FACILITATING ACCESS TO ENVIRONMENTAL INFORMATION WITHIN THE USEPA

A STRATEGIC WHITE PAPER

Developed by:
Dr. Robert F. Shepanek
National Center for Environmental Assessment
Office of Research and Development
U.S. Environmental Protection Agency
Washington, DC

March 17, 1997

EXECUTIVE SUMMARY

"EPA's data resources represent one of the Agency's greatest assets. As a national Federal source of reliable and comprehensive statistical information on the state of public health and the environment, EPA is uniquely equipped to provide the public with critical tools to pursue responsible policies. "

Browner and Hansen, EPA Reorganization Memorandum, February, 1997

One of the preeminent challenges to EPA today is to provide a uniform and easy way for scientists, decision makers and the public to identify, share and use the critical tools and information that it holds. What is needed are inventories of these items, easily accessible through publicly available and inexpensive software. These inventories should be kept up-to-date by EPA program offices, regions, laboratories, states, tribes, communities and other EPA partners who generate and need to share environmental information and tools. The inventories need to be linked, so that an individual looking for data finds it with a single request, regardless of where the information resides.

A system that demonstrates these features is the Environmental Information Management System (EIMS) developed by ORD's National Center for Environmental Assessment (NCEA). It is currently operating in Region 10 supporting environmental analysis in the Pacific Northwest, and in ORD's National Exposure Research Laboratory (NERL) supporting environmental analysis in the Mid-Atlantic states.

The approach recommended in this paper is to integrate the EIMS with other region and program office systems. As part of this deployment, software will be made available that will enable a user to answer questions like, "What data sets are available throughout EPA that contain information on contaminants in sediments?" In addition, software will be provided that will allow EPA and its partners to directly update their inventory database over the Internet. Because of investments made by ORD and Region 10 in the development of the system, the cost of implementing in additional programs and regions is low, while the benefits to EPA of a searchable directory to support initiatives like "Right to Know" and the Center for Environmental Information and Statistics are enormous.

The technical impediments to deployment have for the most part been overcome. The major hurdles that remain are management issues. It is envisioned that the systems will be managed through the EPA Executive Steering Committee for Information Resources Management to ensure that EPA-wide search capabilities are not compromised over time. ORD, Region 10, and the Office of Waters "Surf Your Watershed Program" are working collaboratively on development of policies, standards and the implementation of the distributed inventory. The purpose of this paper is to initiate discussion among the full range of EPA entities that must work together to make the vision of an EPA-wide inventory a reality.

DEFINING THE PROBLEM

The discipline of environmental science is changing. The analyses and assessments needed to address environmental problems now require integrating multiple kinds of data at multiple scales. These analyses and assessments also require the concerted effort of teams of investigators who often are separated geographically, yet need to share data and information. These changes are being driven by the needs of assessment scientists and managers, as well as by the needs of programs within the Agency such as the Office of Water's Watershed Assessment Program and ORD regional environmental assessments.

As the kinds of environmental analyses being performed change, the management of environmental information must also change. Large volumes of remotely sensed data need to be managed and placed in context with an overwhelming volume of data from traditional sources. Assessment teams require access to these data and often various models needed to complete analyses. More importantly, these teams require access to complete descriptions of data and models to evaluate their applicability for specific analyses and assessments. With increasing emphasis on the distribution of federal information resources to the public, the implementation of community Right-to-Know initiatives, and the formation of the Center for Environmental Information and Statistics, providing public access to descriptive information and data has become a priority.

NEEDED TOOL

Interagency Access

Direct Access By Agency Staff

INVENTORY OF EPA ENVIRONMENTAL RESOURCES

Public Access Via Internet

Stakeholder Access And Input

Currently, there is no single point of entry to identify and find environmental data sets, data bases and models developed or held by the Agency. This situation has contributed to the inaccessibility and loss of some environmental data, and the duplicate storage of other data. In recognition of this problem, information management systems developed within the past five years have emphasized the need for inventories of environmental data similar to a library catalog for published media. For example, the Government Information Locator System (GILS) provides an index of EPA information management systems and Envirofacts is a warehouse for EPA's large data bases related to regulatory functions. These systems, and other information management systems developed for specific programs or offices, are not linked to provide an integrated inventory of the Agency's environmental information resources. Such an integrated inventory is needed by Agency staff conducting analyses and assessments, other federal scientists needing access to environmental information held by the Agency, regional stakeholders and the public interested in environmental information in general.
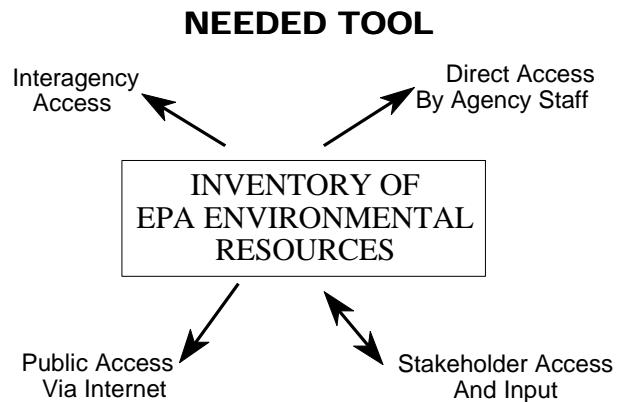
PROPOSED SOLUTION

The National Center for Environmental Assessment (NCEA) has developed an environmental information management system (EIMS) that facilitates the organization and maintenance of descriptive information (metadata) about data sets, data bases, projects, models, and documents, thus providing the design for an inventory of environmental information. The inventory is accessed, revised, and added to using readily available Internet World Wide Web browser software. This pilot system is being used by ORD's Regional Vulnerability Assessment (ReVA) Program and Region 10 to provide information about environmental resources being collected and used for ongoing analyses and assessments.

> **Metadata**
>
> The term metadata refers to information about data. Common components of metadata include summary descriptions (abstract), methods, quality assurance information, contacts, temporal information, and keywords describing contents and geographic extent. For this project, metadata also refers to information describing projects, models, and documents.

Inventory information is stored and maintained in a relational data base management system (Oracle). The inventory can point to data either stored within the system or as distributed external files. This provides the capability to track remote sensing data, GIS coverages, and other kinds of data for which entry into a relational data base structure is not appropriate. Inventory information within the relational data base is consistent with the Federal Geographic Data Committee (FGDC) metadata content standards for spatial data; however, a significant enhancement of these standards, is the addition of a hierarchical metadata framework that is used to organize detailed scientific documentation into sections reflecting the organization of a scientific paper or report. This facilitates the review of descriptive information by users.

The NCEA EIMS design provides a repository for scientific documentation that can be accessed easily with standard Web browser software. This capability places a virtual library on the desktop of EPA staff and others with Internet access. Using Web forms, users can complete searches within the virtual library of environmental information of interest based upon defined criteria related to spatial and temporal attributes, kinds of environmental resources, data origin, or contact person. In this respect, user-defined searches are more efficient than currently used Internet World Wide Web search engines.

In addition to being a principal means of access to inventory information, Web forms can be employed to assist users with providing information for new inventory entries. Using this approach, regional stakeholders can participate in the growth of the inventory and, as in the case of Region 10, use the inventory as a way of providing information about their own data holdings.

To realize the vision of an Agency-wide inventory of environmental information, NCEA proposes that the metadata component of the EIMS be used as the prototype design for the management of descriptive information about data, projects, and models developed or held by the Agency. Direct

links between systems based upon the NCEA design (such as the ReVA and Region 10 systems) are immediately possible to form a virtual inventory of environmental information; however, integration with other data systems is required to complete the Agency-wide inventory. One of the principal tasks identified as part of this project is to refine the technical approach and develop the management approach to support this integration of various information management systems.

IMPLEMENTATION AND RESOURCES

The NCEA EIMS design was based upon that used for the first phase of ORD's Environmental Monitoring and Assessment Program. Since then, significant improvements have been made and the design implemented for one research program (ReVA), one regional office (Region 10), and an office within ORD (NCEA). The implementation process has highlighted how similar the needs of the participants are in the area of metadata management. Despite slightly different requirements articulated by the users associated with each program or office, the resources expended for each implementation have decreased dramatically as a result of the technological development and experience gained from each. Within the next month, the EIMS design will be implemented for the Officer of Water's Surf Your Watershed Program; it is expected that the resources needed for this implementation will continue to show this drecreasing trend.
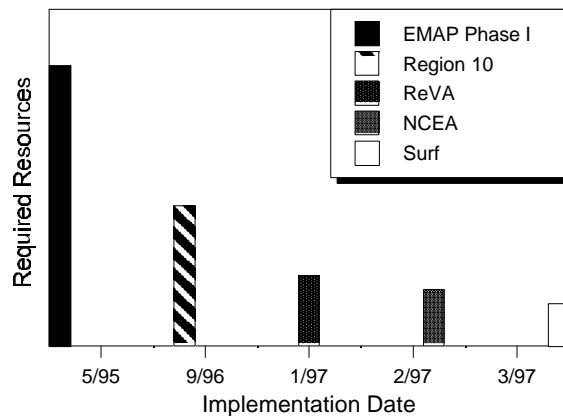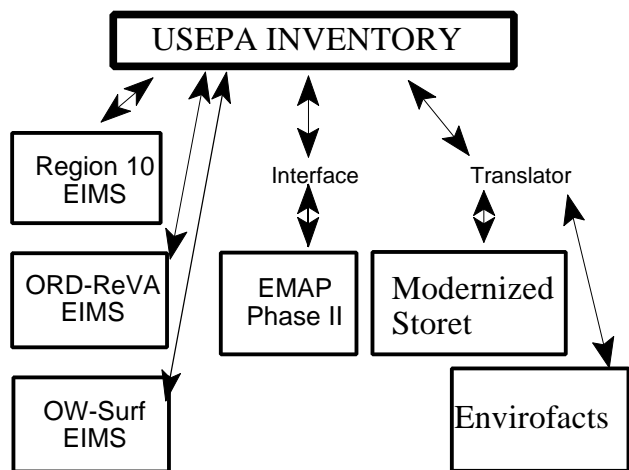
Because of their common design, the inventory information in the ReVA, Region 10, NCEA, and Surf Your Watershed environmental information management systems can be integrated immediately to form an inventory of environmental information. Integrating inventory information residing in existing systems (i.e., STORET, EnviroFacts, GILS, etc.) will require a "fit analysis" to identify differences in the way information used in the inventory is managed and stored and optimal solutions for integration. This fit analysis will result in further refinement of the NCEA EIMS design; however, the results of the fit analysis will also be used to develop a series of interfaces or translators that will link existing systems with the inventory. These translators will reflect emerging technologies for integrating relational data bases, simultaneous query of multiple data bases, and access to tabular and graphical data using the Internet.

Resources will be needed to conduct the "fit analysis", refine the core EIMS design, and implement the emerging technologies that will enable simultaneous query of multiple systems using common Web browsers. Resources are also needed to develop the interfaces and translators needed to link to existing systems. Guidelines will also need to be developed for the content and organization of

inventory information to facilitate integration across the Agency. These guidelines will be used to maintain consistency among various scientific data systems held by the Agency and will establish standards for metadata content. Initially, metadata content standards will be limited to descriptive data; however, eventually standards will be developed for the naming of individual data elements, thus facilitating the linkage across systems of both metadata and data.

The EPA's inventory of environmental resources will be both dynamic and distributed. The development of EIMS and various guidelines will be an evolving process to meet new requirements defined by the scientific community, emerging interagency federal standards, and the needs of the public in search of environmental information.

USEPA INVENTORY

Region 10 EIMS

ORD-ReVA EIMS

OW-Surf EIMS

Interface

EMAP Phase II

Translator

Modernized Storet

Envirofacts

MANAGEMENT

The greatest challenge will be to implement a management structure that can coordinate the resources needed and maintain consistency between multiple scientific information management systems. The prototype systems currently implemented in Region 10 and for the ORD ReVA Project demonstrate that the technical challenges can be overcome with readily available existing technology. The management challenges to EPA-wide implementation remain. The following solution is proposed.

- ! Use the existing EPA-wide Information Resources Management (IRM) policy infrastructure by creating a subgroup of the Executive Steering Committee for IRM. This will provide a means of governance that will ensure that an EPA-wide distributed inventory of scientific environmental data and information will evolve in a coordinated way that is consistent and responsive to the needs of top management, scientists, program and regional managers.

- ! Develop an approach under the Working Capital Fund that will promote involvement and enable the purchase of services by participating program offices. A working capital manager would be designated in ORD. Examples of the type of services to be provided might include: implementation of the EIMS for additional programs or offices; integration of additional systems; compilation of inventory entries; or, assistance with the implementation of data administration programs.

! Develop and implement an outreach program that would inform EPA and other potential customers of EPA's scientific data, information and tools.

The management approach needed emphasizes collaboration among OPPE, ORD, OIRM, Regions and the program offices that will facilitate sharing of environmental information within EPA and with other agencies, states, tribes and communities.

BENEFITS

The benefits of an Agency-wide inventory of environmental information are enormous. For the first time, USEPA staff and others will have the ability to find descriptive information about data, models, and projects within the Agency. The strategic approach described in this white paper coupled with the EIMS provides :

! a framework and system design for the management of metadata for environmental resources held by the Agency;
! a means to integrate environmental inventory efforts sponsored by individual offices and projects;
! a consistent data base design to support watershed and regional assessments;
! standards for the content and organization of descriptive information (metadata) about environmental resources;
! minimization of redundant storage of environmental information across various systems maintained by the Agency.

Implementation of the inventory as described in the plan and demonstrated by the EIMS will bring to the desktop of Agency staff a virtual library of descriptive information about environmental data resources, greatly reducing the time it takes to find, access, and assemble information needed to perform analyses and assessments. This capability is essential for EPA to meet the challenge of emerging environmental problems that are both multi-media in nature and of greatly increasing spatial and temporal scale. If adopted, this approach can greatly enhance the value of one of EPA's greatest assets, environmental data and tools, by making them a resource that is shared throughout the Agency and with our partners in environmental protection.