

Environmental Data Warehouse Integrates Analytical Data, GIS, and the Web

Jeff N. Stovall, Ph.D., E.I.T. (BWXT Pantex, L.L.C.), David Korns (Science Applications International Corporation), and Clarence Lamb (BWXT Pantex, L.L.C.)

Abstract

An environmental data warehouse (EDW), developed for USDOE/NNSA's Pantex Plant, brings environmental analytical data from nearly 2 million legacy environmental records and spatial information about the site together into a single database accessible through an intranet web page or ArcGIS software. The web page includes a tabular analytical data query interface, a GIS viewer to display spatial site information and analytical data, online access to well lithologic and geophysical logs and site photographs, and a simple data trending tool. ArcSDE provides linked GIS and analytical data for complex data analysis and cartography by individual ArcGIS users throughout the department.

INTRODUCTION

The Pantex Plant is a USDOE/NNSA-owned, contractor-operated facility, currently managed by BWXT Pantex, L.L.C. Several areas at the Pantex Plant are undergoing RCRA facility investigations (RFI) because past operations at the plant, dating back to World War II, resulted in contamination of environmental media at numerous sites across the plant. In addition, the waste permits for the plant contain a number of environmental monitoring and reporting requirements.

The Environmental Remediation Services (ERS) and Regulatory Compliance Departments (RCD), responsible for the investigation, cleanup, and

closure of those sites, must manage an ever-growing set of environmental information that requires a systematic approach for retrieval and display of analytical data, maps, logs, and photographs. The information includes analytical sample results from various environmental media, environmental restoration reports and supporting documents, maps, facility drawings, well and boring logs, and photography for sites currently and formerly managed by Pantex. The ERS project managers and scientists require daily access to recent sampling data collected from almost 7,000 locations, including over 150 groundwater and soil gas investigation wells, soil borings, playa lakes, solid waste management units (SWMUs), and other environmental sample sites. To meet this need, an intranet-accessible web site was designed to provide managers and scientists access to the data to support quarterly monitoring and reporting and to provide RFI data for fate and transport modeling and risk assessment.

SYSTEM COMPONENTS

The environmental data warehouse (EDW) consists of a number of linked components as shown in Figure 1. The components include the Integrated Environmental Database (IED), the analytical data warehouse (ADW), the geodatabase, an intranet web site, and a set of queries and batch files for updating the data.

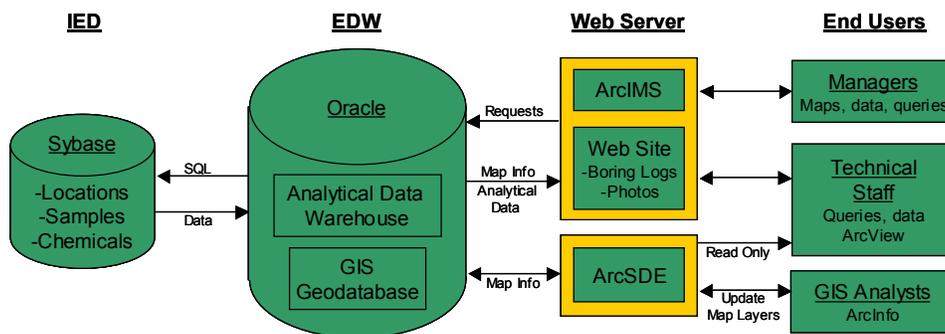


Figure 1. Environmental Data Warehouse and GIS

Integrated Environmental Database

Analytical data from environmental samples collected at Pantex are stored in an Integrated Environmental Database (IED). Data stored in the IED includes chemical sampling data from soil, groundwater, surface water, air, and other environmental media, geotechnical data collected from subsurface samples, and water level measurements from wells. The IED contains nearly 2 million sample records collected from almost 7,000 locations.

This database was developed in Sybase in 1996 and represents the second iteration of an environmental database for the Pantex Plant. New data is added to the IED daily from either electronic data deliverables provided by the contract laboratories or by loading data packages received from contractors. In addition, ERS staff continually check data stored in the IED for quality purposes, and corrections are made when appropriate.

Analytical Data Warehouse

The ADW is a separate Oracle database that stores data queried from the IED tables. The data warehouse was developed to simplify the many IED tables into a common standardized repository for both web based queries and desktop and web-based GIS requirements. Additionally, as new, non-IED environmental data sources become available, they can be integrated into the ADW without making changes to any of the existing applications. Data stored in the ADW has been verified against a set of rules to ensure data validation and consistency. Analytical results are linked to a master station file that allows spatial representation of the data.

Because the data in the IED is updated daily with new records and corrections, the ADW must be regularly updated. To begin the update process, a set of six tables are copied from the IED to the ADW. Because the tables must be copied from the Sybase IED to the Oracle ADW, the Datastage software is used to copy the tables.

In the second step, a SQL script is used to build the tables of the ADW. This SQL script performs four major functions. The first part of the script takes

data from the IED tables to the ADW and creates lookup tables for web use. The second part of the script generates data to be used in tables to show the relationship of a sample result to the applicable risk standard. The third part of the script creates the information used for trending of groundwater detections on the web page. The final part of the script creates a table that identifies data that failed to pass the validation test and the reason the data failed. This table can later be examined to correct information in the IED. After the execution of the script is completed, the information stored in the ADW matches the information stored in the copied IED tables.

Geodatabase

GIS datasets from the Pantex environmental GIS were integrated into the Oracle data warehouse using ESRI ArcSDE software. ArcSDE enables ArcGIS and ArcIMS applications to store, manage, and access spatial data directly in an Oracle database. The ArcSDE geodatabase provides a number of beneficial features, including:

- centralized management of all GIS data;
- layer versioning;
- multi-user access from ArcGIS and ArcIMS;
- dynamic linking of features to the analytical data warehouse; and
- raster data storage.

The first three features allow the GIS staff to maintain all GIS information in a single location. As GIS staff update the Pantex ArcSDE geodatabase from their desktops, all web and ArcGIS users see the most up-to-date map features. In addition, Pantex Information Technology Management (ITM) personnel maintain and backup the database, allowing the GIS staff to focus on maintaining the data. Because the geodatabase is centralized, it can be shared with other organizations at the Plant as they begin to take advantage of GIS applications.

Dynamic linking of features to the ADW allows sample results for groundwater, soil, surface water, and wastewater samples to be converted into ArcSDE point features from the coordinate information in the IED. When a sample location is queried in a GIS application, all of the sampling

results for that location are returned. This feature provides automatic spatial representation of sampling results.

The dynamically-linked ArcSDE point features are updated via a set of three batch files. The first batch file runs the sdetable command to delete the existing feature-linked tables. The second batch file runs a custom ArcObjects program that creates an ArcSDE point feature class from the IED sample location table. The third batch file runs the sdetable command to join the ArcSDE point features to the sampling results. A separate SDE table is created for each sample media type.

Intranet Web Site

The ERS intranet web site allows management and technical staff to interact with the information in the EDW. The main web page provides links to the five functions currently available: analytical data query, GIS viewer, well and boring logs, photos, and temporal data plots. The web page does not require any additional software other than a web browser.

Analytical Data Query

The analytical data query page contains a query tool that allows the user to pick sample locations, sample media, individual analytes, analyte groups, sample depths, and sample dates. The user also selects whether to return results with detections, non-detections, rejected samples, or any combination. The results of the query are returned as a table in the web browser or as a Microsoft Excel spreadsheet file that is saved to the user's computer. The query parameters that the user picks from are generated from the lookup tables created by the SQL update scripts. The pick lists are therefore

dynamic based on the data in the IED. Figure 2 is a screenshot of the analytical data query page.

The screenshot displays the Analytical Data Query Page interface. It features several dropdown menus and selection options:

- LOCATIONS:** A dropdown menu with options: 001, 001-A, 002, 003, 004.
- ANALYTES:** A dropdown menu with options: 1,1,1,2-TETRACHLOROETHANE, 1,1,1-TRICHLOROETHANE, 1,1,2-TETRACHLOROETHANE, 1,1,2-TRICHLOROETHANE, 1,1-DICHLOROETHANE.
- MEDIA:** A dropdown menu with options: Air, Fauna, Groundwater, Other, Quality Control.
- FRACTION:** A dropdown menu with options: Explosives, Herbicides, Metals, Miscellaneous, PCBs.
- DETECTS:** A dropdown menu with options: NON-DETECTS, REJECTED.
- DEPTHS:** STARTING and ENDING dropdown menus, both currently set to 0.
- SAMPLE DATES:** STARTING and ENDING dropdown menus, currently set to 01-02-1992 and 02-05-2003 respectively.
- ORDER BY:** Primary and Secondary dropdown menus, currently set to Location and Analyte respectively.

At the bottom, there are two buttons: "Submit Query to View Data in Table" and "Submit Query to Excel Spreadsheet".

Figure 2. Analytical Data Query Page

The analytical data query also contains a second query page that allows the user to pick sample location, analyte, and a reference standard for comparison, and the query returns all sample results the exceed the specified standard. The standard can be the groundwater maximum contaminant level (MCL), the residential risk reduction standard, or the industrial risk reduction standard.

GIS Viewer

The GIS viewer is a customized ArcIMS web page that is linked to the EDW. The GIS viewer displays a base map of the site with a number of map layers that the user can turn on or off. The viewer displays the sampling locations on the map. The viewer has the basic GIS tools available such as zoom, pan, query, and select, and includes an overview map of the site. Figure 3 is a screenshot of the GIS viewer window zoomed in on a small area of the plant. The "Official Use Only" notifications appear because some of the data that can be displayed by the viewer carries this designation.

The map layers that can be displayed include all sample locations, sample locations by media type, various map features such as roads, fences, property boundaries, topographic contours, and building locations, and environmental units such as SWMUs and waste management groupings (WMGs). The user can also display a digital orthophoto of the site on the map. The viewer dynamically displays labels for sampling locations, buildings, SWMUs, and WMGs based on the scale of the display.

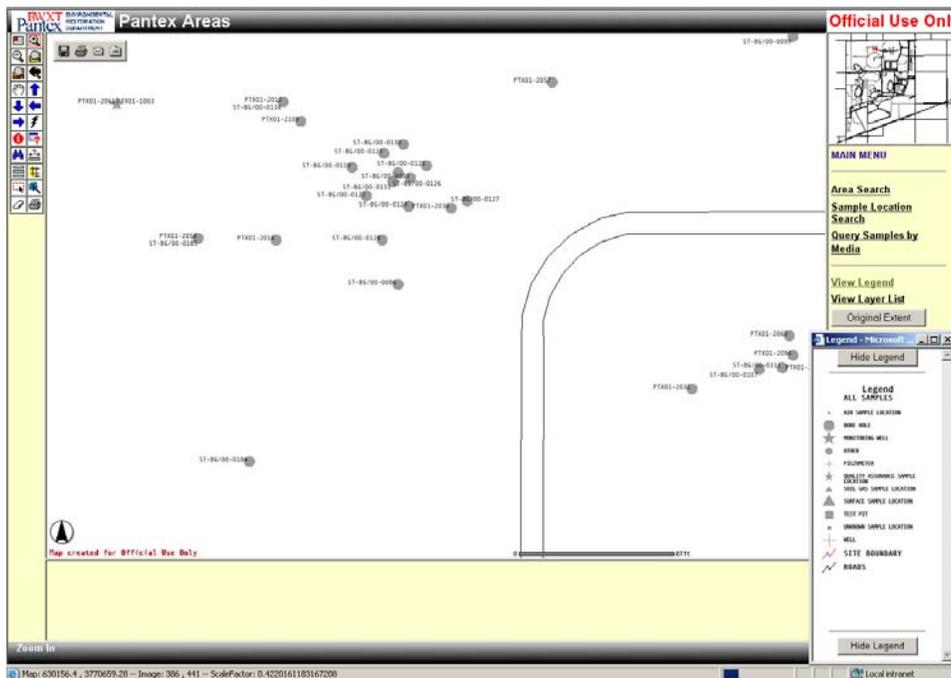


Figure 3. GIS Viewer Window

The GIS viewer also contains a set of powerful query tools. The first tool allows the user to search for a specific area of the plant by SWMU, WMG, RFI report, or building number. The user selects the search area from a pick list that is dynamically generated from the database. The second tool is a sample location search that allows the user to select a sample location from the pick list and displays that location on the map. The third tool allows the user to query samples by media. This tool is similar to the analytical data query page, except that the sample results are returned on the map. The user selects a media type, then an analyte. Depending on the type of query selected, the map displays either sample locations with detections of the selected analyte, sample locations with non-detections of the selected analyte, or sample locations with detections of the selected analyte exceeding a user-specified concentration.

Well and Bore Logs

The well and bore log viewer allows the user to pick a well or boring location from a pick list, and an Adobe Acrobat PDF containing the lithologic and geophysical logs is displayed in a separate window. The pick list of locations is dynamically generated

from the locations in the IED, but the file location reference must be manually added to a table in the EDW as logs are added.

Photos

The site photo viewer is similar to the log viewer. The user pick a location from

a pick list, and a web page of photographs from that site is displayed. The pick list of locations is generated from a table in the EDW, but the locations with photos must be updated manually when new photos are available. The page that displays the photos is dynamically generated based on the number of photos stored in that location's folder. Thus, if more photos of a particular location are added, the photos are automatically displayed when the location is selected.

Temporal Plots

The temporal plot viewer uses Visual Mining's NetCharts software. The user selects a location and analyte from a dynamically-generated pick list, and the software generates a temporal plot of the data including both detections and non-detections. Figure 4 is a screenshot of the temporal plot window.

The data used for the graph is created as part of the SQL script that creates the EDW tables, so the graph always displays the most current data.

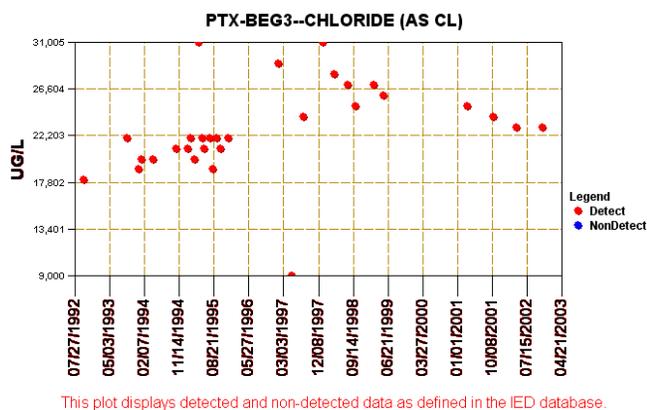


Figure 4. Temporal Plot Window

Operating Environment

The environmental data warehouse was designed to take advantage of leading GIS and database technologies and to meet the requirements of Pantex ITM. The existing IED is a Sybase database; however, Oracle 8i was selected for the EDW because it is now the standard database system at Pantex Plant. The ESRI ArcGIS system was selected to provide GIS functionality, and the system uses ArcSDE, ArcInfo, and ArcView at the 8.2 release and ArcIMS at the 4.0 release. The intranet web site was created using Java Server Pages and operated under the IBM WebSphere web server. All of the server applications run on AIX UNIX servers.

A fully-functioning copy of the system has been set up on a laptop computer, allowing access to the environmental data and GIS for offsite meetings and presentations.

SUMMARY AND CONCLUSIONS

An environmental data warehouse was developed to bring environmental analytical data and spatial information about the site together into a single database accessible through an intranet web page or ArcGIS software. The EDW provides access to useful environmental data through an easy-to-use

intranet web site that allows users to query analytical data, display spatial site information and analytical data together in a GIS viewer, view well lithologic and geophysical logs and site photographs, and trend temporal data. Using ESRI ArcSDE, the system also provides linked GIS and analytical data for complex data analysis and cartography by individual ArcGIS users throughout the department.

The EDW has proven to be useful for both managers and technical staff. While the system does not replace SQL-based database query tools, it allows for more widespread access to the most commonly used environmental data. Because the system also allows data to be viewed in its spatial context, project managers can now assess site data without having to wait for others to prepare data visualizations. The ArcSDE geodatabase has also made GIS readily available to other organizations at the site.

AUTHOR CONTACT INFORMATION

Jeff N. Stovall, Ph.D., E.I.T.
 BWXT Pantex, L.L.C.
 Environmental Remediation Services
 P.O. Box 30020
 Amarillo, TX 79120
 jstovall@pantex.com

David Korns
 Science Applications International Corporation
 4900 Blazier Parkway
 Dublin, OH 43017
 kornsd@saic.com

Clarence Lamb
 BWXT Pantex, L.L.C.
 Environmental Remediation Services
 P.O. Box 30020
 Amarillo, TX 79120
 cllamb@pantex.com