Prospective study of a computerized system for nematode collection records.
B. Proposals for a central computerized system.

1. INTRODUCTION

Among other duties, curators of nematode collections answer requests for slides of a particular species. These slides generally are arranged in boxes within a slide cabinet. Usually, each box contains slides belonging to a particular genus or family. This traditional arrangement of slides, according to the taxonomic position of the contained specimens, is not compatible with computerized records. Results of electronic searches identify slides by their serial accession number, and the slides must be arranged by this accession number, regardless of classifications.

Many curators believe that arrangement based on systematic classification is an easier way to locate a slide. This may be true for a small collection including only a few taxonomic groups. However, in general and particularly for larger collections, this arrangement presents many problems.

Each change in classification--and such changes are particularly frequent with plant nematodes--would require that slides are rearranged into new boxes corresponding to the newly created genera and families. This is not likely to be done routinely, but only for genera or familes of particular interest for a scientist. Year after year, the collection progressively becomes a motley assemblage of boxes representing various concepts of nematode classification, differing from group to group according to the prevailing classification at the time when scientists started studying each of them.

Nontaxonomic considerations may also be apparent in the arrangement of boxes. In a collection, there may be a box marked "Tylenchus", but another box marked "Tylenchus from Brazil" may be stored in another drawer with an extensive collection of Brazilian specimens. A student of the genus Tylenchus will have to research the entire collection, box by box, to locate all relevant slides.

Slides with several specimens belonging to different genera can be stored under only one of these genera. There is no easy way to locate the other genera on the slide.

Slides are often put back into the wrong box. Subsequent failure to locate a slide is very frustrating - is the slide missing or destroyed, or is it hidden somewhere among the thousands or the tens of thousands of slides that consitute the collection?

By contrast with the traditional, taxonomic arrangement, slides arranged by their serial accession number are easily located and easily replaced in their correct locations. Though it is inconvenient to do a computer search to locate slides with a particular taxon, a computer printout of

the collection records arranged by taxonomic categories easily can be prepared. Such a printout would immediately give the accession number of the slides needed.

Electronic searches are needed, not only to locate slides belonging to a particular taxon, but also for most kinds of studies. For example, a nematologist may want to establish the list of hosts for a particular nematode species S. The usual approach to such a question is to search the literature. However, a cautious compiler should check the accuracy of some of the published data. Also, additional hosts may have been identified for species S, but never published. A computerized general system would provide a list of all slides with species S on various hosts in the nematode collections in the world.

Modern taxonomists have moved away from the typological concept of the species (a species is completely defined by its type), to a statistical one (a species is the statistical distribution--described by mean and standard deviation--of all the individuals in all the populations that constitute it). Description of a species would then require the study of several representative samples. A computerized central system would facilitate localization of suitable sized samples of the species to be redescribed.

There should be no doubt that centralized computerization of collection records would be most beneficial for nematology. At the same time, the survey of the current situation in nematode collections revealed the obstacles such a project would have to overcome. The system proposed below is believed to make the best use of the limited resources in time, personnel, and money available to most curators. The project is divided into several steps. Each step is easy to achieve and yields immediate results. This will boost the morale and confidence of all persons involved, and also make it easier to request additional fundings.

2. OUTLINE OF THE PROJECT

Step 1: Creation of individual minimal in-house database.

Step 2: Use of any available records to complete the information stored in each database.

Step 3: Creation of a common central database (SLIDE) by downloading each individual database as soon as it is completed into a central computer.

Step 4: Splitting of SLIDE into two databases: SLIDE and SAMPLE each with selected fields of data. SLIDE and SAMPLE will be connected to each other and to other databases (NOMEN, REFERENCE) by a relational database management system (UNIFY).

3. DETAILED DESCRIPTION OF EACH STEP OF THE PROJECT

3.1. CREATION OF MINIMAL IN-HOUSE DATABASES

This first step will have to be taken separately and independently by each collection curator who wishes to computerize his/her records. As a matter of fact, several curators have already started or are planning to start the computerization of their data in the near future. For the moment, the individual methods used are not compatible. Only if curators are willing to adopt a common method for data entry will they later be able to download their data into a common central system.

It is important that the first stage of the project be as simple and as short as possible while yielding immediate results. For Step 1, it is proposed to enter only the minimum amount of information. Ideally, only the information available from the slide labels should be used. In this way, the minimal database is created directly from the existing slides, not from records that may be obsolete or erroneous. For large collections, it may be necessary to subdivide Step 1 into smaller endeavours--type material database, voucher specimen's database; or plant nematode's database, free-living nematode's database, etc.

Purposes of minimal in-house databases.

1. Create an updated and complete catalogue of slides in the entire collection (or in a defined section of the collection);

2. Provide answers to some cross searches;

3. Create a basis for further developments (Steps 2 to 4); and

4. Provide easy records of slides on loan.

Definition of a common structure for the individual databases.

It is proposed that all individual databases be established as ASCII files and with structure of fields of data defined as follows (note that the fields of data are designed primarily for plant parasites. They may have to be modified to accommodate insect-parasite,free-living or marine nematodes):

Field #	Field Name	Field Length	Comments
1	COLLECTN	10	Abbreviated name of the nematode collection (Ex: UCDNC for "University of California, Davis, Nematode Collection). For personal collection use owner's name plus "PC" (Ex: ANDRASSYPC). Eventually, ask central project manager whether a name is available (i.e., not already in use for another collection).

Field #	Field Name	Field Length	Comments
2	SLIDENBR	6	Slide accession number (1, 2, 3, etc.) or slide number using Thorne's system (27e, 9g, 10h, etc.) or any system currently in use in your collection.
3	FEMALES	2	Enter number of specimens: females, males, juvenile, or other (cysts, perineal patterns, cross sections, face views, etc.) in the slide.
4	MALES	2	
5	JUVENILS	2	
6	OTHERS	2	
7	SPNAME	75	Enter the species name (binomen, trinomen) exactly as noted on the slide label. If there are several species on the same slide, create as many records as necessary.
8	GENUS	25	Enter known information only if species is not identified. If nothing is known at all, enter Y for yes into field 11 (UNKNOWN).
9	FAMILY	25	
10	ORDER	20	
11	UNKNOWN	1	
12	IDENTIED	20	Enter name of person who identified species, and date of identification, as in "Smith, 15 NOV 1984". If the identification is later contested, enter name and date of new identification, or reference to a publication (see field #13 on how to enter references). If a new specific name is proposed, open a new record for it (see note below).
13	REFERNCE	8	Enter reference of article where the species is described, redescribed, named, or used for a specialized study. Use Helminthological Abstracts for a coded access to a complete reference: "HA50 816" means "see Helminth. Abstr. (B) volume 50, abstract number 816".

Note: Fields 3 to 13 refer to the specimens of one species on the slide. If a slide includes specimens of several species, these fields will have to be duplicated as many times as necessary.

Field #	Field Name	Field Length	Comments
14	SLDATE	11	Enter date slide was made. Use day-month-year format as in "15 NOV 1984".
15	PURPOSE	8	Enter: TYPE for holotype, paratypes, neotype, topotypes, etc.; VOUCHER for specimens used for an article (description, survey, host, distribution, ecology, control, etc.); TEACHING for slides to be used by students; STOCK for slides in the general collection without specific purpose; and OTHER describes any other usage.
16	FIXATION	10	Enter method for fixation and mounting.
17	LOAN	20	Use this field to indicate that the slide is on loan. Enter name of person who loaned the slide and date of loan. Delete the field when slide is returned.
	Note: Fields 1, 2 and 14 to 17 refer to the slide itself.		
18	SAMPLNBR	10	Enter field sample accession number as used in your collection.
19	COORDINT	15	Enter geographical coordinates as in "038.33N 121.30W".
20	UTM	10	Enter locality using UTM grid system.
21	LANDMARK	50	Describe locality using landmark references as noted on slide.
22	AREA 1	15	Enter locality using administrative and geographical subdivisions at various levels. For example: AREA 1 = North America, AREA 2 =
23	AREA 2	15	
24	AREA 3	15	
25	AREA 4	15	

Field #	Field Name	Field Length	Comments
			USA, AREA 3 = California; AREA 4 = Yolo County.
26	HOSTLATN	30	Enter host plant's Latin name.
27	HOSTCOMN	20	Enter host plant's common name.
28	HOSTRMRK	20	Enter any other information about identity of host: cultivar, variety, hybrid, etc.
29	PLANTPRT	11	Enter part of plant sampled (ROOTS, TUBER, LEAVES, FRUIT, etc.) or enter RHIZOSPHERE (vicinity of plant roots) or SOIL (bare soil).
30	SADATE	11	Enter date sample was collected.
31	COLLECTR	10	Enter name of collector or name of person who donated the slide.
32	VIAL	1	Enter Y (yes) or N (no) to indicate if a wet collection has been preserved.

Note: Fields 18 to 32 refer to the sample from where the specimens on the slide were extracted.

Problems and Remarks

Any project for computerization of existing records is bound to encounter snags and problems. These problems will differ from collection to collection and each curator will have to identify and solve them on his own. Curators may want to be aware of the following difficulties. For easier use, various possible problems are arranged under the fields where they would occur.

1. COLLECTN

The name of the collection must be unique. For example, 24 curators identify their collection as "Nematode Collection". It is evident that qualificatives must be used for a centralized database. The name choosen for a collection must be checked against the names listed in Nematology Newsletter, March 1983, 29(1):5-10.

2. SLIDENBR

All slides must have an individual number. A numbering system may have to be created as the first step of the project. If no system exists, it would be best to use

		serial accession numbers (slides marked 1, 2, 3, etc.).
		If a numbering system already exists, occasional slides may be found without a number. This number may be recovered from existing records, or a new number may have to be attributed to the slide. This can be done either at the end of the project or during data entry, depending on the proportion of such unnumbered slides estimated during the "problem evaluation" stage (see below).
		At the end of the project, SLIDENBR will be checked automatically for duplicate slide numbers.
3. to 6.	(Contents)	The actual number of specimens mounted in the slide may differ from the number on the slide label. The proportion of such discrepancies must be estimated during the problem evaluation stage. If the proportion of erroneous label is high, each slide may have to be checked before data entry.
7.	SPNAME	Enter the name exactly as it appears on the slide label. All names will be checked automatically during Step 4 against information in database NOMEN.
8.-10.	(Higher Categories)	Use these fields only if a specific identification is lacking. When specific identification has been made, these higher taxonomic categories will be obtained automatically from a separate database CLASSIFICATION that will perform as a taxonomic dictionary.
12.	IDENTIED	If a specific identification is contested and a different name is proposed for the specimens in a slide, a new record needs to be opened with information on fields 3 to 13.
16.	FIXATION	Fixation and processing methods used for the specimens on the slide need to be identified by the reference to a publication describing the method. A dictionary of these methods will later be created as a separate database.

18.	SAMPLNBR	If field samples are not routinely numbered for a collection, a numbering system will have to be created before Step 4 (creation of SAMPLE database).
19.-25.	(Locality)	The geographical origin of the sample must be described from the information available on the slide label. Categories 22 to 25 will later be obtained automatically from a geographical dictionary available at Step 4.
26.-27.	(Host Name)	Host name must be entered as it appears on the slide label. It will later be checked against a botanical dictionary during Step 4 of the project. Vegetal associations (forest, prairie, desert, grasses, etc.) can be entered in the field HOSTCOMN. If sample comes from bare soil fields 26, 27 must be left blank and SOIL is entered in field 29.

Realization of Step 1

The implementation of Step 1 will better be explained later when the records of a first nematode collection have been actually entered into a computer. The first individual project is supposed to be conducted in 1986-87 at the University of California, Riverside. Once the project is completed, a report will be published to serve as an example of which methods were used, which problems were encountered, and which solutions were adopted.

Very generally speaking, the following points will have to be considered for each individual project:

1. Preparation.

   - Decide on a name for the collection, check for duplicate names.
   - Check if slides have serial accession numbers. Eventually number the slides.
   - Check if the field sample records have accession number.
   - Gather all slides. Check for slides on loan or in use.
   - Gather all records about slides and field samples.
   - Decide which hardware and software will be used to create and later search the inhouse database. Check with the central project manager for future compatibility with the central computer and database management system (UNIX-UNIFY). Any dbms (for example dbase) will probably do, as long as it has a field structure and the field definitions proposed are followed.

2. Estimate nature and size of potential problems.

   - Examine about 150 slides taken randomly from the collection. Note what kind of information can be obtained from the slide labels.

   - Check for broken slides, illegible data, errors on slide labels (contents, species identification). Check for missing data, particularly in the fields:

     2 SLIDENBR

     18 SAMPLENBR

     7 SPNAME

     14 SLDATE

     19 to 25 (Locality)

     26 to 27 (Host)

   - Note what are the percentages of "problem slides" with missing or erroneous information.

   - Decide if these problems must be solved before, during, or after data entry.

3. Test run.

   (1) Take box, empty contents into tray.

   (2) Take first slide.

   (3) Enter data.

   (4) Put slide back into original box.

   Repeat steps 2 to 4 for all slides in the tray.

   (5) Replace box into collection cabinet.

   - Do an actual data entry with 200 slides.

   - Check for problems.

   - Send a copy of this small database to the central project manager for compatibility testing.

4. Enter all data.

   Conclusion of Step 1.

   When Step 1 is completed, the database created can be used immediately by the collection curator. On-line searches can be made on any

subject according to search strategies available with the dbms used. Printout can be created with the information arranged in several ways:

- Slides records arranged by slide accession numbers.
- Taxonomic catalogue with all slides available for each species/ genus.
- Host cross index.
- Etc.

3.2. STEP 2: ENTRY OF ADDITIONAL INFORMATION

Step 1 creates a catalogue of the collection. This catalogue is accurate because it includes all the slides present in the collection and no more than these slides. It may be incomplete because only the information available from the slide labels was used in order to reach a quick and easy completion of Step 1.

Purpose

Other sources of information, such as log-books, index cards, publications, etc., can then be used to check the accuracy of the data and eventually enter additional information. It is also possible to check for missing slides for which records exist but that were not found in the collection cabinet. Step 2 will result in complete and accurate records that can be used for more elaborate searches and that are ready for downloading into the central system (Step 3).

Realisation

Step 2 with verification and additional entries will vary enormously from collection to collection according to the nature, size, and scope of existing collection records.

Difficulties may occur when matching a slide record as entered into the computer with the corresponding sample record unless the sample accession number was noted on the slide label. It will be necessary to decide for each collection if this matching can be done automatically by computer search, or if it is better done manually.

There will be duplication of information when several slides have been made from the same field sample. This will be taken care of during Step 4 (see below).

New fields of data may be needed to enter new information about the slide or its field sample. A list of the data fields needed must be sent for approval to the central project manager.

3.3. STEP 3. CREATION OF A CENTRAL DATABASE

As soon as steps 1 and 2 have been completed for a particular collection, the individual database can be downloaded into the central computer to become a part of the general SLIDE database.

Purpose

SLIDE will include data on more and more collections. It will be easier to locate the slides needed for a study as many records can be searched at the same time.

Realisation

Downloading is easily realized by sending a tape (or a diskette) with a copy of the collection database to the central project manager.

Individual databases will be maintained for local use by their curator. They will be continuously updated with new slide records and new information on old records. The updates must be regularly sent to the central project manager for update of the general SLIDE database.

3.4. STEP 4. CREATION OF RELATED DATABASES

SLIDE will be created in a computer that already stores other nematological databases such as NOMEN (ex NEMAS) and REFERENCE. Other databases and taxonomic and geographical dictionaries are planned in the near future. SLIDE itself will be split into at least two databases: SLIDE and SAMPLE.

Purposes

- Creation of small databases, each gathering information about one small subject, therefore easier to manage;
- Suppression of duplicate information; and
- Creation of a relational structure where all databases will be related, enabling cross searches.

Realisation

The central project manager will take care of step 4.

The original SLIDE database will be split into SLIDE and SAMPLE databases, each being in turn composed of two subdatabases according to the following scheme:

SLIDE				SAMPLE			
Nematode		Slide		Sample		Collection	
1.	COLLECTN	1.	COLLECTN	1.	COLLECTN	1.	COLLECTN
2.	SLIDENBR	2.	SLIDENBR	18.	SAMPLENBR	18.	SAMPLENBR
		18.	SAMPLENBR	26.-29.	HOST	19.-25.	LOCALITY
3.-6.	(Contents)	14.	SLDATE	32.	VIAL	30.	SADATE
7.	SPNAME	15.	PURPOSE			31.	COLLECTOR
8.-11.	(Classification)	16.	FIXATION				
12.	IDENTIED	17.	LOAN				
13.	REFERENCE						

The splitting of either SLIDE or SAMPLE into two subdatabases is needed when there are several nematode species on the same slide, or when several samples have been collected in the same conditions (locality, date, collection), but with different position (multiple sampling of a field) or host.

If enough information is available, SAMPLE-collection may later have to be further subdivided into more databases: LOCALITY, ECOLOGY, ect.

Fields of data in common to two databases (fields: 1. Collection Name, 2. Slide Number, 18. Sample Number) are used to relate each database to the others. Duplicate information, created when several nematode species were mounted on the same slide, or when several slides were made from the same field sample, will be deleted.

SLIDE and SAMPLE will then be connected to other related databases through the relational dbms UNIFY:

NOMEN: Information on current and past nomenclature and taxonomic status of all plant nematodes (already exists).

REFERENCE: Complete references (authors, date, title, journal, pages) to all nematology articles. REFERENCE is being created for taxonomic papers. It is hoped that it will be expanded later to other articles, if possible by downloading Helminthological Abstracts files from DIALOG databank.

Taxonomic dictionary: A dictionary of the classification of genera and families of plant nematodes will have to be created later.

Geographical dictionary: Hierarchical listing of administrative and natural divisions will also be created.

These related databases and dictionaries will provide continually updated information on species name, classification, geographical location, and article references to the SLIDE database. Conversely, a user of the database NOMEN will have access to the information stored in SLIDE. After finding about the name of a nematode species, the user will have direct

access to existing slides with specimens of this species. The relational database system will allow extensive cross searches of many subjects.

Access

It is planned to store all databases into a single computer a Vax-Unix at UC Davis will be used during the period of development of NOMEN and REFERENCE. The final resting place of the system will be decided later. The use of a central computer poses the question on how to access the information.

The best solution would be to allow all nematologists to directly access the computer and do themselves the searches they are interested in. This access would have to be through a communication network such as TELENET to allow faraway scientists to gain access to the system for a reasonable fee. No such on-line access is currently available at UC Davis for worldwide usage.

As a temporary solution, it is proposed that nematologists send requests for specific searches by mail to the central project manager. Forms will be available to define exactly the subject of the search. Requests can be sent by mail or by electronic mail using for example BITNET. The searches will be made at UC Davis by a student for a fixed fee. Print out with results of the search will be mailed back to the scientist.

While this system is far from perfect, it can be used for a while until a high enough number of nematologists are using the system. If the volume of searches becomes high enough, it can be hoped that commercial data-banks, such as DIALOG, will become interested in adding the nematological databases to the services they provide.

4. CONCLUSION

The project presented here is very ambitious. It will take many years to be completed for a significant number of the world's nematode collections. At the same time, each step of the project (particularly Step 1) is a small task that can yield its own rewards for a limited amount of time and money.

Considering the immense potentialities of the completed project with many interrelated databases providing instant, up to date, and complete information on all aspects of nematology, it is definitely worthwhile to take the first steps on the road to electronic data processing.

Prepared by Renaud Fortuner
For Systematic Resources Committee
Society of Nematologists, April 1986

SURVEY

Opinions About the Proposed System

In the attached report, it is proposed that:

1. All nematode slide collection records should be computerized.

2. Individual collections should be computerized independently, but according to a common format.

3. After completion, each individual database should be downloaded into a unique computer to create a central database SLIDE.

4. SLIDE should be connected to other related databases via a relational database management system.

Even if you have no intention to computerize your collection, or to use computerized access to other collections, would you please answer the following questions:

Your Name: Size of your collection: _____ Slide _____ Vials

1. Do you like the idea of computerizing nematode collection records: yes or no, and why?

2. Do you intend to computerize your own collection records?

   a. They already are totally/partly computerized.

   b. I am in the process of computerizing my collection.

   d. I am planning to computerize in _____ months _____years.

   e. I hope to computerize within _____ years, but I am stopped by the following problem:

   e. No, I'll never computerize because:

3. If you like the idea of computerized collection records, do you like the system proposed: yes or no, and why?

4. If you do not like the system proposed, explain what other systems you advocate (either an existing system or an ideal system that would have to be defined):

5. If you have not already started on a project, and if and when you start computerizing your collections, will you use the system outlined in the attached report?

   a. Yes, I'll definitely use it.

   b. Yes, I'll use the format proposed, but I'll not download my records into a central system because:

   c. I may decide to use it, but I want to see it works first.

   d. I don't like it too much, but I 'll use it for want of a better system.

   e. No, I'll never use it under any circumstances.

6. How can the system proposed be ameliorated? Other fields of data needed; other amelioration; general comments.

Please return as soon as possible to:

Renaud Fortuner
California Department of Food and Agriculture
Analysis and Identification, Room 340
1220 N Street
Sacramento, California 95814 USA