KNOWLEDGE-BASED DECISION SUPPORT FOR ENVIRONMENTAL ASSESSMENT

Keith Reynolds, Scott Murray, Michael Saunders, John Slade, and Bruce Miller[1]

ABSTRACT.—EMDS is a knowledge-based decision support system for the design and conduct of ecological assessments at any geographic scale. ArcView provides the primary system interface to the EMDS system. The NetWeaver development system incorporated into EMDS provides a knowledge base development environment in which an assessment team assembles a knowledge base that describes the logical relations (dependencies) among topics of interest to an assessment, underlying ecosystem states and processes, and data needed to evaluate topics, states, and processes. The Assessment system includes two subsystems: 1) the Analysis subsystem provides an interface to the knowledge base performing analyses. Knowledge-based reasoning as implemented in the NetWeaver knowledge base engine provides powerful analytical capabilities for environmental assessment; 2) the Data Acquisition Manager uses information about the influence of missing data, and information gathered from the user about the ease of acquiring missing data to prioritize missing data as an aid to planning new data collection to improve an assessment.

INTRODUCTION

The objective of the Ecosystem Management Decision Support (EMDS) project is to improve the quality of environmental assessments and the efficiency with which they are performed. EMDS version 1.0 is an application framework for knowledge-based decision support of environmental assessments that integrates state-of-the-art geographic information system (GIS) and knowledge-based reasoning technologies in the Microsoft Windows® environment to conduct analyses at any geographic scale.[2] We describe the EMDS system as an application framework, in part at least, because it does not come "ready to run out of the box." Instead, the system provides a very general solution method (e.g., a framework) for conducting environmental assessments.

Perhaps sometime in the future, environmental assessment teams will be able to assemble a list of all topics they want to include in an assessment, assemble a list of data requirements needed to address those topics, and find they have all the required data. In the meantime, however, assessments routinely need to deal with incomplete data. There may be some missing observations for several to many data types. There may be no data at all for (possibly many) others.

One solution to the problem of missing data is to tailor the assessment to existing data. However, this is a poor way to conduct an assessment of environmental phenomena with complex inter-relations because it is extremely difficult, if not impossible, in such an approach to infer what is not known about the data that might be relevant to the assessment and how influential omitted data might be for determining observed ecosystem states and processes.

[1]Keith Reynolds, USDA Forest Service, Pacific Northwest Research Station, Corvallis, OR 97331. Phone: 541-750-7434. FAX: 541-750-7329. Email: reynoldsk@fsl.orst.edu. Scott Murray, Environmental Systems Research Institute, 380 New York Street, Redlands, CA 92373. Phone: 909-793-2853. Email: smurray@esri.com. Michael Saunders, Entomology Dept., 501 ASI Bldg, Penn State University, University Park, PA 16802. Phone: 814-863-2979. Email: michael_saunders@agcs.cas.psu.edu. John Slade, Knowledge Garden, Inc., 5116 Hartwick Lane, West Palm Beach, FL 33415. Phone: 561-615-8209. Email: jslade@kgarden.com. Bruce Miller, Rules of Thumb, Inc., P.O. Box 406, North East, PA 16428. Phone: 814-725-2312. Email: bjmiller@psupen.psu.edu

[2]The use of trade or firm names in this publication is for reader information and does not imply endorsement by the U.S. Department of Agriculture of any product or service.

Instead, a basic premise underlying the development of the EMDS system is that it is preferable to construct a logical model that reflects our best understanding of how all states and processes relevant to the assessment are inter-related, regardless of whether relevant data are currently available.

The EMDS inference engine uses a simple algorithm to compute the influence of missing data based on how many states and processes use the information, and at what level the information enters a knowledge base structure. The EMDS system's Data Acquisition Manager component uses information about data influence to assist users with prioritizing new data acquisition needs.

KNOWLEDGE BASE CONCEPTS

Knowledge-based reasoning is a form of knowledge representation in which phenomena are described in terms of abstract entities and their logical relations to one another (Schmoldt and Rauscher 1995). Two basic reasons for using knowledge-based reasoning are:

1. The entities or relations involved in the problem to be solved are inherently abstract so that mathematical models of the problem are difficult or impossible to formulate.
2. A mathematical solution is possible in principle, but current knowledge is too imprecise to formulate an accurate mathematical model.

A knowledge base embodies knowledge about how to solve a problem in some domain of interest. Several different formalisms for knowledge-based representation have been developed over about the last 20 years (Schmoldt and Rauscher 1995). EMDS incorporates the NetWeaver™ knowledge base engine for knowledge-based reasoning (Saunders *et al.* 1989, 1990). This system implements a knowledge base as networks of logical dependencies (Stone *et al.* 1986) which represents knowledge about how to solve a problem (in our case, how to evaluate topics of interest in an assessment) in terms of the topics of interest in the problem domain, and logical relations among these topics.

EMDS DESIGN AND OPERATION

The ArcView®, Assessment, and NetWeaver systems are the three primary subsystems of EMDS (Figure 1). Communication between ArcView and the Assessment systems in particular is implemented with a dynamic data exchange (DDE) client-server architecture in which ArcView functions as the server application. Reynolds et al. (1996) give an indepth account of system design and operation. EMDS version 1.0 provides support for:

- constructing knowledge bases that describe logical relations among ecosystem states and processes of interest in an assessment;
- processing the resulting knowledge base to get a geographic assessment of ecosystem states and processes at the level of landscape features (**Error! Reference source not found.**);
- evaluating the influence of missing data;
- prioritizing new data acquisition in light of data influence; and
- basic project management.

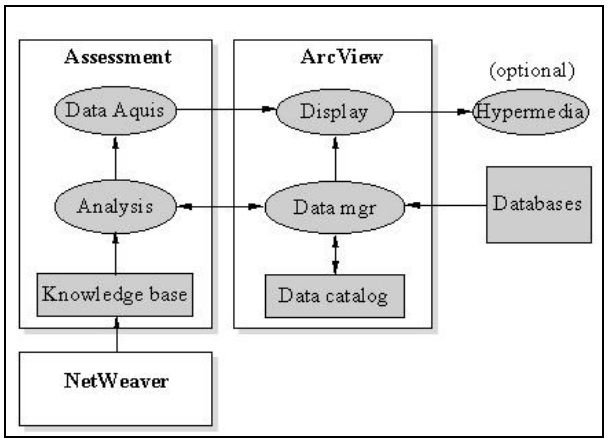

Figure 1.—Basic architecture of the EMDS system.

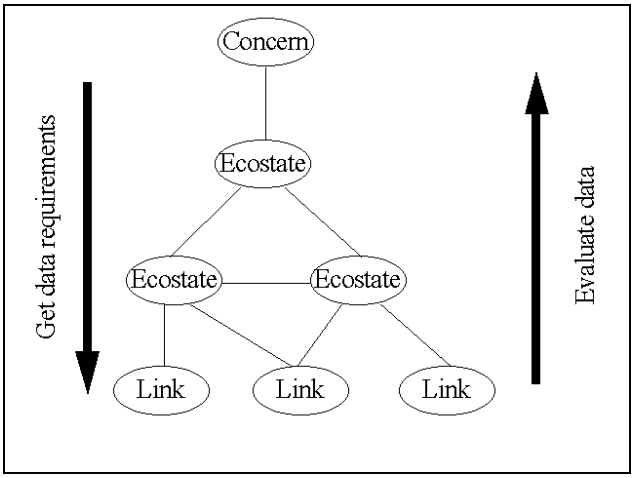

Figure 2.—NetWeaver knowledge base processing

To conduct an assessment with EMDS, an assessment team:

1. Constructs a data catalog that identifies the sources of all GIS themes that can potentially enter into an assessment.
2. Constructs a knowledge base that describes relations among all ecosystem states and processes of interest to the assessment.
3. Selects topics on which to run an assessment, view results, assess data requirements, and perform what-if scenarios.

EMDS includes components that manage all three activities. The customized ArcView project template in EMDS includes utilities for setting up an EMDS ArcView project, creating and editing the data catalog, and displaying analysis results.

NetWeaver provides the knowledge base development environment. NetWeaver knowledge bases are highly modular. As a consequence, the system lends itself very well to the incremental evolution of knowledge bases in the sense that one can start with a simple initial representation and gradually evolve the knowledge base into a larger, more complex representation in a series of small increments. NetWeaver also fully integrates a complete implementation of fuzzy calculus that overcomes a basic knowledge representation problem of traditional rule-based systems which are inherently bivalent.

Although the EMDS framework approach requires users to build the data catalog and the knowledge base, there are also two significant advantages to a framework solution. Topics to be included in an assessment, relations among ecosystem states and processes in the knowledge base, and sources of data are completely determined by the user, so the user has a very high degree of control over problem specification. Also, assessment applications can be developed somewhat generically for a larger region, and then easily customized for specific subregions.

EMDS OUTPUT

The EMDS Assessment system can display a variety of maps in ArcView after an analysis has been run. Map options include display of network truth values, indicators for absence of data, and influence rank values of missing data. All values are mapped to specific landscape features within their respective GIS themes in the assessment area.

Once the DAM process has been completed for prioritizing data types, several graphs summarizing attributes of missing data also can be displayed in ArcView:

- priority value for data links (synthesis of ease of acquisition and data influence);
- ease of acquisition (supplied by the user in the Data Acquisition Manager);
- influence (computed by the NetWeaver knowledge base engine);
- number of rows (feature records) with influence (for each data type);
- percent of rows with influence (for each data type);
- number of no data values (for each data type); and
- percent of no data values (for each data type).

A data type may have missing data (that is, item 6, number of no data values, is positive), but not have any influence on improving the completeness of an assessment, because the missing data associated with the data type are not relevant given the availability of other data. The inherent ability of NetWeaver to provide this type of information has important implications for improving the efficiency of assessments.

EFFICIENT USE OF THE EMDS FRAMEWORK

The EMDS framework is a flexible environment for designing and performing assessments. It is perfectly acceptable, for example, to conduct an assessment in the conventional way:

1. perform an information needs assessment;
2. acquire the data that has been identified;
3. enter the information in a GIS system;
4. analyze the data, and
5. produce a report.

However, an assessment conducted in the usual manner ignores much of the power inherent in a knowledge-based approach to assessment. The dependency network knowledge base encapsulates important information about inter-relations of topics and data that can be used to improve the efficiency with which an assessment is conducted. Given a knowledge base, the NetWeaver engine automatically generates valuable information about the influence of missing data. This is true even in the extreme case in which there are no data at all.

Because of the size and complexity of ecological assessments, thinking about all the implications of all the relations is virtually impossible even for a room full of technical experts. Although dependency networks can be extremely large, they are incrementally built out of small, simple elements, and can be built fairly quickly. The result of knowledge base construction with networks is, in effect, a road map of the dependencies that the system can use to provide useful information to guide the assessment process. With this in mind, we suggest an alternative

model for the assessment process that can achieve significant savings in time, funds, and use of personnel (**Error! Reference source not found.**).

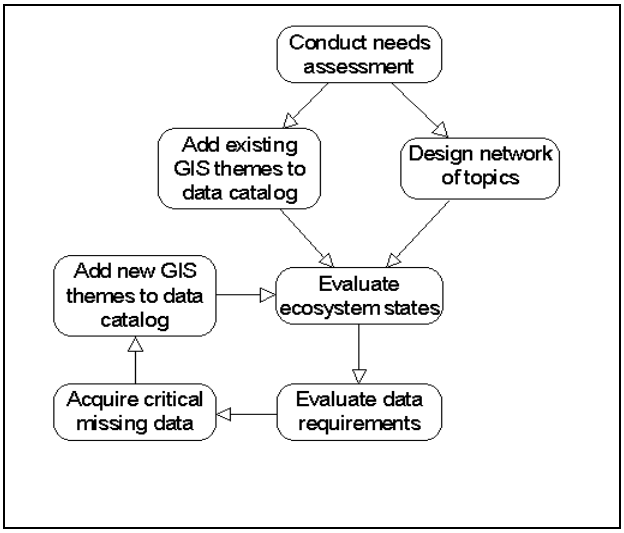

Figure 3.—Iterative cycle of data acquisition based on evaluation of data influence.

PRELIMINARY EXPERIENCE WITH EMDS

Beta testing of the system began in December 1996 with a late-version prototype. Field trials are currently being conducted on three multiagency ecoregional assessments in the United States (Northern Great Plains, Great Lakes, and Ozark/Ouachita Highlands Assessments). Although it is too soon to provide a complete evaluation of system performance in these tests, initial results have been very encouraging. Land managers and resource specialists have demonstrated a reasonable level of proficiency in the art and science of knowledge base development with only 1 day of training. Similarly, the basics of system setup and maintenance have been mastered by GIS staff with three to four hours of training.

DISCUSSION

Information about the influence of missing data can be as important to an assessment as information that can be derived from available data. There are two basic questions about missing data that determine its priority for collection:

- What missing data should be obtained to best improve completeness of an assessment (considering their influence)?
- How easy is it to obtain the missing data?

The EMDS inference engine automatically answers the first question in the course of processing a knowledge base. The EMDS Data Acquisition Manager determines the priority of missing data by synthesizing a priority score from each data type's influence score and a measure of its "ease of acquisition."

Data requirements associated with assessments often make the process expensive, time consuming, and labor intensive. The EMDS system has the potential to significantly improve the efficiency of an assessment process by identifying data that are no longer important by virtue of being redundant to, or rendered irrelevant by, information that is already available. The cost associated with this benefit is, of course, that one must first construct a knowledge base, which itself requires time and effort. However, we believe that managers conducting assessments

will find the investment in constructing a formal logical inferencing model yields returns in time and effort saved that substantially outweigh the investment in knowledge base development.

EMDS allows construction and evaluation of arbitrarily large knowledge bases that can process attribute data from a diverse collection of landscape features within a single analysis, making it relatively easy to integrate many different types of concerns into an analysis, and account for their inter-relations, at least to the extent that we can understand and represent them symbolically.

ACKNOWLEDGMENTS

Cooperators participating in systems analysis and design to support development of EMDS version 1.0 are: Dr. Michael Foster (Penn State University, University Park, PA), Dr. Richard Olson (Agricultural Research Service, Mississippi State, MP), Dr. Donald Latham (USDA Forest Service, Intermountain Research Station, Missoula, MT), and Dr. Daniel Schmoldt, USDA Forest Service, Southern Research Station, Blacksburg, VA.

REFERENCES

REYNOLDS, K., CUNNINGHAM, P., BEDNAR, L., SAUNDERS, M., FOSTER, M., OLSON, R., SCHMOLDT, D., LATHAM, D., MILLER, B., AND STEFFENSON, J. 1996. A Design Framework for a Knowledge-based Information Management System for Watershed Analysis in the Pacific Northwest U.S. *AI Applications*, **10**, 9-22.

SAUNDERS, M.C., RAJOTTE, E.G., AND TRAVIS, J.W. 1989. The use and significance of expert systems in small fruit IPM. In N.J. Bostanian, L.T. Wilson, and T. J. Dennehy (eds), *Monitoring and Integrated Management of Arthropod Pests of Small Fruit Crops.* Intercept, London.

SAUNDERS, M.C., COULSON, R.N., AND FOLSE, J. 1990. Applications of artificial intelligence in agriculture and natural resource management. pp. 1-14 in A. Kent and J. Willeams (eds) *Encyclopedia of Computer Science and Technology.* Volume 25, supplement 10. Marcel Dekker Inc., New York.

SCHMOLDT, D.L., AND RAUSCHER, H.M. 1995. *Building Knowledge-Based Systems for Natural Resource Management.* New York: Chapman & Hall.

STONE, N. D., COULSON, R.N., FRISBIE, R.E., AND LOH, D.K. 1986. Expert systems in entomology: three approaches to problem solving. *Bulletin of the Entomological Society of America,* **32**, 161-66.