Making Environmental Data Accessible for Public Health Aims

The Massachusetts Environmental Database Project.

CHRISTOPHER A. PAULU
DAVID M. OZONOFF, MD
PATRICIA COOGAN, MPH
DANIEL WARTENBERG, PHD

The following scenario will be distressingly familiar to many State and local public health officials. A citizen calls from a small town to report four neighbors with a rare connective tissue disease. The citizen wishes to bring this to the attention of the health department and to inquire if there is a special reason for this seemingly unlikely event. One of the most important steps, of course, is to verify the diagnoses, but this can take time, and in the interim some basic environmental data about the community can be obtained immediately from existing information. Or can it?

Health officials routinely collect a wide array of geographically specified vital events and other health outcome data, and similarly, environmental protection officials routinely compile location specific environmental data. Yet while the environmental data exist and are theoretically available to public health officials and researchers, the number of databases and their scattered locations make them inaccessible in a practical sense (*1*). To be of any practical utility, these data must be accessible without cumbersome data processing requests to disparate agencies, and they should not require extensive training with complex computer software.

We developed the Massachusetts Environmental Database, a microcomputer-based data management system that accesses and integrates routinely collected environmental data. By "environmental data," we mean any data that describes or measures contaminants and constituents of the air, water, or land, or any potential source thereof. As might be expected, this is a project more complex in execution than in conception. We intend this as a practical guide for those engaged in similar endeavors, with a discussion of the potential uses of the resulting information system.

Characterizing Environmental Quality

An element in community diagnosis. Public health as a profession has traditionally been preoccupied with the problems of populations, in contrast to clinical medicine which has focused on the problems of the individual person. Despite this difference, there remain important similarities beyond the obvious one of commitment to a common science and knowledge base. In particular, the community stands in much the same relationship to the public health official as does the patient to physician. Recently, the practices of both professions have changed in ways that bring them even closer together. Specifically, clinical medicine has become more prevention-oriented while public health officers more and more frequently are called upon by citizens to engage in what might be called community diagnosis and treatment.

Unfortunately, one of the most important tools available to the physician, the medical history of a patient, is not available in an analogous form to the public health official. There is no readily available community history. Unlike a physician who almost always has at least some rudimentary social and clinical history available, our hypothetical public health official probably knows little about this small town with a perceived disease cluster, except perhaps its location. This official is no better off than a physician making important clinical decisions without the benefit of a medical history.

Sources of information about a community do not reside in the memory or records of an individual person but rather in institutional arrangements made for many different purposes. Using an environmental database, it is possible to "interrogate" a community and obtain a "history" that can be useful, if not in making a definitive diagnosis, at least in narrowing

Some Agencies and Organizations Interviewed to Obtain Data for the Massachusetts Environmental Database

- United States Environmental Protection Agency
- United States Geological Survey
- United States Department of Fish & Wildlife
- Bureau of the Census
- Massachusetts Department of Environmental Quality Engineering
- Massachusetts Department of Fisheries & Wildlife
- Massachusetts Department of Public Health
- Massachusetts Special Legislative Commission on Water Supply
- Massachusetts Department of Environmental Management
- Massachusetts Department of Food & Agriculture
- Massachusetts State Data Center
- New England Interstate Water Pollution Control Commission
- Tufts University Center for Environmental Management
- Massachusetts Public Interest Research Group

down the possibilities and perhaps suggesting new avenues of inquiry. As a desktop system, our data base often can supply information immediately, potentially even during the course of a citizen inquiry. Furthermore, from a broader standpoint, we may be concerned with information that describes not just a single community, but many communities. With consistent environmental data that describes the whole State, we can begin to explore geographic patterns, that is, to map the regional environmental history.

Environmental data in environmental epidemiology. The impetus for constructing the Environmental Database was the observation that both routine environmental and health data are often mappable. Perhaps the simplest of all ideas in environmental epidemiology is to make two maps, one depicting exposure, the other depicting outcome, and superimposing them. However, the simple process of superimposing maps, as is commonly practiced with Geographic Information Systems (GIS), is inadequate for research purposes. Spatial information about outcomes is not easy to summarize, and once summarized not easy to analyze and interpret. If the spatial data is a case series located as a set of points, for example, what data should one use as a reference population (the denominator)? If one has chosen a reference data set (for example, the census population), what boundaries should be used to aggregate the data in order to calculate rates? Moreover, the relationship between the exposure and the outcome is rarely as simple as the idea of superimposed maps suggests—the maps themselves are just snapshots of dynamic processes; spatial coincidence (or lack of coincidence) does not necessarily imply (or contradict) effective exposure; and the accuracy of the maps is often in question.

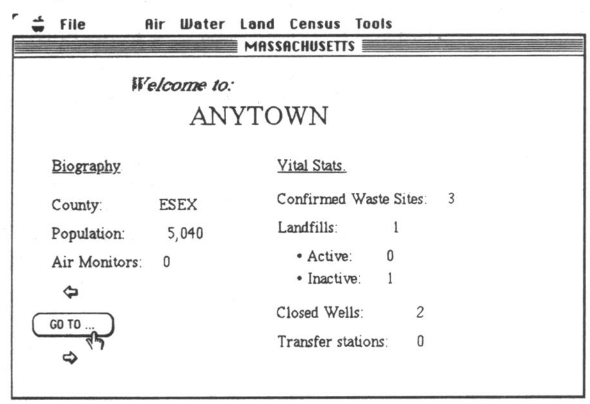
On the other hand, routinely collected environmental monitoring data represent an under utilized resource for epidemiologic study (*2,3*). One reason given for not using environmental monitoring data is that these data are of such poor quality or idiosyncratic structure as to preclude meaningful study and interpretation. However, one could direct the same criticisms against many health outcome data sets, such as death certificate data, yet these data have yielded much valuable information. Indeed epidemiologic studies have frequently used routinely collected environmental data. For example, a series of studies have compared drinking or ground water quality to the incidence of cancer (*4, 5*), Alzheimer's disease (*6*), and birth defects (*4, 7-9*).

Similarly, a series of studies link air quality with mortality (*10-12*), respiratory effects (*13-16*), and birth defects (*18*). Data on pesticide use have been linked with birth defects (*17,18*), and proximity to toxic waste sites has been linked with mortality and cancer incidence (*19,20*). Routinely collected data have thus been a source of useful information in epidemiologic studies.

In the case of Massachusetts, environmental data could be used in a variety of ways (*2*). For example, averaged data for each town can be used to provide exposure estimates for traditional ecological studies. Despite all the caveats and limitations of ecological study design (*21-23*), such assessments are often useful for generating hypotheses and evaluating existing ones. Griffith and coworkers (*20*) have looked at cancer rates in United States counties with hazardous waste sites on the National Priorities List. Similar studies can be done for various health outcomes in the 351 cities and towns of Massachusetts. Thus site specific cancer incidence rates for all towns with water contaminated above a certain level with volatile organics could be compared to towns without contamination. Although such studies are not definitive, they can make rapid evaluations and provide useful insights.

Another epidemiologic use of the Environmental Database is to provide information about exposure for town residents in an individual-level study. When constructing a questionnaire around a waste site, for example, it is helpful to know about the principal industries in the area and the existence and location of environmental hazards apart from the one under study.

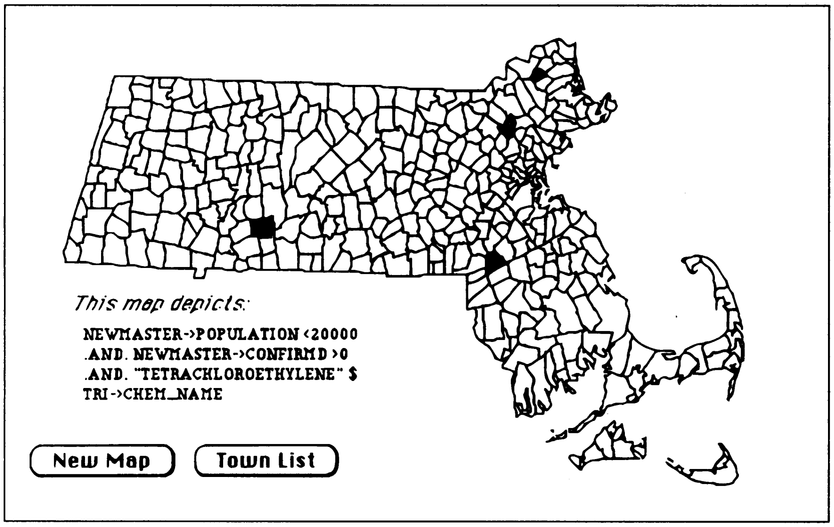
Figure 1. Sample screens reveal general environmental characteristics of hypothetical "Anytown"



Tri

Town	Chem_name	eport_	Fugitive	Stack_	Other_off
ANYTOWN	TETRACHLOROETHYLENE	87	270	2700	26000
ANYTOWN	1,1,1-TRICHLOROETHANE	87	670	6700	25000
ANYTOWN	TETRACHLOROETHYLENE	88	290	2893	750
ANYTOWN	1,1,1-TRICHLOROETHANE	88	1608	1536	5056
ANYTOWN	GLYCOL ETHERS	89	100	420	0
ANYTOWN	1,1,1-TRICHLOROETHANE	89	8900	4100	3600
ANYTOWN	FREON 113	89	330	1300	0
ANYTOWN	TETRACHLOROETHYLENE	89	380	800	450
ANYTOWN	XYLENE (MIXED ISOMERS	89	1500	600	240
ANYTOWN	PHOSPHORIC ACID	89	40	160	0

Figure 2. A map of four towns with similar characteristics to hypothetical "Anytown



A desktop environmental database can provide features beyond that of an eletronic filing cabinet, such as the ability to query data, summarize it, and display it. As a public health official getting familiar with the characteristics of one town, you might want to know "what other towns are similar?" The screen above is a simple map generated by a desktop environmental database, in answer to this sort of question. In this example, you have selected criteria from three different data sets—the census, the list of confirmed hazardous waste sites, and the Toxic Release Inventory—that seemed to distinquish the town in question. Indeed, the map reveals that there are only three other towns in the state with similar characteristics.

Another use would be to study the relationship among birth weight, demographic characteristics, and ambient air quality. One could reasonably assume, as a first approximation, that residents of the same town have similar exposures that differ from residents of other towns. (Variations and refinements are possible, including bringing in information about neighboring towns.) Where ambient monitoring data are available, one could use the data as an exposure variable for individual residents in a traditional regression analysis. In a more detailed analysis, one might model air exposures using facility-specific emissions data from the database as input data to an atmospheric dispersion model.

Clearly, these do not exhaust the possibilities for using these data for epidemiologic studies and are given for the sake of example. The key point is that *accessibility* is a prerequisite for the data's use.

Environmental data on your desktop: a hypothetical example. Returning to the hypothetical example that motivated our discussion, imagine yourself as the health official who has just promised to return a call to the concerned citizen within a half hour with more information. Let us also assume that the disease in question is known to be associated with exposures to volatile organics. At this stage, however, you will be thinking more in terms of ascertaining potentially important environmental factors in the community than undertaking environmental epidemiology. You now turn to the Environmental Database running on your desktop computer (fig. 1).

An inquiry in the master file (fig. 1) reveals the town has a population of slightly more than 5,000, has three confirmed hazardous waste sites, and has had its public water supply well closed at some time in the past (due to contamination from fuel storage tanks). There is also an inactive land-fill. Although there are no air monitoring stations in this small town, the master file also indicates (not shown) that the statewide emissions inventory lists an air source of volatile organics that emitted 73 tons per year for the last five years. The fact that these data sets came from four different agencies—the Bureau of the Census, the U.S. Environmental Protection Agency, the Massachusetts Department of Environmental Quality Engineering, and the Massachusetts Department of Public Health—is completely transparent (as it ought to be).

If this were all that the system could do, it would be useful, but the fact that the actual primary data, which is only summarized in the master file, is also accessible makes it more so. Routine inorganic analyses of the town water supply show three samples from the tap in the town hall, dating from 1978, 1979, and 1985. The earlier samples contained arsenic, barium, and mercury; the later sample contained lead. No data were available from the State's survey of purgeable organics in water supplies carried out in 1981 and 1984, the reason being that in 1979 the two public wells were closed due to 118 parts per billion of trichloroethylene contamination in one and 11 parts in the other, according to the inventory of closed wells.

The confirmed hazardous waste site inventory lists two petroleum sites and a third non-petroleum site that is on the National Priorities List. The agricultural file, based on a State aerial photography survey, revealed only small amounts of farmland (355 acres), most given over to hay and pasture. However, herbs, orchards, and blueberries were also listed along with the more usual crops in the remainder. The census file reveals that 80 percent of the population is classified as urban rather than rural, almost entirely white, and almost 75 percent are younger than age 44.

Opening the air-related files immediately shows the name, location, and output of the principal emitting facility in the town. In this case we have data from the National Emissions Data System (NEDS), the Volatile Organics Compound Inventory (VOCinfo), and the Toxic Release Inventory (TRI). The diversity of data collection allows us to synthesize different perspectives on the same facility. The NEDS file informs us that the facility changed SIC codes (and hence production) between 1979 and 1981, and that its emissions were below reportable limits at that time.

The VOCinfo file indicates steady VOC emissions of 37 tons per year (TPY) between 1981 and 1985, dropping to five in 1986 and 12 TPY in 1987. The TRI database provides specific data on the quantity and quality of more recent emissions, indicating, for example, a decrease in combined fugitive and stack emissions of tetrachloroethylene, and an increase in emissions of 1,1,1-trichloroethane, between 1987 and 1989. Curiously, the VOC file lists modeled VOC emissions of 12 TPY (or 24,000 pounds) for 1987, while the TRI reports only 10,040 pounds of total fugitive and stack VOC emissions. Some of this information may need to be corrected after further inquiry at the facilities (for example, because they represent reporting artifacts or misunderstandings), but they provide an unusually detailed starting point.

In addition to the routine interrogations illustrated so far, relational databases allow complex queries and reports, both within and across databases. For example, we can pose a query to find all towns with similar characteristics to our hypothetical Anytown—for example, having at

Summary of Data Collection for Massachusetts Environmental Database

Environmental media	*Data type*	*Geography*	*Substances*
Air			
National Emissions Data System	Modelled	Statewide	Criteria pollutants
SAROAD[1]	Measurement	Statewide	Criteria pollutants
Volatile Organic Compounds (VOC) inventory	Emissions	Statewide	VOCs combined
Toxic Release Inventory (TRI)[2]	Emissions	Statewide	Diverse
Water			
List of water supply wells closed due to contamination	Inventory and measurement	Statewide	Varies—mostly organic chemicals
Interagency pesticide monitoring program	Measurement	27 towns in Connecticut River Valley	Eight agricultural chemicals
State Purgeable Organic Testing Program	Measurement	Statewide	VOCs
Inorganics monitoring	Measurement	Statewide	10 inorganics
Hazardous waste			
Active landfills	Inventory	Statewide	N/A
Inactive landfills	Inventory	Statewide	N/A
Transfer stations	Inventory	Statewide	N/A
Confirmed hazardous wastes sites	Inventory	Statewide	N/A
Manifested hazardous waste sites	Inventory	Statewide	N/A
Other			
Census data	Quantified	Statewide	(Population data)
Agricultural land use	Quantified (acres)	Statewide	(Crops)

[1]Storage and retrieval of aerometric data.

[2]Toxic Release Inventory includes categories for air (stack and fugitive), land, water (treatment works and wells), and off-site transfers.)

N/A = Not available.

least one confirmed waste site, releases of tetrachloroethylene reported in TRI, and a population of less than 20,000. Only four towns satisfy these characteristics, and we find that they are scattered about the State (fig. 2).

Although there is nothing diagnostic about any of this information, it is striking how much better a picture of the community is available after less than 10 minutes of interrogation of the routinely collected data. In response to the citizen that motivated the inquiry, you will be able to ask about the presence of nearby facilities, chemical odors, and so forth—and do so in a way that applies the specific knowledge about local characteristics obtained from the Environmental Database. Furthermore, you now have enough information to form the basis of a more thorough followup. You could contact the principal emitting facility and inquire about the change in production operations a decade ago, as well as the discrepancy between modelled and self-reported emissions of volatile organics. You could also query relevant officials about the type of contamination and status of clean up operations at the three hazardous waste sites.

This example, while hypothetical, demonstrates how useful routinely collected environmental data can be at a very fundamental level—the level of communication between a concerned citizen and a public health official. Not only did the Environmental Database give the official a better information base with which to respond to the citizen in a timely manner, the information suggested further avenues of inquiry that might otherwise have been overlooked.

Easily used and accessible hardware and software now allow the integration of a variety of routine environmental data into everyday public health practice and research endeavors. In this paper, we provide some practical advice on building an integrated environmental data system for a State or locality, describe such a system, and suggest some of the uses for it. We believe that similar systems will provide new opportunities to take account of environmental data for public health purposes.

Mr. Paulu, Dr. Ozonoff, and Ms. Coogan are with the Department of Environmental Health, Boston University School of Public Health. Dr. Wartenberg is with the Department of Environmental and Community Medicine, University of Medicine and Dentistry of New Jersey-Robert Wood Johnson Medical School (UMDNJ).

Tearsheet requests to Dr. Wartenberg, Department of Environmental and Community Medicine, UMDNJ-Robert Wood Johnson Medical School, 75 Hoes Lane, Piscataway, NJ 08854-5635; tel. 908-445-0197; fax 908-445-0784; e-mail <<dew@cohsi.rutgers.edu>>

References

1. Sexton, K., et al.: Estimating human exposures to environmental pollutants: availability and utility of existing data. Arch Environ Health 47: 398–407 (1992).
2. Frisch, J. D., Shaw, G. M., and Harris, J. A.: Epidemiologic research using existing databases of environmental measures. Arch Environ Health 45: 303–307 (1990).
3. Matanoski, G., et al.: Role of exposure databases in epidemiology. Arch Environ Health 47: 439–446 (1992).
4. Lagakos, S. W., Wesson, B. J., and Zelen, M.: An analysis of contaminated well water and health effects in Woburn Massachusetts. J Am Stat Soc 81: 583–596 (1986).
5. Fagliano, J., et al.: Drinking water contamination and the incidence of leukemia: an ecologic study. Am J Public Health 80: 1209–1212 (1990).
6. Martyn, C. N., et al.: Geographical relation between Alzheimer's Disease and aluminum in drinking water. Lancet 8629: 59–62 (1989).
7. Zierler, S., et al.: Chemical quality of maternal drinking water and congenital heart disease. Int J Epidemiol 17: 589–594 (1988).
8. Aschengrau, A., Zierler, S., and Cohen, A.: Quality of community drinking water and the occurrence of late adverse pregnancy outcomes. Arch Environ Health 44: 105–113 (1993).
9. Kramer, M. D., et al.: The association of waterborne chloroform with intrauterine growth retardation. Epidemiology 3: 407–413 (1992).
10. Lave, L. B., and Seskin, E. P.: Air pollution and human health. Science 947: 723–733 (1970).
11. Hatzakis, A., et al.: Short-term effects of air pollution on mortality in Athens. Int J Epidemiol 15: 73–81 (1986).
12. Buffler, P. A., et al.: Air pollution and lung cancer mortality in Harris County, Texas, 1979–1981. Amer J Epidemiol 128: 683–699 (1988).
13. Sterling, T. D., et al.: Urban morbidity and air pollution. Arch Environ Health 13: 158–170 (1966).
14. Levy, D., Gent, M., and Newhouse, M. T.: Relationship between acute respiratory illness and air pollution levels in an industrial city. Amer Rev Respir Dis 116: 167–173 (1977).
15. Whittemore, A. S., and Korn, E. L.: Asthma and air pollution in the Los Angeles area. Am J Public Health: 687–696 (1980).
16. Holberg, C. J., O'Rourke, M. K., and Lebowitz, M. D.: Multivariate analysis of ambient environmental factors and respiratory effects. Int J Epidemiol 16: 399–410 (1987).
17. Schwartz, D. A., and LoGerfo, J. P.: Congenital limb reduction defects in the agricultural setting. Am J Public Health 78: 654–659 (1988).
18. Grether, J. K., et al.: Exposure to aerial Malathion application and the occurrence of congenital anomalies and low birthweight. Am J Public Health 77: 1009–1010 (1987).
19. Najem, G. R., and Cappadona, J. L.: Health effects of hazardous chemical waste disposal sites in New Jersey and in the United States: a review. Am J Prev Med 7: 352–362 (1991).
20. Griffith, J., et al.: Cancer mortality in the U.S. counties with hazardous waste sites and ground water pollution. Arch Environ Health 44: 69–74 (1989).
21. Greenland, S., and Morgenstern, H.: Neither within-region nor cross-regional independence of exposure and covariates prevents ecological bias. Int J Epidemiol 20: 816–818 (1991).
22. Susser, M.: The logic in ecological: I. The logic of analysis]. Am J Public Health 84: 825–829 (1994).
23. Susser, M.: The logic in ecological: II. The logic of design. Am J Public Health 84: 830–5 (1994).
24. Paulu, C., et al.: Massachusetts Environmental Database: User's Manual. School of Public Health, Boston University, 1992, pp. 1–33.

Developing Your Own Environmental Database

Routinely collected environmental data are considerably more various and fragmented than vital events data, owing to the disparate agencies that collect the data, the variety of environmental media covered, and the diverse reasons for which the data are collected. Many Federal, State and even local agencies collect environmental data, either on a routine basis or in response to a specific need or incident. We set down here some of the lessons we learned in creating our system.

Lesson One

Public access does not equal easy access. Although we knew there was no existing inventory for all of these databases, it came as a surprise that individual agencies themselves usually did not appreciate the breadth and extent of the information they collected. We therefore contacted the major subdivisions of each agency and asked to speak with the person in charge of data. We also contacted major user groups, such as environmental organizations, planning departments, and university-based researchers. With each contact we asked not only about data sets the person worked on or used, but other data sets known to them. We identified sources of environmental data in 14 Federal, State and private agencies (first box), comprising 34 different environmental databases.

It was usually possible to obtain preliminary documentation of the data sets from these initial contacts, but acquiring the data itself was more difficult. In the typical case, this involved at least three people, the program person whose job it is to collect and maintain the data and who is familiar with the data set, its quality and its problems; a programmer, whose job it is to get data into the system and to get it out again (this is the person who supplies the actual data set on paper, disc, or tape); and the supervisor or manager whose division collects the data. Acquisition of a data set involved meeting with all of these people and took one or more separate meetings with the supervisors in the relevant department to enlist their support. We have found that personnel will not take time to fulfill a data request without the approval of a supervisor, and that approval requires a certain amount of perseverance. Sometimes more than one level of management had to be traversed for a single data set.

The lesson: sufficient time—often on the scale of months—must be set aside for what appears to be a trivial task, acquiring an identified public information data set.

Lesson Two

Don't start with the goal of comprehensiveness. We know that others have tried to construct an integrated environmental database and failed. We believe that most of these previous attempts foundered by trying to include all available databases in their environmental database management systems. Indeed, the effort would be overwhelming, given the complexity of identifying and obtaining all data sets. We therefore limited the scope of the Massachusetts Environmental Database to include only data with a favorable assessment in relation to these five criteria:

1. Quality and completeness,
2. Geographic coverage,
3. Time coverage,
4. Whether substances signaled potential health effects, and
5. Human exposure potential.

It quickly became apparent that few data sets were completely satisfactory with respect to all or even most of these considerations. Thus we included many data sets that were potentially useful despite some restrictions of geographic or temporal coverage or other limitations. Ease of access, machine-readable form and relatively uncomplicated structure emerged as important secondary criteria. Ultimately, we incorporated 15 environmental and demographic data sets into the Massachusetts Environmental Database (second box). These included routinely collected data, data collected for special projects and site assessments, compliance data, ambient and emissions data, and facility inventory information.

The lesson: forego comprehensiveness in exchange for expediency, at least in the initial stages of the project.

Lesson Three

Build in flexibility. Collecting data sets is a time-consuming process. We acquired the smaller, simpler, machine-readable data sets first, which allowed us to erect the framework for the system, debug it with actual data sets, and avoid becoming mired in the attempt to acquire all possibly relevant data before building the system. For this strategy to work, the system itself must be capable of easy expansion.

To accommodate these needs, our system maintains data in a common database file format (DBF), allowing straightforward updates and additions of new data. An additional benefit of this approach is that we were able to maintain much of the data in exactly the same form that it was recorded, and therefore retained data source integrity.

In addition, we used an applications generator (FoxBase+/Mac™) that has a moderately simple but powerful programming language. This software (now available in a cross-platform package called FoxPro™) offers a variety of interface and data management tools that make programming much more efficient than constructing a system *de novo* using a programming language such as C. Use of an applications generator also eases a transition should personnel change in the course of the project. Our system's program code can be learned or read by a large community of programmers and, if necessary, modified easily.

The lesson: use reliable, widely used software technology, and construct a prototype system with a flexible database architecture.

Case-in-Point

The Massachusetts Environmental Database. For the purposes of illustration, we will provide some details about the features we incorporated into the Massachusetts Environmental Database. Our intent is to identify the tools that we found most useful and show what can be accomplished within the context of conventional relational database software. All operations in the Massachusetts Environmental Database are accomplished behind a user-courteous interface compatible with any MacIntosh OS-equipped computer. (For a look at how the program appears to the user, see figures in the accompanying article.) Implementation on a DOS platform is equally possible, since similar software is available for either operating system and the program logic is the same.

The Massachusetts Environmental Database has four major functions—Browse, Query, Calculate, and Report.

The Browse feature enables the investigator to view each database in its entirety, line by line. This is essentially the same as reviewing a paper version of the data. The Query feature enables the investigator to pose a question about the data and view the results instantaneously. For example, the user could ask which towns had measurable trichloroethylene (TCE) in drinking water samples. The queries are entered by using a mouse to select field names and logical or other operators (available on pull down menus). The result of the query is to have one or more databases filtered to display data only for those towns that satisfy the query. Queries can also be compound, relating information from multiple databases. For instance, the investigator could ask "which towns have greater than 10,000 residents and TCE in their drinking water and at least one confirmed hazardous waste site?" Each of the three databases can be browsed immediately for their role in the query (for example, the census database will now be restricted to towns with a total population greater than 10,000). The user may view the full query as a list or as a map.

The Calculate feature allows the investigator to compute simple statistics on any available numeric data: for example, the average concentration of TCE in sampled water supplies. The result can be computed for each town or across all towns, and further modified by a query specification, (such as samples taken in 1988 only). A report of selected variables (from any database, query, or calculation) can be output to screen, printer, or computer file. This last feature is particularly useful for creating input files to statistical packages or other computer platforms to produce health or exposure assessments or to create presentation quality maps and charts.

Detailed instructions for all four features are described in a user's manual (*24*) available from the authors. The features we have outlined represent the core functionality we would expect of a desktop environmental database, which is just a fraction of what is possible with current software.

Getting the Job Done

Relational database or GIS? What organizational structure? A question we hear often is "what sort of software should I use?" A conspicuous alternative to the relational database approach that we have taken is that of a GIS. Our choice of a relational database system reflects our own particular needs, in terms of system functionality, the format of our data sets, and software availability, all at the time the Massachusetts Environmental Database was completed (circa 1990). Our decision to use a relational database is by no means a general endorsement of this approach over the GIS option. The choice between these two approaches will depend on how geographically-oriented the source data is, and how important it is to have interactive maps or to produce maps as output. Of course, these two approaches, as we have identified them, are not mutually exclusive. A GIS may serve as the front-end to a database server, and a database front-end may gain spatial display and analysis functionality by communicating with a GIS.

Perhaps an even more basic question that we encounter is "how should the data sets be organized?" The answer to this will depend on the data one has to work with and even the region of interest. In Massachusetts, the entire land mass is incorporated into one or another of 351 towns. While "town" is a fairly large unit of analysis, it was an attribute we found contained in most environmental data in Massachusetts. Demographic data likewise could be obtained at the town level. The town identifier thus became an expedient way to relate different data sets. An added benefit was heuristic: most users were immediately comfortable with data organized by town. Outside of Massachusetts, environmental data sets may dictate interrelations at a larger level of geography (county, for example), a smaller level, such census tract or zip code, or not based on geo-referenced entities at all: perhaps by type of contaminant, media of contamination, or by relevance to community concerns. In the end, choosing an organizational structure for an environmental database is a matter of mediating the investigator's wishes and the realities of the data sets at hand.

The reader's own needs will determine the specifics of organizational structure, interface, software, and hardware to use. The bottom line, as we have indicated already, is getting the job done, which frequently necessitates a flexible approach. We hope this has been an enlightening—and encouraging—excursion into our own experience.