AN INTERACTIVE MAPPING AND ANALYSIS SYSTEM FOR USE WITH ENVIRONMENTAL AND HEALTH DATA

Laurance J. Milask
and
Carol G. Graves
Information and Scientific Applications Division
Sigma Data Computing Corp.
2021 K Street, N.W.
Washington, D.C. 20006

I. Introduction and Background

The President's Council on Environmental Quality (CEQ) was established under the National Environmental Policy Act (NEPA) to provide national environmental oversight, policy, and decision support analysis for the Executive Office of the President. This work includes the monitoring and general administration of NEPA's guidelines and regulations, the environmental impact statement process and a yearly and comprehensive assessment of environmental and related conditions and trends. This last activity involves the production of a series of issue documents and a yearly report titled the "President's Report on Environmental Quality" which reviews the state of the environment, and relationships between environmental quality and energy, land use, conservation, national health and other factors.

CEQ's early efforts to meet its decision support and general assessment needs clearly demonstrated the problems involving the integration of diverse large scale data bases. Moreover, it became evident that new analysis and graphic display tools were needed which would permit rapid access to established data bases by policy analysts and other end users.

Early in 1975, CEQ began the development of the UPGRADE (User Prompted Graphic Data Evaluation) system, and the establishment of several high quality environmental, health, and demographic data bases. The objectives of UPGRADE are:

* Provide a highly user oriented computer based tool for the analysis and display of environmental conditions, trends and relationships.

* Establish and refine decision support analysis methodologies which utilize computer based tools and data bases.

* Define and integrate data bases from multiple disciplines which will support both immediate issues and long term trends and conditions analysis requirements.

CEQ, under the NEPA enabling legislation, is also mandated to promote the development and dissemination of tools and resources necessary to the study and control of environmental quality. As such, early in the development of the UPGRADE facility other key federal and state agencies were contacted and associated with the project. These agenices, which include the Environmental Protection Agency (Office of Research and Development and Office of Toxic Substances), Department of Energy, U.S. Geological Survey, National Cancer Institute and the New Jersey Department of Environmental Protection have contributed substantial sums of funding, technical support, data bases, computing facilities and enthusiastic encouragement.

II. Analysis Objectives and UPGRADE Data Bases

The basic analysis requirements underlying the UPGRADE facility (analysis tools and data bases) are:

* Periodic national and regional assessment of environmental, health and related conditions and trends.

* Rapid response capability to support special issues and decision making analyses.

Trends and conditions analyses

CEQ's national and regional assessment of environmental, health and related conditions are usually performed in support of the Annual Report or other large scale reporting activities. These studies tend to be repeated from year to year and may take several months of planning, data base enhancement and analysis to complete. The analysts are usually well trained in the use of the UPGRADE analysis system and are thoroughly familiar with the data bases in use. Finally, a large variety of computer based tools are used to manipulate, condition, analyze and display the data. As will be briefly described in the final section of this paper, UPGRADE's analytical system provides a totally integrated and uniform access to all of these capabilities.

The data base requirements necessary to meet trends and conditions analysis are:

* National in scope, with emphasis on high quality and stable sources which may be conveniently updated on a yearly basis.

* Both micro data (such as site specific monitoring data) and macro level data (such as statistically aggregated census, land use, economics and health data) may be used.

Micro level data must often be aggregated so as to represent either geopolitical or natural geographic areas to be suitable for spacial analysis and display. This involves a series of special data conditioning and quality assurance activities which are heavily dependent upon UPGRADE's analytical support staff and data base management capabilities. For instance, UPGRADE's Mapping Procedure permits the display of water quality data according to USGS established water basin boundaries. Micro level time series data from the USGS NASQAN data base is employed via the UPGRADE data base extraction system (EXTRACT) which permits the automatic aggregation of time series and multiple site observations for each water basin area. Year-to-year variations in physical monitoring site locations, NASQAN system location codes, data quality errors and similar problems must often be worked out manually by the

analytical support staff before the data base is ready for use. In fact, while the effort expended to develop UPGRADE's analytical and data handling tools has been substantial, it has represented only 10% - 20% of the total effort needed to collect, quality assure and integrate the available data bases.

The data bases currently available to support conditions and trends studies are given in Table 1. They include special subsets or complete copies of nationally available air and water quality, mortality and demographic data. These macro level environmental health and demographic data are integrated on the FIPS County Code level. This data base, titled CLIDE, is under continual expansion. Such new data as soil types, land use, wildlife, new mortality, morbidity, drinking water, emissions and industrial toxics chemicals production data among others are currently planned for inclusion. Moreover, other geopolitical aggregations as census tract, SMSA, and the like will be accommodated as need and appropriate data arise.

Special Issues and Decision Support Analyses.

Special issues and decision support analysis entail short response times and macro level data to accomplish. Often senior staff who are not familiar with UPGRADE or the available data bases are involved, and special data not available within the established data bases must be obtained or simulated. The analysis tools most often used are summary tables, line and bar charts and regional or national shaded maps. The conflict between the use of macro level data and summary graphic outputs on the one hand, and the need for high confidence in results to support decision making places an extremely high burden upon the accuracy of data and analysis tools and the quality of the analytical support team.

The analysis process must draw upon the same data bases established for the conditions and trends studies. In addition, special data bases must be quickly obtained and integrated. For instance, assume that an issue arises concerning the impact of steel production on water quality and possible health hazards on the Monongahela River in Pennsylvania. An examination of CEQ's available data might reveal the need to obtain additional water quality data from STORET, and the need to

pinpoint drinking water sources for the area of interest. The analysis support team must be able to obtain this data, integrate it into the data base facility and, when necessary, condition or convert data to simulate drinking water quality. Once this is done, the decision maker must be able to quickly gain contact or direct access with the data to produce the final summary graphs and maps. The burden at this point resides upon the UPGRADE analytical capabilities.

III. UPGRADE Analytical and Display System

The UPGRADE analytical system is characterized by the following features and capabilities:

* Interactive operation employing straightforward prompts and simple user responses. The user interface or prompting procedure is combined with a "HELP" feature which enables users to obtain immediate information about a prompt or analysis procedure.

* An extensive collection of analytical procedures ranging from simple scatter and bar charts and basic summary statistical tables up to a completely interactive procedural interface to the SAS statistical system (SAS Institute, Raleigh, North Carolina), as well as national and regional shaded mapping procedures.

The above combination of an extremely simplified user interface to the system and the availability of either summary statistical analysis and basic graphic display, or advanced statistical analysis and data manipulation routines make UPGRADE suitable for the full range of analysis requirements needed by CEQ. CEQ's experience has shown that new users of UPGRADE require one or two days of hands-on experience to begin effective use of the system. Figure 1 shows a typical example of an UPGRADE prompting sequence which demonstrates the straightforward step-by-step operation of the system.

IV. Future Directions for UPGRADE

CEQ and the supporting agencies are continuing the development of the UPGRADE system in several ways.

* The UPGRADE mapping procedures will be enhanced to include spot mapping, spatial analysis overlaying procedures and color single and bivariant thematic maps. Since UPGRADE is designed to interface to other established computer code, it is quite likely that packages like the Harvard Odessy system will be obtained to meet many of the advanced mapping needs.

* A new version of the UPGRADE system is currently under design for implementation on a DEC VAX 11/780 minicomputer. The current version runs only on IBM 370 compatible mainframes under the TSO time sharing system. CEQ will continue to maintain the IBM version.

* Specialized procedures and models will be developed or obtained to support various analysis requirements. One special procedure under development is designed to estimate air quality levels in urban areas and to assess possible health risk for various air pollution concentrations.

* A commercial installation for the UPGRADE system and associated support services is now being established on a nationally accessible time-sharing center.

TABLE 1 MAJOR DATABASES

NAME	SOURCE	DESCRIPTION	ACCESS TO UPGRADE
STORET	EPA	Water quality data 69 million observations 200 thousand sites monitoring done by many agencies 2000 parameters	WATER QUALITY INTERFACE
SAROAD	EPA	Air quality data 250 million observations 12 thousand stations	AIR QUALITY INTERFACE
NASQAN	USGS	Water quality data 500 stations-one for each hydrologic accounting unit	NASQAN EXTRACT and INTERFACE
CLIDE	NCHS	Mortality rates for 1968-78 age-adjusted, race/sex specific	INTEGRATED DATABASE EXTRACT and INTERFACE
	EPA	Water quality data	
	Census	Demographic, socio-economic and migration variables	
	VA	Veterans data	

```
                    *** REGRESSION PLOTTING ***


ENTER Y-AXIS VARIABLE KEYCODE(S)
? 30 35

ENTER X-AXIS VARIABLE KEYCODE(S)
? 11 11

VERIFY?
? yes

YOU HAVE REQUESTED THE FOLLOWING 2 PLOT(S)

      Y-AXIS                        X-AXIS
ENTRY  KEY  VARIABLE           KEY  VARIABLE
  1     30  DISS: OXYGEN       11  WATER TEMP
  2     35  FECAL COLIFORM     11  WATER TEMP

AUTO SEQUENCE?
? yes

PLOT OPTIONS?
? no
```

Figure 1. The prompting sequence for setting up two regression plots in UPGRADE. Two graphs will be produced on the terminal screen and automatically copied. Each graph will have all points plotted and the regression line imposed. Regression statistics will also be given.