

Data Analysis Management In Remote Fields

By Robert Stottlemeyer

Data management I define as a process for making data accessible and usable to others. If one accepts my definition, it logically follows that good data management should also lead to better data analysis by the researcher.

Four years ago I left the plush security of the bureaucracy and re-entered the research business in an academic community. The array of information I found routinely collected in field ecological study, which I had been familiar with years before was a startling change. But so too was the technology now available to handle such data. The question was, and is, how to match the two, for data generated in remote field locations.

Severe fiscal constraints, a long-term habit of composing on the typewriter, and faculty frustration with peripheral services on the university mainframe made the decision to buy my first microcomputer with word processing software easy. This was four years ago. Sure one can type faster on a wordprocessor, but it is the library buildup that really is the time saver. It is truly amazing how much of what we do on an annual basis is repetitive. That carefully prepared abstract for a national meeting can, with editing, serve as a progress report. Updating annual reports, vitas, follow suit. Fortunately, I purchased equipment from one of the three major manufacturers so a formidable array of software is available at modest cost. This also was a big help when I had to decide on an SPS package to handle the mounting data set being generated by a laboratory analyzing several thousand samples a year, each with over 20 variables. I could easily have gone with the university's mainframe. But faculty colleagues encouraged me to go "micro" all the way until file size dictated otherwise. Hardware advances have moved much faster than my file sizes and analytical demands.

Data organization and structure all too often do not get enough thought early in the game. I was fortunate in having to address this question at the beginning, when we developed a "universal" field data form. Two questions arose: "What are the types of entities about which we have data?", and "What data do we have about each type of entity?" Our spatial and temporal field data requirements were similar to those of a soils scientist, and a much modified computer-based soil science field form, which includes a "variable" for data documentation, has served us well.

The conduct of field research in especially remote sites has its own set of quality assurance problems. Data documentation serves to record those "field contingencies" which invariably occur with remotely-placed field recorders. But our work involves fairly sophisticated chemical analyses of field samples, and this must be done in a facility much removed from the sampling sites. While one can easily develop protocols for sample handling and data recording how can their effectiveness be checked?

The key is a vertically integrated quality assurance program which can respond in a timeframe sufficient to correct procedural errors before one or two additional sampling dates take place. For our work the need is to merge certain field data variables with laboratory analyses for "outlier" detection by regression.

Data Analysis, Cont'd.

This sounds simple (*the heck it does!* — Ed.) except when dealing with remote sites where environmental, logistical, and shipping factors prevent routine field data transmission by mail.

To reduce data handling and obtain field data on a timely basis for merging, we have replaced the field form with a truly portable microcomputer spread sheet. The unit we use for all remote sites has simple word processing capability allowing complete weekly report generation. Unlike a datalogger it should not be carried out in the rain. But almost any shelter will house it for it needs no external power source. Its standard keyboard and number pad, computer flexibility, and price tag of less than one half the datalogger makes it an attractive alternative. With the built-in modem, weekly transfer of field files across the phone to the lab or office micro files is fairly straight forward.

This solves our problem of timely data transmission, but concern about initial manual data entry remains. Very few research programs yet have fully automated data handling for all aspects of the study. There is an overwhelming urge to delegate the responsibility of data entry. I think this is very risky. Data entry still is best done by the person familiar with the data, and it is best done during the data collection process. With the low cost, efficiency and portability

of microcomputers, data entry by the researcher is often the method of choice.

I now have four micros at work most of the time. All reports and data analyses done by graduate students or research technicians are done on one of these machines. The hardware is fully compatible, allowing easy editing of text files and analyses on machines in different physical locations. Machine time, especially in field stations, is increasing with more installation of solid-state recording instruments. Perhaps most important in the short term has been the improved quality assurance, which is so dependent on timely data transmission from remote research sites. We are taking the easy initial steps in developing a data management system. This is considerably helped by the competitive nature of the micro software market.

But there are caution flags. While hardware is the thing that really seems to be pushed on the consumer, I have always found software, even with a major brand computer, the limiting factor. I'm a user, not a computer buff, so I don't ever expect to be a hot programmer. I think you have to take a hard look at your data analyses and management needs before you can separate limiting factors due to hardware from those due to software. Once done, you might find yourself still typing on an old antique eight-bit processor such as this was composed on.

Stottlmyer is a Research Scientist with the NPS/CPSU at Michigan Technological University.