Prospective study of a computerized system
for nematode collection records.
A. Survey of current methods

--

In its national plan, proposed in 1973, the Association of
Systematics Collections recognized the importance of Electronic
Data Processing (EDP):

"EDP offers the only realistic mechanism, now and in the
future, by which the enormous data resources of systematics
Collections may be acquired, stored and queried."[1]

The Council on Standards for Systematics Collections of ASC
was established and its goals defined as follows:

"The highest priority of this Council is to develop appropri-
ate standards for collecting and recording specimen-related
data. Without such standards, work of the Council on Elec-
tronic Data Processing cannot go forward."[2]

ASC can certainly provide very useful guidelines, such as the
report compiled by L. Sarasan and A. M. Neuner in 1983, Museum
Collections and Computers. However, it is up to the profession
to adapt these guidelines to the particular needs of nematology
and make them accessible to all nematologists.

Currently, computers are hardly used by nematode collection
curators.   Some curators know that data recording on tradi-
tional supports (such as ledgers or index cards) is time consum-
ing, error prone, and difficult to edit and update. They know
that retrieval of stored information is difficult except when
using a cumbersome cross-index card system. They are aware of
the advantages of EDP, e.g., data entered only once in inter-
related databases, word processing for easy corrections, automa-
tic spelling check, and infinite possibilities of cross searches
on any subject. A few curators are planning or are actually
working on computerization of their collection records.

However, very few collections have been actually computerized
and some attempts at computerization have failed.   Investi-
gation and evaluation of available EDP systems, and prepara-
tion of guidelines for the particular requirements of a nematode
collection is a difficult task for which most curators have no
time or experience. It would benefit nematology in general to
identify the problems raised by computerization of nematode
collections and to offer solutions to these problems through
appropriate guidelines.

---

[1] & [2] America's systematics collections: A national plan.
Irwin, H. S.; Payne, W. W.; Bates, D. M; and Humphrey, P. S.
(Eds.), Assn. Syst. Coll. (Publ.), Dec. 1983, xiii + 63 p.

The Systematic Resources Committee of Society of Nematologists has initiated a prospective study of a computerized system for nematode collection records to: (i) describe the various systems currently used to record and search data on nematode collections, (ii) identify the problems attached to these systems, (iii) define the utilization of nematode collection records by nematode systematists and other categories of users, and (iv) propose an EDP system that will solve these problems and answer the needs of the various users of nematode collections.

The study of the Systematic Resources Committee will take several steps. The first step was a survey of the collection record systems currently in use. The survey ended May 31, 1985 and the answers received are analyzed below.

1. The questionnaire.

    A questionnaire was designed by the Committee and sent to 66 curators of nematode collections in the world, as identified in Nematology Newsletter 29(1):5-10. The questionnaire was in two parts (a fact-finding survey and an opinion survey). The former was to provide answers on points (i) and (ii) as described above. The latter was to give the Committee some hints on how people felt about this project and what would be the major blocks to computerization.

    Answers received uncovered some shortcomings of the questionnaires, i.e., questions badly worded and misunderstood, and use of terms unknown by the surveyed scientists unfamiliar with computer jargon. Also, the first questionnaire was oriented towards plant nematology, and curators of marine collections had difficulty answering it. In spite of these limitations, the questionnaire fulfilled its purpose and provided the information needed.

2. The answers.

    Forty-one of the sixty-six questionnaires sent out to collection curators were returned in time to be included in the present report. Half of the questionnaires sent in the USA and Canada were returned (13/26), but only 35 to 38 percent from other industrialized countries (10/26), and from third world and eastern bloc countries (5/14). The results of the survey are heavily biased in favor of North America, probably because scientists are more familiar with computers in this region. Answers came from small collections (a few hundred slides), medium-sized collections (a few thousand slides), and large collections (several tens-of-thousand slides including the 75,000 slides of the largest nematode collection in the world, USDA, Beltsville).

Below is given a list of the questions asked and the number of affirmative answers for each question. Because of the biases discovered in the geographical origin of the answers, no percentage is given. These answers do no represent the situation in all nematode collections in the world.

a. Fact-finding survey (27 answers).

(1) Where do you keep your collection records?

- No records (other than what is written on the slides) are kept.	3
- Ledgers.	9
- Index card system.	20
- Computerized system.	2
- Original reports, correspondance, etc.	1
- In a transition phase to computerized system.	2

(2) How good is your record keeping?

- No records are kept.	2
- Poor and out of date.	5
- Poor, but up to date.	3
- Good, but out of date.	6
- Excellent: good and up dated continously.	9

(3) What kind of data is recorded?

Record Identification

- Name of the collection.	12
- Reference number of the record.	19
- Type; original number	1

Description of the Collection Item

- Nature of the support (slide/vial/SEM stub/other).	14
- Processing (killing/fixing/mounting).	12
- Total number of specimens.	19
- Number of specimens by stage/sex.	19
- Quality of specimens.	8
- Date of mounting.	12
- Name of slide preparer	1

Origin of Specimens

- Locality.	26
- Host.	22
- Parts of the plant sampled.	14
- Crop and farming system.	7
- Climate, soil, and ecological conditions.	10
- Date of sampling.	21
- Collector's name.	22
- Extraction method.	4

- Date of extraction.   3
- Donor.   1
- Habitat (marine nematodes).   2
- Date received (for donated specimens).   1

Identification of Specimens

- Family name.   7
- Genus name.   27
- Species name.   27
- Change in nomenclature.   8
- Identifier's name.   16
- Date of identification.   4
- Multiple identifications (question misunderstood:
  answers deleted).   5
- Other taxonomic categories (phylum, class, order, etc.).   1

References

- Reference to articles where the specimens are described
  (or utilized if voucher specimens).   8

(4) What kind of information can you search in your
     current records.

- Search of a species-type material.   19
- Search of all populations of a species.   15
- Search of all slides of a genus/family.   18
- Search by host plant.   9
- Search by locality.   12
- Search by size of sample.   2

Cross Searches

- Can you cross search by 2 items (ex: species S AND host H)? 11
- Can you cross search by 3 items or more (ex: S AND H AND
  sample size > S)?   5
- Can you use bolean operators?   1

(5) What kind of problems do you experience in recording
     the data?

How do you handle misidentified species?
- Change only the slide label.   2
- Add correct name to old record.   1
- Make new record; save old one.   4
- Make new record; delete old one   5

How do you record specimens not yet identified?
- Keep out of collection without any record.   5
- Kept in separate collection.   7
- Included in collection under known taxonomic categories
  (family, genus) or under an ID number.   13

- Included in collection under tentative identification. 1

How do you treat nomenclature changes?
- Not considered: original name is kept. 3
- New name is added to old record ("remarks"). 6
- Make new record, save old one. 4
- Make new record, delete old one. 4

How do you record the name of host?
- Always common name. 3
- Always Latin name. 4
- Either or both. 13
- May also record plant associations (grass, prairie, forest, etc.). 1
- None apply (marine nematode). 2

How do you record slides with specimens belonging to several species/genera?
- Several records are made. 9
- Main record and secondary records cross referenced to main record (or single card with several perforations). 9
- None apply: slides kept monospecific. 3

How do you record the geographical location?
- Landmark (city, roadmarkers, rivers, etc.). 9
- Administrative divisions (county, state, province, etc.). 2
- Either or both the above, with or without geographical coordinates. 13
- Marine nematodes (depth, etc.). 1

What kind of reference number are you using?
- Accession number (1, 2, 3, etc.). 11
- Thorne's system of code numbers and letters for genus and species (15f, 33g, etc.). 5
- Accession number plus code letters for geographical origin (ex: ADK101 for slide 101 from the Adirondak collection). 6

(6) What kind of problems do you experience when searching the records?

- Records incomplete or out dated. 12
- Cross searches difficult. 10

(7) Loan of specimens.

How often do you loan slides?
- Daily. 0
- Weekly. 4
- Once a month. 8
- Once a year. 8
- Almost never. 5

Do you place restrictions on the loan of certain
kinds of slides?
- No restrictions.   7
- Restriction on loan of holotype.   13
- Restriction on loan of paratypes.   7
- Restriction on loan of nontype material.   6

If you have some restrictions, what kind of
restriction do you place on slide loan?
- Loans limited to recognized organizations or scientists.   16
- Specimens not sent outside but available for study at the
  laboratory housing the collection.   3
- Slide not available under any circumstances.   0
- None apply.   4

(8) What kind of problems do you experience with loan
     of specimens?

- Problems associated with record keeping (difficulty in
  locating the requested slide).   0
- Difficulty in answering specific requests.   0
- Slides lost or damaged by users (indicate how often such
  accidents occur).
  Often.   4
  Very seldom.   14

(9) What is the one major drawback of your current system
     of record keeping?

- No drawback.   6
- Cross searches difficult.   6
- Records out of date.   5
- Lack of time/personnel for data entry.   6
- Lack of time to keep track of loans.   1
- Lack of knowledge of systematics.   1
- Lack of funds.   1

  Other drawbacks not quite related to record keeping
  included:
- Safety: only one copy of the records exists.   1
- Slides said to have been deposited in the collection
  but never sent by authors.   2

b. Opinion survey.

   The collection curators were asked to consider the following
   statements and to indicate if they agree, disagree, or have
   no opinion about each one. A total of 26 answers were
   received.

	Agree	Disagree	Don't Know or no Answer
The current record keeping system works fine; I feel we don't need to use a computer at this time.	9	10	7
Computerized records would be limited in their usefulness, because many labs have no access to computers and could not use our system.	8	12	6

I would like to computerize my collection records, but I don't have:

	Agree	Disagree	Don't Know or no Answer
The money to do it.	11	4	11
The time to do it.	14	6	6
The personnel to do it.	15	5	6

If you don't know much about computers, give your opinion about the following two statements.

	Agree	Disagree	Don't Know or no Answer
I am not qualified to manage the computerization of my collection.	10	8	8
I can try to manage the computerization of my collection, with a lot of technical help.	18	1	7

If, due to circumstances beyond your control, the collection records are not well kept, give your opinion about the following three statements.

	Agree	Disagree	Don't Know or no Answer
The records would have to be put in order before I can think of computerizing them.	6	7	13
I can put everything I have into the computer and it will sort it out for me.	3	8	15
Both above statements are too extreme. I will have to put some order in my records and the computer can help me to do that.	6	3	17
I don't see the point in making such an elaborate feasibility survey as the one you are conducting.	1	11	14
I know a data management program that will do the job just fine. Let's use it.	7	4	15

	Agree	Disagree	Don't Know or no Answer

Collection records can be used for purposes other than just locating a slide (e.g., survey, host list, etc.). With this in mind, give your opinion about the following two statements:

	Agree	Disagree	Don't Know or no Answer
I do not want my collection records to be made public because I am afraid people would use them for their own research without my knowledge and authorization.	3	11	12
I like the idea of computerizing my records because I will be able to do some interesting searches that will help my own research.	9	4	13

What are your computer skills?

	True
I have no experience with computers.	3
I have limited knowledge of computers.	17
I am computer literate.	4
I am a computer wizard, a real hacker.	0

What kind of material do you dispose?

I do not have access to a computer.	8
I do have access to a microcomputer.	9
I have access to a minicomputer.	6
I have access to a mainframe computer.	12
I have a terminal and a modem.	7

When you are using a computer, where is the keyboard of the material actually located (micro, mini, dumb terminal, etc.); if you have access to several kinds of material, check for the one you are using the most.

I am working at my own desk.	2
I have to go to another room in my lab.	9
I have to go to a different floor/building.	6

3. Analysis of the answers received.

As explained above, the comments below are valid only for half of the nematode collections in North America and for about one-third of the collections in the rest of the world.

In about three-fourths of these collections, records are kept on some sort of index card system that authorizes some cross searching. Ledgers is the second most used system. EDP is used in only 2 out of 22 collections, but two others are in the process of being computerized. Cross searches

are said to be difficult in almost half of the collections, and are the major problem in over a fourth of the collections.

Record keeping is described as excellent by about a third of the curators, but other curators complain about lack of time and personnel and are unable to keep data entry up to date. In about half the collections, curators experience some problems when searching their records because of incomplete data; this is the major drawback in 5 out of 22 collections.

The kind of data recorded varies widely among collections.

There is generally some sort of reference number which is either an accession number (one-third of the collections where slides are numbered starting with number 1 and continuing infinitely as each slide is received), or what is often called Thorne's system (one-third of the collections where genera and species receive a code name-- generally a numeral for genus and a letter for species). A composite system is used in the rest of the collections.

Users of Thorne's system complain that the recent splitting of many genera makes it difficult to continue using code numbers. The questionnaire did not ask whether the same code name always refers to the same taxon in different collections.

The name of the collection is not always indicated on the slide record, which is understandable when the records of each collection are kept separately. This name will have to be added to each record if several slide collections are placed in a central computerized system.

The collection item (slide, vial, etc.) is generally well described. In some collections, processing is made by a standard method and needs not to be recorded. There again, this method will have to be described in case of a central system.

The origin of specimens (= field sample record) is described in many different ways in the various collections.

In some cases, references is made to a field sample record number. It is this other record that contains data about the origin of the specimens. In other instances, the slide records include all field data. In both systems locality is recorded either by landmark (nearest town, road marker, river, etc.) or by administrative subdivision (township, range, county, province, state, county, etc.). Often both kinds of information co-exist. When recorded, geographical

coordinates are always associated with either or both landmark or administrative subdivision records. This variety of systems will create problems for computer searchs of slides from a particular geographical origin.

Host is often recorded under both common and Latin name. Again, this will have to be considered in the design of a computerized system because a search on either kind of names must find all slides recorded under both.

Data from marine nematode collections will be difficult to fit in a system designed primarily for plant nematodes. A separate system may well be in order.

Identification of specimens found on each slide always includes names of genus and species. These names pose many problems. Sometimes, specific identification later prove to be erroneous. Often the name of a species changes with advances in systematics. The various curators surveyed tackled these questions in a number of ways. Either they do nothing and keep the old name as it was, or they add a note to the old record. Some curators create a new record in addition or in place of the old one. In case of nomenclatural changes, some curators wait to see if the new name is accepted by the scientific community before they change the corresponding record. It is not known how the curators handle requests made under a new name not yet included in the collection records. Some curators lack the time and/or the knowledge of nematode systematics and nomenclature to make such changes. This is one area where a central computerized system would be most helpful by connecting collection data files to a nomenclatural file keeping track of changes in names.

Specimens not yet identified are either placed in a separate collection with or without their own records or kept in the main collection under whatever taxonomic category is known for them (genus, family, etc.).

About one-third of the collections record articles where collection material is described, or articles where voucher specimens are used.

In most collections, searches are made only for slides of a particular species or for slides of all species of a particular genus. In such cases, searches are generally described as easy. Electronic data processing is certainly not needed for such tasks, particularly for small and medium sized collections. In fact, no records are needed at all if the slides are arranged in boxes under genus and species names.

In about half of the cases, cross searches are often or sometimes made, mostly by geographical origins and by hosts. These kinds of searches were said to be possible or easy in the collections with index cards or sorter systems. However, it was noted that a sorter card has only four sides and a limited number of perforations. The number of subjects that can be searched and the number of categories in each subject are, therefore, limited. For example, there are over 300 names of plant nematode genera. They cannot all be included on a sorter card. Other index systems can become very bulky. In a particular collection, four records are created for each slide received, to be filed in four indexes by nematode, by host, by origin, and by date of receipt. Not surprizingly, the major drawback of this collection is said to be "typing and filing." When no indexes are kept, and sometimes even when they are, cross searches are said to be tedious, difficult, or impossible. Use of Bolean operators (AND, OR, and NOT) are naturally restricted to the few computerized collections.

Loans are made at least once a month in half the collections surveyed. Most curators restrict loans of type material to recognized organizations and scientists. They have no difficulty locating requested material, but sometimes material is lost because the curators have no time to keep track of loans.

Some curators will not make public slide records that contain sensitive information (e.g., origin and owner of plants intercepted under quarantine laws) or records with possibly erroneous species indentifications. Such records will have to be either kept out of a central EDP system or be given restricted access security. It may be that the information will become available in the future, e.g., after definitive identification has been made or when quarantine information becomes public domaine. In this event, temporary in-house records can be easily downloaded into the central system if they follow the same format.

Finally, loan policies or data recorded often differ between records of type or nontype material in the same collection. These records may have to be kept separate.

Opinion survey - Half the persons who answered the questionnaires think that current record keeping systems are unsatisfactory and that they need to be computerized. However, this answer is probably biased because most of those curators not interested in computers did not answer the survey. Among those who did answer, most have at least a limited knowledge of computers and have access to a computer (micro, mini, or mainframe). However, the majority must go out of their room, and sometimes out of their lab, to use it.

Some curators expressed concern that a computerized system would be unaccessible to labs that don't have access to a computer. This number would certainly have been greater if all curators had answered. This question will have to be addressed in the final proposal.

Even if a majority of collection curators would want to computerize their collection, this would still be difficult or impossible because most lack personnel, time, and money to do the job; and many do not feel competent enough to manage the computerization of their collection. It should be noted, however, that the majority is willing to try with a lot of help.

Only half the persons who answered the survey are convinced of the interest of the present study. Another third would rather use an existing program (packaged database management system, programs already in use in other collections, or programs recommended by ASC).

4. Conclusions.

Computerization of nematode slide collections will progress slowly and will have to overcome a lot of resistance from many people that do not see its advantages. At the same time, some labs have already switched to EDP and more are planning to do so in the near future. It is imperative and urgent to offer a common system before too many incompatible one are created.

The Committee will have to work along two lines. First, act as an advocate of EDP by highlighting its advantages over other systems; and second, propose a system that is truly advantageous and takes into account the limitations of the curators (time, money, expertise) and the real needs of the users. It is only if we can propose a system that is both more effective than current systems, and cheaper and easier to maintain, that EDP has a chance to be accepted in the nematology community.

Prepared by Renaud Fortuner for the Systematics Resources Committee of the Society of Nematologists, November 1985.