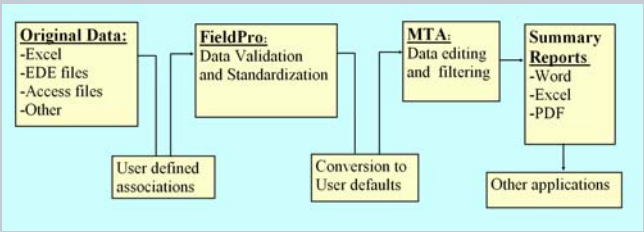
Utilization of a Relational Database System for Summarization of Global Research Data on Diseases such as Soybean Rust.

Bruce A. Fulling and Wayne Olson
Heartland Technologies, Inc., Fishers, Indiana 46038

Abstract

For worldwide diseases such as Soybean Rust, the development of software with the ability to summarize data from numerous sources can provide valuable insights into the effectiveness of disease management strategies. The software system has to address three major areas of concern; 1) the importing data from various sources and geographic locals to create homogenous databases, 2) the ability to edit and combine data category items, and 3) providing data filters to generate specific summary reports. The FieldPro software has been developed as a relational database for importing data. The data is standardized during this process with the extensive use of dictionaries based on internationally recognized sources such as the ACPA SART codes, BBCH growth stage codes, and the EPPO Bayer species codes. The speed of trial import and standardization depends greatly on the data quality of the original trials. After importing and standardization, the data is sent into the MTA (Multi-Trial Analysis) module, which allows the data to be edited and summarized without affecting the original data points. To date the MTA summarization module has been tested on databases of over 6000 trials originating from numerous countries.



Setting Filter ranges and Combining Data in MTA

Filter ranges allow the user to group data into categories that can then be averaged together. Ranges can be set for several soil categories such as Organic matter, CEC and pH. These filters are used to include or exclude trials from MTA summaries. Ranges can also be set up according to date.

After all the edits are made and filters ranges are set, then the data is combined. This process merges the individual trial lists and produces a unique list of evaluation headers and treatments for the entire database. These are the lists that are used for filtering the data and producing summaries.

Data Standardization

Data summarization involves the combining of data from numerous sources. This can involve multiple researchers from multiple countries. A primary pitfall in data summarization can be the lack of universal standards among the data sources. A single species may have several scientific and common names by which it has been described among different trials. Add in the numerous freeform names and abbreviations researchers may have used and the process becomes very complex.

Any system that is used for data summarization needs to incorporate codes and names that have as universal consensus as possible. A relational database allows the effective use of standardized coding dictionaries. In the FieldPro software, the 100 point BBCH system for pest/crop stages and the Bayer Codes for naming the different species have been used. The SART codes (Standard Agricultural Research Terminology) have been used, wherever possible, in the data entry fields.

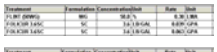
MTA Structure

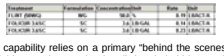
The MTA (Multi Trial Analysis) is designed to import data from FieldPro and summarize the data. The data management (FieldPro) and data summarization (MTA) were set up as separate systems so that edits made in the summarization process in MTA does not affect the original trial data in FieldPro.

MTA allows the same FieldPro trial to be imported into several different databases. This allows the MTA databases to be focused in specific areas to enhance the efficiency of the summarization process. For example, one database may consist of trials that include evaluations on soybean rust. Some of the same trials can be imported into a second database that focuses on Septoria leaf spot.

Default settings in FieldPro determine how data will be displayed in some general aspects in MTA. For example, treatment rates can be imported in metric or U.S. units. Species common names can be imported in English, German, French, or Spanish.

Additional default settings within MTA allow the researcher to change the display as needed. For example, rates can be displayed as active ingredient per area or in product units as shown below.





This capability relies on a primary "behind the scenes" feature of FieldPro and MTA, that is, all data is held in standard base units of measure. Data is converted through a Unit Code dictionary to be displayed or output to a default format of the user's choice.

Report Generation

Filters are used to include or exclude data from a summary report. After the researcher has selected which evaluation headers and treatments to summarize (bottom row of buttons), the MTA filter dialog has 12 filter selections to refine what data is used in the summary. The button changes color depending on what is selected: Green indicates all available items in that category are selected, yellow indicates some items are selected, and red indicates no items are selected.

After the filters are set, a report can be generated. A researcher can then change the filter settings to create a different summary report. A large number of various summary reports can be output in a short amount of time in this manner.

INTRODUCTION

The past 25 years has seen an explosion in the usage of computers for field research management and data storage. Most field trial management programs were initially designed to make the researcher's job easier with little thought to data summarization. These programs usually created data structures in memory to hold the various components of a trial but stored this data in a flat file structure in much the same way as a spreadsheet or word processor. The real limitation of these programs has been the inability to summarize and analyze data from multiple sources. In order to perform any multi-trial analysis, the flat data files must be re-assembled into some sort of ordered structure again. Bringing these trials into a database is complicated by the fact that coding systems are not always uniform.

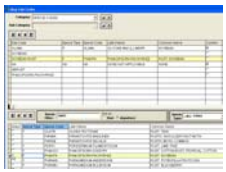
FieldPro and MTA were designed by merging the best features of both types of systems. FieldPro stores data directly into a relational database using standard units of measure. This allows trials from various sources to be immediately analyzed and eases the task of uploading data to a summarization program such as MTA. Visual FoxPro™ by Microsoft was chosen as the underlying database because of its speed, reliability, and rich programming language.

The need for a relational database system is most apparent when summarizing research data for a disease of global importance such as Soybean rust. The types of data coding and data management can vary widely and only a relational database would have the capability of resolving these differences and bringing the data into a single standardized system.

Data Import into FieldPro

When summarizing data from numerous sources, it is almost certain that the data will be in different formats. Common formats can involve spreadsheets such as excel, other database systems such as Access, word documents, or other data management software such as ARM. Both FieldPro and ARM utilize EDE (Electronic Data Exchange) file format tools to allow the exchange of data.

It is during the data import process that the standardization of the data takes place. For example, in the import of the EDE files, a window comes up that lists all of the codes in the file that don't match the codes used in FieldPro. In the example shown, there are several codes for soybean rust in the EDE file (top part of the screen). The codes are matched to the correct code in FieldPro. After all of the codes are matched, the process of importing the EDE file continues. The matches that have been made are remembered in FieldPro so that the same matches don't have to be repeated for each EDE file that is imported. Thus, the import process becomes progressively quicker as additional trials are imported.



Editing in MTA

While the data has been standardized during the import process into FieldPro, it most likely requires further editing so that useful summaries can be made. Many common edits that are made are to treatment rates, treatment names, variety names, etc.

In the example below, several rates of the products imported into MTA are only slightly different (lower list). For purposes of summarization, the user would want to combine them. Two of the Quadris treatments below have been tagged to be combined into one.

Discussion

The FieldPro and MTA software has been developed and utilized the past 15 years in both industry and academia. The Foxpro database on which FieldPro and MTA are based on has proved to be versatile and robust. MTA has been shown to be capable of handling databases of over 6000 trials though subject specific databases of smaller size are usually more effective and efficient for the user.

Data import and standardization from other sources can be fairly quick, especially if the data is in electronic format such as EDE files or Access databases. The speed of import of data from these sources can reach several trials a minute. However, the quality of the data is only as good as the original research and documentation and the time required for import may be affected by the quality of the trial information.

A key factor in designing a relational database for data summarization is making the system acceptable for users around the globe. FieldPro and MTA has been used at the international level for over 15 years. Significant effort has been made by several groups to develop coding and dictionary standards to represent the needs of all researchers.