

Creating an Environmental Database to Facilitate Groundwater Contamination Analysis

Jimmy Bowden
Daniel B. Stephens and Associates

Abstract

Assigning responsibility for the cleanup of groundwater contamination requires the analysis of large amounts of complicated data. Daniel B. Stephens and Associates team of geologists and hydrologists are charged with investigating various potential responsible parties and determining the sources of groundwater contamination.

GIS has become a powerful tool to allow volumes of data to be quickly and accurately displayed. ArcView GIS is used to create base maps consisting of building locations, underground storage tanks, well locations, and other features. A Microsoft Access database is developed for the storage of environmental data consisting of groundwater, soil, and soil gas chemistry analysis, groundwater depth, and sampling locations. From this database, queries are preformed and linked to ArcView for spatial analysis. Custom visualization tools were developed to allow for complicated analysis of the data.

Introduction

Analyzing large amounts of data groundwater and soil contaminant data generated from numerous site investigations can be an overwhelming task. Especially when that data is in a variety of formats, produced by numerous consultants, and have questionable levels of accuracy. The use of GIS software like ArcView and a Microsoft Access database can make the visualization of data much simpler. Before visualization and analysis can be done an analytical database and GIS basemap must be developed.

Microsoft Access is the preferred format for storing analytical data because of its simplicity and widespread use. ArcView is used to compile, edit, and display GIS feature and raster data. ArcView's ability to retrieve data from an Access database is a benefit, in that the GIS data and database tables can be displayed together quickly and efficiently.

Because data may exist in various digital or hardcopy formats a database must be developed. This database must be robust enough to handle various types of data but simple enough for novice Access users to query data. It is also important that the data is accurate and has source information. Therefore quality assurance/quality control methods are established and followed.

A basemap is created in ArcView combining aerial photography, shapefiles, CAD drawings, and digitizing off of existing figures. Data from the analytical database is queried and displayed on this basemap using a variety of techniques that allow for hydrogeological analysis.

Developing a Database

The first step in developing a database is understanding the types of data that will be used and how they relate to each other. There are four basic types of data that are created when determining groundwater contamination. They are field or sampling data, analytical data, geographic data, and interpretive data. Field data is data that is collected at the site, such as groundwater level, temperature, or samples taken that are shipped to a lab for processing. A laboratory will then analyze the sample for various parameters and produce analytical data. The location where the samples are gathered and the surrounding physical features at the site are the geographic data. The final type of data is interpretive data or data that is created by analyzing the other types of data. These types data can be represented by individually but are most effective when combined by the relationships that exist between the data types.

Creating an Access database is often the most difficult and time-consuming aspect in preparing for analysis. Numerous challenges arise when developing a database, including incomplete, inconsistent, and inaccurate data. A contaminated groundwater area may include several potentially responsible sites, who each have environmental consultants performing investigations and using different reporting methods. Some of these

investigations were done decades ago and may have been sampled and analyzed using outdated methods and might not include accurate maps. Other investigations are done in the last few months these investigations may include the use of advanced sampling methods and GPS gathered locations.

The data presented in these investigation reports can come in many formats such as databases, tabular data files, summary tables, and laboratory analysis sheets. The advantages and drawbacks of these types of data inputs are outlined in Table 1. Data that is already in digital format usually requires extensive formatting and error checking before being merged with existing data. Often times data received in digital formats uses different naming conventions than existing data, these items need to be matched up before the import process. Another issue is that of duplicate data, which is handled quite well by using Access primary keys. However it is necessary to determine if the duplicates that reported by Access are truly duplicated or just errors or similar records of data.

Table 1 Advantages and Drawbacks of Various Analytical Data Sources

Data source	Advantages	Drawbacks
Existing database, Access, dbase, other	Already digital, easily imported into existing database.	May not be sourced or quality checked. Need another source to check validity.
Tabular data file, excel, delimited text, laboratory electronic data delivery (edd) files	Already digital, good for viewing and printing.	May have multiple column headers. Requires formatting and/or running a macro before it can be imported in the database. May not be sourced or quality checked.
Summary tables	Simple to read and concise. Can be scanned in with OCR.	Not digital. Is only a summary of data therefore other data may exist and not be represented.
Laboratory Analytical Sheets	Typically the most complete and accurate source of data.	Not digital. Must be hand entered in most cases. May not be organized in a fashion that is easy to read by data entry personnel.

Other data may not exist digitally and only are available in hardcopy reports. These data will need to be hand entered or scanned in using optical character recognition (OCR). Hand entering data is time consuming and can be inconsistent if multiple people are entering the data. OCR is only good for large tables of data that are clear and consistent. In all cases data needs to be thoroughly checked for accuracy against the most reliable source available.

Once all the available data has been imported or entered and gone through the qa/qc process the database is ready to be queried and connected to ArcView for display on the basemap.

Developing a Basemap

While the Access database is being developed a GIS basemap is also being created in ArcView. The first step in developing a database is to access existing and available digital data. Internet GIS data sites are have digital raster graphics (DRG), digital ortho quarter quads (DOQQ), digital elevation models (DEM), shapefiles, and aerial photographs available for download. Because a site or group of sites are often viewed at scales less than 3,600 these data may not be scaled appropriately for the final basemap, but it is a good starting point.

Other existing data can be shapefiles or CAD drawings that were created by consultants performing the site investigations. Often older CAD files were developed in internal

coordinates systems that do not reflect any projection. These files may be referenced using a world file two-point transformation if common points between the CAD file and other existing data exist. Many shapefiles created in ArcView 3.2 and earlier do not indicate what projection the data is in. In this case testing against data with known projection is the best way of determining the projection.

In most cases basic site features and sampling locations will not have been in any of the above data sources and must be created for consultant report figures. The initial step in this process is to go through numerous documents to find maps and figures that represent the sites. Features such as buildings, property boundaries, aboveground and underground storage tanks, container storage areas, sampling points, and roadways are gotten from these figures. Changing conditions at a site, receiving better information, or differing collection methods, can cause large discrepancies in feature location.

Sampling points are the most difficult type of features to obtain accurately. Sampling points are a single point where data is collected. These can be groundwater monitoring wells, soil borings, soil gas probes, or soil excavation samples. The same monitoring well may be represented by different locations on separate figures. It is also possible that soil excavation samples are only on hand drawn figures where collection locations were approximated. Even in the best-case scenario problems can arise, such as when a table of well coordinates is presented as a table in a report. The data may be in decimal degrees with a limited number of decimal precision or in a coordinate system that is not defined. Another potential problem with sampling locations is entering different names in the GIS data than in the analytical database so that the data doesn't join. The best way to assure that data does join is to use a consistent naming convention even if it means editing the name given with the data and on the figure.

A figure that was hand drawn twenty years ago may not look the same as a figure created from a survey last month. The newer figure is obviously the preferred data source, however the older figure may have some information that is not and cannot be presented on any other figure. For example a leaking drum that is next to a building in the hand drawn figure may have been removed from the site prior to any other investigation being done. Some features are underground which can make it hard to determine their exact location. In this case it is common to see the same consultant edit the location or shape of these features to represent newer more accurate data. Another issue is that adjacent sites have diverging interests and may be unwilling or unable to share accurate information. Because of these issues it is good to determine best-fit figures that may be different for each set of features.

Once the best fit has been selected features are then digitized either by setting up a digitizing board or scanning and geo-referencing those figures. With the easy to use image geo-referencing tool in ArcView 8.x the heads down digitizer has less appeal. Once the features are digitized a site layout map is created using symbology and colors that highlight the most important features. This site map is then presented to the project technical lead to perform qa/qc on the work. Since this is the basis for all future figures the project manager should approve it before creating other figures based on it.

Visualization

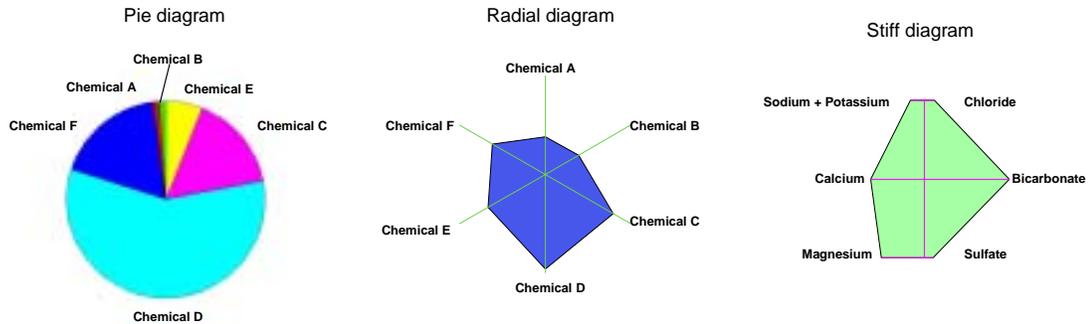
The analytical data and the basemap are combined in ArcView and visualizing data can begin. There are four dimensions that are used in analyzing groundwater contamination data; they are x and y coordinate, depth or elevation the z component, and time.

The simplest method for displaying data is having points representing sampling locations and labeling values next to the point. Typically these values are a sampling location designation, a specific chemical analysis result or a groundwater level and the date or the depth the sample was taken. A simple variation on this idea is to use colored or graduated symbols to indicate smaller to larger values.

Three other methods of visualization are pie, radial, or stiff diagrams. See figure 1 for an example of these diagrams. To make these types of diagrams a query in crosstab format is necessary. Crosstab format is where different chemicals are reported as fields based on a grouping of sampling point and date. A pie diagram is created using the ArcView chart symbology option, with different colors denoting different chemicals and values represented

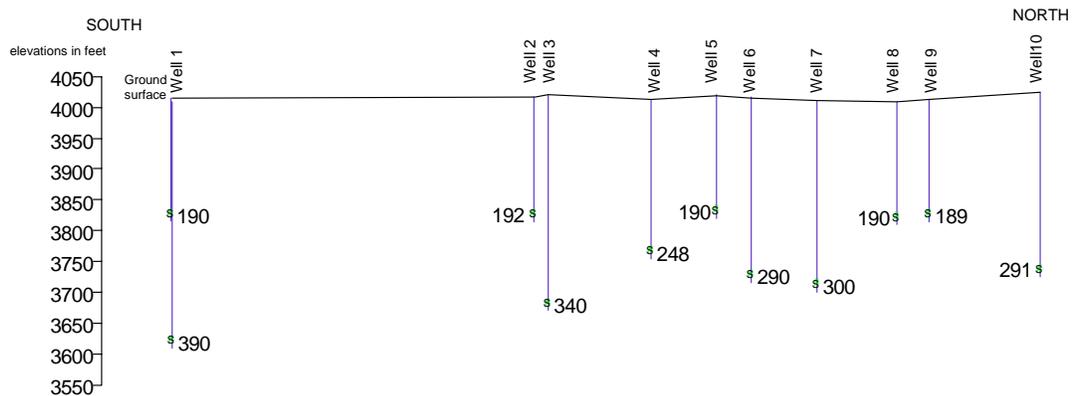
by percentages of the whole pie. A radial diagram is similar except that values are represented along varied directional axis' that designate different chemicals. The values are then connected as a polygon thereby giving a distinct shape that allows for seeing chemical signatures. Radial diagrams are created in ArcView 3.2 by a custom script that was developed in house. An ArcView 3.2 avenue script also generates stiff diagrams. They are similar to radial diagrams but use six or eight horizontal axis' using specified ion, cations, and metals. Again a polygon is created that has the same shape in regions where groundwater conditions are similar.

Figure 1



Another way to view data is as a simple cross section, where sampling points along a line are drawn with a sequential x coordinate and sampling depth as the y coordinate. Again Avenue in ArcView 3.2 is used to create a cross section like the one shown in figure 2. Once a cross section is created the above display methods can be used to see chemical data system by joining data to the points created by the cross section routine.

Figure 2



Conclusions

The use of database management systems like Microsoft Access allows for large amounts of data to be organized in a standard format. These data can come from a variety of sources and formats, each with potential advantages and drawbacks. The use of a database to store groundwater data allows for various types of users to query the data without the need for extensive GIS or database training.

Creating and integrating GIS feature and raster data into ArcView is a common task. The ability to use GIS created data along with data queried from an analytical database is beneficial for contamination analysis. The ability of ArcView to display data queried directly from Access in a spatial context is extremely valuable.

Jimmy Bowden
GIS Analyst
Daniel B. Stephens and Associates
6020 Academy Rd NE Suite 100
Albuquerque, NM 87111
(505) 822-9400
jbowden@dbstephens.com