

THE USE OF DATA PROCESSING METHODS IN THE HERBARIUM

JAMES H. SOPER*

ABSTRACT

This paper describes a system developed in the Herbarium of Vascular Plants (TRT) at the Botany Department, University of Toronto, during the years 1963-1967 and now being introduced and expanded at the National Herbarium of Canada (CAN) in Ottawa. Much of the information has already been published (Soper & Perring, 1967) but changes are reported in some of the procedures and formats previously outlined.

Descriptions are given of the application of data-processing techniques to routine operations in a herbarium such as the preparation of *a*) catalogue records or other index entries; *b*) labels for herbarium specimens; *c*) lists of exchange and loan material; *d*) inventories; *e*) distribution maps. A discussion is included of some aspects of a general computerized search program being developed for retrieving data from the information system. Illustrations are provided to show the formats of the catalogue record form, herbarium label and tab-card as well as samples of distribution maps plotted by machine.

RESUMEN

Este trabajo describe un sistema desarrollado en el herbario de plantas vasculares (TRT) en el Departamento de Botánica de la Universidad de Toronto durante los años 1963-1967 y ahora introducido y ampliado en el Herbario Nacional de Canadá (CAN) en Ottawa. Mucha de esta información ya fue publicada (Soper & Perring, 1967) pero muchos cambios fueron introducidos en algunos de los procedimientos y formatos previamente dibujados.

Se presenta la descripción de la aplicación de técnicas de procesamiento de datos a operaciones de rutina en un herbario, así como la preparación de:

- a*) Registro de catálogos o de otros índices de registro.
- b*) Etiquetas para ejemplares de herbario.
- c*) Listas de canje y donación de materiales.
- d*) Inventarios.
- e*) Mapas de distribución.

Se incluye una discusión de varios aspectos de un programa de investigación general que está siendo desarrollado para recuperar datos a partir del sistema de información.

Se presentan ilustraciones para mostrar el equipo utilizado en el herbario, así como las formas de registro de catálogos de datos, etiquetas de herbario, mapas de distribución y diversas listas obtenidas por medio de las computadoras y equipo periférico.

* National Museum of Canada.

All herbarium curators and most plant taxonomists are thoroughly familiar with the various operations which are carried out in a herbarium from the time a botanical specimen is received until it is inserted in the collection or sent out on loan or exchange. These operations may begin with the preparation of dried pressed specimens from freshly picked samples but more frequently specimens are received already dried and pressed. The usual sources of specimens are the collections made in the field by professional or amateur botanists associated with the herbarium, material received on an exchange basis or by gift or purchase, and samples sent in for identification. The initial treatment may involve assigning an identification number to a specimen or a "lot" of specimens, routine fumigation, and temporary storage with or without a preliminary sorting of the specimens within a "lot" or collection.

Up to this point the operations are mostly manual ones, usually including the entry in a record-book or card file of the relevant identification number assigned upon receipt of the material. There is little need for the introduction of data-processing techniques at this stage, although the records of incoming material might be transferred to tab-cards and thence to computer storage for the production of a general inventory which could be up-dated periodically and searched when information was required.

The next step in processing specimens for the herbarium is usually identification or verification of the names supplied by the collector. An important exception is exchange material received with labels carrying identifications, these identifications in some cases being accepted at least for the purpose of a primary distribution of specimens within the herbarium. It is to the subsequent operations that certain data-processing methods may be applied

with various benefits accruing directly and indirectly to curators, taxonomists, and phytogeographers. These operations include the preparation of: *a*) catalogue records or other index entries; *b*) labels for the master set and for duplicate sets (replicates); *c*) lists of exchange or loan material; *d*) inventories, and *e*) distribution maps.

Catalogue Records. Some herbaria attempt to keep a catalogue or index of their specimens but frequently this is abandoned as the collection grows, partly due to the time required to produce the catalogue but more often, perhaps, because of the difficulties of correcting the records when identifications of specimens are changed by annotation.

Although some botanists have argued that a herbarium is an index, there is an important difference between the two. An index is a listing of certain items (entries) with a reference to the location where that item exists or has been described or discussed. From that standpoint an index is not usually needed to find a specimen in a herbarium since the specimens are filed either in an alphabetical sequence or by some known system of classification reflecting relationships. The specimens, however, are normally accompanied by certain data (on the labels) which refer to their origin or former location as well as to other pertinent information of interest to the taxonomist or phytogeographer.

Furthermore, the specimens themselves represent considerable additional information in the form of their inherent characteristics. Thus a specimen is much more than an "entry" with a cross-reference, it is a primary source of data. The true purpose of a catalogue or index, then, should be to make available the information on the label and, if feasible, some of the characteristics of the specimen itself. It has been said that one prominent museum di-

rector defined a museum as "a collection of accurate labels adequately supported by good specimens".

With modern data-processing equipment such as the Friden "Flexowriter", the SCM "Typetronic" and the IBM "870" Document Writing System, it is possible to automate the recording of herbarium specimens to produce a catalogue in almost any desired form such as cards (or record slips) for an index file, printed lists which can be bound in book form, or punched tab-cards which can be sorted mechanically before listing or read into computer storage for subsequent sorting to produce lists by taxon, by locality, by collector and so forth.

In a small herbarium (i.e. one with less than 50,000 sheets), it should be possible to record the complete collection over a period of a few years. A complete catalogue, which could be periodically updated, would be a useful reference in such a herbarium. At the other extreme, it would be a tremendous job to record all the material in a herbarium having a million or more specimens. There would be some value, however, in recording parts of such a large collection, for example, families or genera under taxonomic study, material being sent out on loan, or specimens forming the basis for distribution maps.

Preparation of Labels. The disadvantages and dangers in the use of handwritten labels have been discussed by Soper & Perring (1967) and by others. Typewritten labels or labels prepared by letterpress or by some method of copying the printed or typewritten characters are preferred by most curators. Some of the modern machines currently used in business operations are ideal for label preparation. (Scheinvar & Gómez-Pompa, 1969) At the University of Toronto Herbarium (TRT) a Friden Flexowriter is used to prepare labels for specimens after they have been identified and are ready to be mounted. At the herbarium the prepara-

tion of labels is combined with the recording of specimens for a Catalogue of the Vascular Plants of Ontario and with the preparation of a tab-card for inventory and mapping purposes. A brief description of this system follows.

After having been identified, specimens are laid out on herbarium sheets each of which bears the herbarium stamp and an accession number. The specimens may be sorted according to the number of labels required for the corresponding series of replicates (or duplicates). Each specimen is recorded on the Flexowriter which produces a catalogue entry form (fig. 1) for the file —on a slip of paper 3 1/2 inches by 5 inches (appr. 9 cm by 12.5 cms). The recording is done by a typist operating a keyboard essentially like that of an electric typewriter. The Flexowriter is controlled by a "Master Tape" which makes machine operations such as skipping and carriage returns automatic. The control tape can also be made so that fixed data such as dates or names of localities or collectors are automatically entered in the correct positions on the form. During the typing of the data pertaining to a single specimen, a complete record is obtained in the form of a punched paper tape generated by the Flexowriter. Errors found in the catalogue entry can be corrected in the paper tape.

The error-free or corrected tape can be processed later by the Flexowriter on an appropriate label form (fig. 1) to produce the necessary number of duplicate labels. The two main advantages are that the labels can be produced automatically by the machine with a minimum of supervision and that the labels for duplicates are error-free and exactly alike. Normally a batch of specimens would be processed on the same tape so that it is an advantage to have the material pre-sorted into lots which require the same number of labels.

Exchange and Loan Records. In the system developed at the University of To-

ronto, the operation of recording a specimen also produces an 80-column punched tab-card. This is accomplished by having an IBM 026 Printing Punch attached to the Flexowriter by means of a TCPC unit (Tab-Card-Punch-Control). The tab-card (fig. 2) carries a code number for the taxon (genus, species and one sub-specific category), a geographic code for the area, the latitude and longitude for the collection locality, the name of the locality, the date of collection, the name of the collector and his field number (collection number), the code for the herbarium and the accession number of the sheet on which the specimen is mounted. Thus these 80-column cards can form the basis for lists of specimens sent on exchange or on loan to other institutions. The cards can be sorted mechanically by taxon, by area, by locality, by date, by collector, by collector's number, by herbarium or by accession number.

Inventories. The recording-labelling system just described provides the basis for two types of inventory. The tab-cards produced as a by-product of recording specimens can be used to prepare lists of specimens by taxon, by locality or by collector. A complete inventory would, of course, require the recording of all the specimens in the herbarium, in this case (at TRT) some 150,000 sheets, a task which could well take several years unless a second Flexowriter were available for full-time use. In recording specimens already mounted and labelled, the program is simplified to eliminate the production of labels from the by-product paper tape. It has already been suggested that a partial or selective inventory would be of considerable value, for example in monographic studies or in the collection of distribution records for mapping.

A second type of inventory could be produced from the paper tape generated during recording of specimens on the Flexowriter. This tape contains more in-

formation than is punched on the 80-column tab-cards and all this information could be stored in a computer for the preparation of a more detailed inventory. It is questionable whether the additional data would really be required for the purpose of a general herbarium inventory but it would be useful in a program for the storage and retrieval of information relating to the specimens themselves.

Distribution Maps. The original purpose of using data-processing methods at the University of Toronto was to convert from hand-plotting to machine-plotting techniques. Details of a series of tests carried out on a variety of different mechanical plotters have been reported in two previous papers (Soper, 1964, 1966). During the summer of 1967 a program was developed at Toronto which plotted records on a base map of Southern Ontario. The input for this program was 80-column tab-cards produced as a by-product of the recording of specimens on the Flexowriter. The plotter used was a CalComp Plotter (Model 565) and examples of the output are shown in figures 3, 4 and 5.

These maps are not entirely satisfactory for direct use in publication. Some of the lines of the base map are not smooth lines and the symbols should be firmer, darker and neater in outline. These are technical difficulties which may be overcome by modifications in the equipment and in the procedures used. For example, the base-map can be preprinted on the plotting paper from a high-quality original with smoothed lines. Some improvement could probably be made in the shape of the symbols and in their size relative to the scale of the map. Alternatively, there are now available other types of plotter which could print the symbols instead of drawing them on the map. Thus future developments should result in maps which could be used directly as "final copy" for publication.

There are only two items on the tab-card which are essential for plotting the record. These are the code numbers for the taxon (to allow selection of records for a single taxon) and the latitude and longitude of the locality to be plotted. The mapping program converts the latitude and longitude (given in degrees and minutes) to the corresponding X and Y co-ordinates for the relevant base map. With appropriate computer programs, it is possible to plot the same records (i.e. for a particular vascular plant) on a map of Southern Ontario, on a map of Canada or on a map of North America. Since latitude and longitude are used in the plotting program, it is helpful when collectors have included this information on the labels. When it has not been provided, it must be found from a gazetteer or map before the record is processed on the Flexowriter. Fortunately the latitude-longitude system is the most familiar one in world-wide use today and it is available on both large — and small — scale maps all over the world. Geologists who have made exhaustive studies of the application of data-processing techniques to the storage and retrieval of their data have also concluded that the designation of a locality by its latitude and longitude is the most feasible and desirable system (Robinson *et al.*, 1967). If other systems are used, conversion of the geographic co-ordinates is necessary before the records can be plotted by the existing program.

One additional refinement in the program is the selection of different plotting symbols which is accomplished by inserting an appropriate single-digit code in the tab-card. Thus specimen records can be distinguished from literature reports and from sight records or other reports. A change in the use of one column on the tab-card format shown in figure 2 allows the insertion of the appropriate symbol-identification code during the recording process. Column 11, labelled "F" for taxonomic form, is used as format for the

map symbol. Taxa are now coded only as trinomials.

The recording-labelling-mapping program developed at the University of Toronto is being initiated at the National Herbarium of Canada (CAN) in Ottawa, where it is planned to make two further extensions of the system. The first is the preparation of a suitable base map of Canada and the pertinent computer program to plot distribution records for the vascular flora of Canada. Preliminary tests of a program to plot records on a simplified base-map of Canada were carried out at the University of Toronto. The second extension concerns the development of an information retrieval program.

Information Retrieval. The production of 80-column tab-cards is considered a temporary measure. Initially there are some advantages such as the accumulation of a file which can actually be seen and handled. Cards can be added and deleted or replaced if they contain errors. They are also accessible for sorting, listing and for input to the mapping programs.

Ultimately the need for tab-cards will disappear as a result of the development of an information-retrieval program. In such a program all or most of the information relating to herbarium specimens and other types of distribution records can be placed in computer storage under a set of standard headings. With a suitable program, information can be found by a computer-search and printed out in any desired format. It will also be possible to extract geographical data and prepare a magnetic tape to operate a mechanical plotter.

In terms of the present system based on the Friden Flexowriter, it will be necessary to add certain control codes to the punched paper tape so that all the data may be transferred to computer storage. The paper tape records will be converted to a magnetic tape file and stored either on magnetic tape, on disks

or in some other form. In the near future a conversion to a more versatile type of equipment may result in a more efficient method of producing labels together with the direct production of magnetic tapes for input of the data into computer sto-

rage. The present system is flexible and can form a suitable basis both for the development of an information-retrieval program and for a conversion to operation on a system using more sophisticated equipment.

LITERATURE

- ROBINSON, S. C., *et al.*, 1967. A National System for Storage and Retrieval of Geological Data in Canada. v + 175 pp. Pub. by the *National Advisory Committee on Research in the Geological Sciences*. Available from the Geological Survey of Canada, Ottawa.
- SCHEINVAR, L. & A. GÓMEZ-POMPA, 1969. Algunos métodos automáticos para la elaboración de etiquetas de herbario. *Bol. Soc. Bot. México* 30: 73-93.
- SOPER, J. H., 1964. Mapping the Distribution of Plants by Machine. *Canad. J. Bot.* 42: 1087-1100.
- SOPER, J. H., 1966. Machine-plotting of Phytogeographical Data. *Canad. Geogr.* 10: 15-26.
- SOPER, J. H. & PERRING, F. H., 1967. Data Processing in the Herbarium and Museum. *Taxon* 16: 13-19.

2541 003 2 2

Anemone multifida Poir. var. hudsoniana DC.
forma sanguinea (Pursh) Fern.

PROVINCE **ONTARIO**

COUNTY **MANITOULIN**

LATITUDE
45° 39' N.

LONGITUDE
82° 57' W.

HABITAT **In clearing of gravelly wooded ridge.**

LOCALITY **GREAT DUCK IS. Loc. No. 552**
Woods north of the main E-W trail to the lighthouse.

DATE **28 June 1959** COLLECTOR & NO. **J.H.Soper** | **8695**
& C.E.Heidenreich

ACC. NO. & HERBARIUM **TRT 117909** DET. BY **JHS**

DATE SEEN **06/01/67** FLOWER **X** FRUIT

NOTES **See Kodachrome JHS No. 59-3-34** NO. OF REPLICATES **2**

PLANTS OF ONTARIO

UNIVERSITY OF TORONTO (TRT), TORONTO 5, ONTARIO, CANADA

REF. NO. 2541 003 2 2

Anemone multifida Poir. var. hudsoniana DC.
forma sanguinea (Pursh) Fern.

COUNTY **MANITOULIN** 45° 39' N. 82° 57' W.

HABITAT **In clearing of gravelly wooded ridge.**

LOCALITY **GREAT DUCK IS. Loc. No. 552**
Woods north of the main E-W trail to the lighthouse.

DATE **28 June 1959** COLLECTOR & NO. **J.H.Soper** | **8695**
& C.E.Heidenreich

DET. BY **JHS** NO. OF REPLICATES **2**

NOTES **See Kodachrome JHS No. 59-3-34**

Fig. 1. Examples of a Catalogue Record Form (above) and the corresponding Herbarium Specimen label (below) prepared on a Flexowriter.

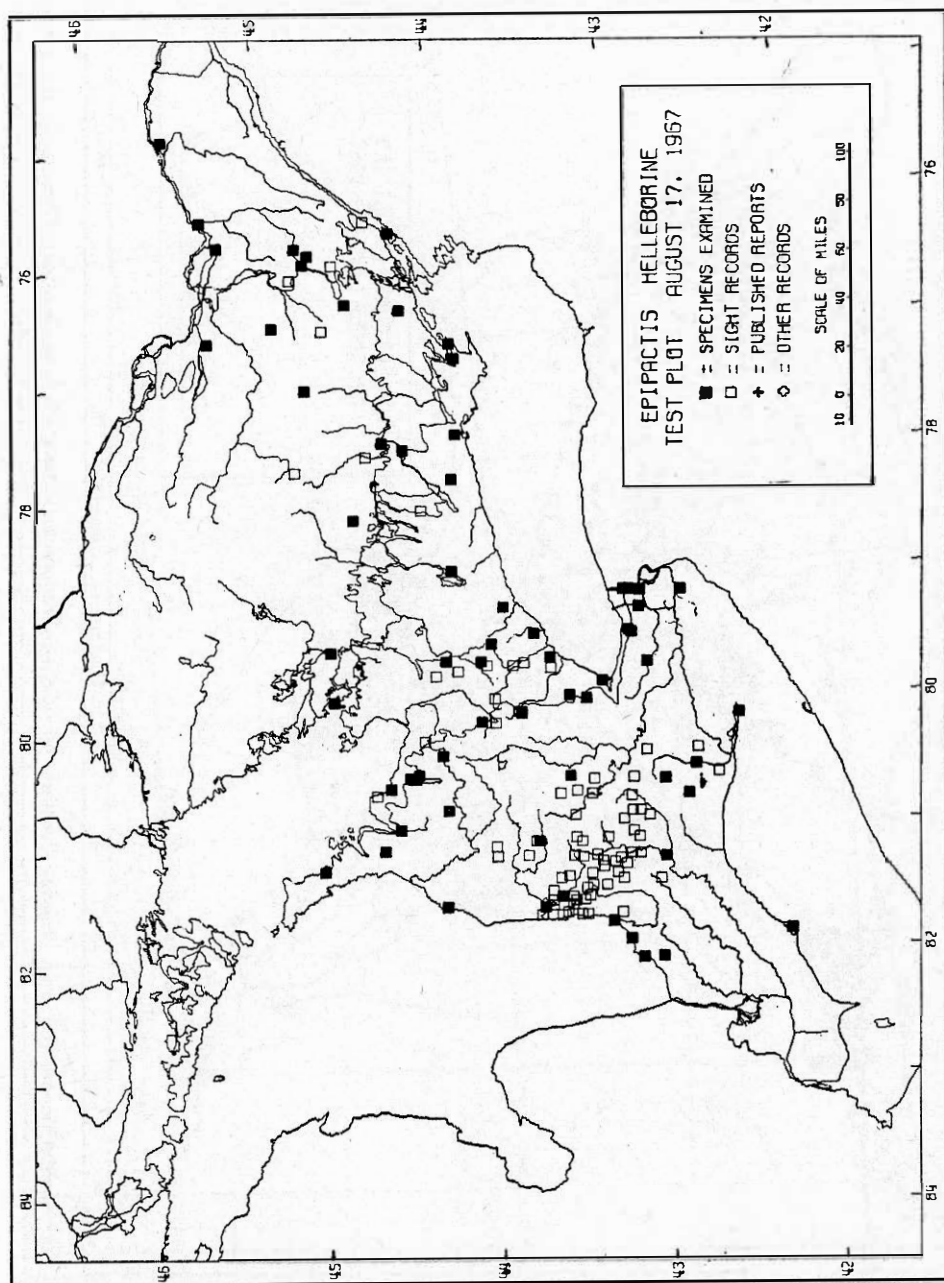


Fig. 3. Machine-plotted map showing distribution records, for an introduced orchid (*Epipactis helleborine*), in Southern Ontario, Canada.

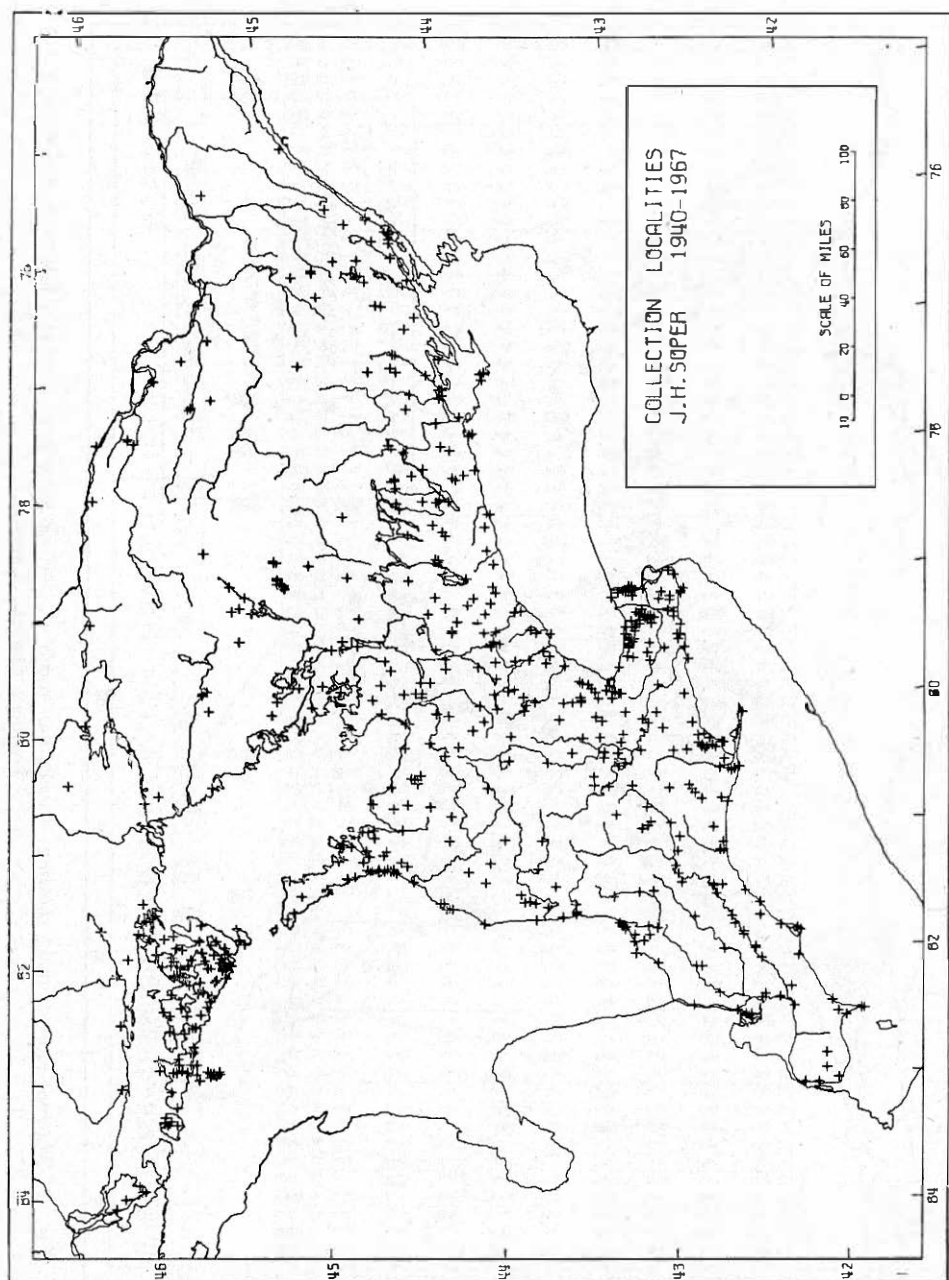


Fig. 4. Machine-plotted map showing localities at which collections have been made in Southern Ontario by the author.

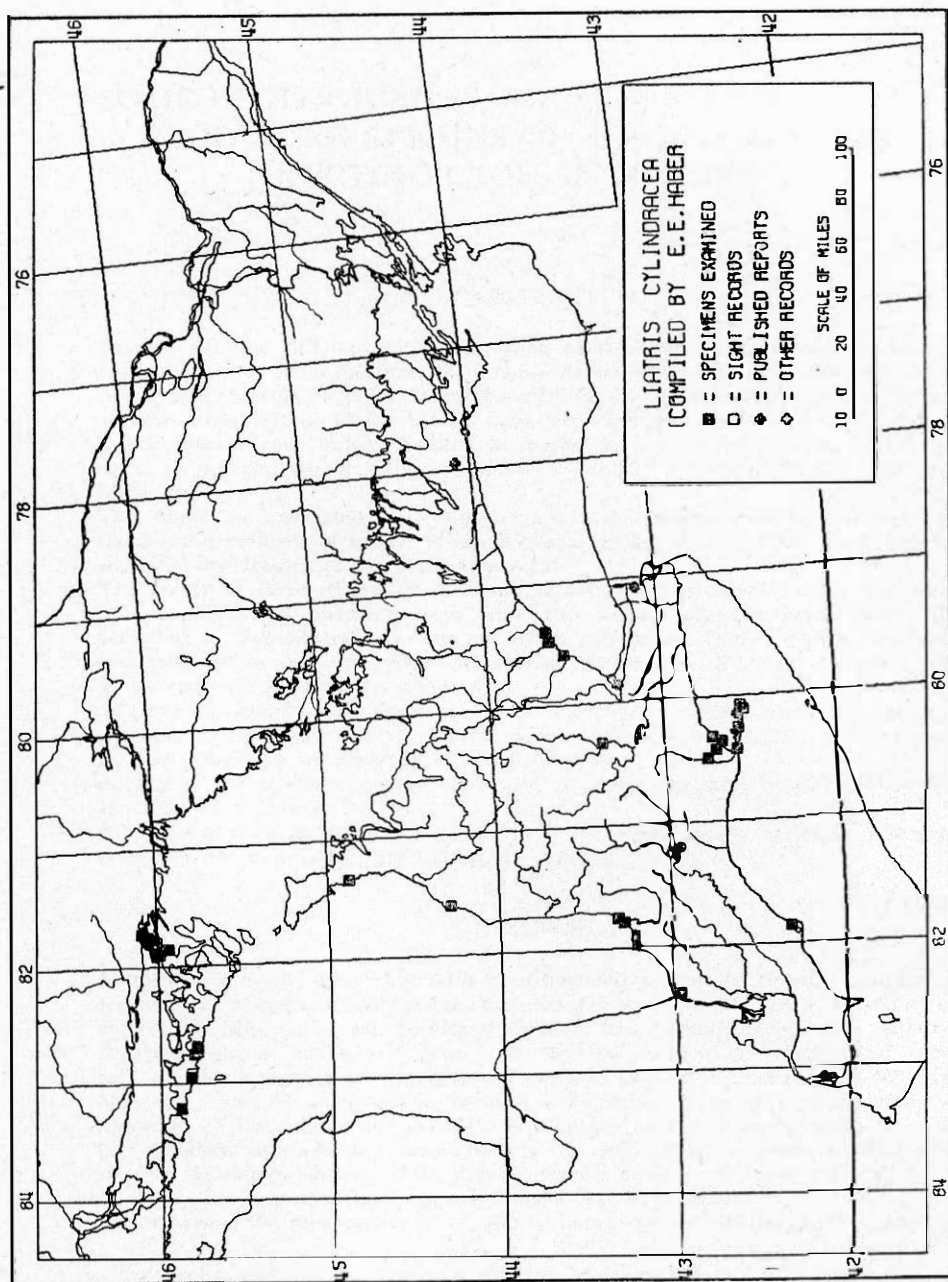


Fig. 5. Machine-plotted map showing distribution records for a prairie composite (*Liatris cylindracea*) in Southern Ontario. (The mapping program allows the choice of adding or omitting the latitude-longitude grid.)