PLANT MATERIALS DATABASE DEVELOPMENT
AT THE SOUTHERN INSTITUTE OF FOREST GENETICS[1]

M.E. Mason, C.D. Nelson, T.D. Caldwell, and W.L. Nance [2]

Abstract.-- The Southern Institute of Forest Genetics (USDA Forest Service, Gulfport, MS) is working with a large number of plant materials in a variety of research programs. A database has been developed using Borland's PARADOX 4.0[Tm] to track plant materials and record pertinent information about them. Through the databases primary and secondary tables a large amount of information can be efficiently managed. A menu-driven application is being developed to make the data easily accessible.

Keywords: Database management, relational databases, pedigrees, tree breeding, clonal propagation, record keeping.

INTRODUCTION

Since the 1950s, researchers at the Southern Institute of Forest Genetics have selected and maintained a large number of pine genotypes for use in genetic research (Synder et al. 1968). Funding and personnel reductions during the 1970s and 1980s slowed the development of these materials and jeopardized their maintenance at Gulfport and many other labs throughout the South. Because genetic research depends heavily on the availability of diverse and well-characterized genotypes and pedigrees, the Institute has recently expanded its program of collecting and generating these types of research materials. Efforts are primarily focused on producing multi-generation pedigrees for genetic analysis of various traits and preserving diverse materials from native provenances.

To efficiently manage the information associated with the acquired and developed materials, a database using Borland's PARADOX v4.0[TM] has been created. The purpose of the database is to maintain all pertinent information about the materials in an organized manner and in a form that can be easily accessed by project scientists and technicians as well as non-project personnel. A custom application written in Paradox Application Language (PAL) is being prepared for use in performing all the basic functions, such as data protection, data entry and editing, interactive queries (searches and calculations), and report writing. The objectives of this paper are to describe the database's overall structure, including relationships between the tables, and demonstrate its functionality by presenting a detailed example of a complicated query.

DATABASE STRUCTURE
Overview

A database is a series of related (linked) tables (data sets) that are used to store information in an organized way. Tables consist of records (rows) and fields (columns). A table to store height and

diameter data for a clonal field planting might include the following fields: CLONE, ROW, COLUMN, HEIGHT, and DIAMETER. The data for each tree (ramet) comprises a record, and each field contains a certain type of data. The linkages between tables are designed to minimize the amount of data redundancy, thus maximizing storage and retrieval efficiency. For example, to keep all the data for a clone in one table would be cumbersome, especially if different kinds of data were being kept. A "normalized" database would contain data for the clone in several tables, with each table containing different kinds of information (Borland International, 1992). Thus, when searching (querying) for specific information about the clone (e.g., species, provenance, pedigree, rooting data, field growth data, etc.), only the pertinent tables need to be accessed. During the query, the database program "links" tables through fields included in both tables to properly associate (merge) the corresponding records and return the information requested.

Primary tables. The central table of our database is the *clone*[3] table. Information specific to the clone (genotype) is entered into the *clone* table. Each ramet of a clone is entered into the *ramet* table along with a code indicating its physical location. A table for each possible location lists the included ramets and other information about the ramets at that location. Pedigree information is stored in the *pedigree* table and provenance information in the *provenance* table. Because clones are part of active or closed experiments, critical information on these experiments is stored in the *experiment* table.

Secondary tables. The key to all the possible physical locations for a ramet is given in the *location* table. A series of tables *(ab, cb, co,* and field) contain information on these locations. Pollination records and pollen and seed storage records are kept in the *flowers, pollen,* and *seed* tables, respectively. A large group of tables contains data (such as rooting or field growth measurements) on ramets or seedlings that are involved in various experiments. Seedlings selected from these will be given a clone number and then vegetatively propagated for use in future research. The names of these tables are usually prefixed with the experiment code such as *SD18-27, SD3,* or *SD4.*

Clones and Ramets

Information unique to a clone is maintained in the *clone* table. *Clone* includes the following fields: CLONE, ALIAS, SPECIES, EXPERIMENT, and ENTRY. CLONE is used to record the clone number. A clone number is assigned to a plant when it is vegetativley propagated or if it is part of another clone's pedigree. ALIAS is a record of past identification codes. SPECIES is a two-letter to four-letter code based on the Latin specific name. EXPERIMENT is a two-letter code for the experiment in which the clone was originally entered. ENTRY is generally synonymous with family. Entries within an experiment are numbered "1" to "n". EXPERIMENT+ENTRY defines a specific family (all clones with the same EXPERIMENT and ENTRY are full or half siblings), with the exception of wild trees (trees of unknown parentage). Wild trees are entered as experiment "ZZ". The entry here refers to the provenance of the tree, not the family. Thus, two trees of entry "2" in experiment "ZZ" are wild trees from the same provenance.

The *ramet* table contains information specific to a particular ramet of a clone. *Ramet* includes the following fields: CLONE, RAMET, and LOCATION. The seedling ramet (ortet) is always assigned RAMET "0". A key to location codes is found in the *location* table. Some examples of locations are accelerated breeding greenhouse, clone bank, cutting orchard, or dead. A table for each location (e.g., *AB,*

---

[3] To clarify references between tables, fields, and values, the following conventions are used throughout this paper: table names are italicized *(clone);* field names are in all capitals (CLONE); and values are in double quotes (clone "10").

CB, CO) gives the specific information about each ramet, including its coordinates within the location. Figure 1 shows the *clone* and *ramet* tables.

Figure 1. An image of a computer screen displaying the *clone* and *ramet* tables.   *Clone* contains information unique to the genotype. *Ramet* contains information on individual ramets. Table fields are explained in the text.

Pedigrees and Provenances

As explained in the previous section, each EXPERIMENT+ENTRY combination defines a family or a provenance. The *pedigree* table records all pedigree information for each clone and family. Its four fields are EXPERIMENT, ENTRY, FEMALE (clone number of female parent), and MALE (clone number of male parent). Wild trees are assigned to experiment "ZZ". Provenance information is stored in the *provenance* table. Trees of unknown provenance are assigned entry "0". Known provenances are coded "1" to "n". Figure 2 displays a sample of the *pedigree* and *provenance* tables.

Linking Clones. Pedigrees, and Provenances

One of the most frequently performed database tasks is researching a clone's pedigree or provenance. The link between a clone and its parents is EXPERIMENT+ENTRY. Figures 1 and 2 show views of the *clone, ramet, pedigree,* and *provenance* tables, which may be consulted to follow the examples given. Assume a user wishes to find out about clone "10" and find all of its full siblings. Clone "10" the first clone listed in the *clone* table, was originally in experiment "SD" and entry "25". Experiment "SD" and entry "25" are found in record 8 of the *pedigree* table. The parents of clone "10" are clone "1002" as female and clone "1005" as male. To further examine the pedigree, the parents of clones "1002" and "1005" can be found-- first by finding their EXPERIMENT and ENTRY in the *clone*

table and second by linking to the *pedigree* table. In the *pedigree* table, clone "1002" can be found in record 5 and clone "1005" in record 7. Both trees originated in experiment "ZZ" and entry "1" informing the user that both are wild trees and of the same provenance. To identify the provenance, the *provenance* table is searched for entry "1". Record 2 of the *provenance* table shows entry "1" to be Harrison County, MS. Thus, clone "10" is a control-cross offspring of two wild trees both from Harrison County, MS.

```
 View  Ask  Report  Create  Modify  Image  Forms  Tools  Scripts  Exit
                                   Pedigree
PEDIGREE|Expt |Entry | Female   | Male
      1 | AA  |  23  |  1002    | 1003
      2 | AA  |  25  |  1002    | 1005
      3 | AD  |  19  |  1266    | 1271
      4 | AD  |  25  |  1258    | 1277
      5 | AD  |   7  |  1260    | 1276
      6 | SD  |  22  |  1002    | 1002
      7 | SD  |  23  |  1002    | 1003
      8 | SD  |  25  |  1002    | 1005

=[↙]============================ Provenan ==============================[↑]=
PROVENAN|Expt |Entry |              |    Provenance
      1 | ZZ  |  1   | Harrison co. MS
      2 | ZZ  |  0   | UNKNOWN



====== 1 of 2 ======◄░░░░░░░░░░░░░░░░░░░░░░░░░░░░░░░░░►
F1 Help   F7 Form   Alt-F9 CoEdit              ..              | Main  |
```

Figure 2. *Pedigree* and *provenance* tables. *Pedigree* contains parent information for trees of known parentage. *Provenance* contains provenance information for wild trees. Table fields are explained in the text.

To find all the full siblings of clone "10" the *pedigree* table is first searched for all records having "1002" and "1005" as parents (either female "1002" and male "1005" or female "1005" and male "1002"). Record 2, experiment "AA" and entry "25", and record 8, experiment "SD" and entry "25", are the only matches. Next, to find the clone numbers of the full siblings, the *clone* table is searched for clones matching experiment "AA" and entry "25" or experiment "SD" and entry "25". Clone 10 (the original clone of interest) and clone "463" are the only matches. Finally, to locate all ramets of clones "463" and "10" the *ramet* table is searched. Four ramets are found-- clone "10" ramet "0" (i.e., the original stock plant of clone "10") at location "GG" and clone "463" ramets "0", "A", and "B" at locations "FI ", "AB", and "AB", respectively.

Locations and Associated Tables

The *location* table is a key to the location codes found in the *ramet* table. It includes a brief description of the location, the names of associated tables, and the name of a technician or scientist to contact for further information. To continue the example from the previous section, clone "463" ramets "A" and "B" are found at location "AB". The location table (Figure 3) shows that "AB" refers to the

accelerated breeding greenhouse, the associated table is *ab,* and the contact person is "Mary Mason". By searching the *ab* table (see Figure 3) for clone "463" ramets "A" and "B", information on the greenhouse location (block), the potting media (pot mix), and the grafting rootstock of ramets "A" and "B" can be found (without bothering Mary). Most locations have an associated table. The location tables' first two fields are always CLONE and RAMET. The remaining fields contain data pertinent to that location. For example, the fields for the *cb* (clone bank) table are: CLONE, RAMET, GC (propagation information), ROW, COLUMN, YEAR_IN, and ROOT_STOCK. Currently, location tables are available for the accelerated breeding greenhouse, the clone bank, the potted and field cutting orchards, and several field plantings.

```
  View  Ask  Report  Create  Modify  Image  Forms  Tools  Scripts  Exit
                                      Ab
AB         Clone |RAMET | GC  |   BLOCK  |   YR IN |POT MIX|ROOT STOCK |
     26      462    B     0G         5       1989      1
     27      462    C     0G        15       1990      2
     28      462    D     0G         1       1989      1
     29      463    A     0G         1       1989      1    W-1-20
     30      463    B     0G         5       1989      1    W-1-20
     31      464    A     0G         9       1990      2    8-7
     32      466    A     0G         9       1990      2    8-7
     33      467    A     0G         7       1990      2    W-1-20

=[↙]========================== Location ==========================[↑]=
LOCATION|Location code|     Location name     |Assoc tables|    Contact
     1      AB         accelerated breeding ghs      AB       Mary Mason
     2      CB         clone bank                    CB       Mary Mason
     3      GG         Gulfport greenhouses                   Tom Caldwel
     4      HO         hedging orchard                        Tom Caldwel
     5      FI         Field planting             FIELD
     6      HG         HEF Greenhouses

===== 1 of 6 =====◀/▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▶
 F1 Help  F7 Form  Alt-F9 CoEdit                        |   Main   |
```

Figure 3. *AB* and *location* tables. Information on trees in the accelerated breeding facility is found in *ab.* *Location* contains information about possible locations of ramets. Table fields are explained in the text.

Other Tables

        Records on pollinations completed and pollen and seed storage are also included in the database. The *flowers* table records information from the field pollination records including: year, female clone, female ramet, tag number, number of conelets pollinated, average conelet maturity, date of pollination, male clone, male ramet, and pollen age. The table is updated with 6-month cone counts, cone collection counts, and pollination or cone collection remarks. Seed lots are entered into the *seed* table, containing information such as year collected, female and male parents (clone and ramet), 100-seed weight, number of seeds, and storage location. Open-pollinated and pollen mix seed lots are assigned to male parent clone "0". A separate table, *Male_0,* records information on specific pollen mixes (open-pollinated seed lots are assigned pollen mix "0").

The final part of the database is a group of tables containing information on experiments. *Experiment* is a master table of experiments similar to the *location* table. It gives a key to experiment codes, a brief description of the experiment, and a contact person. Tables for individual experiments generally take one of three forms depending on the propagation history of the trees. The first form contains ramets of established clones, the second seedlings, and the third a combination of ramets and seedlings. All of these tables include pertinent experimental data such as cutting production and rooting percentage for a rooting experiment, or planting location (row, column) and measurements for a field planting. Pedigrees for clones are accessed by linking to CLONE in the *clone* table to find EXPERIMENT and ENTRY information, and then linking to *pedigree*. Seedlings are given a plant number within the experiment. EXPERIMENT and ENTRY information for each plant (seedling) is recorded in the data table for the experiment. The parent information is recorded in the *pedigree* table. Upon selection for clonal propagation or breeding work, seedlings are assigned a clone number, the seedling is assigned RAMET = 0, and the plant number is recorded as ALIAS. Some of these tables of the second and third form may be "hidden" from the user because they are not based entirely on clone number. An indication of their existence may include one of the following: an entry in the *pedigree* table to which no clones belong, an entry in the *experiment* table, or a group of small seed lots (remnant seed) in the *seed* table.

USING THE DATABASE

Query-By-Example

To understand how to use the database, it is important to know about several features of relational databases. Tables can be viewed, edited, or queried. When in view or edit mode, the zoom feature can be used. Zoom searches the table for the first record in a chosen field matching an input value and advances the cursor to that record. It is generally quick to use zoom to locate records in one table pertaining to one or two clones. To obtain information involving a larger number of records or from more than one table, it is generally faster to construct a query. Paradox uses "query-by-example," which involves filling out a "query form." Fields to be included in the "answer" table are "check marked," and selection criteria are typed into the appropriate fields. Links between tables are established by typing the same "example element" (series of characters) into the common fields in each table involved in the query. The example element serves only to identify the common fields to Paradox, so the choice of characters is arbitrary.

The best way to illustrate how to use the database is to return to the example begun in Linking Clones. Pedigrees. and Provenances. To begin to find the pedigree and full siblings of clone "10" the user would view the *clone* table to find clone "10" and note its experiment and entry information ("SD" and "25"). Next, the user would browse the *pedigree* table to find the parents of "SD"+"25" (clones "1002" and "1005"). Knowing the parent clone numbers, the user can return to the *clone* table to find their experiment and entry information, then browse *provenance* to find their provenance.

Since up to this point only one or two clones are being considered, it is efficient to simply browse the tables. To find all the full siblings of clone "10" it is better to use a query. To perform the query, the user would fill in the query form as shown in Figure 4. The first line of the form specifies a search for the cross "1002" x "1005", and the second for the cross "1005" x "1002". The answer table returned for this query is also shown in Figure 4. A more complex query involving a link between *clone* and *ramet* will return all ramets of the clones of interest in one step. Figure 5 shows the necessary query and the answer table it returns. The example element "abc" allows Paradox to find the ramets of the desired

clones. The answer table can be renamed and saved for future reference or printed out under a standard or a custom report format.
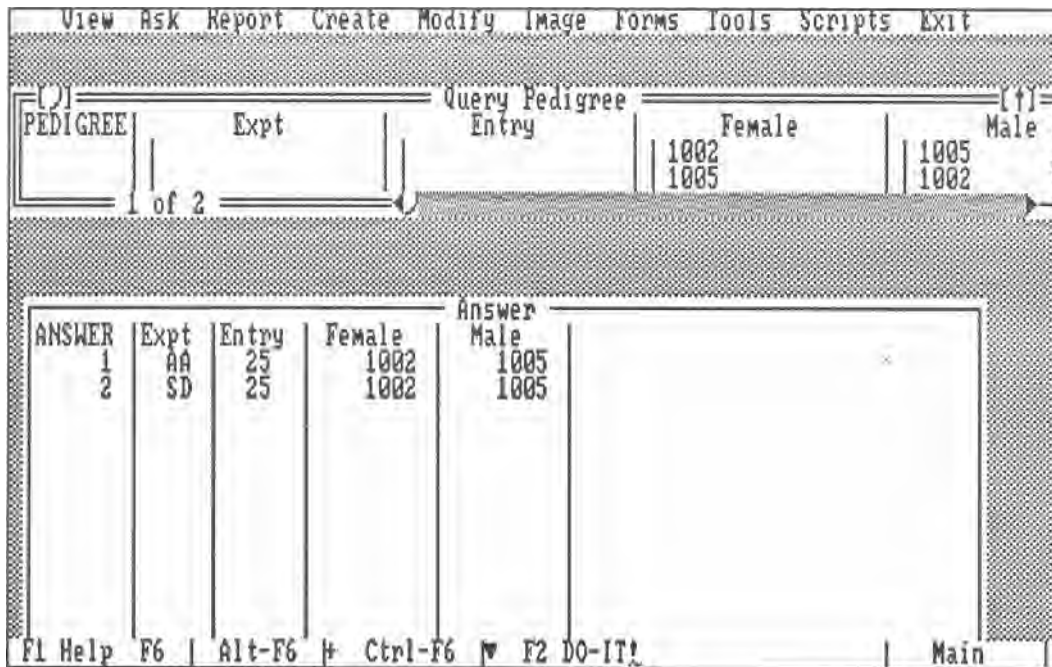

Figure 4. A query to return all clones that are full siblings of clone "10", a control-cross offspring of clones' "1002" and "1005". The first line of the query returns entries in the cross "1002" x "1005". The second line returns cross "1005" x "1002". The entries of interest are shown in the *answer* table.

Custom Application

A custom application is being developed to facilitate access to the database for those who are unfamiliar with it. A custom application offers several advantages over interactive use. First, it provides better security against accidental loss or corruption of data. Second, it allows tasks to be programmed so that a user can get information without having to know the details of the database program. For example, a pedigree can be researched without the user having to fill out a query form with complex links between tables. The application can be menu-driven so that a user can select a task from the menu, fill in a pop-up "dialogue box" to provide the application with needed details to complete the task, and view the desired information on a form with supplemental or explanatory text.

Application modules for use in a complete custom application for managing the database are currently being developed. The completed application should allow a user to see an overview of the database, enter data into various tables, ask a wide range of questions about the data (searches and calculations), and receive printed reports detailing the results of the queries. Extensive interactive use is required to identify the queries most often constructed and thus most necessary for inclusion in the custom application. Anticipated menu-driven queries include the following: constructing complete multi-generation pedigrees; searching for clones, ramets, seed, and pollen lots, etc.; and calculating numbers of clones per family, ramets per clone, and number of seeds per cross, etc.

```
    View  Ask  Report  Create  Modify  Image  Forms  Tools  Scripts  Exit
                                    Query Clone
CLONE  |      Clone      |        Alias      |      Species      |   Experimen
       | abc,463 or 10   |                   |                   | |

                                    Query Ramet
RAMET  |      Clone      |        Ramet      |   Location code   |
       |      abc        | |                 | |                 |

=[·]=============================== Answer ==============================[↑]=
ANSWER |   Clone  |Experiment|Entry |Ramet |Location code|                  ▲
   1   |    10    |   SD     |  25  |  0   |    GG       |
   2   |   463    |   AA     |  25  | ·0   |    FI       |
   3   |   463    |   AA     |  25  |  A   |    AB       |
   4   |   463    |   AA     |  25  |  B   |    AB       |
   5   |   463    |   AA     |  25  |  C   |    AB       |

F1 Help  F7 Form  Alt-F9 CoEdit                              |   Main
```

Figure 5. Query to find all ramets of clones "10" and "463", and their locations. "abc" is an example element linking the CLONE field in the two tables *clone* and *ramet*. "463" and "10" are selection criteria for the CLONE field. The *answer* table shows the ramets of interest.

SUMMARY

A database has been developed to track plant materials maintained and developed at the Southern Institute of Forest Genetics. Through several primary tables, the database maintains clonal identification, pedigrees, provenances, location of ramets, and pertinent details about ramets at each location. Secondary tables contain pollination and seed records and experimental data. The database allows reliable, quick, and easy access to a large amount of information, resulting in more research time and less hard copy search time. A menu-driven custom application is being developed to ensure these advantages are not lost for those who only occasionally use the database.

LITERATURE CITED

Borland International. 1992. Paradox Version 4.0 User's Guide. Borland International, Scott's Valley, CA. 832 p.

Snyder, E.B., J.W. Beland, and R.C. Schmidtling. 1968. A breed tree information card with automatic data processing and needle-sorting provisions. Unpublished document on file with FS-SO-4153, Gulfport, MS. 12 p.