



PDHonline Course C423 (15 PDH)

GIS Applications in Water Resources and Environmental Engineering

Instructor: Steve Ramroop, Ph.D.

2020

PDH Online | PDH Center

5272 Meadow Estates Drive
Fairfax, VA 22030-6658
Phone: 703-988-0088
www.PDHonline.com

An Approved Continuing Education Provider

COURSE CONTENT – Lecture 2

Course C423 is a practical GIS course with particular reference to applications in Water Resources and Environmental Engineering, addressing some of the basic, yet common GIS needs in these two fields. The course describes the availability of free databases and public domain GIS software, and their use in performing common GIS processing tasks that are relevant to Water Resources and Environmental Engineering. The ability to search, download and make use of such data sets and software is presented within the context of this course. The course consists of 5 lectures, which start off with introductory level material in Lecture 1, and progressively builds up through the remaining four lectures. A basic awareness of the concept of GIS is desirable but not required as a pre-requisite for this course.

The following topics are covered in this 2nd lecture of Course C423:

- ❖ A discussion of common GIS data types such as:
 - Shapefiles;
 - GeoMedia;
 - ArcInfo coverage;
 - ArcInfo Interchange export format (E00);
 - Geodatabase; and,
 - AutoCAD, Microstation, etc.

GIS data types

GIS is a commonly used to capture, store, manage, process, and output spatial data to assist in making better decisions by various users. For this reason, there is a variety of spatial data types depending upon the software being used and the computer operating platform. This lecture will address some of the commonly used data types.

Some Common data types are as follows:

- ❖ Shapefiles (.shp)
- ❖ GeoTIFF or TIFF with world files (.tif, .tfw)
- ❖ JPEG with world files (.jpg, .jpw)
- ❖ GPS data (.gpx)

These are widely used data types, and a lot can be accomplished with just these vector data formats. Other vector formats that are common are:

- ❖ ArcInfo Binary Coverage
- ❖ ArcInfo Interchange File (.e00)
- ❖ MapInfo (.tab, .mid, .mif)
- ❖ SDTS, a data transfer standard for both vector and raster data
- ❖ Topologically Integrated Geographic Encoding and Referencing (TIGER) data, used and distributed by the U.S. Census Bureau
- ❖ Digital Line Graphics (DLG)

There are a lot of raster formats available, including the following:

- ❖ ERMapper Compressed Wavelets (.ecw)
- ❖ Erdas Imagine (.img)
- ❖ Digital Elevation Models (.dem)
- ❖ JPEG 2000 (.jp2, .j2k)
- ❖ Multi-Resolution Seamless Image Database (MrSID) (.sid)
- ❖ GTOPO30, a global digital elevation model (DEM) derived from a number of raster and vector sources

Some of the formats are open, meaning that they have a published specification and users can use it to write applications and utilities that work with the format. Others are closed, requiring you to use the vendor provided Application Program Interface (API). Of course, this is a concern only if users want to write their own applications and utilities. However there are free services called Open Source GIS Services (**OSGIS**) which specializes in providing cost-effective GIS products and services to small businesses using open source and inexpensive commercial software. If users are content with using the OSGIS applications available, someone else has done the hard work.

1. Shapefiles ¹

The shapefile is a vector data storage format for storing the location, shape, and attributes of geographic features. A shapefile is stored in a folder and contains one feature class.

The shapefile is the proprietary data format developed by the Environmental Systems Research Institute (ESRI)². It is developed and regulated by ESRI as a (mostly) open specification for data interoperability among ESRI and other software products.

There are other descriptions of what is a shapefile. For example:

- ❖ Shapefiles as a set of files that contain a set of points, arcs, or polygons (or features) that hold tabular data and a spatial location.
- ❖ A digital vector (non-topological) storage format for storing geometric location and associated attribute information.

In shapefiles the geometry for a feature is stored as a shape comprising a set of vector coordinates. GIS data has “intelligence” built into its structure. The intelligence is linked to the topology. Topology is a mathematical approach that allows us to structure data based on the principles of feature adjacency and feature connectivity. It is in fact the mathematical method used to define spatial relationships. Shapefiles do not have the processing overhead of a topological data structure. They have advantages over other data sources such as faster drawing speed and edit ability. Shapefiles handle single features that overlap or that is noncontiguous. They also typically require less disk space and are easier to read and write.

¹ <http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf>

² <http://www.esri.com/>

1.1 How Shapefiles Can Be Created

Shapefiles can be created with the following four general methods:

- ❖ Export -- Shapefiles can be created by exporting any data source to a shapefile using ARC/INFO®, PC ARC/INFO®, Spatial Database Engine™ (SDE™), ArcView GIS, or *BusinessMAP™* software.
- ❖ Digitize -- Shapefiles can be created directly by digitizing shapes using ArcView GIS feature creation tools.
- ❖ Programming -- Using Avenue™ (ArcView GIS), MapObjects™, ARC Macro Language (AML™) (ARC/INFO), or Simple Macro Language (SML™)

An ESRI shapefile consists of a main file, an index file, and a dBASE table. The main file is a direct access, variable-record-length file in which each record describes a shape with a list of its vertices. In the index file, each record contains the offset of the corresponding main file record from the beginning of the main file. The dBASE table contains feature attributes with one record per feature. The one-to-one relationship between geometry and attributes is based on record number. Attribute records in the dBASE file must be in the same order as records in the main file.

The shapefile consists of at least 3 parts that exist as files with the same primary name but a different extension:

- ❖ **.shp** - shape format which contains the feature geometry of lines, or points or polygons but never any mixture of geometry.
- ❖ **.shx** - shape index format which is a positional index.
- ❖ **.dbf** - an old school file format that is a remnant of the DBase system which is an old database system. This file holds attribute data such as names, addresses and descriptions.

There are a number of other files that may exist depending on the nature of your shapefile:

- ❖ **.prj** - a projection format will be present if a projection is defined for the shapefile. Projections are important to create for data because they reflect real world coordinates and as such can be reused and exported often.
- ❖ **.sbn** and **.sbx** - spatial indexes of the features in the shapefile.
- ❖ **.shp.xml** - metadata in XML format. Metadata is data about data.

One common type of shapefile which is available for free is the TIGER/Line shapefile. The TIGER/Line Shapefiles are extracts containing selected geographic and cartographic information from the Census Bureau's MAF/TIGER database. The MAF/TIGER database was developed at the Census Bureau to support a variety of geographic programs and operations including functions such as mapping, geocoding,

and geographic reference files that are used in decennial and economic censuses and sample survey programs.

Spatial data for geographic features such as roads, railroads, rivers, and lakes, as well as legal and statistical geographic areas are included in the product. Other information about these features, such as the name, the type of feature, address ranges, and the geographic relationship to other features, also are included. The TIGER/Line Shapefiles and associated relationship files are offered in a compressed format. One zipped file is available for each layer; it has a file extension of .zip. The TIGER/Line shapefile can be downloaded at:

<http://www.census.gov/geo/www/tiger/tgrshp2010/tgrshp2010.html>

Figure 1 shows an example of a TIGER/Line shapefile of the New York area.



Figure 1: An example of a TIGER/Line shapefile of the New York

1.2 GeoMedia

GeoMedia is the technology and suite of software components in Intergraph³ Corporation's GIS and family of software products that include: GeoMedia Objects, GeoMedia Viewer (free), GeoMedia, GeoMedia Professional, GeoMedia WebMap, GeoMedia WebMap Professional and others. It was developed as client or server software specifically for the Microsoft Windows environment.

³ <http://www.intergraph.com/>

There is no single proprietary format for GeoMedia. Its uses all formats. The core technology of GeoMedia makes it possible to simultaneously read data directly from multiple GIS data sources, which are presented in Table 1 below.

GIS Organization	GIS data type
ESRI	Shapefiles
ESRI	ESRI Coverage
ESRI	ESRI SDE via FME plug-in for GeoMedia
AutoDesk	AutoCAD DWG, AutoCAD DXF
Bentley Systems	MicroStation DGN
Oracle Corporation	Oracle Spatial
MapInfo	MapInfo
Intergraph	FRAMME
Intergraph	Modular GIS Environment (MGE)
Intergraph	GeoMedia Warehouses Microsoft Access Microsoft SQL Server

Table 1: Common GIS types and their developers

Along with the flexibility GeoMedia provides in accessing a wide range of "source" data comes the complexity of providing the information about each data type and the software needs to correctly process and present data in an integrated workspace. Setting up data servers to access and properly process warehouse data is one of the most complex tasks performed in GeoMedia. GeoMedia is delivered with the following servers:

- ❖ Access
- ❖ ARC/INFO
- ❖ ArcView
- ❖ CAD
- ❖ FRAMME
- ❖ MGE
- ❖ MGE Data Manager (MGDM)
- ❖ MGE Segment Manager (MGSM)
- ❖ Oracle

An example of a GeoMedia layout is shown in Figure 2. The area shown is South Huntsville, Alabama.⁴ The basic map features are shown together with the topographic features depicting the planning map theme.

⁴ <http://dominoc925.blogspot.com/2009/10/geomedia-61s-new-export-layout-to.html>

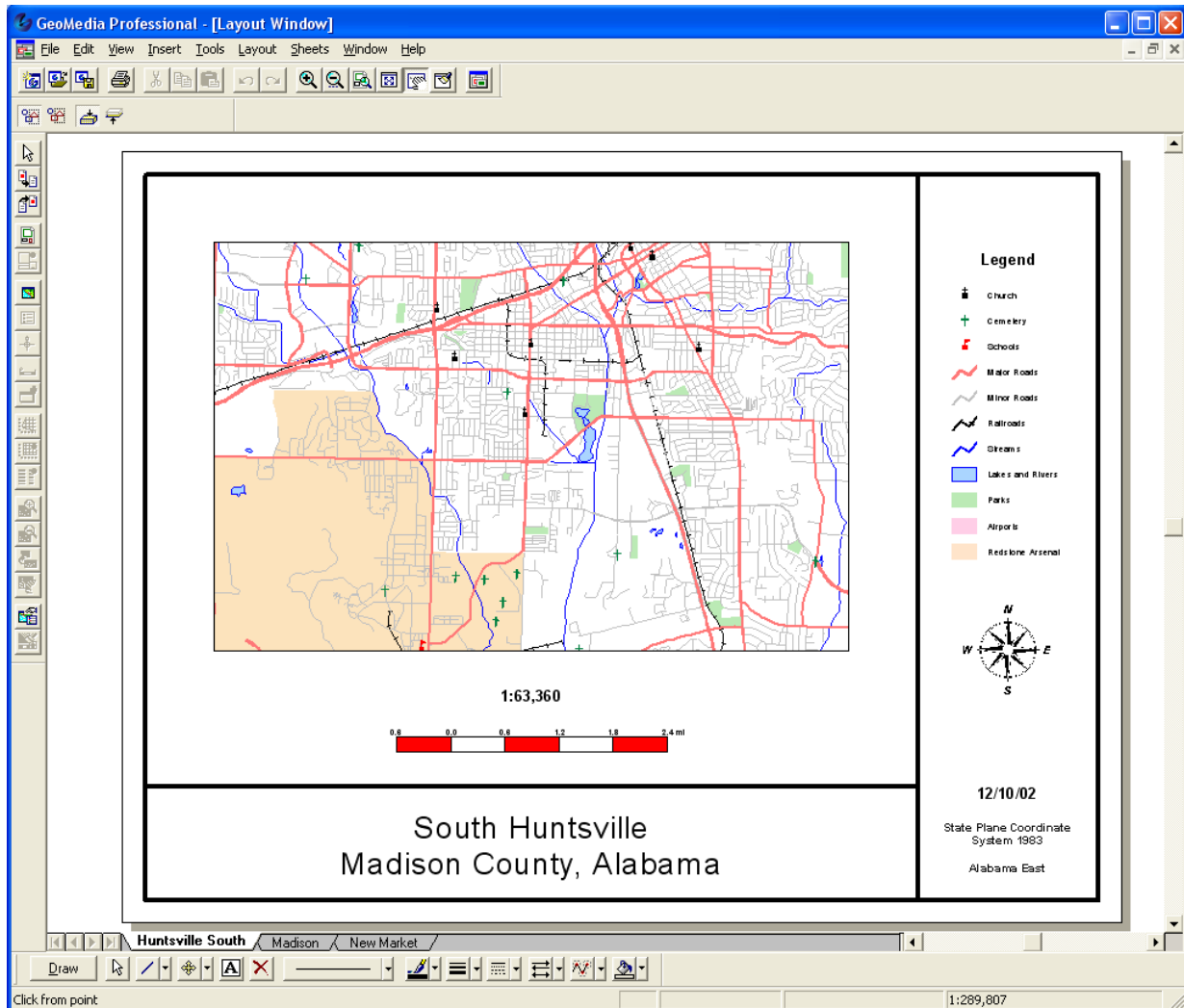


Figure 2: An example of planning data using GeoMedia for South Huntsville, Alabama

1.3 ArcInfo Coverage

ArcInfo (formerly called **ARC/INFO**) is a full-featured GIS produced by ESRI, and is the highest level of licensing (and therefore functionality) in the ArcGIS Desktop product line. It was originally a command-line based system. The command-line processing abilities are now available through the graphical user interface of the ArcGIS Desktop product.

The first version of ARC/INFO was launched in 1982 on minicomputers. The early releases of ARC/INFO were a set of FORTRAN programs linked and accessed through a command-line interface built with the scripting language of the minicomputer. The software was built under a paradigm of tools that could be used together within a command-line interface to perform GIS database development, geoprocessing, and output functions. After ARC/INFO version 7, ESRI underwent a major change in its GIS product family when it released ArcGIS 8.0 late in 1999. With this release, ARC/INFO was discontinued and its code base largely frozen. However,

there is a large amount of data sets which were collected using ArcInfo and is shared via the Internet.

ArcInfo coverages appear as folders on the computer. Folders contain many files with names such as *arc.adf*, *arx.adf*, *pal.adf*, *pat.adf* etc. The files inside the coverage folder are binary files and can not be viewed in any word processing program or ASCII editor. A folder that contains ArcInfo coverages is called a workspace. A workspace always contains a folder named *info*. The *info* folder contains attributes to all ArcInfo coverages within the workspace.

ArcInfo coverages **CAN NOT** be copied in the windows environment by 'copy-paste' or 'drag and drop' from one folder to the other. By copying in windows the data in the **info** directory will be lost and the coverage will be irrecoverably corrupted and lost forever. However users can move the entire workspace using the windows 'copy-paste' or 'drag and drop'.

ArcInfo coverages are of three types:

- ❖ **Point Coverages**: If a lab.adf file exists with point attributes (eg. COVER.PAT in info) then this will also be imported and attached to the points. If no point attribute table exists, the points will only have one attribute, called USER-ID, containing what would be the COVER-ID field in a point attribute table.
- ❖ **Arc Coverage**: The arcs (arc.adf) are imported as a single vector layer. If an arc attribute file exists (eg. aat.adf or COVER.AAT in info), then it will also be imported, as well as the attributes associated with each of the arcs. Whether or not the aat.adf file exists, each arc will have attributes FNODE (from node), TNODE (to node), LPOLY (left polygon), RPOLY (right polygon) which are related to topology. If no aat.adf file exists, the user attribute will also be attached to the arcs with the name USER-ID.
- ❖ **Polygon Coverage**: A polygon coverage is treated as two layers: an arc coverage containing the polygon boundary arcs, and a point coverage containing the polygon centroids and any associated polygon attributes.

If the ArcInfo coverage has a defined projection (prj.adf) then this will also be read and associated with the vector layers.

An example of an ArcInfo coverage data set is shown in Figure 3, of the Columbia River Basin ecological provinces and sub-basins.⁵

⁵ <http://nwhi.org/images/nwhi/crbecosb.jpg>



Figure 3: An example of ArcInfo coverage data

1.4 ArcInfo Interchange export format (E00)

The ArcInfo interchange export format of coverage data is called .E00 (pronounced e-o-o or e-zero-zero). The E00 files are used to conveniently copy and move ArcInfo GIS coverages and grid data types of ArcInfo. An .e00 file must be "imported" in order to use the data in ArcView or other GIS software.

The file extension which is used by ESRI's ArcInfo software was called coverages, which are typically graphic data (and its related files such as map projection, tabular data, ground control files, etc) are stored in subdirectories. In order to export coverages, an interchange export file is created which had an ".E00" file extension. E00 is a transfer format available both as ASCII and binary form. E00 is a common format for GIS data found on the Internet.

Description	Zipped shapefiles	File Size	ArcInfo .e00 files	File size
ArcView project and all shapefiles compressed	pw_ss.zip	17.6 Mb		
Age dates	age_dates.zip	4 Kb	age_dates.e00	4 Kb
County lines	county_lines.zip	4 Kb	county_lines.e00	8 Kb
Drill holes	drill_holes.zip	4 Kb	drill_holes.e00	12 Kb
Faults	faults.zip	12 Kb	faults.e00	36 Kb
Aeromagnetic form lines	fomlines.zip	12 Kb	fomlines.e00	52 Kb
Geology	geology.zip	40 Kb	geology.e00	216 Kb
Outcrops	outcrops.zip	116 Kb	outcrops.e00	904 Kb
Rock quarries	rock_quarries.zip	4 Kb	rock_quarries.e00	16 Kb
Structure	structure.zip	8 Kb	structure.e00	28 Kb
Cross section geology	xsect_geology.zip	8 Kb	xsect_geology.e00	28 Kb
Cross section line	xsection_line.zip	4 Kb	xsection_line.e00	4 Kb
All .e00 files compressed			all_e00.zip	288 Kb

Figure 4: An example of E00 data from USGS

There are many data sets that are of the E00 data type. To the inexperienced GIS user, the E00 data sets can often be overlooked. Figure 4 is an example of a website at the USGS⁶ where a tremendous amount of E00 data sets is available for Port Wing, Solon Springs, and parts of the Duluth in Wisconsin.

1.5 Geodatabase

The geodatabase is a collection of geographic datasets of various types, with the most basic types being vector, raster, and tabular data. There are three types of geodatabases: file, personal, and ArcSDE. Geodatabases are the native data format for ESRI's ArcGIS.

The *geodatabase* supports a model of topologically integrated *feature classes*, similar to the *coverage* model. It also extends the coverage model with support for complex networks, *relationships* among feature classes, and other object-oriented *features*. The ESRI ArcGIS Applications (ArcMap., ArcCatalog., and ArcToolbox.) work with geodatabases as well as with coverages.

⁶ <http://pubs.usgs.gov/of/2004/1303/>

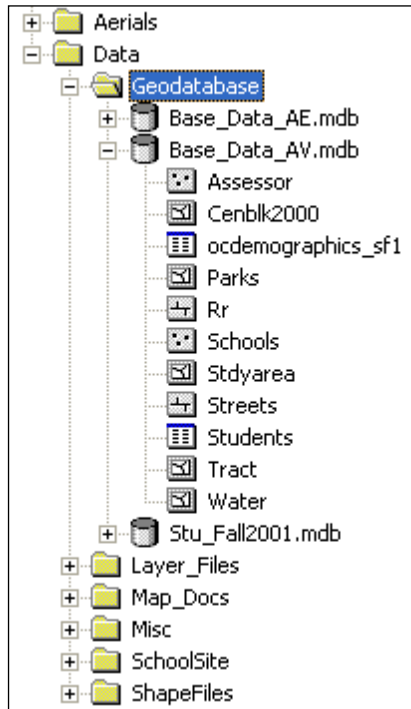
Within a spatial database, spatial data is treated as any other data type. Vector data can be stored as point, line or polygon data types, and may have an associated spatial reference system. A geodatabase record can use a geometry data type to represent the location of an object in the physical world and other standard database data types to store the object's associated attributes. Some geodatabases, such as Rasterlite and those used by ESRI in their ArcGIS software, also include support for storing raster data.

Many geodatabases have custom functions that allow the spatial data to be manipulated and queried using Structure Query Language (SQL), for example to find all the residents of an area within an exposure zone for a potential environmental hazard. However the spatial data in some geodatabases can only be accessed by using specialized client software.

Like other spatial data formats geodatabases can also be used to serve data directly to web map server software, such as ESRI's ArcGIS Internet Map Server, MapServer and Google's mapping API.

To summarize the geodatabase offers you the ability to:

- ❖ **Store** a rich collection of spatial data in a centralized location.
- ❖ Apply **sophisticated rules and relationships** to the data.
- ❖ Define **advanced geospatial relational models** (e.g., topologies, networks).
- ❖ **Maintain integrity** of spatial data with a consistent, accurate database.
- ❖ Work within a **multiuser access and editing** environment.
- ❖ **Integrate spatial data** with other IT databases.
- ❖ Easily **scale** your storage solution.
- ❖ Support **custom features** and behavior.
- ❖ **Leverage** your spatial data to its full potential.



In Figure 5 is shown an example of a geodatabase. There are three geodatabases shown using ArcCatalog in the ArcGIS software. Under the geodatabase called "Base_Data_AV.mdb" there is a set of data sets which includes point, line and polygon feature classes; and access tabular data.

Figure 5: A Geodatabase

1.6 AutoCAD

AutoCAD is a CAD (Computer Aided Design or Computer Aided Drafting) software application for 2D and 3D design and drafting. It was developed and sold by Autodesk, Inc. First released in December 1982, AutoCAD was one of the first CAD programs to run on personal computers, notably the IBM PC. At that time, most other CAD programs ran on mainframe computers or mini-computers which were connected to a graphics computer terminal for each user.

DWG ("drawing") is a file format used for storing two and three dimensional design data and metadata. It is the native format for several CAD packages including AutoCAD, and IntelliCAD (and its variants). In addition, DWG is supported non-natively by many other CAD applications. The .bak (drawing backup), .dws (drawing standards), .dwt (drawing template) and .sv\$ (temporary automatic save) files are also DWG files.

The native DWG drawing file format for AutoCAD 2010 is not compatible with previous AutoCAD releases. AutoCAD 2010 can open drawing file formats from previous releases dating back to the 1980's with AutoCAD 2.x, In order to open AutoCAD 2010

files in previous releases, you will need to use the SAVEAS command and save it to the appropriate DWG file format or set your default Save format version in the Options.

AutoCAD DXF (Drawing Interchange Format, or Drawing Exchange Format) is a CAD data file format developed by Autodesk for enabling data interoperability between AutoCAD and other programs.

An example of an AutoCAD drawing is shown in Figure 6. It represents a viewport layout with its measurements and dimensions shown. This is a typical DWG file type.

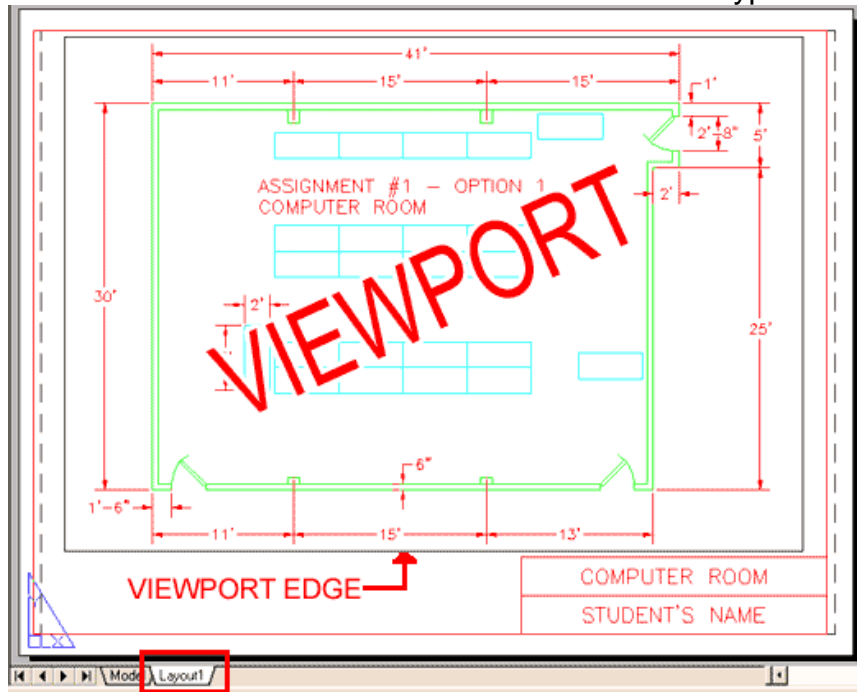


Figure 6: An example of AutoCAD drawing

1.7 Mircostation files

MicroStation is a CAD software product for 2- and 3-dimensional design and drafting, developed and sold by Bentley Systems⁷. The latest versions of the software are released solely for Microsoft Windows operating systems, but historically MicroStation was available for Macintosh platforms and a number of Unix-like operating systems.

MicroStation is the platform architectural and engineering software package developed by Bentley Systems, Incorporated. Among a number of things, it generates 2D/3D vector graphic objects and elements. Current version is MicroStation V8i.

The DGN extension in a MicroStation file is short for "design." An example of a DGN file is shown in Figure 7. It depicts a slotted drain spillway assembly that was designed using MicroStation.

⁷ <http://www.bentley.com/en-US/>

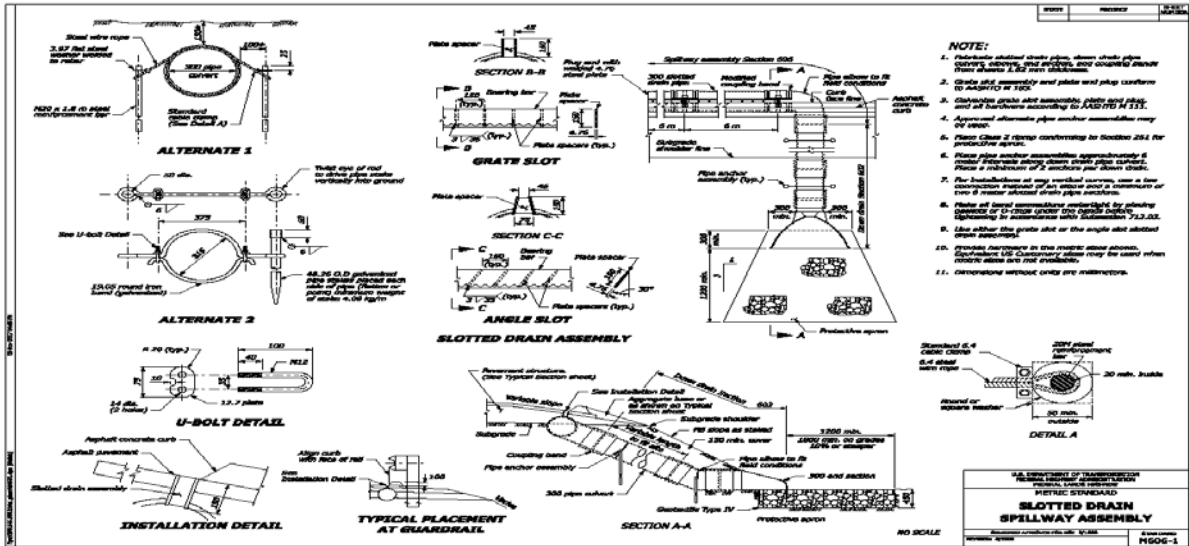


Figure 7: Example of a MicroStation DGN file

1.8 Other data types

There are other data types available, and a presentation of all of them is beyond the scope of this lecture. Discussed already are some of the more popular data types being used in the GIS world.