals Stamens Pistils Sepals Bracts Pedicel Fruit Buds	0-1	I
	Color of Stamens	All listed colors	0-1+	II
	Number of stamens	All integers from 1-999	0-1	III
	Color ofPistils	All listed colors	0-1	III

FLOWERS

	Data Element	Possible Values	Occurrence	Prior-ity
	Color of buds	Similar to color of flowers	0-1+	II
	Duration of readily visible buds	All integers from 1-999 days	0-1+	III
	Appearance of readily visible buds under natural conditions, first Repeat once for each of the following 25°S - 25°N latitude 26° - 35° N/S 36°+ N/S	Spring Summer Fall Winter All months in Northern Hemisphere All months in Southern Hemisphere All year	0-1+	III
	Second appearance of readily visible buds under natural conditions	Spring Summer Fall Winter All months Northern Hemisphere All months Southern Hemisphere	0-1	III
	First appearance of full blooms under natural conditions 25°S - 25°N latitude 26° - 35° N/S 36°+ N/S	All year Spring Summer Fall Winter All months Northern Hemisphere All months Southern Hemisphere	0-1+	I
	Second appearance of full blooms under natural conditions	Spring Summer Fall Winter All months Northern Hemisphere All months Southern Hemisphere	0-1	III
	Duration of full bloom under natural conditions-each inflorescence	All integers from 1-99 days	0-1+	II
	Duration of full bloom under natural conditions-whole plant	All integers from 1-99 days	0-1+	II

	Data Element	Possible Values	Occurrence	Prior-ity
cont'd		Long bloom period Short bloom period		
	Color changes after full bloom but before abscission	None Same color faded All listed colors	0-1+	I
	Persistent spent flowers	No Yes Yes, attractive Yes, unattractive	0-1	II
	Duration of persistence of spent flowers	Throughout dormant season All integers from 1-99 days	0-1	III
	Wildlife attracted by flowers	Bees Butterflies Wasps Hummingbirds	0-1+	III
	Sexual behavior of flowers	Sterile Monoecious=Stamens & pistils on different flowers, same plant Dioecious=Stamens & pistils on different plants Bisexual=Stamens & pistils on same flowers	0-1	III
	Age at which plant first flowers	All integers from 1-99 years	0-1	III
	Inflorescence is	Determinant Indeterminant	0-1	

	Data Element	Possible Values	Occurrence	Priority
	Pollenation vector, primary	None Wind Insects Bees Flies Wasps Butterflies Hummingbirds Birds Bats Snails &/or slugs Water Specialized plant organ	0-1	
	Species of animal vector	All scientific names of animal vectors This is probably a list that will be compiled over time	0-1+	
	Size of pollen	Large Medium Small	0-1	
	Pollen coating	None Gummy Waxy Powdery	0-1	
	Relative amount of pollen per inflorescence or stamen	Abundant Average Little	0-1	
	Plant is a known allergen	Yes No Maybe	0-1	
	Season of pollenation	Spring Summer Fall	0-1	

continued next page

	Data Element	Possible Values	Occurrence	Prior-ity
	continued from previous page	Winter Depends on climatic conditions		
	Duration of pollenation season	All integers from 1-12 months	0-1	
	Location of pollen in flower	Deep Shallow Surface	0-1	

FLOWERS

	Data Element	Possible Values	Occurrence	Priority
	Fruit is	Conspicuous Inconspicuous Ornamental Profuse Sparse None	0-1+	I
	Fruit is readily identifiableas such under natural conditions 25°S - 25°N latitude 26° - 35° N/S 36°+ N/S	All year Spring Summer Fall Winter All months Northern Hemisphere All months Southern Hemisphere	0-1+	II
	Fruit is ripe (ready to be eaten or to foster seed germination) under natural conditions	All year Spring Summer Fall Winter All months Northern Hemisphere All months Southern Hemisphere	0-1+	II
	Period from first appearance of fruit to full ripening naturally	All integers from 1-999 days	0-1	III
	Period fruit remains attached to plant after full ripening naturally	Separates from plant when fully ripe Persists indefinitely Persists through dormant season Persists for a short time	0-1	III
	Color of fruit at first appearance	All listed colors	0-1+	III
	Color of ripe fruit	All listed colors	0-1+	I
	Color of persistent fruit	All listed colors	0-1+	III
	Fruit type when fully ripe	Fleshy Dry w/fleshy part	0-1	II

	Data Element	Possible Values	Occurrence	Priority
		Dry Grain Achene Samara Nut Pod Capsule Drupe Aggregate/Multiple Pome Berry Berry-like Cone		
	Length of individual fruit	All integers from 1-999mm	0-1+	III
	Diameter of individual fruit	All integers from 1-999mm		
	Number of fruits in clusters	No clusters All integers from 1-999	0-1	III
	Length of clusters	Short Long All integers from 1-9999mm	0-1	II
	Diameter of clusters	Fat Narrow All integers from 1-999mm	0-1	II
	Shape of clusters		0-1	II
	Method by which fruit leaves plant	Natural abscission, drops to ground Wind Rain Animal agent	0-1+	III

FRUIT

	Data Element	Possible Values	Occurrence	Prior-ity
	Texture of epidermis of individual fruits	Smooth Glossy Tomentose Scaly Spiny Plumose Sclerotic	0-1	III
	Age at which plant bears fruit, first	All integers from 1-99 years	0-1	III
	Size of individual seeds	Tiny Small Medium Large Huge	0-1	
	Number of seeds per kilogram	All integers from 1-999,999,999	0-1	
	Number of seeds per fruit	All integers from 1-999	0-1	
	Color of seeds	All listed colors Must be able to express combinations of two colors	0-1+	

	Data Element	Possible Values	Occurrence	Prior -ity
	Fragrance of wood, time	Evening Warm weather Cool weather After rain In dry weather In place When crushed live When crushed dry	0-1+	II
	Fragrance of wood, general	Mild, pleasant Mild, unpleasant Strong, pleasant Strong, unpleasant	0-1	II
	Fragrance of wood is like	Lemon Orange Sagebrush Pine Cinnamon Mint Strawberry Hay Sweet Herbal Onion Garlic Banana Acrid	0-1+	II

	Data Element	Possible Values	Occurrence	Prior-ity
	Fragrance of fruit, time	Evening Warm weather Cool weather After rain In dry weather In place When crushed live When crushed dry	0-1+	II
	Fragrance of fruit, general	Mild, pleasant Mild, unpleasant Strong, pleasant Strong, unpleasant	0-1	II
	Fragrance of fruit is like	Lemon Orange Sagebrush Pine Cinnamon Mint Strawberry Hay Sweet Herbal Onion Garlic Banana Acrid	0-1+	II

	Data Element	Possible Values	Occurrence	Priority
	Fragrance of flowers, time	Evening Warm weather Cool weather After rain In dry weather In place When crushed live When crushed dry	0-1+	II
	Fragrance of flowers, general	Mild, pleasant Mild, unpleasant Strong, pleasant Strong, unpleasant	0-1	II
	Fragrance of flowers is like	Lemon Orange Sagebrush Pine Cinnamon Mint Strawberry Hay Sweet Herbal Onion Garlic Banana Acrid	0-1+	II

	Data Element	Possible Values	Occurrence	Prior -ity
	Fragrance of leaves, time	Evening Warm weather Cool weather After rain In dry weather In place When crushed live When crushed dry	0-1+	II
	Fragrance of leaves, general	Mild, pleasant Mild, unpleasant Strong, pleasant Strong, unpleasant	0-1	II
	Fragrance of leaves is like	Lemon Orange Sagebrush Pine Cinnamon Mint Strawberry Hay Sweet Herbal Onion Garlic Banana Acrid	0-1+	II

	Data Element	Possible Values	Occurrence	Prior-ity
	Fragrance of bark, time	Evening Warm weather Cool weather After rain In dry weather In place When crushed live When crushed dry	0-1+	II
	Fragrance of bark, general	Mild, pleasant Mild, unpleasant Strong, pleasant Strong, unpleasant	0-1	II
	Fragrance of bark is like	Lemon Orange Sagebrush Pine Cinnamon Mint Strawberry Hay Sweet Herbal Onion Garlic Banana Acrid	0-1+	II

	Data Element	Possible Values	Occurrence	Prior-ity
	Fragrance of roots, time	Evening Warm weather Cool weather After rain In dry weather In place When crushed live When crushed dry	0-1+	II
	Fragrance of roots, general	Mild, pleasant Mild, unpleasant Strong, pleasant Strong, unpleasant	0-1	II
	Fragrance of roots is like	Lemon Orange Sagebrush Pine Cinnamon Mint Strawberry Hay Sweet Herbal Onion Garlic Banana Acrid	0-1+	II

	Data Element	Possible Values	Occurrence	Prior -ity
	Height at maturity	All integers from 1-9999cm	0-1	II
	Diameter at maturity	All integers from 1-9999cm	0-1	II
	Area within dripline at maturity	All integers from 1-99,999cm-2	0-1	III
	Growth rate, general description	Fast Medium Slow	0-1	I
	% growth in first year	All integers from 1-100%	0-1	III
	% growth in first five years	All integers from 1-100%	0-1	III
	Total height after first year	All integers from 1-9999cm	0-1	II
	Total height after first five years	All integers from 1-9999cm	0-1	II
	Total diameter after first year	All integers from 1-999cm	0-1	III
	Total diameter after first five years	All integers from 1-9999cm	0-1	II
	Total height at maturity	All integers from 1-9999cm	0-1	I
	Total diameter at maturity	All integers from 1-9999cm	0-1	I
	Years to mature size	All integers from 1-99 years	0-1	II
ALL OF ABOVE FOR NATURAL CONDITIONS				

SIZE, ETC.

	Data Element	Possible Values	Occurrence	Prior -ity
	Vegetative reproduction, naturally	None Runners Rooting branches Plantlets Stolons	0-1+	II
	Life expectancy	Less than one year All integers from 1-99 years More than 99 years	1-0	II

SIZE, ETC.

SELECTED BIBLIOGRAPHY

Atchley, Allen, Economic Botany Laboratory, USDA, Beltsville, Maryland. 1982. Personal Communication

Atkin, R.H. 1974. Mathematical Structure in Human Affairs.Heinemann Educational Books. London.

Brenan, J.P.M., R.Ross and J.T.Williams, Eds. 1975. Computers in Botanical Collections. Plenum Press. New York.

Cooper, Connie, ComCorps, Washington, D.C. 1981. Personal Communication.

Date, C.J. 1981. Introduction to Database Systems. Addison-Wesley Publishing Compant. Reading, Massachusetts.

Dexler, J. 1981. "Optical Memory Media for Laser (sic) Recording and Image Storage," Computer Graphics World 4/9, Sepetmber:60-62.

Dittberner, Philip, Botanist, Fish and Wildlife Service, Western Energy and Land Use Team, USDI, Fort Collins, Colorado. 1982 Personal communication.

Doszkocs, Tamas E., Barbara A. Rapp and Harold M. Schoolman. 1980. "Automated Information Retrieval in Science and Technology," Science 208, April: 25-30.

Duke, James A. and Edward E.Terrell. 1974. "Crop Diversification Matrix: Introduction," Taxon 23(5/6) November: 759-799.

--------. 1978. "Chapter 4. Ecosystematics," in Beltsville Symposium on Agricultural Research 2. Biosystematics in Agriculture. Allenheld Osmun & Co. Montclair, New Jersey.

--------. 1981. "Lab Review. Economic Botany Laboratory, October 21, 1981." Unpublished.

Freeman, Linton C. 1980. "Q-analysis and the structure of friendship networks," International Journal of Man-Machine Studies 12:367-378.

Gould, Peter. 1980. "Is it necessary to choose? Some technical, hermeneutic and emancipatory thoughts on inquiry," Conference on Philosophy and Geography: A Search for Common Ground, Bellagio, Italy, August 1-5. Pion, London.

--------.1981. "A structural language of relations," Future Trends in Geomathematics. R.G.Craig and M.L.Labovitz, Eds. Pion, London.

Holland, Marianna. 1979a. A New Source of Plant Information for the Southwest. M.L.A.Thesis. University of Arizona.

--------.1979b. "AVIS: A Prototype Arid Vegetation Information System," Desert Plants 1(2) November:71-76.

--------.1982. "A Data Bank for Arid Plants," Landscape Architecture 72(2) March: 91.

Ingerman, Peter Z., Systems Consultant, Willingboro, New Jersey. 1977-1982. Personal Communications.

Johnson, Craig and Joyce Harsh. 1981. "Computer Assisted Plant Selection." NEA Grant Progress Report. The Ohio State University. October.

--------. 1981. "Computer Assisted Plant Selection Addendum." NEA Grant Technical Report. The Ohio State University. December.

Kelly, Kathleen. 1980. "Lexicon for a Data Base on Desert Plants." Presented at the Annual Meeting, New Jersey Academy of Sciences, Computer Sciences Section. New Brunswick. March.

--------. 1981. "Applications of Video Discs to Landscape Architecture" Presented in Educational Session, ASLA Annual Meeting, Washington, D.C. November.

--------. and R.T.Schnadelbach. 1976. Landscaping the Saudi Arabian Desert. The Delancey Press. New York (OOP).

Kroenke, David. 1978. Database: A professional's primer. Science Research Associates. Chicago.

Lauzzana, Roy. 1981. "Video Disc: A Necessary Revolution," Computer Graphics World 4/9 September: 55-57.

Lodwick, Weldon, Michigan State University, East Lansing, Michigan, 1982. Personal communication.

Lomax, J.D. 1977. Data Dictionary Systems. NCC Publications Ltd., Manchester, England.

McCracken, Daniel D. 1980. A Guide to NOMAD for Applications Development. National CSS. Wilton, Connecticut.

Mount, William, Technical Service, DiscoVision, Costa Mesa, California. 1981. Personal communications.

Neely, Joel and Steve Stewart. 1981. "Fundamentals of Relational Data Organization," BYTE 6(11) November: 46ff.

Pitt, David.1982 "Survey of Microcomputer Applications in Landscape Architecture." Unpublished.

Polk, Warren J. and Kerry Byrd. 1981. "Managing the Very Large Database," Datamation 27(10) September 115-124.

Renard, K.G., J.R.Simanton and H.B.Osborn. 1974. "Applicability of the Universal Soil Loss Equation to Semiarid Rangeland Conditions in the Southwest," Volume 4 Hydrology and Water Resources in Arizona and the Southwest. Proceedings of the 1974 Meeting of the Arizona Section-American Water Resources Association and the Hydrology Section-Arizona Academy of Science. April 19-20, 1974. Flagstaff, Arizona.

Sorkow, Janice. 1981. "Archival Storage of Visual Resources on a Videodisc: A Pilot Project," Videodisc News March: 6-7.

Zohary, Michael. 1966. Flora Palaestina: Part One - Volume I. Israel Academy of Sciences and Humanities. Jerusalem.