

TAMING ENVIRONMENTAL DATA

NENO DUPLANCIC
GREGORY BUCKLE

The main problem with environmental management is the vast quantity of data generated for and by analysis. The parameters for any given investigation can be numerous: There are 127 on the EPA list of priority pollutants alone. A new computer system moves data from the database to commercially available graphics software without operator intervention.

Closing a landfill requires far more than layering it with dirt and a weatherproof cap. The owner must prove that there's no danger of contamination, especially to local aquifers, by monitoring wells, analyzing the data and reporting results to one or more regulatory agencies.

In California, one landfill-closure program involves several sites operated by International Technology Corp. (IT) of Torrance, Calif. The company is preparing to monitor more than 200 wells over a 30-year period. The California Regional Water Quality Control Board requires a semi-annual *x-y* plot of changes in concentration over time for every parameter tested.

The engineers and scientists will be able to supply those plots because of the IT Environmental Database Management System (ITEMS), a computer system for managing vast amounts of environmental data generated during site characterization and monitoring studies. The system can move

data from the database to graphics software, download analytical results from different laboratories electronically and format output in any number of ways.

On the East Coast, IT Corp. is investigating leachate distribution at the Fresh Kills Landfill on Staten Island for the New York City Department of Sanitation. The landfill is the largest in the world, and an exceptionally large amount of data is being developed in the hydrogeological, surface-water and sediment investigations. The project involves analyzing 288 parameters in samples from about 180 ground-water wells and 17 surface-water stations each quarter. Some 200 soil borings have been taken, and all of the analytical data, boring data and the well construction data is stored and managed by the ITEMS system.

Outputs vary from standard to complex. Tables for submission to the regulators can be formatted in almost any sequence, reporting only those parameters specified.

Well construction and boring logs, on the other hand, include all of the geologic descriptions produced through the interface with AutoCAD. Backup quality-assurance data is automatically generated for each chemical.

Other outputs through standard software packages such as SURFER and GRAPHER by Golden Software, Inc., Golden, Colo., generate both two- and three-dimensional concentration contour plots based on a coordinate grid system that locates each well. The system also works with the geographic information system GEO/SQL-5. Users work directly on the ITEMS system and manipulate the data directly. Other interfaces, including Lotus 1-2-3, allow them to download data to their own computers for analysis and manipulation related to the project. The major time-saver is downloading the data from the analytical laboratory directly into ITEMS.

The system was designed in-house by



ENGINEERS ARE MONITORING LEACHATE DISTRIBUTION AT NEW YORK CITY'S FRESH KILLS LANDFILL, THE LARGEST LANDFILL IN THE WORLD. THE IT ENVIRONMENTAL DATABASE-MANAGEMENT SYSTEM WILL STORE AND MANAGE ALMOST 300 PARAMETERS FOR THIS PROJECT. PHOTO COURTESY IT CORP.

engineers and technicians for the landfill projects and other environmental projects associated with soil and ground-water contamination. Deciding that it made no sense to duplicate readily available software, they kept costs within bounds by incorporating existing graphics packages. This also allows future users to reap the benefits of new graphics capabilities as they become available.

ITEMS consists of standard forms for entering and editing geotechnical, hydrogeological, sampling and chemical data; various tables for storing the data; a secondary database containing information, such as the Chemical Abstract Series (CAS) numbers, names, chemical/physical properties and regulatory limits; a report generator; and a variety of application programs that link the central database to graphics and

contouring packages.

We developed the central database—consisting of several levels of menus, input screens and the report generator—using the Oracle Relational Data Base Management System (RDBMS by Oracle Corp., Belmont, Calif.) and C programming language. To generate graphics, the system relies on commercially available programs such as Autocad, SURFER and GRAPHIER. The links to these and other software packages operate in such a way that data are moved transparently without the need to exit from one program and load another.

THE DATABASE

There are two ways to enter data into the central database: (1) manually input field, analytical and mapping data; or (2) import files containing these data. More than 50

entry forms can record stratigraphic and lithologic information from soil and rock borings; monitoring-well construction details; geotechnical data; well development records; hydraulic conductivity and water-level measurements; survey data; and ground-water, surface-water, sediment, soil and waste sampling information. Users select the forms they need.

Although they appear as simple screen displays for entering and modifying data, the forms are quite complex. Many have triggers that are executed when data are entered or the cursor is moved into a field. These are used to validate entries or display related data from another form. ITEMS provides a standard file format for interchanging analytical data. For instance, laboratory results can be imported to the database tables via modem or floppy disk,

speeding the task and reducing errors.

The central database provides a means to enter, store, maintain, update and retrieve data. Menus control all data processing (aside from entry)—gaining access to entry forms, selecting what standard report is to be printed, specifying what information is to be included and linking to other application programs. Users do not have to be familiar with any computer language or database-management system software to perform routine data processing. This not only makes the system easy to use but also easy to learn.

The on-line help screens contain information on the purpose and structure of the form, identity of required fields, type of information to be entered in specific fields, and special keystrokes that speed data entry and cursor movement within the form. Certain data fields also have their own help screens. Of the several methods of querying the database, one of the simplest is to enter queries on the same forms used to initially enter the data. Search criteria are expressed in terms of the values found in certain fields. Queries can also be initiated via Oracle's database add-in for Lotus 1-2-3 or via other Oracle products such as Easy*SQL.

One of the major weaknesses of many database-management systems is the inability to extract specific data sets for a printed report. With ITEMS, users select from several reporting formats and can specify which information is to be included. For analytical data reports, for example, selections can be based on sample number, sampling date, analytical parameter and/or whether the result was above the detection limit. Users can also choose to print only those results that exceed project-specific limits or regulatory limits, which can be retrieved from the secondary chemical database. Other retrieval and printout choices can be made by well location or number, well screen elevation and/or geological monitoring zone. An equally important capability is generating ad hoc reports, and this can be done via programs of the Oracle RDBMS family.

The secondary database stores regulatory and chemistry information (names, CAS numbers, chemical/physical properties and regulatory limits) on more than 2,500 commonly tested substances. When they set up a project database, users choose information for their project tables after se-

lecting the chemicals that will be tested. One set of menus lists parameters cited in federal regulations and guidelines, which range from the well-known (appendix 9 of *Primary and Secondary Drinking Water Parameters*) to the obscure (Department of Energy's *Derived Concentration Guides for Members of the Public*, as cited in DOE 5400).

Another set of menus groups chemicals according to any of the several hundred methods of analysis. A third set allows selections from common chemical groupings such as metals, volatiles, radiological or general chemistry parameters.

As parameters are selected, ITEMS copies their CAS numbers and names to the project-specific database. This eliminates the need to enter names of substances manually. Although compound names are printed in reports, CAS numbers identify re-

ITEMS has standard forms for entering and editing a variety of data, tables for storage, a secondary database, a report generator, and various application programs that link the central database to graphics and contouring packages.

sults in the electronic files, as they are generally easier to input, sort and query. Furthermore, where substances are uniquely identified by their CAS numbers, each can have many synonyms.

Maximum concentration limits cited in federal regulations and EPA publications can also be copied from the secondary to the project-specific database. The report generator can compare concentrations to these established limits and print which samples exceed the limits.

PRESENTING THE DATA

One of the biggest bottlenecks of environmental databases is between the central database and graphics. With ITEMS, all boring logs, well logs and well location maps can be generated automatically from the database via AutoCAD. Interfaces also exist to GRAPHER (for x - y graphs of changes over time) and SURFER or Quicksurf (for creat-

ing contour maps).

Users select data from menus within ITEMS, then export them to the graphics packages via translator programs without exiting one program and loading another. Instead, the application program is automatically loaded into the computer's memory. In some cases, the graphics display is generated without further input; in others, the user completes the generation with selections from menu options. Because users have access to all the features of each program, they may make additions and alterations at will.

THE STATISTICAL MODULE

A statistical module is divided into three parts: summary statistics, ground-water statistics and trend statistics. The summary program calculates sample means, standard deviations, medians, coefficients of kurtosis and skewness, interquartile ranges, quantiles and confidence intervals.

The user has a choice of 17 methods of handling nondetects, from simple methods such as using one-half the detection limit to more sophisticated methods based on either the normal and lognormal distribution or the maximum likelihood method. (Many of these methods are also available in the other two programs.)

The ground-water monitoring program closely follows recommendations given in EPA's 1989 *Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities*. Users select any of three ways of analyzing ground-water sampling data. One set of tests can be used to compare differences between upgradient and downgradient wells. A second set determines if concentration limits are exceeded, and the third relies on Shewhard-CUSUM control charts to track level changes of a constituent in a single well.

The trend program uses the Mann-Kendall test to determine upward or downward trends in a well or to evaluate trends in different seasons. The user can test for similarity in trends and create an x - y plot of concentration vs. time for each well and parameter. ▽

Neno Duplancic, M.ASCE, now principal designer with Canonie Environmental Services Corp., Pleasanton, Calif., was director of engineering during development of ITEMS at IT Corp., Martinez, Calif. Gregory Buckle, an IT staff consultant, designed the database structure and statistics module.