



DELTA Newsletter

Number 2, October 1988

Note from the Editor — The DELTA Newsletter is designed to promote communication among scientists developing and applying computer technology in the collection, storage, analysis, and presentation of taxonomic data for the production of descriptions, keys, interactive identification, and information retrieval. To achieve this goal the DN will be issued in April and October of each year. At this point the features to appear in each issue are yet to be determined, however Michael J. Dallwitz and Richard J. Pankhurst are expected to contribute to each issue. Readers are encouraged to suggest which features they feel should appear regularly. Contributions in the form of short comments or long discussions and explanations are encouraged from all developers and users of DELTA format and similar systems. Comments on application, suggestions for improvements, or criticisms of current technology are encouraged. These will be routed to the appropriate person for reply, and the ones of more general interest will be published in subsequent issues. — Robert D. Webster, SBM&NL, Bldg. 265, BARC-East, Beltsville, MD 20705, USA.

Notes on DELTA and Confor

Mike Dallwitz

Users' suggestions

The development of the DELTA format and associated programs has always been strongly influenced by suggestions from users. Most such suggestions are implemented, but, because our programming resources are limited, we have to weigh the benefits against the difficulty of implementation. Many ideas which were too difficult to implement in CONFOR will be incorporated in the interactive database system which we are developing to replace CONFOR.

Some suggestions, and my comments on them, are given below.

Punctuation in natural-language descriptions

In natural-language descriptions produced by CONFOR, it should be possible to specify a comma instead of a semicolon between characters.

CONFOR uses a semicolon rather than a comma to avoid possible wording difficulties when other commas are present. (Commas may come directly from the character list, and are also inserted by CONFOR between different states of the same character.) The consistent use of the semicolon in this way also makes it easier to identify the contribution of each individual character to the natural-language text. However, in response to many requests to allow the use of commas, we have recently added a REPLACE SEMICOLON BY COMMA directive. The substitution can be restricted to specified sets of characters.

Coding of 'large' numeric values

It would sometimes be useful to be able to code numeric values as 'indeterminately large or many'.
(C. Glasby, L. Watson)

We plan to make provision for this in the new system.

Representation of accuracy by number of significant figures

The number of significant figures in a real numeric value is sometimes intended to convey a rough measure of the accuracy of the value. For example, writing a value as 1.20 might imply greater accuracy than if it were written 1.2. However, a value entered as 1.20 in DELTA format is converted to 1.2 in output produced by CONFOR. The problem cannot always be solved by using the DECIMAL PLACES directive, because different numbers of decimal places may be appropriate for different taxa. (A. Gibbs, C. Glasby)

The normal method of storing numeric values in computers loses the information about trailing zeros after the decimal point. To avoid this, it would be necessary either to store the information in its original 'text' form (which would be awkward and inefficient for doing arithmetic on it), or the information about the number of zeros would need to be stored separately. Either of these options would greatly complicate the programs. In any case, the number of significant figures is an extremely rough method of conveying accuracy. If the accuracy is considered an important part of the information, it would probably be worth stating explicitly, for example, as a comment.

Position of comments in attributes

It should be possible to place comments at any positions within an attribute

I agree. At present, comments may appear after the character number, or after a value unless the value is followed by '-' or '&'. (For example, 1<a>,2/2&3<c> is allowed, but not 1,2<a>&3.) It would be very difficult to remove these restrictions from CONFOR (drastic changes in the internal data representation would be required). However, the restrictions will be removed in the new system.

Formulas

Space can be saved in natural-language descriptions by using a formula like '6-10x2-3 cm' instead of '6-10 cm long; 2-3 cm wide'. This could be implemented by a directive similar to LINK CHARACTERS. (J. Kirkbride)

This will be done in the new system.

Combining taxa

It should be possible to generate a combined description of several taxa, e.g., generate a genus description from species descriptions.

The new system will automatically pass information up and down the taxonomic hierarchy. Currently, Pankhurst's PANKEY package can generate combined natural-language descriptions of sets of taxa. The SUMMARY commands in CONFOR and INTKEY indicate the numbers of taxa which have each character state, and the means and ranges of numeric characters. The 'DELTA' command of INTKEY outputs a DELTA-format description of any specified set of taxa,

Lower-triangle output from DIST

The program DIST generates a distance matrix only in upper-triangle form. Lower-triangle form should be available as an option. (J. Kirkbride, L. Lefkovitch, B. Simon)

This will be added when time permits. Many clustering programs will accept input in either form.

Notes on INTKEY**Mike Dallwitz**

In 1982, Richard Pankhurst visited Canberra, and brought with him Version 3 of his interactive identification program, ONLINE. Since then, we have been distributing a version of this program with our other taxonomic programs. Over the years, Richard has released new versions of the program (it is now in Version 6), while we have independently made our own modifications based on Richard's

Version 3. Toni Paine and I recently decided that further modifications of that program were impractical, so we wrote a new program, which we called INTKEY. Although the program itself is completely new, it owes a great deal to the experience gained with ONLINE. ONLINE and INTKEY have evolved in rather different directions, but their common ancestry will be apparent to users.

INTKEY does not yet have printed documentation, but it has extensive internal documentation. Entering '?' at any prompt displays information about the possible responses to that prompt. Brief descriptions of the program's commands are available in three menus, which may be displayed automatically or on request. Fuller descriptions of the commands are available via a HELP command. HELP also provides a general introduction, and explains how to use the program for identification or information retrieval. It is also possible to produce a file containing all the HELP information.

When used for identification, the program progressively eliminates taxa whose character values do not match those of the specimen. The basic procedure is as follows.

- (1) Enter RESTART to start a new identification.
- (2) Choose an appropriate character to describe the specimen. To help the choice, there is a BEST command, which displays characters giving the best separation of the remaining taxa; and a CHARACTERS command, which displays specified sets of characters.
- (3) Enter the number of the chosen character.
- (4) Enter the character value or values exhibited by the specimen.
- (5) Repeat from step (2) until only one taxon remains. If no taxa remain, or if, for any other reason, you suspect an error, the TOLERANCE command can be used to increase the error-tolerance level.
- (6) Check the specimen against a description of the one remaining taxon. The DESCRIBE command can be used to display a full description of the taxon, or the DIAGNOSE command to display a diagnostic description in terms of the characters not used in the identification.
- (7) For other specimens, repeat from step (1).

In order to refer easily to subsets of the characters and taxa, 'keywords' can be defined. For example, a character keyword 'inflorescence' could be defined to mean the set of characters relating to the inflorescence, and a taxon keyword 'Australia' to mean those taxa occurring in Australia. CHARACTERS INFLORESCENCE (or any unambiguous abbreviation such as CHAR INF) would then display the inflorescence characters, and DESCRIBE AUS INF would describe the Australian taxa in terms of the inflorescence characters. INCLUDE TAXA AUS would confine the operation of all subsequent commands to the Australian taxa. This could be used to facilitate the identification of specimens known to come from Australia.

Although INTKEY was designed primarily for identification, it can also be used for information retrieval. By entering characters and values as though carrying out an identification, a list of taxa having the specified attributes can be obtained. It is also possible to list the taxa not having those attributes. There are DIFFERENCES and SIMILARITIES commands to display the differences or similarities between two taxa.

The MATCH command allows character values which are not identical to be treated as matching. MATCH I and MATCH U specify respectively that 'inapplicable' and 'unknown' match any value. MATCH S specifies that two sets of values match if one set is a subset of the other. (E.g. 1/2 matches 1/2/4 but not 2/3; 2-5 matches 1-6 but not 4-10). MATCH O specifies that two sets of values match if they overlap, i.e. if they have any values in common (e.g. 1/2 matches 2/3; 2-5 matches 4-10).

The default setting is MATCH O U, which is usually the most appropriate for identification. Under these conditions, taxa for which a specified character is inapplicable will be eliminated. For example, if leaf length is specified, taxa without leaves will be eliminated. Taxa for which a specified character is unknown will not be eliminated (to eliminate them would risk eliminating the correct answer). For information retrieval, the most appropriate setting is usually MATCH O, so that taxa for which the specified characters are either inapplicable or unknown are eliminated.

Notes on PANKEY

Richard Pankhurst

Hercules graphics for ONLIN6

In the first newsletter I gave details of the expert identification program version 6 with EGA colour graphics images for characters. There is now a 'black-and-white' version for those IBM/PC users who have monochrome graphics (Hercules adapter). This means practically any of the more recent machines. You will still need a Microsoft Mouse with the Paintbrush program in order to prepare your own images for your characters. Why did I do the colour version before the black-and-white? Partly because it was easier; Hercules is not much supported by MSDOS, and also because I have only recently seen any details on the Hercules hardware.

Data capture for DELTA

I have written a pair of programs for data collection which I use a lot myself, but which are rather crude. They are a temporary stopgap while we are waiting for a proper DELTA editor, but some of you may find them useful, and I am willing to distribute them free of charge for IBM/PC on receipt of a disc.

Program PREP reads DELTA data, checks it and makes an intermediate file from it. The data file has to have a complete character list and a list of dependent characters, and be syntactically correct and complete, but only one taxon need be present, and that can be just a name without any data. Program COND4 reads the file made by PREP and asks whether you already have a file with taxon names and possibly descriptive data on it. If there is no such file, you are asked for a new taxon name. If the file exists and has just taxon name(s) in it, the next taxon name is taken. If there is an (incomplete) description present as well, then this is read too. COND4 puts out a short menu of the states for each unscored character in turn, and expects you to give states for them. The usual range of characters, variability and comments are all accepted. Characters whose states are unknown can be passed over. Characters which depend on other characters which are not yet scored, or which are inapplicable because of the states already scored, are not asked. The program runs in cycles; when it has asked for states for all currently scorable characters, it will ask whether you want to go round again and score some more. You repeat this until either all characters are scored or you can't score any more. You are then asked whether you want a printed list of missing characters for the last taxon, and then it passes on to the next taxon.

COND4 does not (currently) accept IMPLICIT CHARACTERS and does have any means to change states which have already been scored. You can take the output file from COND4, edit it somehow and pass the data back into COND4 again, but if you break DELTA rules (especially the dependencies) it will crash. If you want to change the character list, edit the DELTA file and start again with PREP. If changes in the character list imply changes in the descriptions of taxa already described, you will need to make the corrections 'by hand'.

The outline of DELTA editor (I call it DEDIT) already exists as a C program which reads DELTA data and will output it again after deleting and/or reordering characters, and it will then output ASCII and/or binary files. As it stands, it is rather like the TIDY option from CONFOR. The central section, which does the editing, does not yet exist.

Development of ONLIN6

I am working on a version which has windows and is menu-driven. The key to British species of *Carex* (see below) has some new features.

- searching for names of taxa and characters by text strings, e.g. 'CHAR =Lea' will find all characters with 'Leaf or 'leaves'
- character weighting is possible. This only affects the order in which characters are listed in the commands for finding 'best' or diagnostic characters; characters in the group with the highest weight are listed first

- character help is provided for some characters. On pressing a function key addition explanations for that character come up. This information could be taken from the comments attached to character names in DELTA, but I've not done it that way.

PANKEY versions for other computers (VAX, Macintosh, PRIME)

PANKEY originated on mainframes but most users now run it on IBM/PCs. I have recently made a version for VAX (without graphics) and a version for Macintosh is imminent. A version exists for the PRIME minicomputer.

ATARI Is there anyone who would like to have PANKEY on ATARI micros? XENIX Would anyone like PANKEY for Microsoft or SCO XEXIX?

Are there any other machines that you would like me to convert to?

Corrections to interactive key construction (KCONI)

There has been an accumulation of minor corrections and improvements over the past year. It was possible with very wordy and lengthy keys to run out of buffer space, but there is now a means to recover from this. Would any KCONI users who do not already have a recent version like to send me a floppy disc? A paper about the KCONI program is expected to appear in *Taxon* in August 1988.

ADVANCED REVELATION and DELTA

As an alternative to writing a special purpose editor, such as DEDIT, I have long been looking for a database system which is sophisticated enough to accept and manipulate DELTA data. The best candidate would appear to be EMPRESS (Empress Software Inc, Toronto) but that is impossibly expensive.

Also pretty good and more affordable is Revelation from COSMOS, Seattle, especially Advanced Revelation. David French at the BM(NH) has been working on this for me and has managed to prepare data capture screens by using the database screen design features and without having to do special programming. Revelation is unusual in that it allows both variable field lengths and even more important, variable field contents. It is closely related to PICK. Bob Magill at the Missouri Botanical Garden has also had some success with DELTA and Revelation in connection with the Flora North America.

A new book on identification

'Biological identification' by RJP was published in 1978 and is now out-of-print (and out-of-date). I am working on a new book which will bring identification methods up-to-date, and in addition will cover computerised taxonomy in general. There will be chapters on classification methods (phenetics and cladistics) and also on taxonomic databases.

Taxonomic research projects using DELTA

The following are available from the Research School of Biological Sciences, Australian National University.

L. Watson (Taxonomy Lab)

Grass genera of the world. 'Complete' and fully operational, but continuously updated. Set comprising illustrations of characters, INTKEY data on floppy disks for MS-DOS microcomputers and full descriptions on microfiches, A\$40.

The Genera of Caesalpinioideae. Complete to 1982, with some subsequent updating. DELTA and/or working INTKEY set available on floppy disks or MS-DOS microcomputers, free. Accompanying booklet with full, automatically generated descriptions, keys etc., A\$15.

Families of Angiosperms (world treatment). Under development (extensive character list and sample descriptions, complete automated classification available).

Pooideae of Australia. Floristic treatment, funded by Australian Bureau of Flora and Fauna, commencing July 1988.

J. Bruhl (Taxonomy Lab)

World Genera of Cyperaceae. Detailed taxonomic account, nearing completion: available early 1989.

A. J. Gibbs (Plant Molecular Biology)

Plant Viruses. Detailed descriptions complete for those of Legumes and for those of Australian plants; being extended to cover all plants viruses, commencing with those of tropical crops. INTKEY sets available on floppy disks for MS-DOS microcomputers.

R. D. Webster (USDA, Beltsville)

Paniceae of Australia. Complete floristic account, funded by Australian Bureau of Flora and Fauna. INTKEY set available on floppy disks for MS-DOS microcomputers, free from Webster or Watson (Canberra). Book, 'The Australian Paniceae (Poaceae)' (1987) with computer generated descriptions and keys available from J. Cramer (Berlin, Stuttgart).

The following data sets are available (marked * if free) from the British Museum (Natural History).

**Angiosperm families (Hansen and Rahn)*. This is the same data as is used in the original punched card system from Dansk Botanisk Arkiv, plus corrections. The MEKA program also uses this same data set, and help from Tom Duncan in preparing the DELTA version from the MEKA data is acknowledged.

**British grasses, vegetative characters*.

Monocot families. This is the data used in the corresponding BM(NH) publication.

Sedges of the British Isles (Carex, Cyperaceae). This is a publication of the Botanical Society of the British Isles (BSBI Computer Key No. 1), and is available as an IBM floppy disc and a 40 page booklet at £10 from

BSBI Publications
24 Glapthorne Road
Oundle
PETERBOROUGH PE8 4JQ, England

This is ONLIN6 with new features as described above, and is designed to be used in conjunction with the BSBI Handbook No. 1 on 'Sedges of the British Isles', also obtainable from the above address. A new version with monochrome (Hercules) images for the characters is expected shortly.

British orchid species.

DELTA directory by country

[A list of subscribers to the DELTA Newsletter, current at the time of publication of this issue]

Australia

Armstrong, J. A. • Barendrecht, Peter • Barlow, B. A. • Bejsak, Richard • Bostock, P. D. •
Browning, G. P. • Clifford, H. T. • Collier, P. A. • Cookson, Laurie J. • Cranston, P. S. •
Dallwitz, M. J. • George, A. S. • Glasby, Chris • Hallett, Anna • Hooper, John N. A. •
Horning, Woody • Jackes, Betsy • James, Steve • Jones, Rhondda E. • Kelly, Sean • Kim, S. P. •
Lazarides, Michael • Leach, G. • Macfarlane, T. D. • Milward, Norman E. • Morrison, D. A. •
Paine, P. S. • Pitt, J. I. • Price, Ian R. • Ross, E. M. • Ross, Richard • Russell, B. C. • Simon, B. K. •
Spencer, R. D. • Thomson, B. • Watson, Les • Whiffen, Trevor • Williams, D. G.

Belgium

Empain, A. • Goetghebeur, Paul • Robbrecht, Elmar

Botswana

Sydes, Christine L. • Turton, Lilian M.

Brasil

Cavalcanti, Mauro J. • Fontoura, Talita • Cure, Jose Ricardo

Canada

Aiken, S. G. • Argus, George W. • Baum, Bernard R. • Dickinson, T. A. • Lefkovitch, L. P. •
Left, Carolyn • Sharkey, Michael J. • Turnbull, John

Colombia

Palacros, Pablo A.

Czechoslovakia

Sustek, Zdenek

Denmark

Renner, Susanne • Thrane, Ulf

El Salvador

Berendsohn, W.

Finland

Valovirta, Ilmari

France

Lebbe, Jacques

Greece

Damanakis, M. • Legakis, A.

Israel

Weitz, Shulamith

Italy

Cristofolini, Giovanni • Feoli, Enrico • Konopka, Jan

Kenya

Newton, L. E.

Malaysia

Lesslar, Philip

Mexico

Chiang, Fernando • Sanchez Sousa, Mario • Valdez Reyna, Jesus

Netherlands

Gonzales, P. J. Barbeito • Nooteboom, H. P • Pierrot, A. • van der Maesen, L. J. G. • Veldkamp, V. H.

New Zealand

Craw, Robin C. • Crosby, Trevor K. • Garnock-Jones, Phil • Penny, David

Nigeria

Isawumi, M. A. • Jayeola, A. A. • Jayeola, A. J. • Jayeola, S.

Norway

Birks, John

Panama

Correa, Mireya D.

Papua New Guinea

Eddie, William M. M. • Fortune-Hopkins, Helen

Portugal

Almeida, M. T. • Dias, Eduardo • Nogueira, Antonio Jose

South Africa

Balkwill, K. • Barker, N. P. • Gibbs-Russell, Beth • Grierson, D. • Hall, A. V. • Le Roux, A. • Linder, H. P. • Me Devette, K. • McDonald, G. • Oliver, E. H. H. • Roux, C. • Steiner, K. • Stirton, C. H. • Trollope, W. S. W. • Vincent, P. L. D.

Spain

Fernández Casas, Javier • Garcia-Valdecasas, Antonio • Ginoves Acebes, Juan Ramon • Hernandez Bermejo, Esteban • Querra, Arnoldo Santos • Susanna, Alfonso • Valdecasas, A. G.

Sweden

Prentice, Honor

Switzerland

Burdet, Hervé • Cook, C. D. K. • Finger, Jacques • von Arx, Bertrand • Zellweger, Catherine

Turkey

Babac, Mehmet Tiekin

U.S.A.

Barkworth, M. • Bates, Vernon M. • Baumgardner, George • Beach, James H. • Bieler, Rudiger • Boufford, David E. • Brotherson, Jack D. • Bruederle, Leo P. • Chuey, Carl F. Cudney, Dave • D'Arcy, W. G. • Dawson, James • Duke, Jim • Dutton, Bryan A. • Edmiston, James F. • Eilers, Larry • Feuillet, Christian • Fortuner, Renaud • Guala, Gerald F. • Gupta, V. K. • Hanson, W. • Hatch, Stephen L. • Huff, John • Jackman, John • Jarvie, J. • Jensen, Kevin B. • Jones, Gretchen • Jones, Stanley • Kirkbride, Joseph H. • Kopp, Marie • Kwater, Elizabeth • Lang, Frank • MacLean, David B. • Mikkelsen, Paula M. • Miller, Regis B. • Moreno, Nancy • Morin, Nancy R. • Newcombe, Lydia F. • Newton, Jr., Alfred F. • Ostergard, Mina • Pacheco, Victor • Paisley, Cato • Peterson, Paul • Pickering, Jerry L. • Plant, Richard E. • Platnick, Norman • Rhoades, Fred • Rosenow, Elizabeth • Schmid, Rudolf • Seiler, Gerald J. • Shaw, Robert B. • Shevlin, Dennis • Skog, Laurence E. • Starks, Gilbert • Steck, G. J. • Stieber, Michael T. • Straw, Richard •

Strawn, Ann J. • Sturm, Nicholas • Trana, Thomas D. • Weber, Darrell J. • Wheeler, Elisabeth • Wilson, Hugh • Wipff, J. K. • Wujek, Daniel E.

U.S.S.R.

Golovkin, Boris N. • Nekrasov, Valery I. • Parmasto, Erast • Sviridov, A. V.

United Kingdom

Atkinson, Mark • Atkinson, W. D. Atkinson, W. D. • Blackmore, S. • Brayford, David • Brough, D. • Brown, D. S. • Chimonides, J. • Coode, Mye • Copson, Pam • Cornelius, P. • Crawley, M. • Crust, Rick • Curds, C. • Day, M. • Easton, E. • Galloway, D. • Gauld, I. • Goldman, N. • Gowen, J. C. • Hale, Monica • Hardwick, L. W. • Harley, R. M. • Hayworth, C. C. • Heywood, V. H. • Hill, C. • Humphries, C. • Ingrouille, Martin • Jermy, A. C. • John, D. • Kathirithamby, J. • Kidd, Andrew D • Lambshead, P. J. D. • Lane, R.P. • Langman, Carol D. • Legg, Colin J. • Lonie, J. H. • Lyal, C. • McNeill, J. • Minter, Donald D. • Mordan, Peter B. • Moss, Helen • Pankhurst, Richard • Tittley, I. • Tolley, Hilary • Townsend, Bruce • von Hayek, C. • Walpole, M. • Paterson, Mike • Patterson, C • Peake, J. • Pryce, Richard D. • Rhoads, Ann F. • Roberts, D. • Sands, W. A. • Schrire, B. D. • Shaw, K. • Southgate, V. • Spicer, Robert A. • Stork, N. • Sutton, D. • Tebbs, M. • Thompson, R. • Thomson, N. • Tilling, S. M.

West Germany

Frahm, Jan Peter • Lindacher, Roland • Richter, George H.