WEB 2.0 TO THE RESCUE FOR 'DIRTY DATA' IN THE ENVIRONMENTAL INDUSTRY

By Neno Duplancic, President and CEO, Locus Technologies

Data published by the *Environmental Business Journal* indicate that the U.S. environmental industry generated cumulative revenues of $265 billion dollars in 2005. The industry grew 5.4%, its second consecutive year of growth greater than 5% and, by many indicators, one of its best years ever.

Although such numbers would seem to be cause for celebration, some troubling trends persist within the industry. Notable among these is its failure to embrace the information management revolution that has had deep and far-reaching impacts on so many other business sectors. In particular, this failure to adopt the latest technologies for storing, distributing and managing information increases the costs and delays the cleanup of contaminated sites.

The nearby figure shows the amount of money that different business sectors spend per employee on information technology (IT). The top three are, as one would expect, finance, communications, and utilities. Who's at the bottom? Here we find the construction and resource industries (e.g., mining and forestry)—no surprise given the significant role that manual labor and heavy machinery play in these industries.

Joining these IT laggards, however, is the environmental industry. This is an industry that has developed many advanced cleanup technologies in recent years and is heavily populated by engineers and scientists. It is also an industry that our society expects to provide technological leadership in confronting the environmental challenges that we will increasingly face in the coming years. What, then, can account for its apparent aversion to spending money on IT, a reluctance that seemingly makes no sense given the vast amounts of information generated by its activities?

Most companies "own" their financial, human resource, customer relation, and other data. This information typically resides on computers located in the company's facilities, or it may be housed off site in data centers managed by an outside party. Regardless of which option is adopted, both are similar in that 1) information is stored in a consistent and organized manner in central databases, and 2) employees within the company have, to the extent that their privileges permit, continuous and unimpeded access to this data.

The manner in which companies with environmental liabilities manage and store their environmental information stands in marked contrast to the model that they have adopted for all their other key data. Why this is the case is a question worth examining in more detail.

The work involved in investigating and remediating contaminated sites is almost universally performed by outside consulting firms. Companies rarely "put all their eggs in one basket," choosing instead to apportion their environmental work among several to 10, 20, or even more consulting firms. The actual work at a particular site is generally managed and performed by the nearest local office of the firm that has been assigned to the site.

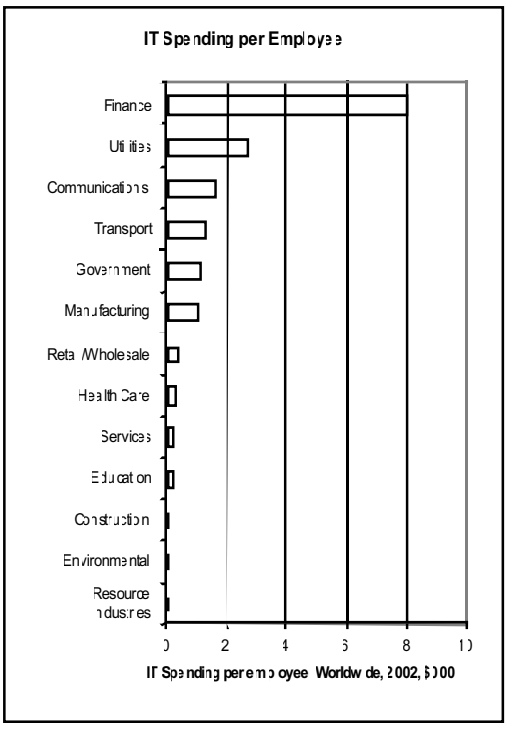
At larger production facilities, or at a Superfund site, the environmental work is likely to span 10, 20, or 30 years, while monitoring may continue even longer. Over this period of time, investigations are planned, samples collected, reports written, remedial designs created and, following agency approval, one or more remedies implemented. Not only is turnover in personnel commonplace, but owing to the rebidding of national contracts, the firm assigned to do the work typically changes multiple times over the life span of a remedial project. A quick look at the list of the top environmental consulting firms published yearly by *Engineering News Record* (ENR) magazine reveals that over 50% of the firms that occupied the top 100 spots only ten years ago are no longer on the list today. They have either merged with or been acquired by other companies, or they simply went out of business.

The investigation of a single large, potentially contaminated site often requires the collection of hundreds or even thousands of environmental samples. A typical sample may be tested for the presence of several hundreds of chemicals, and many locations may be sampled multiple times per year over the course of many years. The end result is an extraordinary amount of information—even just for one site. And of course, large companies with manufacturing and production facilities often have anywhere from a few to several hundred sites. Those that also have a retail component to their operations (e.g., oil companies) can have thousands of sites. Add to this list compliance and reporting data, engineering studies and real-time emissions monitoring, and the amount of environmental data becomes staggering and unmanageable by conventional databases and spreadsheets.

LOW PRIORITY FOR ENVIRO DATA

Given the magnitude and importance of this information, one would expect environmental data management to be a high priority in the overall strategy of any company subject to environmental laws and regulations. This is not so, however; instead, surveys reveal that a large portion of information sits in spreadsheets and home-built databases. In short, you have an entire industry, responsible for billions of dollars in liability, making decisions using tools that are not up to the task. Robust databases are standard tools in other industries, but for whatever reason, the environmental business has failed to fully embrace them.

What happens to the sampling and analytical data generated from the investigations of a company's sites? At most consulting firms, it is typically entered into spreadsheets, a commercial environmental database sys-



tem, or a home-grown Access, SQL Server, or Oracle database, with spreadsheets often being the most popular of these alternatives. If databases are used, they generally can be accessed from within a single office, or across offices, but rarely by outsiders, including employees of the companies who own or run the sites. Some consulting firms may impose a standard data management practice on all its offices; it is just as common, however, for local offices to be given the leeway to choose whatever digital means they want to store the data. In the worst of cases, the data from some field work may never be transformed into an electronic format, but instead is only kept in its original hard copy form.

Now, look at the situation from the point of view of a company with a large number of contaminated sites. Typically, none of the sampling and analytical data associated with the investigations of its sites is in its hands or easily accessible. Rather, it is scattered across the country, in multiple offices of each of its consultants, varying, as we noted, in its manner of storage from one consultant to the next and even one office to the next. In the event that one firm is replaced by another, all the data held by the outgoing consultant must be transferred to the incoming one, with each charging the client for the work. To put it simply, the situation is a mess.

To a significant extent, the same scenario pertains to other forms of environmental information, including reports, compliance data, auditing data, due-diligence data, real-time emission monitoring readings, site photos, and correspondence between the various parties involved in the investigation and remediation process. Consultants generally store these data in their offices, and while a client may be given copies of some of these documents and papers, the rest are inaccessible to it unless a request is submitted and copies of the desired items are made. Of course, billable hours are charged for every inquiry or transaction.

The lack of standards and consistency in information management practices among environmental vendors imposes a significant cost on the client's overall remediation budget. The fact that some firms may use spreadsheets, others their own databases, and still other various commercial applications may appear on the surface to be a benign practice, as each firm's office uses the tools it is most comfortable with.

Consider, however, the results of a study on environmental data management conducted by Christopher French as part of his master's thesis, *Application Of Six Sigma Methods For Improving The Analytical Data Management Process In Environmental Site Remediation* (Rutger's University, 2003). French examined the costs of data management at 19 sites owned by his firm over a three-year period (1999-2001). A range of tools and systems was used to upload and manage the data (N=657,834 records) at these sites. French divided the data into three categories: manual, hybrid, and automated electronic, based upon the primary means used to enter or upload data. His results showed that the per-record data management costs at these sites ranged from a low of $0.07 to a high of $8.57, with a mean of $1.73. I have no reason to believe these numbers are at all atypical.

How do these costs affect the bottom line? Let's suppose a company has 10-million analytical records stored in spreadsheets and databases in its various consultants' offices. Using French's numbers, the average cost to manage this data over a three-year period would be $1.73 million. However, if all the records had been managed using the system that had the fifth lowest costs ($0.30 per records), this number would drop to $300,000, or a savings of $1.43 million over a three-year period.

A key attribute of any system or process is the consistency of the outcome. French's data show that companies are bearing an unnecessary cost by ceding control of the management of their environmental data to their consultants and, as such, introducing substantial variability into this process.

INTERNET UNCHAINED

After the dot.com busts of the early 2000s, interest in the Internet has enjoyed a robust resurgence, owing in large part to sites like www.myspace.com, www.facebook.com, www.yahoo.com, www.ebay.com, www.flickr.com, www.craigslist.com, and—especially—www.google.com. What Google offers is a breathtakingly fast search and retrieval capability. Documents and other information that are stored on untold numbers of computers around the world can be pulled up and viewed in seconds. While you can search for friends on sites like Myspace and Facebook, and photos on Flickr, the main attraction of these sites is to keep one's contacts informed about one's activities and interests. In short, they permit and promote the electronic sharing of information.

Although restrictions on access exist, the fact is that most companies have successfully centralized and digitized their financial, human resource and customer relationship management (CRM) information, thus permitting their employees to retrieve and share records and documents pertaining to these areas. Yet the fact that environmental data is stored in inaccessible silo systems in the offices of its consultants precludes a company's employees from having the same ability to search and retrieve much of the environmental information on its sites. Why?

For consultants, there's no compelling driver to prompt them to change their current practices. In the environmental business, having satisfied clients is important, but so too is racking up billable hours. As such, technologies that would lower the costs of information management, processing, and retrieval—mainly a centralized, web-accessible database management system, but also other tools and practices such as the use of hand-held electronic devices, remote control and automation systems, wireless sensors, and electronic data validation—are often not adopted by firms because billable hours would be lost in making such a change. Firms also fear losing control of their projects. If data is made available over the web, other companies can easily be assigned to perform selected tasks that are dependent on this data.

It is difficult to know how much these factors affect the decisions and directions that a given firm takes. My feeling is that they are always present; I suspect, however, an even more important—and less sinister—reason for the failure to adopt new technologies. It concerns the level within consulting firms at which decisions about environmental data management are made. Because firms do not perceive themselves to be in the information management business, top-level management rarely gets involved in these matters. Instead, decisions are generally placed in the hands of project managers.

Place yourself in the position of a manager charged with completing a project on time and on budget, while at the same time keeping your employees as billable as possible. One option is to adopt the same data management tools and processes that have been successfully used on previous projects. These may not be the most efficient tools

and processes, but you and your staff are comfortable with them, and you are confident that you can meet your goals with them.

The alternative is to implement a new technology that potentially may save time and money. However, there are costs associated with the switch—costs that have to be borne solely by your project, given that the firm has not adopted any corporate-wide policy related to these matters. Despite your concerns, you may give the new technology a test run. The review, however, is placed in the hands of the very individuals whose workloads would be threatened by the new tool or technology. Inevitably, the end result is the maintenance of the status quo, and this may very well be the correct decision for the project. In summary, if a client does not lead, a consultant will not follow.

The client would appear to have every incentive to assume that leadership role. In the mainstream areas of its business, a company will invest money to adopt a new approach or technology if there is a strong likelihood of reducing overall costs in the future. Environmental cleanups are a necessary but unwelcome expense. For the business group overseeing these projects, the overwhelming push is to keep expenditures as low as possible. Yet new technologies and approaches are often treated with skepticism, even if they may deliver huge cost savings, for fear that they may increase costs temporarily.

Furthermore, an individual manager will typically have multiple sites in his or her portfolio. He or she has much to do—reading reports, managing budgets, tracking the progress at each site, interacting with agencies, and reporting to upper management. These individuals depend heavily on their consultants, and though they may be frustrated at times by the need to go through their consultants to get information on their sites, they have too many other things on their plate to make an issue of this problem.

Shouldn't these managers be concerned with the poor and inconsistent data management practices that many firms engage in? One needs to know that a problem exists in order to insist that it be corrected. On many occasions over the past few years, I have met with clients and potential clients that have many contaminated sites. As the meetings progress, I ask to hear more about the general condition of their sites, the monitoring and reporting commitments that exist at each, and the status of the records (both paper and electronic) that document site conditions. Invariably, the people I speak with are quite knowledgeable about current and past site conditions, and the remedial measures that have been undertaken at each site. They are not, however, very well informed about anything pertaining to record-keeping or information management practices, for the reason already described: Most of the records pertaining to their sites and all the sampling and analytical data are kept by the company's environmental consultants.

In short, being so immersed in their daily tasks, and with little knowledge of the internal practices of their consultants, site managers have little opportunity or basis to push for the adoption of tools and technologies that would not only increase productivity, but also make information more searchable and accessible for all parties involved.

THE SOFTWARE PROBLEM

The existence of so many stand-alone, in-house database applications shows that environmental firms have been designing and building tools for managing environmental data for a long time. These tools, built more often by engineers and geologists than IT professionals, are often poorly documented and supported. Over time, they invariably fail to keep up with the latest changes in technology.

There are several reasons for this failure. First, the main or sole users of these applications are the employees of the firms that developed them. Second, firms generally do not charge their clients for the use of these tools. As a result, the monies required to update these applications are an expense item in a company's budget. It's difficult to justify such expense at any time, regardless of the state of the economy, and these items are the first to be stricken during any downturn, especially when there is no external user that has to be appeased or any real competition.

The one upside to in-house databases is that there generally is some central control over development and modifications. This is not the case for spreadsheet-based applications. Today, spreadsheets are so easy to use and ubiquitous that they have sprouted like weeds throughout most companies. Often they hold important technical and financial data. But what if Mary's last-quarter analytical data spreadsheet differs from Tom's and has faulty data or a unit conversion error? What if Tom's spreadsheet is saved only on his C: drive? How can it possibly benefit a client who must weigh how best to spend its remediation budget when its data is stored in dozens of far-flung spreadsheets rather than a comprehensive database?

Various studies report that 47% to 64% of companies use stand-alone spreadsheets for planning and budgeting. Critics of this practice say spreadsheets—invented as a personal productivity tool—aren't well suited to collaboration, data quality or regulatory compliance. "Besides being extremely unwieldy for processes involving large volumes of data and multiple users, spreadsheets often contain substantial, material errors, according to academic research," notes Paul Hamerman, a Forrester Research analyst. Companies are just starting to look at the problems caused by spreadsheet proliferation.

Questioning the desirability of spreadsheets, given the widespread acceptance they have enjoyed over the past two decades, is akin to questioning the presence of turkey on the Thanksgiving menu. However, for a modern corporation looking to consolidate its planning, environmental liability management, and reporting, spreadsheets pose challenges not dreamed of when they first began popping up on PCs across the land. Storing environmental data in a hodgepodge of spreadsheets makes it hard to maintain one version of the truth, which one must do in order to comply with the law. In particular, the Sarbanes-Oxley Act of 2002 requires companies to maintain a good audit trail; generating such a trail is difficult to do with spreadsheets.

A BETTER WAY

Moving data out of spreadsheets and into a robust database is an absolute must for companies that want to control their costs and spend their money wisely. The solution seems obvious: get all the information about sites out of paper files, spreadsheets, and stand-alone or inaccessible databases and into an electronic repository in a structured and formatted form that—and this is the crucial point—any project participant can access, preferably from the web, at any time and from any place. In other words, the solution is not merely to use computers, but to use the web to link the parties involved in a cleanup, and this includes not only site owners and their consultants but also regulators, laboratories, and insurers, thus making them, in current jargon, "interoperable." This may

be obvious, but today it is also a very distant goal.

What would the ideal IT architecture for the environmental industry look like? It would start with wireless data entry by technicians in the field and wireless sensors where feasible. Labs would upload the results of analytical testing directly from their instrumentation and LIMS systems into the web-based database. During the upload process, any necessary error checking and data validation would take place automatically. Consultants would review these uploads and put their stamp of approval on the data before it becomes part of the permanent database. Air monitoring devices and sensors would automatically upload their measurements into the same system.

Behind the scenes, all data would be formatted and stored according to recognized and standard protocols. Contrary to widespread concerns, this does not require a single central repository for all data or any particular hardware architecture. Instead, it relies on common software protocols and formats so that individual computer applications can find and talk to one another across the Internet. The good news is that the most of these standards, such as XML, SOAP, AJAX, and WSDL, already exist and are used by many industries. Others, such as SEDD, EDF, CROMERR, or EDD (spelling them out makes them sound no less obscure), are unique to the environmental industry and govern data interchange among laboratories, consultants, clients and regulatory agencies. On top of these, there needs to be hacker-proof layers of authentication and password protection so that only the right people can access critical or sensitive information.

There is still some work to do to refine these technologies, but the basic building blocks are already readily available and implemented by a few progressive companies and regulatory agencies. The problems that this changed approach would address are many. First, data would be entered or uploaded just once, preferably electronically. Second, data transfer costs would drop, and data quality would improve. No longer would there be a need to transfer data whenever one consulting firm is replaced by another or to maintain multiple databases that must be kept in sync.

Third, the significant amounts of time that engineers, managers and scientists now spend determining whether a particular report is correct or looking up information on a site would dramatically decline. Fourth, by having their data in a consistent electronic format, companies would be in a better position to comply with the emerging demand to upload information on their sites to state or federal agencies and organizations. Several progressive states have already imposed electronic deliverable standards (e.g. California and New Jersey), and EPA is working on its own standards based on XML technology. Last and most significantly, site owners would assume possession of their data and, as such, finally gain ready access to information about their own sites. This would seem particularly beneficial to public companies attempting to comply with Sarbanes-Oxley.

WEB 2.0

"Web 2.0" is a term often applied to the perceived, ongoing transition of the World Wide Web from a collection of websites to a full-fledged computing platform serving up web applications to end users. Tim O'Reilly, a publisher of computer books and a Silicon Valley guru, coined the phrase in 2003. Ultimately, Web 2.0 services are expected to replace desktop computing applications for many purposes. Many Web 2.0 applications are delivered through an on-demand model that is also referred to as "Software as a Service" (SaaS). An integral part of this model is web portal technology. A web portal is a general delivery point for applications and services that are accessed from a web browser. Portals allow applications and services from different sources to be brought together into a seamless user experience, which is exactly the technology needed by the environmental industry in order for information to be accessed from a single location.

Taking advantage of Web 2.0 technologies, Locus's ePortal demonstrates the facility provided by web portal technology. It provides a single platform from which various services and applications, such as analytical data management (EIM), document management (eLibrary), site information management (eSite), collaboration and knowledge management, are offered. Clients pay subscription fees based on the computing power and capacity they need. The client's data and applications are hosted on Locus's servers. Clients and their consultants simply log in to the Internet to access information or process transactions.

Clients with a portfolio of environmentally impacted sites that have developed certain applications in house, but lack the resources to maintain, integrate or expand them, have an option to integrate their applications with "portlets" within the ePortal framework. For example, a customer may have developed a client server application for financial project management, an asbestos liability database, and a database for waste tracking, and may also have purchased a subscription to a third-party service for material safety data sheet (MSDS) management. This customer would be able to integrate these existing applications into ePortal and, at the same time, subscribe to Locus's eSite and EIM applications. An end user with the proper privileges would need to log in only once to access all these applications inside the portal. In short, portal technology allows companies to leverage their existing data and application assets, repositioning them as on-demand services, while at the same time removing the barriers to searching and sharing that are posed by multiple inaccessible silo systems. Using this approach, gigabytes of information sitting in obscure databases can be pushed up and mobilized to be used or integrated with various applications.

The on-demand platform is a relatively new way of delivering software applications over Internet. Clients who subscribe to on-demand software benefit by reducing their administrative and support costs. One of the key applications that is available within ePortal is EIM, Locus' flagship product for analytical data management. This application has been offered as on-demand software since 2000 and today holds over 45-million records for over 5,000 sites. However, even before EIM was released, Locus realized that it alone was not going to solve the industry's environmental data management problems. Instead, we viewed the system as the first component of a much broader platform that would provide the requisite infrastructure to integrate all environmentally related applications into a seamless portal, thus providing companies with the ability to access and manage their entire project, compliance, and other data from a central location.

Web portal technology allows environmental managers to accomplish what until recently had seemed "mission impossible"—that is, to provide access to their companies' vast amounts of environmental information over the web in a structured manner with very little up-front investment. Actually, the use of the word "access" in the prior sentence does not do justice to the portal's capabilities. In fact, users are able to perform

intelligent searches and mine for information across multiple data sources and types, from analytical data and documents to Google maps and collaboration sites.

Companies with environmental liabilities are not the only ones that stand to benefit from portal technology. Environmental consulting firms that adopt such a tool can bring order and coherence to their data management practices, and thus optimize their business practices and enhance their productivity. Through the portal they could offer their clients various specialized applications concerned with health and safety monitoring and statistics, compliance, due diligence, auditing, permit tracking, Title V reporting, MSDS reporting, waste tracking and reporting, SARA Title III, Section 313 (Toxic Release Inventory) reporting, environmental compliance tracking and reporting, emissions allowance management, monitoring laboratory performance, and much more.

THE FUTURE: WEB MASHUPS

Another popular trend in the Web 2.0 world is the emergence of "mashups." Mashup services typically combine different web-based applications into integrated, typically more powerful, end products. A mashup is a website or application that seamlessly combines content from more than one source into an integrated web experience.

One popular mashup is the integration of Google Maps with various sorts of environmental data. As this technology evolves, it offers the real promise to display environmental information from multiple sources in a meaningful manner. Mashups provide an easy solution to integration problems that previously seemed like a daunting amount of work. Ultimately, mashups will allow users to create, share, and combine data components in ways that best help them understand environmental conditions and remediation progress at their sites.

As already noted, economics are the principal driving force behind the development of web-based portals. Like the on-demand computing revolution, the introduction of the enterprise environmental software portal stands to save corporations and businesses of all sizes significant sums that would otherwise be spent on outdated tools and highly labor intensive activities. The benefits of a portal approach go beyond simple cost avoidance, although such benefits may be hard to quantify at the outset. The true economic benefit of an enterprise environmental solution will come from further integrating business processes, ultimately leading to the retirement of the many islands of information and spreadsheet-ware that exist in the offices of consulting firms.

Estimating how much the portal model could save industry depends on the quality of legacy systems that are currently in use. Some of Locus' existing customers report breaking even within months after initial deployment. For additional insight into these matters, it is useful to draw some comparisons with another information management intensive industry—health care—where the potential savings have been quantified.

In January 2005, the journal *Health Affairs* reported that a fully interoperable network of electronic health records would yield $77.8 billion a year in net benefits, or 5% of America's annual health-care spending. This includes savings from faster referrals between doctors, fewer delays in ordering tests and getting analytical results, fewer errors in oral or hand-written reporting, fewer redundant tests, and automatic ordering and re-fills of drugs. It does not include, however, perhaps the biggest potential benefit: better statistics that would allow faster recognition of disease outbreaks (such as SARS or avian flu).

I would argue that similar, if not greater, savings could be achieved by the environmental industry by automating more processes and building a web-based repository for all data. A 5% cost decrease applied to the $265-billion industry produces a staggering $13.25-billion in savings, but I actually believe that even this number is low. I wouldn't be surprised if the actual number is 10% or higher. This beliefs stem from the results of a study conducted by *Chemical Engineering* magazine in 1994. This study concluded that almost half of the consulting time in the United States is spent looking for existing information that cannot be found, or on procedures that duplicate one another or that are inappropriate. Even if the situation has improved somewhat in the past decade, much time is still wasted on unproductive activities that could be greatly reduced by the implementation of the web-centric model that I have described.

As in the health industry, the savings would not stop there. Perhaps one of the biggest potential benefits would be better data, which would allow for faster identification of problems and thus earlier actions that would either prevent the release of chemicals in the first place, or reduce their impacts on man and the environment.

At some point, I am optimistic that the short shrift given by the environmental industry to information management will change. As more contaminated waste sites, after being cleaned up (or better to say, "buttoned up," as very few sites are really cleaned up these days), enter the monitoring phase, information management costs, together with those associated with sample collection and analysis and data evaluation and reporting, are expected to consume over half of the expected annual budget for sites in this phase. Considering that the monitoring phase often lasts for decades and that an estimated two- to five-million contaminated sites exist in the United States alone, it is clear that both industry and government face substantial costs in the years ahead.

Because most of these charges will be related to information management, companies and agencies will be forced to confront and evaluate their practices in this area, and in so doing, come to realize the benefits to be gained from adopting more cost-effective tools and systems to acquire and manage their environmental data. I predict that the current preference to "button up instead of clean" sites will probably speed up the adoption of new real-time technologies. Many of these do not require big investments. Instead, they make existing computer systems more effective, promising a quick return.

Investments in these and other new technologies could be financed from savings gained by changes in data management practices and the deployment of web technologies. Clients that have spent millions on cleaning their sites without any guarantee that their bottom line would benefit are desperate for such savings.

New approaches are sometimes slow to win acceptance, as a multitude of companies in the IT field will attest. However, as organizations come to understand the opportunities and benefits of on-demand computing, of having a single access point to their data, and of being able to integrate the best of existing and new applications, we expect a seismic shift to eventually occur in the manner in which firms manage their environmental data. As word spreads and more case studies become available, few will be able to ignore the combination of improved data quality and access and lower costs that web portal technology can deliver. ■