



# Data requirements for natural-language descriptions and identification

16 July 2010

**M. J. Dallwitz**

## Introduction

Descriptive taxonomic data has three main uses: the generation of descriptions in natural language, identification (including conventional and interactive keys), and classification (cladistic and phenetic analysis). It is advantageous if the same data (or at least large parts of it) can be used for all three purposes. This paper focuses on the requirements of the first two uses.

The DELTA (Description Language for TAXonomy) data format (Dallwitz 1980; Dallwitz, Paine, and Zurcher 1993a,b) was developed as a way of recording multi-purpose descriptive data for computer processing. I will describe and discuss some of the capabilities of the current DELTA format and possible extensions of it.

## Characters

Taxa (or specimens) are described in terms of characters, each of which consists of a set of states, i.e. permitted values (see Colless 1985 for discussion of terminology). DELTA currently supports the following character types: unordered multistate, ordered multistate, integer numeric, and real numeric. A multistate character has a fixed number of states, whereas a numeric character has (in principle) an infinite number of states. 'Cyclic' characters for representing time of year or time of day have been proposed by Dallwitz, Paine, and Zurcher (1993b). These could be either multistate or numeric.

For description and identification, the significance of the distinction between the unordered, ordered, and cyclic types lies in the interpretation of non-adjacent states connected by 'to' or '-'. For example, if '1. red/ 2. blue/ 3. white/' are states of an unordered multistate character, 'red to white' does not include 'blue'. For an ordered multistate character with states '1. sparse/ 2. normal/ 3.dense/', 'sparse to dense' does include 'normal'. For a cyclic character whose states are '1. January/ 2. February/ ... 12. December/', 'November to January' includes 'December'.

Character types are also relevant in classification, e.g. for calculating distances. However, for this purpose, it may be necessary to specify additional information about the relationship between states. This is best done by specifying this information during the analysis, rather than by defining additional types. This avoids having an unlimited number of types, and allows testing the effects of different relationships.

DELTA also allows 'text characters', against which any free-text information can be recorded. This allows blocks of free text to be manipulated, for some purposes, by the same mechanisms used for true characters. In particular, they can be inserted at the appropriate positions in descriptions, and can be searched for information retrieval.

Table 1. Example of a character list.

#1. striated area on maxillary palp <presence>/
1. present/
2. absent/
#2. pronotum <colour>/
1. red/
2. black/
3. yellow/
#3. eyes <size>/
1. of normal size <i.e. less than 0.5mm in diameter>/
2. very large <i.e. more than 0.5mm in diameter>/
#4. frons <setae>/
1. with setae on anterior middle and above eyes/
2. with setae above eyes only/
3. without setae/
#5. number of lamellae in antennal club/
#6. length/ mm/
#7. <comments>/

Table 1 shows an example of a character list in DELTA format. Characters 1, 2, and 3 are unordered multistate, 4 is ordered multistate, 5 is integer numeric, 6 is real numeric, and 7 is text.

Each *character description* starts with a *feature description* (e.g. ‘pronotum <colour>’). For multistate characters, this is followed by *state descriptions* (e.g. ‘red’). For numeric characters, it may be necessary to specify the *units* (e.g. ‘mm’) in which the character is measured. The parts in angle brackets (<...>) are comments, which are omitted when generating natural-language descriptions.

Both feature and state descriptions may be empty, though they would normally contain at least a comment, e.g. ‘#8. <habit>/ 1. herb/ 2. shrub/’.

DELTA allows explanatory notes and illustrations to be associated with characters. There are no limitations on the amount of text in character descriptions and notes.

Character numbers are satisfactory identifiers for characters within a single data set. However, they may change when characters are added, deleted, or reordered. Identifiers that are fixed within a given context (e.g. within an organization, or worldwide), would be useful for purposes such as merging data sets and checking the consistency of wording.

## Coded descriptions

The entities to be described in terms of the character list are ‘items’ (the term used in DELTA) or OTU’s (Operational Taxonomic Units). These are usually taxa, but may be specimens or unnamed parts of taxa (used as a way of expressing intra-taxon variability — see below).

In DELTA, the *item name* was originally intended simply as the identifying text to be output in descriptions, keys, or classifications. (It is also used for that purpose in other descriptive-data formats, e.g. Nexus.) It can contain a comment, which was originally intended primarily to distinguish different items belonging to the same taxon, but is now typically used for the authority (because, as a comment, it is distinguishable by programs, and can be included in or omitted from output as required). Later, it was also used in some contexts as an internal identifier for the item, to link various material associated with the item, but not part of the description proper. All this is rather unsatisfactory, and needs to be re-examined in the light of current requirements.

Basically, the coded description of an item needs to record the character-state values that are shown by the item for some or all of the characters. In DELTA, the state values (and any associated information) corresponding to a given item and character are called an *attribute*. For example, with the characters defined in Table 1, the coding

1,1 2,1/2 5,3/5 6,8.5–10 7<Only males known.>

represents

Striated area on maxillary palp present. Pronotum red; or black. Number of lamellae in antennal club 3; or 5. Length 8.5–10mm. Only males known.

Particularly for natural-language descriptions, it is desirable to be able to record more than the mere occurrence of character states. DELTA allows state values to be connected by ‘–’, meaning ‘to’, and ‘&’, meaning ‘and’. In addition, qualifying information can be recorded in comments of unlimited length. For example,

1,1<usually>/2<rarely> 2,2/2&3<striped> 3,1–2 5,1/4–6

represents

Striated area on maxillary palp present (usually); or absent (rarely). Pronotum black; or black and yellow (striped). Eyes of normal size to very large. Number of lamellae in antennal club 1; or 4 to 6.

Our current DELTA programs have some arbitrary restrictions on the placement of comments, but, at the request of many users, these will be removed in future versions.

The use of ‘&’ to combine state values is controversial. Some people think it should not be allowed, and that extra states should be defined instead (e.g. ‘4. black and yellow’). Others like the flexibility afforded by this mechanism. Definition of extra states to represent ‘&’ is, of course, not precluded by the current DELTA format. The disadvantages are the proliferation of states (which is often regarded as undesirable), and the loss of the relationship between the states for classification (unless this is provided by another mechanism, as mentioned above).

The ability to combine values of multistate characters with ‘–’ (‘to’) is important for natural-language descriptions, as it conveys a meaning different from ‘/’ (‘or’), and, in the case of ordered characters, gives more succinct descriptions (e.g. ‘sparse to dense’ instead of ‘sparse, normal, or dense’). For other applications, it produces the same results as the equivalent coding with ‘/’.

For numeric characters, it is useful to be able to record normal and extreme ranges of values, and a measure of central tendency (mean, median, or mode). For example, in DELTA ‘6,(3–)5–6.7–9(–12)’ means that the lengths of ‘all’ specimens fall in the range 3–12, most fall in the range 5–9, and the mean (or median or mode) is 6.7. The format currently does not define the precise meanings of these values (e.g. whether they are particular percentiles), but the author of a dataset may do so, and convey these meanings informally to the users of the data. It would be useful to have a formal mechanism for this, and to extend the allowed number of values indefinitely, e.g. for any number of percentiles.

The order of state values in an attribute is significant, that is, ‘3,1/2’ is different from ‘3,2/1’. This is an important requirement for generating natural-language descriptions. It can also be used in classification to indicate that one value (usually the first) should be used and the others discarded.

Character dependencies (see below) allow programs to infer that certain characters are not applicable to certain taxa. It is sometimes desirable to be able to express the inapplicability of characters that are *not* covered by a dependency. DELTA uses the ‘pseudovalue’ ‘–’ for this purpose, e.g. ‘2,–’, meaning that character 2 is not applicable to the taxon. There are also pseudovalues ‘V’, meaning ‘variable’ (all states present), and ‘U’, meaning ‘unknown’. The latter is equivalent, for most purposes, to simply omitting the attribute. Its essential use is to override a value that would otherwise be automatically copied from another item. In the current format, this is used in ‘variant items’. We do not intend to support variant items in future versions of our programs, and, as far as I know, no other programs support them. However, more general mechanisms for passing information between items are proposed, and the ‘U’ pseudovalue would still be necessary in that context.

Programs that create and maintain general-purpose descriptive data need to be able to store and manipulate all of the above kinds of attribute. Programs that only use the data must be capable of parsing the general data, but need only store the parts needed for a particular purpose. For example, most applications do not need the comments, and a classification program might use only the mean of a numeric character.

It should be possible to associate illustrations with items, and to annotate these illustrations (for example, with information about the source and content).

## Character dependencies

Character dependencies specify sets of characters — the ‘dependent’ characters — that are inapplicable when certain other characters — the ‘controlling’ characters — take certain values. For example, the attribute ‘leaves absent’ implies that characters describing the nature of the leaves (e.g. length, shape) are inapplicable, and must not be recorded. If any of the recorded values of a given controlling character *do not* make its dependent characters inapplicable, then the dependent characters may be recorded. For example, ‘leaves present or absent’ allows other leaf characters to be recorded. If a given dependent character is dependent on more than one controlling character, then the dependent character can be recorded only if allowed by *all* of its controlling attributes.

In the current DELTA format, the controlling characters must be multistate characters.

Data-maintenance programs should not allow the entry of data inconsistent with the dependencies (some do!).

## Different natural languages

It should be possible to have versions of the character list, and other text material, in different natural languages, e.g. English and Chinese.

In DELTA, the character list and character notes are normally stored in separate files, which are invoked when needed for particular purpose. Thus, it is possible to have versions of these in different languages, and invoke them as required.

In the coded descriptions (and some other contexts), this approach is not possible, because the text is embedded in the coded information. In our DELTA programs, we have been experimenting with tags to indicate the language, e.g. ‘@en’ for English (ISO-639), ‘@de’ for German, and ‘@all’ for language-independent material such as names. We have recently extended this method to all text, including the character list.

## Alternative wordings

Different wordings, particularly of the character list, may be necessary for different purposes. The use of comments in the character list can go some way towards satisfying this requirement, as they may be displayed or omitted depending on the circumstances.

When a character list is displayed for selection of characters (in interactive identification or editing), rapid scanning of the list is facilitated by displaying only the feature descriptions. The comments must be displayed to distinguish the characters, e.g. ‘leaves <length>’, ‘leaves <shape>’.

When the state descriptions are displayed during interactive identification or data entry, the comments would usually be displayed, to help with the interpretation of the character.

In natural-language descriptions and conventional keys, both feature and state comments are normally omitted in the interests of brevity and readability. (In addition, some or all of the non-comment part of the feature description may be omitted, if it can be inferred from the context.)

Even when comments are used in this way, the wording of the non-comment parts of the characters must often be a compromise between the requirements of descriptions and keys. This is because in descriptions, headings and previous characters provide a context that may allow more succinct wording of the characters. (However, caution is necessary, as some of the context may be missing if attributes are unknown or inapplicable. In diagnostic descriptions, most of the context may be missing, and the wording requirements are like those of keys.)

Alternative wordings may also be required to suit different user groups. Output intended for non-specialists could avoid the use of technical terms, and different specialist groups sometimes use different terms for the same concepts.

Alternative wordings for characters have usually been implemented as complete, alternative character lists. This is inefficient, as large proportions of the wordings are usually the same. A better mechanism would be by tagging of the alternative wordings, as described above for alternative languages. If used for both purposes,

the tags would need to be hierarchical, e.g. '@en-k' for English keys, '@en-d' for English descriptions, '@es-k' for Spanish keys.

When numeric characters are converted to multistate for conventional keys or some classification programs, the automatically generated wording of the state descriptions may be unsatisfactory. Therefore, a means of providing new wordings for these is desirable.

## Indexed lists

Numbered and alphabetic lists of entities such as references, countries, and taxonomic names could be used when the same (long) list of states is required for more than one character. For example, a list of countries could be used in a character giving native distributions and one giving actual distributions (including introductions).

These lists could also be used in 'coded comments' (see below).

For examples, see Dallwitz, Paine, and Zurcher (1993b).

## 'Coded comments' — interpretable information within comments

Currently, material within comments is not interpreted by programs. The incorporation of coded information within comments would provide additional functionality, while retaining compatibility with existing programs that discard comments. Dallwitz, Paine, and Zurcher (1993b) proposed that 'coded comments would be embedded in ordinary comments (that is, enclosed in angle brackets), and would comprise the symbol '@', a single-word 'comment identifier', and the coded information. A coded comment will be terminated by the next coded comment, or by the closing angle bracket of the comment. The following 'coded comments' were proposed.

Coded comment	Meaning
<@probability $x$ >	Probability or frequency of a state value.
<@ $x$ %>	Alternative form of 'probability' comment.
<@rarely>	A low probability for a state value. The value of this probability should be settable.
<@only $a_1 a_2 \dots$ >	Specifies that a character value is for use in applications $a_1 a_2 \dots$ only. It would be omitted from other applications.
<@not $a_1 a_2 \dots$ >	Specifies that a character value is not for use in applications $a_1 a_2 \dots$ . It would be omitted from these applications.
<@for $a_1 a_2 \dots$ >	If this comment appears in an attribute, only the values so marked would be used in applications $a_1 a_2 \dots$ . Equivalent to @not $a_1 a_2 \dots$ on the other values
<@about>	Qualifies numeric values. The extent of the uncertainty would be specifiable.
<@?>	Indicates a guessed value.
<@possibly>	Indicates that there is no evidence that the taxon does <i>not</i> have the value.
<@reliability $x$ >	Specifies a reliability for an attribute, to modify the overall reliability of a character. This information would be important for key generation.
<@edit <i>commands</i> >	Apply editing commands to the natural-language description before output, e.g. replace one text string with another.
<@up>	Attribute generated from information passed up the taxonomic hierarchy.
<@down>	Attribute generated from information passed down the taxonomic hierarchy.
<@note <i>text</i> >	Uninterpreted sub-comment.

In addition, any numbered or alphabetic list name could be recognized as a comment identifier. The omission of coded comments from natural-language descriptions, and the format that it should take if include, would be controllable independently for each identifier.

*Examples***\*ITEM DESCRIPTIONS**

. . . 10,1/3<@prob .1> 11,2/<@rarely>4 12,1<@ref 135 322> 13,2/<@en occasionally @de gelegentlich @5% @ref 54>3 14,1<@note check in fresh specimens> 15,<@about>15-<@about>20 16,2/1<@only keys> 17,7<@only keys>-8.5-9<@for classification>-10-12<@only keys> 18,2<@for classification>/3 19,1<@English usually truncate @German gewöhnlich trunquat><@edit d; (;i;, ;d);> 20,1/2<@not Australia>

The editing command for attribute 19 removes the parentheses from the comment, and places a comma before the comment.

A directive would control the inclusion/omission of values marked by @only, @not, and @for. For example,

**\*APPLICATIONS keys**

would use

16,2/1 17,7-12 18,2/3 20,1/2

and

**\*APPLICATIONS classification Australia**

would use

16,2 17,9 18,2 20,1

Ways of handling values marked with several of these comments have not yet been fully considered.

**Indefinite values**

In numeric attributes, it is sometimes necessary to represent an indefinitely large or small number. Dallwitz, Paine, and Zurcher (1993b) proposed using '~' for this purpose.

*Examples*

~ = many

5-~ = 5 or more (or 5 to many)

~-5 = up to 5

Specific, settable numbers could be substituted for the indefinite numbers where necessary (for example, for calculating a mean, or in identification).

**Readability**

Readability by people was an important consideration in the design of the DELTA format, as data entry as text (often on punched cards) was the norm at that time. This is much less important now that data can be entered via a special-purpose interface. However, readability still has advantages. For example, the data can be published on paper in its concise, coded form (e.g. Britton 1986), thus ensuring its long-term accessibility.

The Appendix shows some data recorded in DELTA, Nexus, Lucid Interchange Formats, XDELTA, and SDD.

## References

- Britton, E.B. 1986. A revision of the Australian chafers (Coleoptera: Scarabaeidea: Melolonthinae). Vol. 4. Tribe Liparetrini: genus *Colpochila*. *Aust. J. Zool., Suppl. ser.* 118: 1–135.
- Colless, D.H. 1985. On “character” and related terms. *Syst. Zool.* 34: 229–233.
- Dallwitz, M.J. 1974. A flexible program for generating identification keys. *Sys. Zool.* 23, 50–7.
- Dallwitz, M.J. 1980. A general system for coding taxonomic descriptions. *Taxon* 29, 41–6.
- Dallwitz, M.J. 1984. Automatic typesetting of computer-generated keys and descriptions. In ‘Databases in Systematics’, Systematics Association Special Volume No. 26, pp. 279–90. (Eds R. Allkin and F.A. Bisby.) (Academic Press: London.)
- Dallwitz, M.J., Paine, T.A., and Zurcher, E.J. 1993a onwards. User’s guide to the DELTA system: a general system for processing taxonomic descriptions. 4th edition. <http://delta-intkey.com>
- Dallwitz, M.J., Paine, T.A. and Zurcher, E.J. 1993b onwards. New features for the DELTA system. <http://delta-intkey.com>
- Dallwitz, M.J., Paine, T.A., and Zurcher, E.J. 1995 onwards. User’s guide to Intkey: a program for interactive identification and information retrieval. 1st edition. <http://delta-intkey.com>
- Dallwitz, M.J., Paine, T.A., and Zurcher, E.J. 1998. Interactive keys. In ‘Information Technology, Plant Pathology and Biodiversity’, pp. 201–212. (Eds P. Bridge, P. Jeffries, D. R. Morse, and P. R. Scott.) (CAB International: Wallingford.)
- Dallwitz, M.J. 1999 onwards. A comparison of formats for descriptive data. <http://delta-intkey.com/www/compdata.htm>
- Dodds, L. 1999. XDELTA — Deriving an XML based format for Taxonomic Information. <http://www.ldodds.com/delta/>
- Lucidcentral. 1999 onwards. Lucid home page. <http://lucidcentral.org>
- Maddison, W.P., and Maddison, D.R. 1992. ‘MacClade: analysis of phylogeny and character evolution.’ Version 3. 398pp. (Sinauer Associates: Sunderland, Massachusetts.)
- Maddison, D.R., Swofford, D.L., and Maddison, W.P. 1997. NEXUS: an extensible file format for systematic information. *Syst. Biol.* 46, 590–621.
- Swofford, D.L. 1991. PAUP: phylogenetic analysis using parsimony. Version 3.1. (Illinois Natural History Survey: Champaign.)
- TDWG. 2007. SDD – an introduction and primer. <http://wiki.tdwg.org/twiki/static/index.htm>

## Appendix

The same data recorded in 6 formats: DELTA, Nexus (Maddison and Maddison 1992; Maddison et al. 1997; Swofford 1991), Lucid Interchange Formats V2 and V3 (Lucidcentral 1999), XDELTA (Dodds 1999), and SDD format (TDWG 2007).

### DELTA Format

Note the following features, which make the data more easily interpretable by human readers.

- Both the character list and the items have explicit character and state numbers.
- Each feature description is immediately followed by its associated state descriptions.
- The characters can be in any order, irrespective of character type.
- The identifying name of an item is immediately followed by its attributes.
- The attributes of an item can be in the same order as the corresponding characters in the character list, irrespective of character type.

\*HEADING Beetles

\*NUMBER OF CHARACTERS 5

\*MAXIMUM NUMBER OF STATES 4

\*MAXIMUM NUMBER OF ITEMS 3

\*CHARACTER TYPES 1,RN 2,OM 4,OM 5,RN

\*NUMBERS OF STATES 2,3 4,4

\*CHARACTER LIST

#1. total length/

mm/

#2. body <degree of convexity>/

1. strongly flattened/

2. slightly flattened to moderately convex/

3. strongly convex/

#3. upper surfaces of body <vestiture>/

1. glabrous or subglabrous/

2. clothed with distinct hairs, setae or scales/

#4. antennae when posteriorly extended <extension>/

1. not reaching middle of prothorax/

2. reaching beyond middle of prothorax but not middle of elytra/

3. reaching beyond middle of elytra but not elytral apices/

4. reaching beyond elytral apices/

#5. ratio of pronotal length to greatest pronotal width/

\*ITEM DESCRIPTIONS

# Amphizoidae/

1,10.7-15.5 2,2 3,1 4,1/2 5,0.47-0.49

# Anischiidae/

1,1.8-3.2 2,2 4,2 5,0.65-0.75

# Anobiidae/

1,1-10.5 2,2/3 3,1/2 4,1/2/3/4 5,0.32-1.5

## Nexus Format

Paup and MacClade, the most widely used programs that use Nexus format, restrict the length of text in feature and state descriptions and do not handle numeric characters. Those restrictions are shown in this sample, which was automatically translated by Confor from the above DELTA data. The numeric characters have been converted to multistate.

---

```
#NEXUS
```

```
BEGIN DATA;
```

```
DIMENSIONS NTAX=3 NCHAR=5;
```

```
[!Beetles]
```

```
FORMAT MISSING=? GAP=- SYMBOLS="1234";
```

```
CHARLABELS
```

```
[1] 'total length'
```

```
[2] 'body <degree of convexity>'
```

```
[3] 'upper surfaces of body <vestit>'
```

```
[4] 'antennae when posteriorly exte'
```

```
[5] 'ratio of pronotal length to gr'
```

```
;
```

```
STATELABELS
```

```
1 'up to 5 mm' '5 to 10 mm' '10 mm or more',
```

```
2 'strongly flattened' 'slightly flattened to moderate' 'strongly convex',
```

```
3 'glabrous or subglabrous' 'clothed with distinct hairs, s',
```

```
4 'not reaching middle of prothor' 'reaching beyond middle of prot' 'reaching beyond middle of elyt' 'reaching  
beyond elytral apices',
```

```
5 'up to 0.5' '0.5 to 1' '1 or more',
```

```
;
```

```
MATRIX
```

```
'Amphizoidae'          321(12)1
```

```
'Anischiidae'         12?22
```

```
'Anobiidae'           (123)(23)(12)(1234)(123)
```

```
;
```

```
END;
```

```
BEGIN ASSUMPTIONS;
```

```
TYPESET * untitled = unord: 3, ord: 1-2 4-5;
```

```
END;
```

---

## Lucid Interchange Format V2.1

---

#Lucid Interchange Format File v. 2.1

[..Character List..]

total length

mm

body <degree of convexity>

strongly flattened

slightly flattened to moderately convex

strongly convex

upper surfaces of body <vestiture>

glabrous or subglabrous

clothed with distinct hairs, setae or scales

antennae when posteriorly extended <extension>

not reaching middle of prothorax

reaching beyond middle of prothorax but not middle of elytra

reaching beyond middle of elytra but not elytral apices

reaching beyond elytral apices

ratio of pronotal length to greatest pronotal width

[..Taxon List..]

Amphizoidae

Anischiidae

Anobiidae

[..Main Data (txs)..]

60101011006

60102201006

60111111116

[..Metric Data..]

1	1	10.7	10.7	15.5	15.5
1	11	0.47	0.47	0.49	0.49
2	1	1.8	1.8	3.2	3.2
2	11	0.65	0.65	0.75	0.75
3	1	1	1	10.5	10.5
3	11	0.32	0.32	1.5	1.5

---

## XDELTA Format

---

```

<description> Beetles </description>
<character-list>
<char-group title="All">
<character number="1"> <num type="real" units="mm">
  <feature> total length </feature>
</character>
<character number="2"> <multi type="ordered">
  <feature> body <comment> degree of convexity </comment> </feature>
  <state number="1"> strongly flattened </state>
  <state number="2"> slightly flattened to moderately convex </state>
  <state number="3"> strongly convex </state>
</character>
<character number="3"> <multi type="unordered">
  <feature> upper surfaces of body <comment> vestiture </comment> </feature>
  <state number="1">glabrous or subglabrous</state>
  <state number="2"> clothed with distinct hairs, setae or scales </state>
</character>
<character number="4"> <multi type="ordered">
  <feature> antennae when posteriorly extended <comment> extension </comment> </feature>
  <state number="1"> not reaching middle of prothorax </state>
  <state number="2"> reaching beyond middle of prothorax but not middle of elytra </state>
  <state number="3"> reaching beyond middle of elytra but not elytral apices </state>
  <state number="4"> reaching beyond elytral apices </state>
</character>
<character number="5"> <num type="real" units="mm">
  <feature> ratio of pronotal length to greatest pronotal width </feature>
</character>
</char-group>
</character-list>
<item-list>
<item itemid="1"> <item-name> Amphizoidae </item-name>
<attribute-list>
  <attribute character="1"> <value start="10.7" end="15.5"></value> </attribute>
  <attribute character="2"> <value>2</value> </attribute>
  <attribute character="3"> <value>1</value> </attribute>
  <attribute character="4"> <value>1</value> <value>2</value> </attribute>
  <attribute character="5"> <value start="0.47" end="0.49"></value> </attribute>
</attribute-list>
</item>
<item itemid="2"> <item-name> Anischiidae </item-name>
<attribute-list>
  <attribute character="1"> <value start="1.8" end="3.2"></value> </attribute>
  <attribute character="2"> <value>2</value> </attribute>
  <attribute character="4"> <value>2</value> </attribute>
  <attribute character="5"> <value start="0.65" end="0.75"></value> </attribute>
</attribute-list>

```

</item>

<item itemid="3"> <item-name> Anobiidae </item-name>

<attribute-list>

<attribute character="1"><value start="1" end="10.5"></value> </attribute>

<attribute character="2"><value>2</value> <value>3</value> </attribute>

<attribute character="3"><value>1</value> <value>2</value> </attribute>

<attribute character="4"><value>1</value> <value>2</value> <value>3</value> <value>4</value>  
</attribute>

<attribute character="5"><value start="0.32" end="1.5"></value> </attribute>

</attribute-list>

</item>

</item-list>

---

## Lucid Interchange Format V3

Lucid Interchange Format V3 as output by the Lucid Builder V3.52.

---

```

<?xml version="1.0" encoding="UTF-8"?>
<lif3_key xsi:schemaLocation="http://www.lucidcentral.org/2006/LIF3/
  http://www.lucidcentral.org/2006/LIF3/lif3.xsd" xmlns="http://www.lucidcentral.org/2006/LIF3/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
<properties>
  <property key="key_authors" type="java.lang.String" value=""/>
  <property key="KeyVersion" type="java.lang.String" value="1.20"/>
  <property key="xmlnl_output_file" type="java.lang.String" value="D:\lucid35\t\descdata-
    descriptions.xml"/>
  <property key="key_title" type="java.lang.String" value=""/>
  <property key="key_description" type="java.lang.String" value=""/>
</properties>
<items>
  <item item_id="2" name="total length" item_type="feature" revision="false" score_type="numeric"
    score_weight="1.0" natural_lang_name="total length" base_unit="metre" unit_prefix="milli"/>
  <item item_id="3" name="body (degree of convexity)" item_type="feature" revision="false"
    score_type="normal" score_weight="1.0" natural_lang_name="body" base_unit="none"
    unit_prefix="none"/>
  <item item_id="4" name="strongly flattened" item_type="state" revision="false" score_type="normal"/>
  <item item_id="5" name="slightly flattened to moderately convex" item_type="state" revision="false"
    score_type="normal"/>
  <item item_id="6" name="strongly convex" item_type="state" revision="false" score_type="normal"/>
  <item item_id="7" name="upper surfaces of body (vestiture)" item_type="feature" revision="false"
    score_type="normal" score_weight="1.0" natural_lang_name="upper surfaces of body" base_unit="none"
    unit_prefix="none"/>
  <item item_id="8" name="glabrous or subglabrous" item_type="state" revision="false"
    score_type="normal"/>
  <item item_id="9" name="clothed with distinct hairs, setae or scales" item_type="state" revision="false"
    score_type="normal"/>
  <item item_id="10" name="antennae when posteriorly extended (extension)" item_type="feature"
    revision="false" score_type="normal" score_weight="1.0" natural_lang_name="antennae when posteriorly
    extended" base_unit="none" unit_prefix="none"/>
  <item item_id="11" name="not reaching middle of prothorax" item_type="state" revision="false"
    score_type="normal"/>
  <item item_id="12" name="reaching beyond middle of prothorax but not middle of elytra"
    item_type="state" revision="false" score_type="normal"/>
  <item item_id="13" name="reaching beyond middle of elytra but not elytral apices" item_type="state"
    revision="false" score_type="normal"/>
  <item item_id="14" name="reaching beyond elytral apices" item_type="state" revision="false"
    score_type="normal"/>
  <item item_id="15" name="ratio of pronotal length to greatest pronotal width" item_type="feature"
    revision="false" score_type="numeric" score_weight="1.0" natural_lang_name="ratio of pronotal length
    to greatest pronotal width" base_unit="none" unit_prefix="none"/>
  <item item_id="16" name="Amphizoidae" item_type="entity" revision="false" score_type="normal"
    is_list_view_node="true"/>
  <item item_id="17" name="Anischiidae" item_type="entity" revision="false" score_type="normal"
    is_list_view_node="true"/>
  <item item_id="18" name="Anobiidae" item_type="entity" revision="false" score_type="normal"
    is_list_view_node="true"/>
</items>

```

```

<trees>
  <tree type="feature">
    <treenode item_id="2" parent_id="0"/>
    <treenode item_id="3" parent_id="0"/>
    <treenode item_id="4" parent_id="3"/>
    <treenode item_id="5" parent_id="3"/>
    <treenode item_id="6" parent_id="3"/>
    <treenode item_id="7" parent_id="0"/>
    <treenode item_id="8" parent_id="7"/>
    <treenode item_id="9" parent_id="7"/>
    <treenode item_id="10" parent_id="0"/>
    <treenode item_id="11" parent_id="10"/>
    <treenode item_id="12" parent_id="10"/>
    <treenode item_id="13" parent_id="10"/>
    <treenode item_id="14" parent_id="10"/>
    <treenode item_id="15" parent_id="0"/>
  </tree>
  <tree type="entity">
    <treenode item_id="16" parent_id="0"/>
    <treenode item_id="17" parent_id="0"/>
    <treenode item_id="18" parent_id="0"/>
  </tree>
  <tree type="filter"/>
</trees>
<descriptions>
  <container type="normal">
    <scoring_item item_id="5">
      <scored_item item_id="17" value="1"/>
      <scored_item item_id="16" value="1"/>
      <scored_item item_id="18" value="1"/>
    </scoring_item>
    <scoring_item item_id="6">
      <scored_item item_id="18" value="1"/>
    </scoring_item>
    <scoring_item item_id="8">
      <scored_item item_id="17" value="3"/>
      <scored_item item_id="16" value="1"/>
      <scored_item item_id="18" value="1"/>
    </scoring_item>
    <scoring_item item_id="9">
      <scored_item item_id="17" value="3"/>
      <scored_item item_id="18" value="1"/>
    </scoring_item>
    <scoring_item item_id="11">
      <scored_item item_id="16" value="1"/>
      <scored_item item_id="18" value="1"/>
    </scoring_item>
    <scoring_item item_id="12">
      <scored_item item_id="17" value="1"/>
      <scored_item item_id="16" value="1"/>
      <scored_item item_id="18" value="1"/>
    </scoring_item>
    <scoring_item item_id="13">
      <scored_item item_id="18" value="1"/>
    </scoring_item>
  </container>
</descriptions>

```

```
<scoring_item item_id="14">
  <scored_item item_id="18" value="1"/>
</scoring_item>
</container>
<container type="numeric">
  <scoring_item item_id="2">
    <scored_item item_id="17">
      <scored_data score_type="normal" unit_prefix="unknown" omin="1.8" nmin="1.8" nmax="3.2"
omax="3.2"/>
    </scored_item>
    <scored_item item_id="16">
      <scored_data score_type="normal" unit_prefix="unknown" omin="10.7" nmin="10.7" nmax="15.5"
omax="15.5"/>
    </scored_item>
    <scored_item item_id="18">
      <scored_data score_type="normal" unit_prefix="unknown" omin="1" nmin="1" nmax="10.5"
omax="10.5"/>
    </scored_item>
  </scoring_item>
  <scoring_item item_id="15">
    <scored_item item_id="17">
      <scored_data score_type="normal" unit_prefix="unknown" omin="0.65" nmin="0.65" nmax="0.75"
omax="0.75"/>
    </scored_item>
    <scored_item item_id="16">
      <scored_data score_type="normal" unit_prefix="unknown" omin="0.47" nmin="0.47" nmax="0.49"
omax="0.49"/>
    </scored_item>
    <scored_item item_id="18">
      <scored_data score_type="normal" unit_prefix="unknown" omin="0.32" nmin="0.32" nmax="1.5"
omax="1.5"/>
    </scored_item>
  </scoring_item>
</container>
<container type="dependency"/>
<container type="filter"/>
</descriptions>
<media base_path="D:\lucid35\t\descdata\Media"/>
</lif3_key>
```

---

## SDD format

SDD format (TDWG 2007) as output by the Lucid Builder V3.52 (sic – the version information included by the program is wrong).

---

```

<?xml version="1.0" encoding="UTF-8"?>
<Datasets xsi:schemaLocation="http://rs.tdwg.org/UBIF/2006/
  http://www.lucidcentral.org/2006/SDD/SDD1.1/SDD.xsd" xmlns="http://rs.tdwg.org/UBIF/2006/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
<TechnicalMetadata created="2010-07-15T16:19:28.187+10:00">
  <Generator name="Lucid3 Builder" version="3.4.0_b01"
    routine="com.cbit.lucid.builder.core.io.sdd11.SDDKeyWriter"
    notes="http://www.lucidcentral.org/lucid3/">
  <TechnicalContact literal="http://www.lucidcentral.org" email="contact@lucidcentral.org"/>
</TechnicalMetadata>
<Dataset xml:lang="en-us">
  <Representation>
    <Label>Untitled</Label>
  </Representation>
  <TaxonNames>
    <TaxonName id="t1" debuglabel="16:Amphizoidae">
      <Representation>
        <Label>Amphizoidae</Label>
      </Representation>
    </TaxonName>
    <TaxonName id="t2" debuglabel="17:Anischiidae">
      <Representation>
        <Label>Anischiidae</Label>
      </Representation>
    </TaxonName>
    <TaxonName id="t3" debuglabel="18:Anobiidae">
      <Representation>
        <Label>Anobiidae</Label>
      </Representation>
    </TaxonName>
  </TaxonNames>
  <TaxonHierarchies>
    <TaxonHierarchy id="th1" debuglabel="1:Default Entity Tree">
      <Representation>
        <Label>Default Entity Tree</Label>
      </Representation>
      <TaxonHierarchyType>UnspecifiedTaxonomy</TaxonHierarchyType>
      <Nodes>
        <Node id="tn1" debuglabel="16:Amphizoidae">
          <TaxonName ref="t1" debuglabel="16:Amphizoidae"/>
        </Node>
        <Node id="tn2" debuglabel="17:Anischiidae">
          <TaxonName ref="t2" debuglabel="17:Anischiidae"/>
        </Node>
        <Node id="tn3" debuglabel="18:Anobiidae">
          <TaxonName ref="t3" debuglabel="18:Anobiidae"/>
        </Node>
      </Nodes>
    </TaxonHierarchy>
  </TaxonHierarchies>
</Dataset>

```

```

</TaxonHierarchies>
<DescriptiveConcepts>
<DescriptiveConcept id="dc0">
  <Representation>
    <Label>Fixed set of modifiers supported in Lucid3</Label>
  </Representation>
  <Modifiers>
    <Modifier id="mod1" debuglabel="1:rarely">
      <Representation>
        <Label>rarely</Label>
      </Representation>
      <ModifierClass>Frequency</ModifierClass>
      <ProportionRange lowerestimate="0.0" upperestimate="0.25"/>
    </Modifier>
    <Modifier id="mod2" debuglabel="2:uncertain">
      <Representation>
        <Label>uncertain</Label>
      </Representation>
      <ModifierClass>Certainty</ModifierClass>
      <ProportionRange lowerestimate="0.0" upperestimate="0.5"/>
    </Modifier>
    <Modifier id="mod3" debuglabel="3:misinterpreted">
      <Representation>
        <Label>misinterpreted</Label>
      </Representation>
      <ModifierClass>TreatAsMisinterpretation</ModifierClass>
    </Modifier>
    <Modifier id="mod4" debuglabel="3:unscoped">
      <Representation>
        <Label>unscoped</Label>
      </Representation>
      <ModifierClass>OtherModifierClass</ModifierClass>
    </Modifier>
  </Modifiers>
</DescriptiveConcept>
</DescriptiveConcepts>
<Characters>
<QuantitativeCharacter id="c1" debuglabel="2:total length">
  <Representation>
    <Label>total length</Label>
  </Representation>
  <MeasurementUnit>
    <Label role="Full">metre</Label>
    <Label role="Abbrev">m</Label>
  </MeasurementUnit>
  <Default>
    <MeasurementUnitPrefix>milli</MeasurementUnitPrefix>
  </Default>
</QuantitativeCharacter>
<CategoricalCharacter id="c2" debuglabel="3:body (degree of convexity)">
  <Representation>
    <Label>body (degree of convexity)</Label>
  </Representation>
  <States>
    <StateDefinition id="s1" debuglabel="4:strongly flattened">

```

```

    <Representation>
      <Label>strongly flattened</Label>
    </Representation>
  </StateDefinition>
  <StateDefinition id="s2" debuglabel="5:slightly flattened to moderately convex">
    <Representation>
      <Label>slightly flattened to moderately convex</Label>
    </Representation>
  </StateDefinition>
  <StateDefinition id="s3" debuglabel="6:strongly convex">
    <Representation>
      <Label>strongly convex</Label>
    </Representation>
  </StateDefinition>
</States>
</CategoricalCharacter>
<CategoricalCharacter id="c3" debuglabel="7:upper surfaces of body (vestiture)">
  <Representation>
    <Label>upper surfaces of body (vestiture)</Label>
  </Representation>
  <States>
    <StateDefinition id="s4" debuglabel="8:glabrous or subglabrous">
      <Representation>
        <Label>glabrous or subglabrous</Label>
      </Representation>
    </StateDefinition>
    <StateDefinition id="s5" debuglabel="9:clothed with distinct hairs, setae or scales">
      <Representation>
        <Label>clothed with distinct hairs, setae or scales</Label>
      </Representation>
    </StateDefinition>
  </States>
</CategoricalCharacter>
<CategoricalCharacter id="c4" debuglabel="10:antennae when posteriorly extended (extension)">
  <Representation>
    <Label>antennae when posteriorly extended (extension)</Label>
  </Representation>
  <States>
    <StateDefinition id="s6" debuglabel="11:not reaching middle of prothorax">
      <Representation>
        <Label>not reaching middle of prothorax</Label>
      </Representation>
    </StateDefinition>
    <StateDefinition id="s7" debuglabel="12:reaching beyond middle of prothorax but not middle of
elytra">
      <Representation>
        <Label>reaching beyond middle of prothorax but not middle of elytra</Label>
      </Representation>
    </StateDefinition>
    <StateDefinition id="s8" debuglabel="13:reaching beyond middle of elytra but not elytral apices">
      <Representation>
        <Label>reaching beyond middle of elytra but not elytral apices</Label>
      </Representation>
    </StateDefinition>
    <StateDefinition id="s9" debuglabel="14:reaching beyond elytral apices">

```

```

    <Representation>
      <Label>reaching beyond elytral apices</Label>
    </Representation>
  </StateDefinition>
</States>
</CategoricalCharacter>
<QuantitativeCharacter id="c5" debuglabel="15:ratio of pronotal length to greatest pronotal width">
  <Representation>
    <Label>ratio of pronotal length to greatest pronotal width</Label>
  </Representation>
</QuantitativeCharacter>
</Characters>
<CharacterTrees>
  <CharacterTree id="ct1" debuglabel="1:Default Feature Tree">
    <Representation>
      <Label>Default Feature Tree</Label>
    </Representation>
    <DesignedFor>
      <Role>InteractiveIdentification</Role>
    </DesignedFor>
    <Nodes>
      <CharNode>
        <Character ref="c1" debuglabel="2:total length"/>
      </CharNode>
      <CharNode>
        <Character ref="c2" debuglabel="3:body (degree of convexity)"/>
      </CharNode>
      <CharNode>
        <Character ref="c3" debuglabel="7:upper surfaces of body (vestiture)"/>
      </CharNode>
      <CharNode>
        <Character ref="c4" debuglabel="10:antennae when posteriorly extended (extension)"/>
      </CharNode>
      <CharNode>
        <Character ref="c5" debuglabel="15:ratio of pronotal length to greatest pronotal width"/>
      </CharNode>
    </Nodes>
  </CharacterTree>
</CharacterTrees>
<CodedDescriptions>
  <CodedDescription id="cd1" debuglabel="16:Amphizoidae">
    <Representation>
      <Label>Amphizoidae</Label>
    </Representation>
    <Scope>
      <TaxonName ref="t1" debuglabel="16:Amphizoidae"/>
    </Scope>
    <SummaryData>
      <Categorical ref="c2" debuglabel="3:body (degree of convexity)">
        <State ref="s2" debuglabel="5:slightly flattened to moderately convex"/>
      </Categorical>
      <Categorical ref="c3" debuglabel="7:upper surfaces of body (vestiture)">
        <State ref="s4" debuglabel="8:glabrous or subglabrous"/>
      </Categorical>
      <Categorical ref="c4" debuglabel="10:antennae when posteriorly extended (extension)">

```

```

    <State ref="s6" debuglabel="11:not reaching middle of prothorax"/>
    <State ref="s7" debuglabel="12:reaching beyond middle of prothorax but not middle of elytra"/>
  </Categorical>
  <Quantitative ref="c1" debuglabel="2:total length">
    <Measure type="Min" value="10.7"/>
    <Measure type="UMethLower" value="10.7"/>
    <Measure type="UMethUpper" value="15.5"/>
    <Measure type="Max" value="15.5"/>
    <MeasurementUnitPrefix>unknown</MeasurementUnitPrefix>
  </Quantitative>
  <Quantitative ref="c5" debuglabel="15:ratio of pronotal length to greatest pronotal width">
    <Measure type="Min" value="0.47"/>
    <Measure type="UMethLower" value="0.47"/>
    <Measure type="UMethUpper" value="0.49"/>
    <Measure type="Max" value="0.49"/>
    <MeasurementUnitPrefix>noprefix</MeasurementUnitPrefix>
  </Quantitative>
</SummaryData>
</CodedDescription>
<CodedDescription id="cd2" debuglabel="17:Anischiidae">
  <Representation>
    <Label>Anischiidae</Label>
  </Representation>
  <Scope>
    <TaxonName ref="t2" debuglabel="17:Anischiidae"/>
  </Scope>
  <SummaryData>
    <Categorical ref="c2" debuglabel="3:body (degree of convexity)">
      <State ref="s2" debuglabel="5:slightly flattened to moderately convex"/>
    </Categorical>
    <Categorical ref="c3" debuglabel="7:upper surfaces of body (vestiture)">
      <State ref="s4" debuglabel="8:glabrous or subglabrous">
        <Modifier ref="mod2" debuglabel="2:uncertain"/>
      </State>
      <State ref="s5" debuglabel="9:clothed with distinct hairs, setae or scales">
        <Modifier ref="mod2" debuglabel="2:uncertain"/>
      </State>
    </Categorical>
    <Categorical ref="c4" debuglabel="10:antennae when posteriorly extended (extension)">
      <State ref="s7" debuglabel="12:reaching beyond middle of prothorax but not middle of elytra"/>
    </Categorical>
    <Quantitative ref="c1" debuglabel="2:total length">
      <Measure type="Min" value="1.8"/>
      <Measure type="UMethLower" value="1.8"/>
      <Measure type="UMethUpper" value="3.2"/>
      <Measure type="Max" value="3.2"/>
      <MeasurementUnitPrefix>unknown</MeasurementUnitPrefix>
    </Quantitative>
    <Quantitative ref="c5" debuglabel="15:ratio of pronotal length to greatest pronotal width">
      <Measure type="Min" value="0.65"/>
      <Measure type="UMethLower" value="0.65"/>
      <Measure type="UMethUpper" value="0.75"/>
      <Measure type="Max" value="0.75"/>
      <MeasurementUnitPrefix>noprefix</MeasurementUnitPrefix>
    </Quantitative>
  </SummaryData>
</CodedDescription>

```

```
</SummaryData>
</CodedDescription>
<CodedDescription id="cd3" debuglabel="18:Anobiidae">
  <Representation>
    <Label>Anobiidae</Label>
  </Representation>
  <Scope>
    <TaxonName ref="t3" debuglabel="18:Anobiidae"/>
  </Scope>
  <SummaryData>
    <Categorical ref="c2" debuglabel="3:body (degree of convexity)">
      <State ref="s2" debuglabel="5:slightly flattened to moderately convex"/>
      <State ref="s3" debuglabel="6:strongly convex"/>
    </Categorical>
    <Categorical ref="c3" debuglabel="7:upper surfaces of body (vestiture)">
      <State ref="s4" debuglabel="8:glabrous or subglabrous"/>
      <State ref="s5" debuglabel="9:clothed with distinct hairs, setae or scales"/>
    </Categorical>
    <Categorical ref="c4" debuglabel="10:antennae when posteriorly extended (extension)">
      <State ref="s6" debuglabel="11:not reaching middle of prothorax"/>
      <State ref="s7" debuglabel="12:reaching beyond middle of prothorax but not middle of elytra"/>
      <State ref="s8" debuglabel="13:reaching beyond middle of elytra but not elytral apices"/>
      <State ref="s9" debuglabel="14:reaching beyond elytral apices"/>
    </Categorical>
    <Quantitative ref="c1" debuglabel="2:total length">
      <Measure type="Min" value="1.0"/>
      <Measure type="UMethLower" value="1.0"/>
      <Measure type="UMethUpper" value="10.5"/>
      <Measure type="Max" value="10.5"/>
      <MeasurementUnitPrefix>unknown</MeasurementUnitPrefix>
    </Quantitative>
    <Quantitative ref="c5" debuglabel="15:ratio of pronotal length to greatest pronotal width">
      <Measure type="Min" value="0.32"/>
      <Measure type="UMethLower" value="0.32"/>
      <Measure type="UMethUpper" value="1.5"/>
      <Measure type="Max" value="1.5"/>
      <MeasurementUnitPrefix>noprefix</MeasurementUnitPrefix>
    </Quantitative>
  </SummaryData>
</CodedDescription>
</CodedDescriptions>
</Dataset>
</Datasets>
```

---