Washington's Data Submittal Guide: Building Access to Environmental Data

March 30, 2000

By

Lynn Singleton, WA Department of Ecology
Russ Darr, WA Department of Ecology
Keith Hinman, Ross & Associates
Lewis McCulloch, Ross & Associates

Abstract

Washington's Department of Ecology is in the process of integrating its management and environmental data across traditional programmatic and 'media' boundaries. Integrated data management agency-wide creates wider data access, better functionality for users, and higher data quality, as well as improved priority-setting and more integrated environmental management. As a critical early step in integrating its data, Ecology has built the Environmental Information Management (EIM) System, designed as an agency-wide repository for ambient environmental data. EIM contains monitoring data and integrates tabular and GIS views of air, soil, sediment, water, groundwater and natural resource information.

To realize its potential, EIM must receive and validate data from a variety of sources outside Ecology, including volunteer monitors, consultants, grantees, contract laboratories, and local governments. To meet this need efficiently, Ecology developed a data entry tool, the EIM Data Submittal Guide, which guides users through the steps needed to create complete, EIM-compatible electronic data files. This HTML-based Guide provides a user-friendly interface, easy navigation, and flexibility in formatting, to accommodate the range of different sources from which data will come. By defining, documenting and checking needed data elements, the Guide also serves to communicate data standards for ambient environmental data and to promote data consistency and quality. Finally, through the use of optional run-time Microsoft Access database software, the Guide can itself serve as a data entry and management tool

for data providers In the future, Ecology plans to make EIM available on the Internet, allowing remote users direct access to all the functions of EIM.

INTRODUCTION

The Washington State Department of Ecology (Ecology) is very similar to many other Environmental Management agencies in the nation: most activities have historically focused on a particular regulatory program. However, today's environmental management issues are growing more complex and we require information from a variety of sources. Landscape, cumulative impacts, and Endangered Species Act concerns are changing the ways we perform our duties and fundamentally changing the way we need to access and use information. There is a growing expectation that we provide our information to local governments and the citizenry to support more placed based decision-making.

Ecology was not fully utilizing new information management tools or making the most effective use of data collected. We found that our data collection and information management solutions had evolved/grown from the old mainframe and paper file days to individual desktop spreadsheets and databases. While this was progress, it was not sufficient to meet the environmental management challenges we faced.

Resources are always tight and agencies must take advantage of their historical data collection efforts and the efforts of others. For example, Ecology spends more than $1 million a year on laboratory analyses for environmental samples. This does not include collection or any further data processing or analysis costs. We also provide grants to others e.g. local governments, to collect environmental data for collective use. Historically, much of that information was lost to file drawers when the initial purpose for collection was completed. Secondary use by others was very difficult—effectively impossible. In short, we have invested a lot of time and money in our environmental data and we need to realize the full potential of that investment.

This situation caused Ecology to fundamentally re-evaluate the way we managed all of our information resources. The vision resulting from this strategic information planning is depicted in Figure 1.

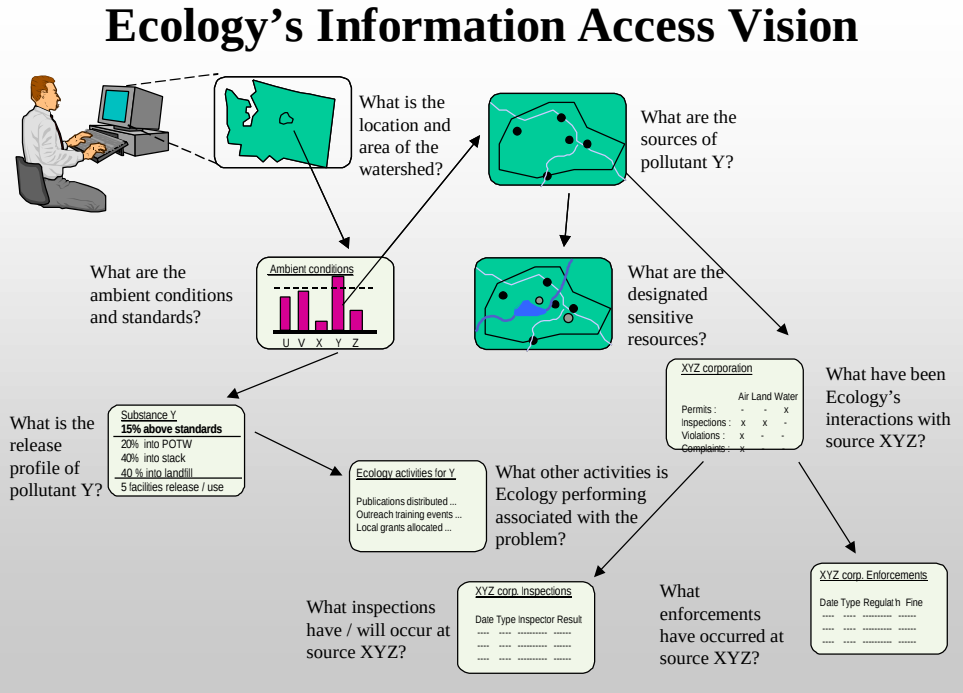

Figure 1: Washington Department of Ecology's Information Access Vision

This vision of environmental information access includes ambient and natural resource data, permitting, releases, compliance, enforcement, and facility related data. Information is accessible by all staff (and ultimately by stakeholders as well) from their desktops. Users can enter the "system" from any of the functional areas mentioned above and conduct their business. Data have a common look and feel, and are adequately documented. Tools such as GIS and database queries are available to all users. This vision drove the creation of The Environmental Information Management System (EIM) and many other integrated components. The Data Submittal Guide (DSG) was built to facilitate the loading of information collected by others into EIM.

THE ENVIRONMENTAL INFORMATION MANAGEMENT (EIM) SYSTEM

The EIM system provides a central repository of Ecology's most important ambient environmental monitoring and natural resource data in a format that all of Ecology's programs can access. Use of the EIM system helps users form a picture of the state of the ambient environment in a particular area.

EIM is designed to include both natural resources and environmental monitoring data. *Natural Resources* are those resources that the agency needs to be aware of or which the agency regulates or influences. Examples include water bodies such as reservoirs, rivers and groundwater; population centers; and sensitive land masses such as arable land, wetlands and shorelands. *Environmental Monitoring* is defined as the collection and analysis of analytical data, which help determine the environmental condition of air, land, and water throughout the state. These activities include ambient
monitoring of background levels of specific chemicals and regulatory sampling of pollutants released in the vicinity.

The EIM system captures information on the *projects* that involved the collection of environmental measurements, *stations* at which measurements were taken, *field activities* (sampling, measurements and observations made in the field), and the *results*, or values of those measurements. *Samples* are representative matter collected in the field and analyzed in a laboratory. *Measurements* are data collected in the field, like temperature. *Observations* are statements such as the sky is partly cloudy.
Together, this information provides specifics about the condition of the environment in a particular area, as well as sufficient documentation to interpret it. The types of ambient information EIM is designed to hold are illustrated in Figure 2.

Why EIM was created

Ecology depends on a solid foundation of high quality, timely, and accessible information to support the full range of its business functions and day-to-day activities. Ecology's long-term information technology goal is to provide agency management, staff and interested stakeholders with timely and direct access to the information they need to answer the questions they face on a daily basis.

EIM was created to meet a number of high priority needs including providing universal access to environmental information across programmatic boundaries, providing a repository for project-specifc and grantee-collected data, and assisting in the sharing of data between Ecology and external data users. It is intended to provide automated information support where none existed, enhance the consistency and documentation of data, and provide all users with access to tools such as GIS and queries made possible through relational database technology. EIM addresses the fragmentation and inaccessibility of databases and spreadsheets within Ecology, and provides for cross-program access to data.

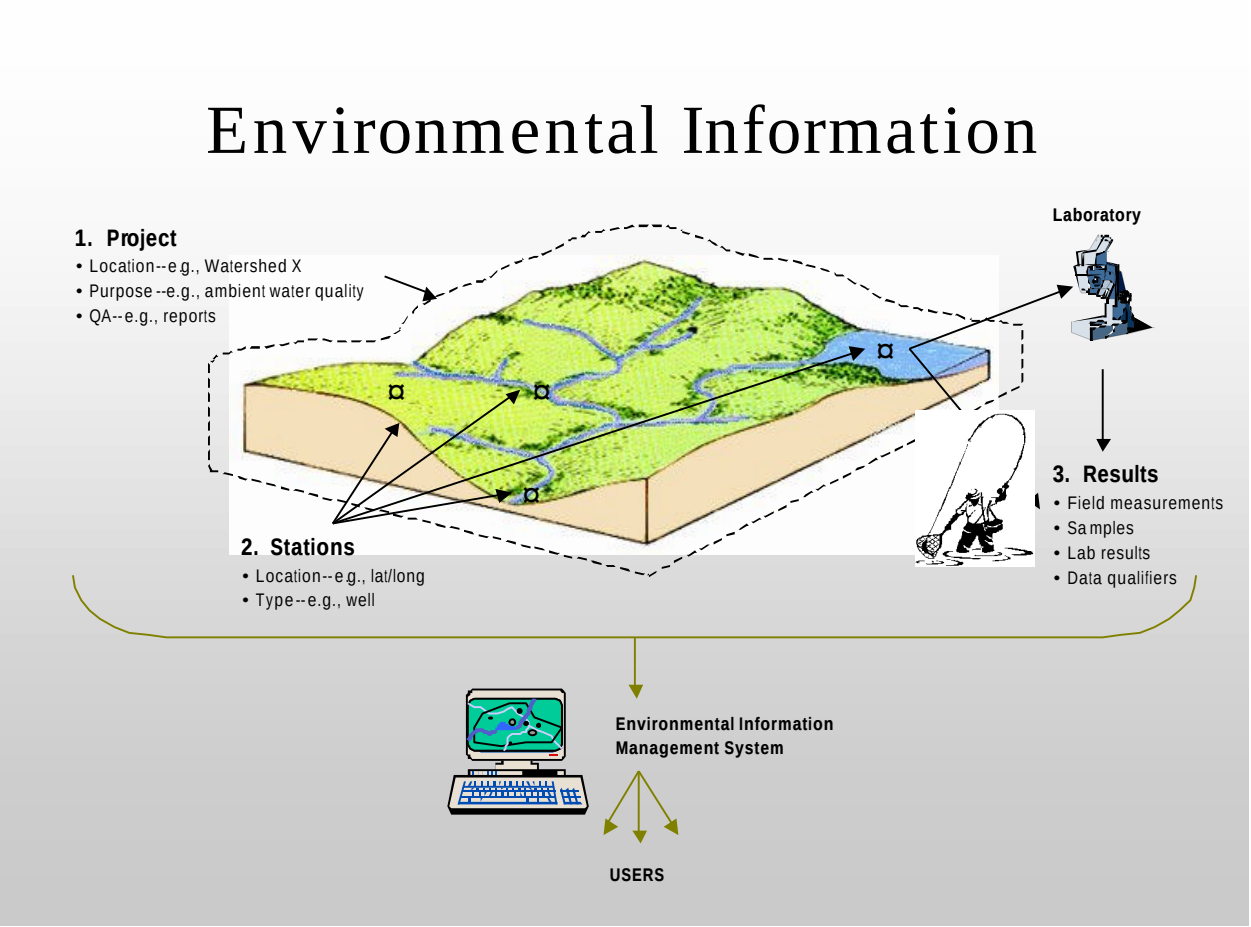

Figure 2: Environmental Information Captured in EIM

Although all data in EIM are associated with specific projects, the data may have value for many uses beyond the original intended purpose. In the past these data were sometimes retyped into a database or spreadsheet when they were needed for further analysis. Even worse, because there was no central repository for this information, duplicative data were sometimes collected and analyzed, because it was not easy for Ecology staff to know what results already existed. It was determined that Ecology could save hundreds of thousands of dollars annually by making all lab data available through EIM.

Because of the desire to accommodate the potential needs of data users and data providers from a wide range of programs, EIM includes a relatively large number of data elements. Users can seek information in a number of ways; for example, one can query the data to find all of the ground water data in a given county, or search for air monitoring data within a radius of a user-provided location. In addition to standard reports, EIM allows users to design their own custom queries to take advantage of the wide range of information available. Since people other than those who initially collected the data can use data, the system requires the documentation of *metadata*, or information about projects, stations, field activities and results.

THE DATA SUBMITTAL GUIDE

The Data Submittal Guide (DSG) refers to a number of files packaged together that the Ecology makes available to any party that wishes to submit data to EIM. The DSG is available on the World Wide Web at http://www.wa.gov/ecology/as/iip/eim/eim_main.html. Environmental data submittal is a requirement of many grants, permits and cleanup actions under Ecology's authority. The DSG was created to assist these individuals or groups in preparing files to send to Ecology for upload into EIM. EIM requires that the input files contain certain mandatory fields in a specified format. The DSG is the tool that helps external groups meet these requirements. Within the DSG there are three main components: a runtime Microsoft Access application, a series of formatted spreadsheets, and the Data Guide, an html document that provides the users with instructions for using the other components. Each one of these components is described below.

The Data Guide provides an overview of the EIM system and the types of data that EIM contains. The Data Guide contains general directions to users to help them evaluate the most effective way for them to use the Data Guide to assist them in preparing their data submittal package. The Data Guide proceeds to describe the three main options for preparing the file and provides users with specific instructions depending on the method selected. The last section of the Data Guide describes how to communicate Ecology if the user runs into problems as well as specifically how to transmit completed files to be uploaded into EIM. The appendix of the Data Guide presents a dictionary of all of the data elements that can be submitted. This dictionary includes the data element name, whether or not the data element is required, required under certain circumstances, or optional, a definition of the data element, specific instructions to users depending on their data entry method, an example of the data element, and the specified format of the data element. The information about the specific data elements can be accessed a number of different ways throughout the data entry process depending on the method of entry selected by the user.

To assist a variety of data providers, Ecology has provided three options for submitting data. Data providers may simply enter their data into a Microsoft Access application via on-screen forms; they may fill in a series of pre-formatted spreadsheets with their data; or they may provide the data as a tab-delimited file from an existing database or spreadsheet program. The Data Guide has instructions for how to use each of these various methods. All three data entry options have the same data element requirements and all seek to aid users in submitting data based on the main components in EIM as depicted in Figure 3. The three options are outlined below:

On-Line Forms: This data entry option provides the users with various screens with windows where data can be entered. It arranges related data together on one screen. This option is probably the most user-friendly as is has all of the permitted value tables available in drop-down lists and provides various search options which aid the user in entering data correctly. There are also direct links between this format and the data element definitions in the Data Guide.

Pre-formatted Spreadsheets: There are two different versions of this data entry option, a color-coded version and a basic version. Each version contains three different spreadsheets, which arrange related data together on a single spreadsheet. The color-coded version contains spreadsheets that have color-coded columns notifying users which data fields are required, required under certain circumstances, or optional. The basic version of the spreadsheet is the same as the color-coded version but without the color. The spreadsheet option may be attractive to those users who have a large number of records in an existing spreadsheet or database, which they would like to "paste in" to the spreadsheets formatted for entry into EIM.

Tab-delimited file: No specific tool is provided for this data entry option. Users may create tab-delimited files for submitting data. The only requirements are that these files follow the order and format of data fields presented in the pre-formatted spreadsheet option. Those transferring data from an existing spreadsheet or database may find this option most convenient.

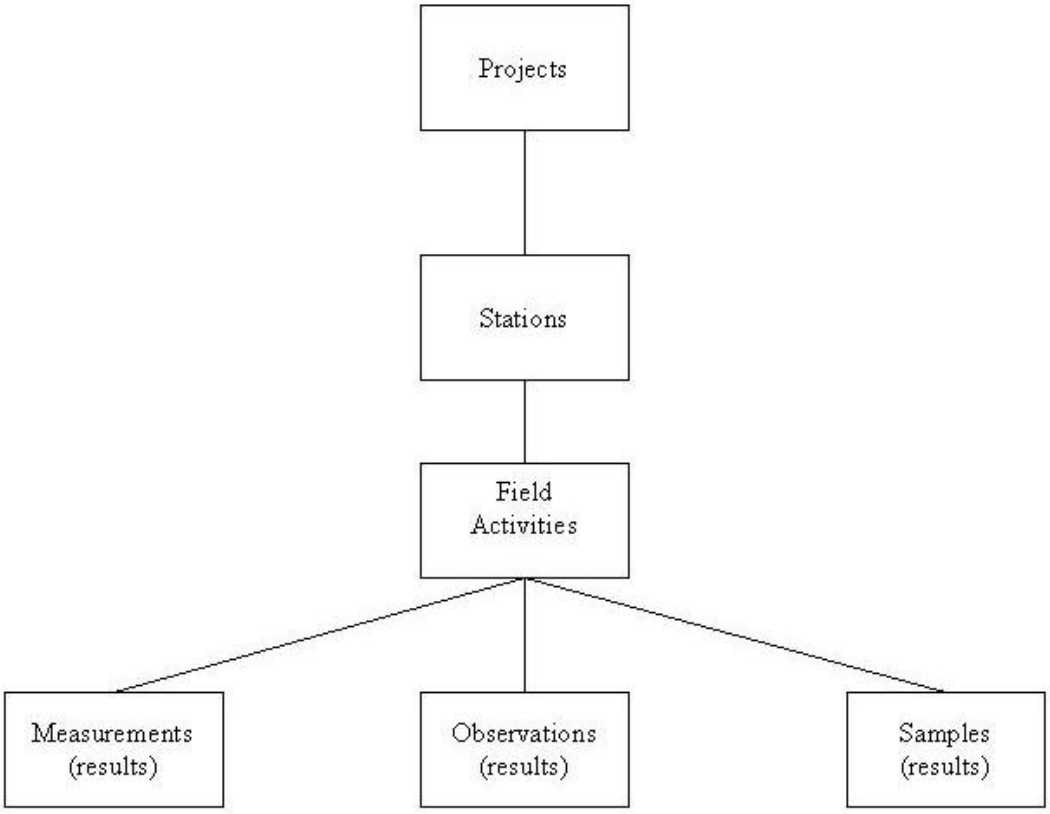

Figure 3: Main Data Components of EIM/DSG

The DSG is designed to mirror the data components in EIM making the transfer of information into EIM much simpler than if the data were being submitted in a variety of formats. Much of the information in EIM is data on the results of field observations or measurements, or samples collected in the field and analyzed by a lab. A majority of the data are result data (such as the concentration of a pollutant in the water.) EIM (and therefore the DSG) requires you to provide certain information about your field activities and results (e.g. the date a sample was taken, any lab qualifiers on the results), in addition to the numerical results themselves. In addition to the data that are recorded about activities and results, information needs to be collected about the project and the station associated with the activities conducted (see Figure 3). For example, when result data are recorded, the information about the purpose of the project and the location of the station(s) and where the data were collected need to be included. These specific data on the project and station usually need only be recorded once for a large set of result data; but are equally important.

The types of documentation described above have been designed to meet the "20 Year Rule:" that data documentation should be sufficient to allow a person to understand the purpose for a data set, methods

used, results obtained, and the quality assurance measure taken 20 years after its collection. In short, data sets need to stand on their own. Since the data are intended for use multiple times, it is critical that they be as accurate and complete as possible. In addition users will search for data about particular chemicals, water bodies, and geographic locations that may originate from different sources. To help ensure consistency, users are required to record answers using standard terms taken from a list of "permitted values" or an appropriate reference table. Without these standardized values, the data would not be consistent enough to permit users of EIM to rely on its quality, run reports, or conduct queries of the database.

The DSG aids users in creating data sets that follow all of these rules. The on-line forms and spreadsheet data entry options aggregate all of the data elements based on projects, stations, and results. All of the required fields are clearly indicated. In the on-line forms direct access to the permitted value tables are provided. Additional spreadsheets of all of the permitted value tables are provided to spreadsheet and tab-delimited data entry users. All of these aids were built into the DSG as a means to make it as easy as possible for users to create complete high quality data sets.

Despite all of these mechanisms for easing the data entry process, the biggest challenge in creating the DSG was still to make it easy for users to submit data. The data in EIM are complex and cover a broad range of activities. This breadth makes EIM a very flexible and valuable tool, but also one that requires some time to learn. The DSG, in order to parallel EIM, is also complex and contains numerous data elements. In user acceptance testing some users commented that most of their data collection activities did not include a lot of the data elements that are a part of EIM/DSG. This comment reflected both the inherent complexity of the system (i.e., the number of possible data entry options and fields) and, in many cases, the lack of documentation information required for a complete data set.

A major challenge in creating the DSG was to make it user-friendly. A broad range of groups with vastly varying technical capabilities submit data to Ecology and therefore the DSG needed to be able to accommodate as much of this range as possible. The DSG simplifies and organizes data entry by providing checklists and question-guided data entry, as well as a user-friendly set of data entry screens. The DSG also helps users identify which information is essential for a complete, documented data set. To provide maximum flexibility, three main data entry options were developed, along with an overarching set of documentation.

IMPLICATIONS AND LESSONS LEARNED

Ecology's experience developing the EIM system and creating the DSG yields a number of lessons learned that may be of interest to other organizations faced with a similar challenge - integrating disparate ambient environmental data from a range of different programs, agencies and other entities. Most of the significant challenges involved in implementing EIM and the DSG flow from two key tensions. First, EIM is intended to make monitoring and other data from specific sources or programs (ranging from groundwater data at Superfund sites to criteria air pollutant data to taxonomic data from surface water bodies) available much more easily and with greater functionality to a much wider range of users than in the past. Thus, its primary benefits accrue to *secondary* users of data (i.e., users other than those who collected and/or "own" the data). Yet it is *primary* data collectors and owners who must bear the cost of new data standards and changed business processes that impose new burdens while providing little direct benefit. Moreover, the wider accessibility of data may not be greeted enthusiastically by data originators for a variety of other reasons - concern about potential misuse or misinterpretation, proprietary attitudes about data, greater scrutiny from environmental or other watchdog groups, etc. Thus, one key challenge is finding ways to

address the disincentives of data originators to take the actions necessary to make data available to the wider audience of secondary users.

A second, related tension is the need to balance the desire for complexity and rigor with the countervailing need for flexibility and ease of use. Throughout the creation of EIM and the DSG, developers (including users aiding in design) needed to make decisions balancing these competing needs. On the one hand, greater complexity and rigor (manifested through such things as greater numbers of data entry fields and options, as well as through mandatory fields and permitted values) help to ensure data are truly comparable, and documented well enough to aid a secondary user in appropriately using and interpreting data. For some purposes (e.g., documenting support for an enforcement action, characterizing a localized, site-specific situation), high quality and clearly documented data are essential. On the other hand, data entry that requires a large number of required fields, the use of specific lists of permitted values, and the review of extensive definitions becomes much more time-consuming and burdensome for data providers. Some users may need technical assistance simply to provide data, and some may not even possess all the required elements of a complete data set. If the process of providing data to EIM becomes too onerous and demanding, fewer data holders will do so and the purpose of the system will have been defeated.

Staff Involvement and Management Commitment

One key implication of these challenges is that involvement by a wide range of staff - future system users - in developing a new system is critical. Ecology involved over 100 staff users in user groups to determine needs, advise on design, and provide advice on data standards and business processes. Development of the DSG involved additional consultation with program staff, particularly those staff involved in the receipt of data from outside the agency (e.g., volunteer data coordinators, grant officers, interagency watershed management coordinators). While this involvement was time-consuming, it paid substantial dividends by: helping to meet the needs of different programs/offices; fostering good decisions in important areas that balanced the need for scientific rigor and documentation with the need for flexibility and ease of use; and building consensus and an influential set of voices in support of the final design and the choices (and procedural changes) it represented. While the development of EIM and the DSG entailed a robust and extensive user involvement process, it would have been desirable to invest even more in interagency and external involvement; however, the realities of limited time and budget constrained the extent of such participation.

Visible leadership and support by management was also critical. It is important for agency managers to convey their sense of the priority and importance of an undertaking requiring such widespread staff involvement. Agency leaders need to convey their vision of the role of integrated, shared data in supporting superior environmental management in the future. At the same time, management must demonstrate discipline when, as inevitably happens in the course of such a project, some individuals and/or offices decide that it really is not worth their effort to continue participating. Without strong leadership reinforcing the vision of universally accessible, integrated information, participants will follow the path of history and least resistance and focus on meeting their own needs, perpetuating a Balkanized system of information management.

Data Standards

The creation of a single information system intended to hold ambient environment data across programs, media, agencies, and private parties posed a significant challenge: the definition of a set of data standards that would work across very diverse sets of data, that were complete enough to allow secondary users to interpret and use information appropriately, yet were flexible enough to accommodate data collected with

different levels of sophistication and documentation (e.g., everything from volunteer-collected water quality monitoring data to heavily documented data with extensive quality assurance and peer review processes). Data standards were important both to assure accurate data entry into EIM and to assure appropriate use and interpretation of those data. At the same time, data standards could not be so extensive as to discourage or disqualify large sets of useful data, or impose overly burdensome procedural changes or data cleaning.

Ecology identified a number of key areas requiring some degree of standardization if data are to be integrated across program and media lines. Key areas where standards were defined or developed included:

Geospatial data and metadata: Locational data are essential to support GIS applications, which are typically viewed as essential in watershed and environmental management. Locational data are the chief means of integrating data sets from different sources. Ecology adopted an approach similar to EPA's, with latitude and longitude (or equivalent) required for all monitoring stations, along with codes identifying source and accuracy. Ecology has also identified a need to "georeference" its ambient data, i.e., tie monitoring (and other) sites to water bodies based upon a consistent indexing system, and has developed such a system in cooperation with other state agencies.

Analytical data and metadata: Standard and agreed-upon ways of referring to chemical, physical, and biological parameters were identified, involving the development and/or adoption of reference lists using Chemical Abstracts Service (CAS) numbers, defining tissue types for tissue analysis, and using the Taxonomic Serial Number (TSN) developed by Integrated Taxonomic Information Service (IT IS) as the reference for taxonomic data. Other standards were defined for units, collection and analysis methods, laboratory qualifiers, sample handling, and the like.

Project data and metadata: A key objective of EIM was to provide essential documentation describing the original purpose behind the collection of a set of data, and to give some insight into data quality. This consensus objective reflected the poor state of documentation of many existing data sets, rendering them unusable to later researchers. In addition to required fields describing project purpose and contact information, EIM involved the development of an innovative system of rating project data sets' quality and level of documentation based upon a four-tiered system of "Quality Assurance Levels," ranging from unknown/undocumented (Level I) to projects with a formal Quality Assurance Project Plan and peer-reviewed Project Report.

In defining data standards in those cases where a need was identified, Ecology followed some simple but important rules.

1. *Where appropriate and possible, take advantage of existing standards, organizations, and authoritative sources.* Examples of this approach include the use of CAS numbers and the ITIS TSN. Simply agreeing on an existing standard can meet the need in many cases, involves far less time and resources than creating (and maintaining) a new one, and will typically result in a better and more widely accepted standard.

2. *Where no suitable existing standard exists, develop a new one in a process involving the experts - including both data users and providers.* Ecology created a single system of lab qualifiers (since laboratories all used different systems to qualify their analytic data), a set of tissue types, a project quality/documentation rating scheme, and (with other state agencies) a water body indexing and georeferencing system.

3. *Ensure that data are accompanied by adequate metadata.* A strong theme in the development of EIM and the DSG was to define sufficient metadata describing and documenting the "raw" data that secondary users would be able to use and understand them.

Other environmental management agencies are already or will likely encounter many of these same needs, since the business of environmental management has many similarities from state to state. Collaboration across states on data standards may be useful to avoid reinventing the wheel in each agency, and to assist in interstate data comparability. In the long run, a common template or some level of data standardization will greatly aid in the free exchange of information among many parties over the Internet. The National Water Quality Monitoring Council (NWQMC) and a new organization, a Data Standards Council being created under the auspices of the Environmental Council of the States (ECOS) and EPA's new Office of Environmental Information (OEI), may be useful organizations for such a multi-state effort to make monitoring and other ambient data more standardized and easily shared.

Business Process Changes

The implementation of EIM and the DSG has set in motion a number of business process changes at Ecology and among its partners in environmental management. Many of these changes are necessary to begin collecting and managing data in a manner consistent with the system's minimum standards and in a form that allows them to be entered into the system. The DSG has documented Ecology's desired standards explicitly, by defining key data fields, setting minimum required data sets, and making available standard reference lists for chemicals, biological parameters, etc., all in a consistent and easily accessible way.

EIM and the DSG have already had an impact on standardizing the collection and management of ambient data, both inside the agency and among its public and private partners. For example, in a significant shift from past practice, all recent grant awards (such as Nonpoint Source management grants under Section 319 of the Clean Water Act) require that data be provided to Ecology electronically, using the DSG, in a form that can be entered into EIM. Business process changes range from changes to collection procedures and forms to use of Global Positioning System (GPS) units to collect locational data to the definition of new rules on who is allowed to edit and change data (and who only to read it).

Responsible Secondary Use of Data

No issue is as crucial to the success of EIM, and to the goal of widespread sharing of and access to environmental data, as ensuring that data are not misused. Many data originators are concerned about possible misinterpretation of their data; these fears must be allayed if data holders are to be persuaded to provide their data to a system, such as EIM, with wide user access. Secondary users themselves want sufficient documentation to be able to understand the purpose and any limitations of data they may wish to review or use.

While most data needs within Ecology, as in most environmental agencies, are primarily within a program or medium, some cross-program data sharing is occurring and is likely to increase in the future. For example, groundwater quality and quantity information is collected from wells in several different programs, including the Water Resources Program (concerned primarily with water supply) and hazardous waste programs, including the Toxics Cleanup Program. These programs each have an interest in augmenting their data with information that may be in the possession of other programs, which heretofore has been largely inaccessible to users outside the originating program area. Watershed management is an increasingly important approach that will drive the sharing of data across programs and governments to

accomplish such program ends as Total Maximum Daily Load (TMDLs) development, Endangered Species Act (ESA) recovery planning, and water resource management.

Ecology's primary means of addressing this issue has been to emphasize the use of metadata, or information documenting data sets, and other data standards that reduce the probability of data misinterpretation. Data providers are encouraged to supply this information by making some of it mandatory, and by appealing to data providers' "better nature" (i.e., the benefit to prospective data users), as well as their fears of possible misinterpretation. The display of data with appropriate caveats or qualifications clearly noted on the same screen is also important; similar considerations apply to the formatting of reports. Of course, not all questions can be anticipated; therefore, one of the most important pieces of metadata about a project data set is contact information. Where a user intends to base analysis or action upon data originally collected by someone else, there is often no substitute for direct communication to exchange information far beyond what is possible to capture via metadata.

Complexity vs. Ease of Use

Balancing complexity with ease of use and minimization of burden was a challenge in every phase of developing EIM and the DSG. This problem is complicated by the great diversity of data types and data management practices encompassed by EIM. Standard data quality assurance and documentation practices in one program area (e.g., written Quality Assurance Project Plans, peer reviewed project reports, validation/verification tests, etc.) may be burdensome overkill for data collected in another program area or for a different purpose (e.g., volunteer monitoring for planning/screening purposes).

Ecology has approached this challenge in several ways. First, a full range of users was involved in system development and implementation, so that EIM represents a balance of needs arrived at by different users after debate and discussion and is not biased in the direction of a particular set of users. Insofar as possible, the system itself is designed to be flexible in accommodating widely varying data sets; for example, 60% of the station-related fields in EIM are specific to groundwater collection from wells, which requires more extensive information on soil types, well construction, etc. than typical surface or air monitoring stations. Similarly, the number of fields involved in entering a "field measurement" is far less, and the data entry procedures simpler, than that for sampling data collected in the field and analyzed in a laboratory, which requires detailed information and options on duplicates, chain of custody, methods, etc.

Another way in which flexibility is built into EIM and the DSG is the use of metadata to describe key aspects of the data, e.g., methods, measurement precision, documentation/quality assurance level, without specifying required data collection standards (e.g., locational data to +/- 10 feet). This allows data of different types to be stored in the system, while giving prospective users enough information to judge its utility for their purposes.

Despite attempts to strike a balance between competing needs and to build a system that's as flexible as possible, EIM is complex enough to require some time to learn; readying and entering data are not trivial exercises. "Cleaning" data so that they are of consistent quality and meet the system's standards can require a substantial investment, depending on the state of a given data set. How then, can users be provided incentives to encourage them to undertake this extra effort, which largely benefits other users? Ecology has adopted a number of approaches to address this issue. First, EIM does provide many users with functionality they previously lacked, including a GIS interface, flexible querying, standard reports, a variety of download formats, and easier exchange of data with colleagues. Some of the users who have gained the most from EIM are those who previously lacked effective automated data systems, or whose needs were outgrowing the tools they were using to manage their data.

Ecology's Information Integration Project has provided resources and/or direct data cleaning/data entry assistance in cases where workload was high or resources were a problem. Ecology has made use of the stick as well as the carrot; grantees are required to provide data in a form that can be entered into EIM, and lab data come directly to EIM, requiring program personnel to "come" to EIM and complete required metadata to obtain their data.

The DSG was developed as a tool to make the creation of EIM-ready data sets both easier and more accurate, thus reducing the burden of providing data. Without such a tool, it would be very difficult to incorporate data from grantees, consultants and others outside Ecology. The DSG has the corollary benefit of providing a user-friendly tool that aids users in entering data and saving it in spreadsheet or database formats for their own data management purposes.

Other environmental agencies seeking to integrate data across programs and agencies will undoubtedly wrestle with the need to strike a balance between complexity and ease of use, and with the challenge of creating sufficient incentives to cause data holders to provide their data. Ecology's experience may help others anticipate these problems and consider strategies for overcoming them.

NEXT STEPS

The EIM system and the DSG have been in production and available for agency use since late 1997. By the end of 2000, Ecology will complete the latest round of minor system enhancements. Migration of the agency's legacy environmental data systems and historical projects will be complete. The DSG will have been in active use by several data providers and contract laboratories and any needed updates will be included. The system will have met Ecology's internal needs for information management very well and this component of the integration effort will have been a success. Additionally the DSG, with some modest enhancements, could be developed into a freestanding Microsoft Access environmental database. We are working with some local and state agencies that have expressed interest in this.

Integrated Management of Ambient Environmental Data Has Significant Benefits

The public and other environmental managers will benefit greatly from the compilation of environmental data and the enhanced access to multiple data sets. The overall data quality will also improve by the inclusion in the system and increased use. EIM enforces standard levels of documentation and metadata for each project, station and result. Once lost historical data are now present and documented for long-term analyses.

Increased access also has a side benefit. Data quality will improve if data are available for public review and use. Errors and inconsistencies will undoubtedly be discovered and corrections made as appropriate. The variety of potential uses of the information also increases as the quantity of information grows. The types of decisions that can be informed by collected data also increases. Data displayed in widely accessible GIS coverages and maps can help regulatory decision makers. For example, a proposed permit in a sensitive area can be determined easily if the data are available and displayed to the permit reviewer. If the data are available via the Internet, the project proponent will be able to determine the conflict before submitting a proposal to the agency. Increased burden reduction is achieved by all concerned parties in this case. This can only happen if we are able to deploy GIS data in a way that is economical and requires minimal user training. The GIS industry is responding to these needs.

Remaining Challenges and Future Plans

At present, data access external to Ecology generally occurs via special request. We want to make the information available on a 7 day X 24 hour basis. This will provide the most benefit to the public and our regulatory partners and will reduce the agency's transaction costs for providing information and responding to custom requests. We are looking at Web-based solutions in this regard.

Improving data access outside Ecology will likely be an area of future growth/interest. Instead of pushing data to our partners, we will have our data posted on Ecology's Web site for direct access. Extensible Markup Language (XML) offers promise in this area and we are looking at ways to pilot this technology. We want to keep one copy of the data current and accessible instead of maintaining several of the unique and maintenance-plagued data feeds to national systems. This might be labeled the "come and get it when you need it" approach to information management. In that way the data are used for the immediate purpose and data maintenance is not a concern for the provider or the receiver. Costs are lower and the best and most current data are used.

The interest in increasing access and cutting costs is driving the need and provides the justification to standardize in a way that agencies have not previously considered. At the time EIM was built, there were very few standards to adopt. As a result, Ecology took a market share approach to the "standards" we used. It was a combination of what was available, who had used it, and what we required. This approach has worked well, but as the needed standards for data exchange are considered and adopted by stewards around the nation, EIM will be modified.

EIM was designed to be an archival system—hold data for the long term. Despite this primary focus, there will be a need to incorporate data analysis tools in the future. This will pose a challenge, since any analytical tools and methods must be understood and accepted throughout the agency. Data that are available from a number of sources must be reviewed before use. The issue of appropriate data use has serious considerations when generic environmental analyses are planned. These factors will cause us to cautiously enter this area of interest.

Ultimately, the need to make information available to meet the growing regulatory, public, and private demands will force us all to reassess the way we manage our information resources. There is much to be gained from looking at what has been done by others in similar instances. EIM borrowed heavily from the good work of others. These learned lessons were tremendously helpful and allowed us to develop a successful product for our agency and the people we serve. We hope that our experience in building and implementing EIM is similarly helpful to others.