ASSET Series A (2003) 3 (4):91-99

ISSN 1595_9694 © UNAAB 2003

A MODEL FOR IMPLEMENTING COMPUTER DATA BASE SYSTEMS FROM MAIZE VARIETY PERFORMANCE DATASETS IN NIGERIA

I. O. DANIEL* AND M. O. AJALA
Department of Plant Breeding and Seed Technology,
University of Agriculture, PMB 2240, Abeokuta, Nigeria.
Corresponding E-mail addresses: <Daniel@unaab.edu.ng><drdayodaniel@yahoo.com>

ABSTRACT

Performance data of maize varieties at different locations need to be accurate and accessible to stimulate the improvement of the Nigerian maize seed system. This paper describes a database model to implement a simple computerized information system for maize varieties and their performance at various locations in Nigeria. Entity-Relationship modelling was used to identify relevant data and their relational structure from subsets of maize germplasm trial data with Microsoft Access™ Relational Database Management System software. The model has the capacity to provide detailed variety performance information in relation to cropping environment with specific search queries when the relevant data is properly documented. Such facility will benefit seed producers who will determine seed materials that meet specific market and production challenges of farmers' with greater precision. Plant breeders can also use the information system for breeding plans and policy makers can use it as a Decision Support System (DSS).

Keywords: Electronic database, maize performance data modelling.

INTRODUCTION

A major limitation to the production of improved seeds for specific crop production problems at different areas in Nigeria is the dearth of robust and accurate information for crop breeders, researchers and seed producers (Usman *et al.* 1992). This dearth of information is a result of data inaccessibility not necessarily unavailability. Data from most of the multilocational trials of different varieties are held in bits by pockets of researchers in the country, rendering the information inaccessible to prospective seed producers who require the information to produce proprietary seeds that meet farmers' production and market needs from time to time. Currently, when such needs were known, much difficulty is encountered in identifying suitable cultivars for different production problems in a timely fashion. Plant breeders may also need to obtain population characteristics for cultivars, breeding lines, or germplasm suited for particular situations, and this can be a time-consuming exercise given the amount of plant genetic resources available. The availability of information resources to serve breeders and seed producers will enhance the development and recommendation of appropriate germplasm for specific seed needs more accurately and precisely. It will also stimulate the development of the seed industry in Nigeria. Such databases have been developed for several crops at different regions of the world *e.g.*. alfalfa in the U.S. (Hannaway *et al.* 1992), world potato collections (Hawkes, 1997), and more recently for some crops in the International agricultural research system *e.g.* the International Crop Information System (ICIS) (McLaren *et al.* 2001).

This paper proposes the application of a computerised information system using the Relational Database Management System (RDBMS). The RDBMS packages are meant for the development of knowledge-based systems, which permit the design of an information system upon the knowledge of the subject matter and database management skills (Darwen and Date. 1998). A relational database uses twodimensional tables (entities) for the storage of information. Entities are made up of attributes or fields (columns), and records (rows), representing individual data entries for each attribute. Each entity holds a segment of related data already identified as a component of business logic and references other entities in the database to provide useful business information. Each table references each other by the use of primary keys (PK) and foreign keys (FK) such that a query can be issued to join information on two or more tables in the database to get needed information. The packages use the Structured Query Language (SQL) to develop database applications. The system serves as a data storage devise, a data warehouse for decision support that permits rapid search and updating of specific information in the database.

A major feature of RDBMS is Entity-Relationship *(ER)* modelling. This involves identifying the key data and its associated data sets and defining their relational pattern for building an information system. *ER* modelling has been used extensively in business computing to aid development of relational database software. Cormack and Moore (1980) discussed the relevance and the importance of using data modelling to analyse land resource data. Tzortzios *et al*. (2000) used the same approach to establish a crop and livestock database. In a similar fashion, *ER* modelling can be used to catalogue plant genetic resources and their performance data. For example, a crop database can hold information on cultivars, traits, growth habit, special qualities,
disease/pest resistance *et tetra* on different entities which can be joined together to make a full crop information. Once the data model is established, the design and implementation of the information system is relatively straightforward.

The purpose of this paper is to present standard data sets for crop researchers in variety performance trials and a logical *ER* model based on the data sets that a software developer can use to implement an information system that manages maize genetic resources and performance data in Nigeria. The model components constitute the modest data to be collected by the crop researchers to elicit crop information maximally from the database system.

MATERIALS AND METHODS

A subset of real data from regional trials of elite maize germplasm of the International Institute of Tropical Agriculture (IITA) (Menkir *et al*. 2002) was used to identify relevant logical data sets for implementing maize variety performance trial database and to formulate the model components. Data components used to form the logic i.e. entities and relationships of the database system were from the variables assessed in the trials. The variables of the data set were .grouped by their relational attributes into database fields of different entities, which are transformable into tables.

The entities and their relationships described in the model were designed and presented on Windows 98 operating system platform with the Microsoft Access database engine, which is a relational database management software that ships with the Microsoft office 2000 suite. SQL scripts were used to create entities and their relationships; the SQL scripts were coded with the database object tools provided in the main window Microsoft Access software.

RESULTS

Data Components

The relevant data identified from the trial data sets are primarily the cultivar identity and the cropping environment that defines the cultivator's performance *i.e.* passport and origin, pedigree, certification and grain data, as well as trial and location information. The secondary data are quantitative performance traits whose expression depends on the Genotype x Environment interaction. Thus, the database model fits data from results of multi-locational trials of breeding populations, breeder, foundation or certified seeds.

The data sets were divided into 5 entities holding a total of 40 fields (Tables 1 and 2); the data type is the form in which each field holds its data. Table 1 shows the information to be recorded in entities that are independent of environmental factors *i.e.* the cultivar and trial entities. The *Cultivar* entity holds information that identify cultivars *i.e.* cultivator's name and/or code for breeder seed materials, pedigree, origin, genotype and seed status and breeder seed accessibility information of individual cultivars. The *Trial* entity contains information on the names, date and owners of trials associated with each cultivar and their dates.

Table 2 shows the performance entity with environment dependent attributes, in other words, the location-specific performance data. On the *Location* entity the *LocationID* field uniquely identifies production or trial areas and describes their ecological attributes including rainfall, elevation, longitude and latitude, which would

Table I. Entities and attributes of cultivar identity

Entity	Attributes	Data type	Information
Cultivar	Cultivar ID	Integer	Unique identifier (PK)
	Cultivar Name	Text	Cultivar names and codes
	Origin	Text	Centre of origin
	Breeder	Text	Breeder's name (institute, company and person
	Pedigree	Text	Parental materials
	Genotype	Text	Breed type (hybrids*, landraces, exotics, wild etc
	Certification Status	Text	Breeder, foundation or certified
Trial	Trail ID	Integer	Unique identifier (PK)
	Trial Name	Text	(Title of trial
	Trial Date	Date/Time	Date of Trial
	Trial Owner	Text	Name and institute of researcher/s

* Single Cross Hybrid, Double Cross Hybrids, Top Cross, Open Pollinated and Breeding population

Table 2. Location and performance trait entities and attributes

Entity	Attributes	Data type	Information
Location	LocationID	Integer	Unique identifier (PK)
	LocationName	I Text i	Name of trial site
	Agroecozone	Text	1 Savannah, Sahel, Rainforest etc. j
	Elevation (m)	Numeric	As applicable
	AnnRainfall (mm)		As applicable As
	Latitude	Numeric	applicable
	Longitude	Numeric	Numeric As applicable
Performance	Perform ID	identifier	Integer Unique
	Yield	Numeric	Seed yields (Kg/ha)
	PlantStand	Numeric	Plant population/Ha
	Day Silk	Numeric	Days to silking
	DayPollen	Numeric	Days to pollen shedding
	PlantHeight	`Numeric	Average plant height (m)
	ShootLodge	Numeric	Rated score or % of shoot lodging
	HuskCover	Numeric	% or number with poor husk
	EarHarvest	Numeric	cover
	Seed Weight	Numeric	Total number of plant harvested 1000 seed weight (g) after drying
		Integer	Unique identifier (PK)
i Resistance	ResistID	Numeric	j Score of plant appeal or damages
	PlantAspect	Numeric	Score of ear appeal or damages
	EarAspect	Numeric	Steak resistance rating scores
	Streak Rust	Numeric	Rust resistance rating scores
	Blight	Numeric	Blight resistance rating scores
	Curvalaria	Numeric	Curvularia resistance rating
	DownyM ildew	Numeric	Scores
	EarRot	Numeric	Number of plant damaged by DM i Number or % affected by ear rots
	StemBorer	Numeric	Resistance rating score for SB
	Striga	Numeric	Resistance rating scores for striga

enhance location based decision support for crop and seed production. Quantitative performance traits including agronomic performances and disease resistance are recorded in the Performance and *Resistance* entities. The *Performance* entity holds crop production data including the yield, maturity period and general morphology of each cultivar under each production environment. *Resistance* entity records the levels of resistance to 6 different maize diseases from disease resistance, incidence and susceptibility scores in evaluation trials. Also recorded in the *Resistance* entity are resistance or susceptibility or incidence scores of insect pests (stem borers) and *Striga*. Resistance to abiotic stresses like heat and drought could be inferred from growth and yield information under the different rainfall patterns associated with the various locations.

ER model of the database

Figure 1 is the logical model that shows relationships among entities in the maize variety performance database. Entity identifier fields are computer generated unique numbers like integers to ensure uniqueness of each record and they bare the post-fix 'ID' after the entity name for example, *CultivarlD*. Primary key constraints (for each entity) and foreign key constraints (that references related entities) were placed on these entity identifier fields of each entity because of the similarities in their data type.

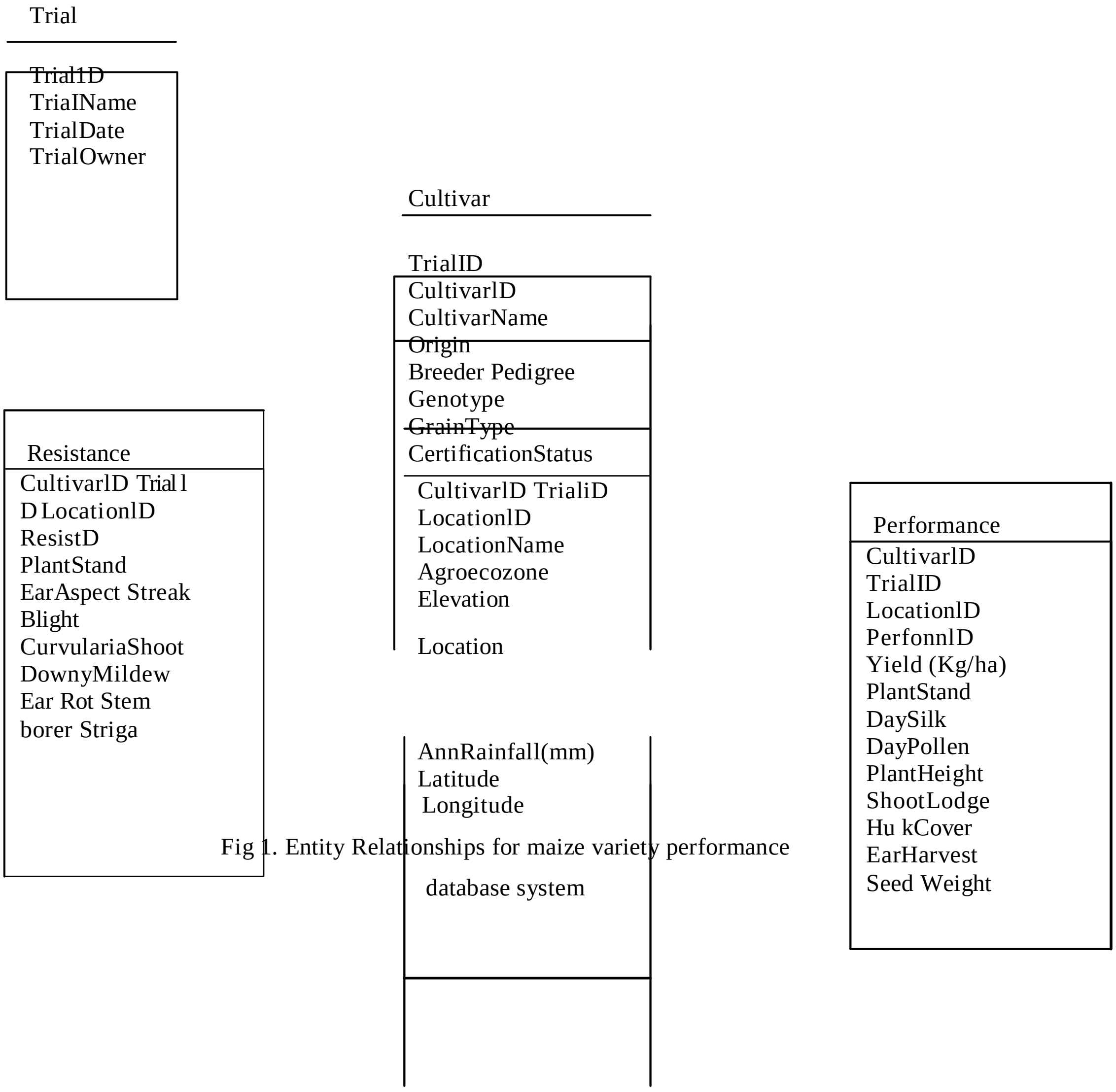

Fig 1. Entity Relationships for maize variety performance database system

The core of the *ER* model of the database consists of the two independent entities, *Cultivar* and *Location* (Fig. 1). The field *CultivarlD* uniquely identify cultivars and *LocationlD* uniquely identify locations in the *Location* entity. The two entities relate to the crop performance attributes on the *Performance* and *Resistance* tables through the *CultivariD* and *LocationlD* fields respectively. Both relationships are one-to-many i. e. each cultivar entry can reference all records in the referenced columns of the two entities. The *Cultivar* entity references *Trial* on the *TrialID* column in a one to

5/06/03

Location

CultivarlD	Trial ID	Location ID	LocationName	Agroecozone
			IBAGAUDA	Lowland dry Savannah

CultivarlD	Trial ID	PerformlD	Yield(Kg/ha)	
	1	1 1	*3154*	
	2	1 2		
	3	1 3	4420	
	4	1 4	3351	
	5	1 5	3844	
	6	1 6	4701	
	7	l 7	4217	
	8	1 8	3696	
	9	1 9	3750	
	1	1	4144	
	II	I	2700	
	1	1	4062	
	1	1	4165	
	1	2	*3736*	
	2	2 15	4764	
	3	2 16	4351	
	4	2	*4696*	
	5	2	4185	
	6	2	4804	
	7	2 20	4720	
	8	2	5257	
	9	2 22	5870	
	10	2 23	3794	
	11	2 24	4801	
	1	2 25	4509	
	1	3 26	0	

2 EGBEMA 3	Lowland	Rainforest
IKENNE 4	Lowland	Rainforest
SAMARU	Lowland Moist Savannah	

Fig 2. Results of a location based selection query to retrieve maize yields in 2 different trials at *Bagauda.* Cultivars

and trial information can be viewed by join queries that look up to *Cultivar* and *Trial* tables many relationships so that it can uniquely identify performance data of each cultivar by a combination of *CultivarlD* and *TrialID* columns since cultivars may differ from trial to trial.

There is no relationship constraints between the *Cultivar* entity and the *Location* entity to allow each performance attribute a unique identification by a combination of *CultivarlD, TrialID, LocationID* and *PerformID* fields and resistance attributes by combination of *CultivarlD, TrialID, LocationlD* and *ResistlD* fields (Fig. 1). The *Location* entity has the *CultivarID* and *TrialID* foreign keys so as to be able to join information on cultivars involved in different trials to the location of the trials (Fig. 1). The primary key *LocationlD* field on the *Location* entity relates to the *Performance* and *Resistance* tables by the foreign key *LocationlD* yielding one to many relationships that can join location information to performance and resistance traits with *CultivarlD, TrialID* and *LocationlD* as unique identifiers.

Data can be loaded and retrieved directly on the MS Access tables presented in the model using the Microsoft Access objects. For example to demonstrate how to load or access information on yields of different varieties at *Bagauda ' i* trials 1 and 2 from this database structure, simply query the *LocationName* field in the *Location* table and click under the *Bagauda* record, then choose the *Performance* table to retrieve the yield attribute data. The result is as shown in Fig. 2 . Advanced search queries to access individual variety performance for instance, selection of a set of varieties with a specified level of yield values at *Bagauda* could be written with SELECT SQL statements to query the affected tables and bring the varieties on the result window.
Such SQL queries will be scripted and coded as visual objects into user-friendly interfaces of the database model at the stage of front-end database development.

DISCUSSION

Implementation

The model described above can be implemented on Windows 98 or higher Operating Systems in the Windows environment. The implementation of the database for deployment as an information system require a set of database programme which manages control of the database, and a set of application programme to manage installation of system software, access to and interfaces of database system. The database programme is executed on Microsoft Access databases as the system's back-end application using the SQL source codes. The visual development tools (e.g. Visual Basic 6.0) (Craig and Webb, 1999) will then be used to develop the front-end application, which integrates with the SQL codes of the MS Access to the database. A user-friendly interface would be ensured through the object-oriented codes of Visual Basic™ used for most SQL queries that control database access and management. The use of Open Database Connectivity (ODBC) interface SQL (Stallings, 1996) will enhance data connectivity with the Internet and other databases *e.g.* the ICIS.

The possibility of implementing the database model on existing database systems makes the model more versatile. For instance, the ICIS implementation is possible because ICIS also uses the Microsoft Access database engine for its logical structure and some of the entities hold marching data types, so that queries can be written to join entities in this data model to ICIS tables. Moreover, ICIS was implemented on two tiers: the central and local databases (McLaren *et al.* 2001), guest users can create many local databases while the only central database holds information on genealogy and location information. Hence, ICIS implementation of the multi-locational performance database for maize varieties in Nigeria will allow link with a vast array of genealogy information as well as a genealogy data analysis module when data types of the data fields are similar. The entry of geographical coordinates in form of latitude and longitude on the *Location* entity of the data model (Table 2) will also enhance integration of variety performance information on maps with Geographical Information System (GIS) facilities.

Application and database programmes can be written to compact discs (CD-ROM) and distributed for installation on stand-alone computer systems, which is a readily acceptable deployment option because of lower investment and overhead costs but data have to be updated regularly on new CD versions. The ultimate option for the deployment of the system is Wide Area Networking (WAN) that will involve a server-client architecture whereby database is held on a dedicated server computer at universities or government agencies and accessed only by registered SQL client computers across the country that can update data in the system. This deployment option is administration and capitalintensive but it has advantage of on-line data control (update/delete) and security settings (access control) according to user authorization.

Potential benefits of the database

The potential benefits of the database shall be examined from the end-users' point of view. The end-users of crop variety performance database are crop breeders, researchers, seed companies. agricultural consultants, universities, research institutes, government service agencies *e.g.* the National Seed Service (NSS), Agricultural Development Projects (ADPs) and the Variety Registration Board, policy makers and agricultural extension agents among others. For instance, crop breeders and seed producers who screen germplasm for desirable traits to develop new varieties in response to production and market problems will find an electronic database a useful tool to reduce the time of selection of a breeding population at the same time give necessary information on seed availability and thus enhance efficiency of crop improvement. This will also result in rapid diversification of the genetic resources base for maize improvement and seed production.

The ability to provide specific information to policy makers, researchers, extension agents and farmers is an important factor in achieving sustainable agricultural development. As each cultivar in the database is related to cropping environment within the country, the database can be queried to identify and recommend cultivars for specific locations. In addition, the database can be used to extract information based on cultivar attributes meeting any combination of selection criteria. For example, a hard copy report can be generated of all cultivars with yields above certain limits for each agroecological zone in the country. This type of specific information can aid decisionmaking in allocation of incentives by government and/or funding agencies.

Moreover, the database can be updated to incorporate new germplasm introduction.

variety registration and additional data e.g. nutritional analysis information. Data can also be retrieved for statistical or mathematical analyses. Finally, the database will add value to data generated by crop researchers from trials and complement data entry at institutional level, which would make crop variety data more functional and consistent.

CONCLUSION

The need to provide crop variety databases to increase efficiency of crop improvement and enable focused decision making for seed production in response to crop production challenges have been emphasized in this paper. The implementation requires a multidisciplinary approach among agricultural and physical scientists within national programmes and universities. For further information about collaboration, the lead author can be contacted at the Department of Plant Breeding and Seed Technology, University of Agriculture. PMB 2240, Abeokuta. E-mail: daniel a unaab.edu.ng or drdayodaniel a.yahoo.com

REFERENCES

Cormak. R. S. and Moore, A. W. 1980. Modelling of land resource data for computer storage and retrieval. In: A. W. Moore, B. G. Cook and L. G. Lynch (Eds.). *Proceedings of the second Australian meeting of the ISSS Working Group on Soil Information Systems.* Pp. 8-24.

Craig, C. J. and Webb, J. 1999. Microsoft Visual Basic 6.0 Programmer's Toolbox. 5T Edition, Microsoft Press. US.

Darwen, H. and Date, J. C. 1998. Foundation for Object/Relational Databases. Addison-Wesley Publishing Co. UK.

Hannaway, D. B., Belie, J. P., Shuler, P. E., Ballerstedt P. J. and Younes, M. B. 1992. ACE: Alfalfa cultivar expert. *J Prod. Agric. 5* :85-88.

Hawkes, J. G. 1997. A database for wild and cultivated potatoes. *Euphytica* 93 :155-161.

McLaren, C., White, J. W. and Fox P. N. 2001. International Crop Information System (ICI S). Technical development manual. IRRI. Phillipines & CIMMYT, Mexico.

Menkir, A., Kling, J. G. and Ajala S. O. 2002. Evaluation of elite maize germplasm in West and Central Africa. Report of the regional trials conducted by IITA and its collaborators, Ibadan, Nigeria. IITA.

Stallings, W. 1996. Computer Organization and Architecture. Prentice-Hall. 4'l' Edition. USA.

Tzortzios, S., Adams, G. and Gitsakis, N. 2000. An interactive biometric-oriented database system for presenting comparative information on conventional and sustainable agriculture. In: S. Nokoe (Ed.). *Biometry and Quality of life. SUSAN-IBS Proceedings.* Pp.205-214.

Usman, 1. A., Maini, N. S., Joshua A. and Falusi, A. O. 1992. Strategies for the national seed subsector development plan. National Seed Quarantine Project, National Seed Service, Federal Department of Agriculture, Federal Republic of Nigeria.