COENOSES 6 (3): 173-180, 1991

AN INSTITUTE FOR THEORETICAL ECOLOGY? PART V: PRACTICAL DATA MANAGEMENT FOR CROSS-SITE ANALYSIS AND SYNTHESIS OF ECOLOGICAL INFORMATION[1]

Walt Conley, Department of Biology, New Mexico State University, Las Cruces, NM 88003 USA
and
James W. Brunt, Department of Biology, University of New Mexico, Albuquerque, NM 87131

Keywords: Research data management, Cross-site data analysis, Theoretical Ecology, Statistical ecology, Mathematical ecology, Inter-disciplinary research

Abstract. Part V of the five part series on Theoretical Ecology presents a description of a practical approach to managing and archiving ecological data. The intent of this discussion is to define a generalized data structure that encourages good ecological data management. This approach facilitates the demands of analysis, synthesis, and modelling, of data across scales that range from local to regional or continental. Part I of this series (Conley 1990) provided a definition of Theoretical Ecology. Part II (Conley 1991a) provided both a critique and a proper design of large inter-disciplinary "Centers" that focus on a broad discipline such as Theoretical Ecology. Part III (Conley 1991b) presented an argument for the desireability of this approach to ecological issues guided by such an "Institute for Theoretical Ecology". Part IV (Conley 1991c), presented a description of a new activity referred to as "Computational Workshops" that offered a fresh approach to the analysis and synthesis of broadly developed ecological and environmental data.

Practical data management for ecology

What's the point?

There is a need for practical solutions to problems of long-term research data management in ecology. Practical solutions are those (i) that work; (ii) that researchers will use; and (iii) that can be afforded. Technological sophistication offers support but not solutions. We require the development of a philosophy for research data management in ecology that is people-oriented, that offers workable solutions to real biologists, that places training and education above misdirected sophistication and complexity of the computing environments, and that offers permanence and security for ecological data.

We need some common agreements concerning a few basic data archiving philosophies and tools that serve the goal of joining the research community in analysis and modelling of ecological data. The complexity of ecological systems requires teams and groups and cooperative attacks on the problems. We need to minimize the busy work of data archiving and manipulation, to maximize the stability and permanence of ecological data, to maximize the potential for cooperation among sites, and to maximize the potential for deep and extensive use of all available ecological information.

The solutions need to be driven by the questions.

The philosophy and methods described below emphasize practical solutions to facilitating communication, and attempt to address ecological questions with a reasonable and balanced approach to cooperative computing in ecological research.

Overview

Data Management in the broad sense involves the keeping and manipulation of research data. This includes analysis and presentation of results. Practical Data Management emphasizes face-to-face interactions among participants in a research team. It stresses the day to day aspects of getting the job done. It avoids unneeded technological sophistication. It concentrates on workable solutions to problems of how research biologists perceive their data and how they choose to work with data. It recognizes that what busy researchers can do and what they will do are not the same thing. It is pragmatic about how researchers really allocate their time commitments. It recognizes the essential dynamic nature of research data bases and the laboratories that generate the data. It understands that researchers want a data management program that works with a minimum of intrusiveness into what are typically considered more important ways to spend limited time and funding. It accepts the fact that

1 This work has been supported by grants from the U.S. National Science Foundation (BSR-8419790 to W. Conley, and DIR-8905886 to W. Conley and Marsha Conley), and by a grant from the Ecological Research Division of the U.S. Department of Energy, through the Idaho National Engineering Laboratory, to W. Conley.

research laboratories are sites of innovation, that innovation implies difference, and that productive heterogeneity across sites is a basic (and desirable) way of research life.

Practical Data Management seeks to focus the perspective of where data management fits in the scheme of research life. Data management is of paramount importance to the success of research projects, but like a good foundation on a tall building, it tends to succeed best in an unseen supportive role.

The pragmatic data management philosophy emphasizes the use of existing hardware and software at both local and national levels. It understands that heterogeneity in computational environments is part of the research solution not part of the problem. Most research sites already have access to varying amounts of computing equipment, and all are within reach of a variety of network access points. The objective of the /.intersite effort is to lead the establishment of a cooperative data management philosophy based on full communication of data and manuscript materials among cooperating groups of biological research sites. A related objective is to make the case for a location to foster intersite communication, training and instructional programs, data analysis and workshops, seminars and symposia, and a formal dialogue concerning practical data management for biological research sites.

Practical Data Management is more an approach and a way of thinking than a technological christmas tree. Practical Data Management emphasizes small-group dynamics and people-oriented solutions and de-emphasizes a technological solution for all perceived problems. Quite aside from the almost evangelical support of hardware and software brands and models and sizes and architectures, what is needed is a program of research, education, development, and coordinated cross-site cooperation. Such a program should concentrate on improving research data management, and should operate on real ecological questions to prove the concept and test the ideas. Such a program should emphasize training and practical solutions to real problems. If the /.intersite model can be implemented, there should be a demonstrable increase in scientific productivity, and an increasing success at addressing complex ecological questions as demonstration sites that specialize in cooperative computing become established.

The topic of the needed research is the proper role of Data Management in ecology in the large sense. What skills should a Data Manager possess? What should a basic computational environment be like in support of Practical Data Management? How can groups of research sites, fully heterogeneous with respect to computational skills and support, work together to address common questions without spending too much time on busy work? Is it possible to increase the depth with which data are analyzed and synthesized? Can the excitement of an active discussion be captured and supported by simultaneous data analysis with all members of a small research group participating fully? Can the results of these workshop sessions be captured in word or graph? What minimal agreements are required to facilitate communication of data and ideas among cooperating research sites?

What needs to be done?

Here are some of the problems to be solved.

- Every ecological research site has their own expertise in their own computing environment; heterogeneity at the local level is further confounded by even more heterogeneity at the national level. Heterogeneity in skills and in computational support is an advantage to creative science that must be exploited - we must learn how.
- The purpose for archiving data is to preserve if for future workers in the field. This implies eventual analysis and interpretation by workers who did not participate in the original collections. This also implies that the data might be used for purposes that were not originally intended. Archiving also requires maximum hardware and software independence to assure permanence and stability. A demonstrated approach to archiving and use of expensive-to-obtain research data must be used to help justify the conduct of critical long-term studies of natural phenomena.
- "Data" all too often means just the numerical part of the information gained from the research project. The misuse of the term "metadata" to denote comments and other documentation misleads both data managers and research administrators into thinking that just the numerical portion of the data are of some use in themselves. The fact is that the numbers are worthless without proper documentation that describes the context within which the numbers were obtained. This fact is critically important when you consider that "long term" research studies contain the implicit assumption that some future scientist will be attempting to analyze and synthesize data that someone else collected. Without proper documentation, there is simply no point to saving the numbers. A means must be found to encourage appropriate documentation, and then to provide data structures for archiving the information that encourages keeping the numbers and the documentation together. The term "metadata" should be reserved to its' proper meaning of "knowledge about the data" in the sense of machine intelligence structures that define higher level knowledge about the contents of the data. The term "data" should refer to both the numbers and the related documentation.
- Biological research sites have a need and a desire to communicate data and manuscript materials via

electronic mail. Such services are good now and growing better. A simple training program can lead the way.

- There are well over 100 data base management packages, and the number of combinations of dbms packages/hardware goes off the graph. Many local sites have the desire, but lack the expertise to get started in a significant computational environment. Others already are well advanced in their data management activities. Some sites require sophisticated computational support, others have more moderate requirements. All require a means to properly manage local data that puts the process in perspective for both local and national needs.
  Geographic Information Systems are data base management systems with a cartography interface (with the more advanced of these also providing some image manipulation tools). At least 37 vendors now offer such systems, and several more reside in the public domain. Costs range from tens of thousands of dollars to free. Biologists require access to this technology, and must be shown what can be done, what is appropriate for local needs, and what the cost/performance considerations really are.
- Every site has its' favorite applications package (s) for conducting statistical analyses and technical writing. Every site has developed local expertise that depends on specific hardware and software environments. Each of these high-level applications have their own user interfacing and data structure requirements that cannot reasonably be changed. The need is to offer cooperating sites maximum local flexibility for moving archived data into applications programs of choice. This would facilitate the cross-site synthesis of complex data sets.
- Graphics files (e.g. those that you can write to a plotter) are often difficult to transmit. They tend to be large, and they typically contain special characters that do not transmit reliably over non-binary network links. Because graphs are an essential component of scientific communication, an approach to transmitting presentation graphics over network lines is needed.
- Bit-map image data has become increasingly important. Files that contain images are large and are typically specific to both hardware and architecture. Transmitting such files among sites requires special consideration. Archiving image files also requires special considerations of space and hardware.

In this paper, we do not presume to solve all of these problems, but rather, choose to address only what we believe are the most significant. The proper archiving of both numbers and documentation is critical to research data management. Similarly, if we are to gain the ability to address landscape and regional questions in ecology, we must develop the means for effective cross-site analysis and synthesis of data.

Both of these problems are solved by the simple expedient of participating sites choosing to make use of a generalized data structure that contains both the numbers and the documentation. Further, these data structures are easily transportable among cooperating sites, and thus, cross-site synthesis of data becomes a matter of posing the hypotheses rather than an exercise in understanding heterogeneous file types, software, and hardware issues.

An /.intersite Archives File Structure

An /.intersite Archives File structure has been defined in order to facilitate the need for an orderly approach to the design and implementation of a data archiving and manipulation capability. Archiving implies permanent storage of the data structures for future use as a research resource. The data structure is the issue here, not the particular convention of storage. The manipulations to be done are alterations on the shape and/or the content of original (archived) data files, and communication of original or descendent files to remote sites.

The /.intersite Archives File is a generalized data structure that contains full documentation and comments. It is intended that the test of adequate documentation is that an archives file should stand alone, and that the file itself should contain sufficient information so that a future investigator who did not participate in collecting the data can use the information for some scientific purpose.

The /.intersite Archives File structure is intended to be used for local data archiving, which then supports communication of information among cooperating research sites that, taken together, represent the ultimate heterogeneous computing environment. The intent is to define a generic data structure that can be useful in any hardware and software environment, and that can be sent on any electronic network or file transfer medium. A companion effort involves the development of an /.intersite ToolKit for obtaining information from the Archives Files. The primary manipulation involves stripping an Archives File of various categories of information to produce descendent files for various purposes. These purposes include development of a data dictionary of attributes and abstracts, and producing stripped files that contain only the rows by columns of numbers which can be read by any application package such as statistics or graphics.

The basic /. intersite data structure is a generic ASCII flat file that contains catagories of information that define the information in the file. /.intersite Archives Files can be of any basic structural type, including statistical data, text data, graphics data (e.g. files that you can write to a graphics plotter), gene sequence data, or bit map image data. Other file types will no doubt be required and are easily accommodated. Note that

```
TOF
\log
A history of the file.

\doc
Any amount of explanatory text in any format.

ABSTRACT Title of the data set followed by a paragraph of
text. You will also want to put the name of the responsible
researcher and a phone number or E-mail address. The abstract
can appear anywhere in the \doc section.

(NOTE: blank line under ABSTRACT allows automatic
extraction.)
\type
statistical
\header
 col1 label  col2 label  ...  coln label
 ..........  ..........  ...  ..........
\data
 DATA ....   in column format as described in the header.
 DATA .....  .....
 DATA .....  .....
{Comments: contained by row within DATA and referring to
           specific portions of data. Any ASCII characters
           are allowed, and no format is imposed other than
           comments occupy an entire line, and must be
           enclosed inside curly brackets. By convention,
           a comment follows the record being described.
}
 DATA .....  .....
 DATA .....  .....
{       Comments may occur anywhere in DATA.
 DATA .....  .....
    .
    .
{the EOF indicator that follows is NOT part of the file}—

EOF
```

Figure 1. General structure of an /.intersite Archives File in Table format.

"file type" in this context refers to the general nature of the data in the file, and not to data typing such as floating point, integer, or character. With some specific exceptions to accommodate graphics or bitmap image files, all of the data in the /.intersite Archives Files are ASCII printing characters. Provision is made in the /.intersite ToolKit for handling files that contain non-printing ASCII characters that make file transfer difficult on some networks and impossible on many of the file transfer protocols.

The general catagories of data in an /.intersite Archives File are as follows.

log: A record of the history of the file; when it was initiated, updating, changes entered, locations and dates of copies of the file. Any ASCII printing characters with any format may be included.

doc: Documentation - as detailed a description as is necessary of the information contained in the file. Any ASCII printing characters with any format may be included. An ABSTRACT may be included here to allow automatic building of data dictionaries from Archives Directories. The ABSTRACT is simply a pragraph beginning with ABSTRACT and ending with a blank line; it may appear anywhere in the documentation section.

type: File type refers to the basic nature of the data. Statistical files are typically rows by columns tables of numeric or character data. Text files include bibliographic data, abstracts, or any prose. Graphics data refers to files that you can write to a plotter or a printer. Genome data refers to long sequences of base pairs that require line delimiters and other embedded information. Image data refers to bit map images. This information is used solely in the context of a data dictionary where descriptions of the contents of the files are maintained.

File typing currently includes statistical, text, graphics, genome, and image. Other file types are possible and can be added as necessary. The only operation anticipated on file type is identification for sorting and further processing.

header: Header refers to labels for the columns of data in a statistical data file, or a list format text file. This header list allows for automatic building of data dictionaries from Archives Directories. Labels in the header are automatically retrieved for the development of data dictionaries in /.intersite Archives data directories. The /.intersite ToolKit provides tools that do this work.

data: "Data" refers to the actual information of the archives file - the numbers, text, etc. The data section may contain embedded comments that further describe individual records of the data.

The /.intersite ToolKit contains programs that manipulate the Archives File data structure, making it ready for applications programs such as relational data management systems, statistical or graphics packages, and reporting systems such as text formatters. Any combination of catagories of information in the Archives Data Files can be extracted for further use. Thus, in a statistical file, it is possible, for example, to extract only the column labels and the table of numbers, only the ABSTRACT, only the documentation section, and so on. The ToolKit also contains an encryption and decryption algorithm (useful for converting files with non-printing characters to files that can be sent over

\log

********** BEGIN CHANGE LOG **********

23 Decembrer 1987. Data entered and documentation established. MAUhl

********** END CHANGE LOG **********

z\doc

ABSTRACT Ant Total Density on the Jornada. THis file, ant total.density, is monthly mean densities of new colonies grouped into zones, pooled for all species. The last 5 columns represent the onthly densities by year, and the first column describes the area ("zone") where the colonies were located. Data were collected by Marsha R. Conley 1982-86.

These 5 species were pooled to create the file:

Code:	Scientific name:	Common name:
PODE	Pogonomyrmex desertorum	Desert Harvester Ant
PORU	Pogonomyrmex rugosus	Red Harvester Ant
MYDE	Myrmecocystus depilis	Honey-pot Ant
MYMI	Myrmecocystus mimicus	Honey-pot Ant
NOCO	Aphaenogaster cockerelli	

\type

statistical

\header

Zone 1982 1983 1984 1985 1986

\data

Playa	0.0	0.0	0.0	0.0	0.0
Mesquite Fringe	2.7	3.0	3.3	3.3	3.3
Basin Slope	6.7	8.1	8.8	7.9	6.8
Bahada	0.5	0.6	0.8	1.3	1.5
Lower Piedmont	2.8	3.1	3.2	2.4	2.6
Upper Piedmont	0.8	0.9	1.1	1.1	0.9

{Only Pogonomyrmex were found in the Upper Piedmont}

Figure 2. Structure of an /.intersite Archives File in Table format.

networks that do not handle binary data or via dial-out modem transfers), and a suite of programs that automatically build and reference a data dictionary that contains various presentations of labels, keywords, and abstracts.

For statistical and text file types, there are 2 data formats that are useful to consider. "Table' format is the typical row X column format of statistical data with a label at the top of the column. "List" format is a transposed table, where the labels are on the left margin, giving unlimited catagory width but with a single column of data. List format is useful for text data such as keyworded bibliographic citations, or any similar kind of text. Note that you can include embedded comments anywhere in the \data section simply by enclosing the comment in curly brackets. The only restriction is that comments and other data cannot be mixed on the same line. (This preserves the positioning of tabular data, and serves the goal of keeping these files "readable" by humans.) There are no formatting requirements other than some white space between columns of data, and of the occurrence of blank lines in the list file format.

The general structure of an /.intersite Archives File (type is "statistical") in Table format is shown in Fig. 1. Note that the category indicators (*\log, \doc, \type, \header, \data*) occupy a separate line but do not need to begin in any particular column. The suggested catagories are optional, although deletion of any catagory limits the usefulness of the file and the use of the /.intersite ToolKit for manipulating the files.

The structure of an actual /.intersite Archives File in Table format is shown in Fig. 2.

The general structure of an /.intersite Archives File in List format is shown in Fig. 3.

The structure of an actual /.intersite Archives File in List format is shown in Fig. 4.

In the *log* and *doc* sections, there are no format requirements, and free-form text can be entered as you choose. In the *header* section and the data section, some structure is necessary. In the Table format, the *header* labels provide searching tags for the data file manipulations (and serve as handy reminders), and the dashed lines indicate the maximum width of each

```
TOF
\log
  Records of the history of the datafile. When it was initiated,
  changes entered, locations and dates of copies of the file.
  Any SCII characters with any format may be included.

\doc
  Documentation: As detailed a description as is necessary of
  the data contained in the file. Any characters with any
  format may be included. An ABSTRACT of 1 paragraph may
  be included anywhere in this section.

\type
  Typically List files are of type {\tt text}.

  \header
  Nothing needed here for List format. Note that the /.intersite
  data dictionary tools will pick up the Labels at the left margin
  of the first record and will automatically treat them similarly
  to the column labels from the Table format.

\data
  \n
{This is a comment. Note that the newline below \data is
required to automatically identify the List format.}

label1 line of text  ....
label2 line of text    :
label3 line of text    : → record 1
    .                  :
labeln line of text  ....
\n
label1 line of text  ....
label2 line of text    :
label3 line of text    : → record 2
    .                  :
labeln line of text  ...
  \n
      .
      .
  \n

EOF
```

Figure 3. General structure of /.intersite Archives File in List format. Note that the Labels are simply the first unbroken string of characters in each line.

```
\log
21 feb 89: file updated

12 mar 74: file updated
03 feb 74: file initiated, wconley
\doc
ABSTRACT A file of keyworded citations that point to a reprint
         library.

Note that this file had several thousand entries and was
screened to a few for this example.
\type
text
\header

\data

author      Fisher, R.A., A. Steven, and C.B. Williams
date        1943
title       The relation between the number of species and
            the number of individuals in a random sample of
            an animal Population.
page nos    42-58
ref         Journal Animal Ecology
keywords    model mathematical population sampling
            estimation

author      Forrester, J.W.
date        1970
title       Counterintuitive behavior of social systems.
page nos    13 (3): 1-16
ref         Technology Review
keywords    model behavior theory systems

autor       Edwards, A.W.F., and L.L. Cavalli-Sforza
date        1965
title       A method for cluster analysis.
page nos    362-375
ref         Biometrics
keywords    model method matrix analysis statistics cluster
```

Figure 4. Structure of an actual /.intersite Archives File in List format. The labels on the left margin allow retrieval of specific citations by keyword, date, author name, or any combination of labels. This structure also allows automatic assembly of citations in a Literature Cited section of a manuscript following common formats for ecological journals.

column of data (which is used for subsequent manipulation of the data columns[2]). In the List format, labels appear at the left of the field, and the dashed-lines indicator for column width is not necessary. In Tables, data columns conform to the *labels* in that they are in the same order, and in the Table format, the data must fit within the number of columns indicated by the dashed lines.

A Table format has 1 or more columns, and a List format has only a single column. Columns may be of arbitrary width. The labels in each case provide for data abstraction in good applications packages in that the researcher may refer to variables by name (i.e. *labels*) rather than, for example, as "column 3". Archives Files are specifically intended to be browsed by human researchers who want to become familiar with the data and the circumstances involved in the collection of the data. Once converted to the descendent files that will be manipulated via available relational operators (etc.),

2. The dashed lines are not necessary for many applications, however, they are useful for providing information for manipulation routines. To include them requires little, and adds considerably to the potential for development of tools to enhance cross-site data manipulation.

data files are not designed to be read by humans, and will be confusing to look at.

In practice, any numerical data set can be put into a rows by columns table format, and the only restriction is that the columns have some white space between them. This is the format that is typically used when recording data in the field, or when reporting data, and the /.intersite data structure simply provides a computerized version of what you probably do anyway. There is a utility in the /.intersite ToolKit called *extract* that can subset the standard /.intersite Archives File structure. This utility can create a new file with any combination of the various elements of an Archives File stripped from the original. The original is of course preserved intact. Other programs in the /.intersite ToolKit provide manipulation and screening of the /.intersite Archives Files, building of a data dictionary based on labels and keywords, extracting and sorting Abstracts, and generally obtaining information from the Archives directories.

Descendent files

Once the documentation has been stripped from the choosen archive files, and they ready for further manipulation or analysis work, the descendent files can be read into any applications package of your choice. A next obvious choice is entering the filtered data into a data base system for further manipulation. If you use a relational data base system you can make use of the labels without further change. Some statistical packages also can use such labels. If your application can make use of short explanations of the labels (e.g. SAS), you can include such information in the *doc* section. If you simply want a table of numbers to read onto a graphics or statistics package extract only the data and not the header or the comments embedded in the data.

The /.intersite ToolKit currently contains a suite of utilities that can operate on the Archives Files data structure. To serve the purpose of brevity, only one of these tools is described below. Other tools in the ToolKit include filters for conversion of the Archives Files into relational data base formats for at least one applications package, a data dictionary capability, a set of encryption-decryption filters to facilitate communicating graphics files (and others containing non-printing ASCII characters), and a large variety of general tools that have been developed under the auspices of the NMSU Science Workbench Project (Conley et al. 1986, Slator et al. 1986, Conley Ms.).

The /.intersite ToolKit

The /.intersite ToolKit is a package of utilities that operate on the basic /.intersite Archives File data structure. The ToolKit sets up archives directories, and provides a basic data dictionary for data files. Additional utilities in the ToolKit provide various file shaping operations that can reform the Archives files into data structures for applications such as data base systems, statistics and graphics, and text formatting. The /.intersite ToolKit currently contains examples of filters that work in all of these alternatives. The ToolKit is fully modular and extendsible to include additional cases.

The ToolKit consists of utilities that can reasonably be divided into several catagories based on the operations that are performed.

Directory Structure and Data Dictionary places the Archives properly, and provides information about the Archives Files and their contents.

Data Base Set-up: Given a directory location containing a data base of /.intersite Archives Files, establish a basic directory structure, place the files properly, and build the basic dictionary lists of keywords from the archives files. Set-up utilities can be invoked at request, automatically at login, or by the system clock, thus providing a means to keep the dictionary current.

Dictionary Lists: Given a basic /.intersite directory structure, some Archives Files, and the lists provided by *setUp*, this package, called *keylist*, establishes a series of "little files" that contain the basic keyword lists for the data dictionary.

Manipulating the Dictionary: Given the listings provided by *keylist* this package, called *searchit*, presents the data dictionary to the user in various ways, and provides *and/or* logical operations for searching among the Archives Files based on keywords. Repeated searching is offered until the proper set of Archives Files are assembled for further work. Once the proper Archives Files are identified, they are written into a workspace where further operations to manipulate the files can be performed.

Building Subsets of the Archives Files: This utility selects categories of information from the Archives Files. It is thus possible to extract any combination of the data catagories from an archive file for further use.

Manipulating the Data for Further Use: This package of tools provides for data encryption/decryption (for electronic mailing of files with non-printing characters over networks that do not handle binary data), compression/decompression (for reducing the size of files for transmission or storage), and filters that convert the basic Archives Files data structure into descendent data structures that are useful for a variety of applications packages. These filters currently include conversion to "/rdb" files (a common relational data base system that runs in UNIX and MS-DOS environments), conversion to generic RxC tables of data (to be read into any

application that reads ASCII rows by columns of data), and filters that convert statistical data into image data and vice versa.

The only filter from the /.intersite ToolKit that will be described in this paper is *extract*. Additional descriptions of the Science Workbench environments are presented in Conley (Ms.), and because of the volume of material will not be repeated here.

Manual Pages for extract

Extract

EXTRACT (0) */.intersite ToolKit* EXTRACT (0)

NAME

extract - retrive sections from /.intersite archive datafiles

SYNOPSIS

extract [*—ldAthDc*] [*<inputfile*] [*>outputfile*]

DESCRIPTION

extract retrieves sections from files in the /.intersite Archives File format based on options indicated. Fields that can be retrieved are *log, doc, type, header,* and *data*.

An ABSTRACT can be retrieved from the documentation section and comments can be stripped from the data section. Extract reads and writes to standard in and standard out.

Section indicators are also stripped.

In all cases, the original file is simply read, and thus preserved intact.

OPTIONS

l retrieves the change log delineated by *log* at the top and the next field indicator *e.g.* *doc*, at the bottom.

d retrieves the documentation delineated by *doc* at the top and the next field indicator eg. *header* at the bottom.

A retrieves the paragraph beginning with *ABSTRACT* and ending with a blank line from within the documentation section.

t retrieves the file type specified for the Archives File as indicated in the *type* section.

h retrieves the Archives File label header delineated by *header* at the top and *data* at the bottom.

D retrieves the data delineated by *data* and EOF (end of file).

c retrieves comments from within the data field enclosed by curly brackets. Comments may be more than one line, but occupy entire lines.

EXAMPLES

— > extract —hD < datafile.in > datafile.out

The header labels or keywords and the data sections are extracted from datafile.in and copied into datafile.out. The input file is left untouched. Note that this strips the comments from the data section.

— > extract —A < datafile.in

The "ABSTRACT" paragraph from the documentation section of datafile.in will be displayed on the terminal screen.

FILES

extract.c
extract.tex (this manual entry)
extract.1 (unix manual entry)

DIAGNOSTICS and BUGS

extract has no way to know whether input files are indeed in /.intersite Archives File format. *extract* looks for section delineators, if they are absent it will do nothing, if it finds something that looks appropriate, results may be unpredictable. In no case is the original file affected.

REFERENCES

CONLEY, W. Manuscript. Computational Ecology: The Management and Synthesis of Ecological Data. NMSU Mimeo., Version 4.03. September 1990. 21 Chapters, pp. xiii+224.

CONLEY, W. 1990. An Institute for Theoretical Ecology - Part I: What is "Theoretical Ecology" and why do we need it? Coenoses 5.

CONLEY, W. 1991a. An Institute for Theoretical Ecology? - Part II: The concept of "Center" - critique and a new design. Coenoses 6.

CONLEY, W. 1991b. An Institute for Theoretical Ecology? - Part III: Why we need it and what it should be like. Coenoses 6.

CONLEY, W. 1991c. An Institute for Theoretical Ecology? - Part IV: "Computational Workshops": A planned activity for Theoretical Ecology. Coenoses 6.

CONLEY, W., B. SLATOR, M. ANDERSON and R. SITZE. 1986. Designing and Prototyping a Scientific Problem Solving Environment: The NMSU Science Workbench. pp. 384-409. *In:* W. Michener, (ed.) Research Data Management in the Ecological Sciences. The Belle W. Baruch Library in Marine Science, Number 16. University of South Carolina Press.

SLATOR, B., M. ANDERSON and W. CONLEY. 1986. Pygmalion at the Interface. Communications of the Association for Computing Machinery, Vol. 29, No. 7, 599-604.

Manuscript received: November 1990